UNIVERSITY OF LIVERPOOL

WATER, GOVERNANCE AND HUMAN DEVELOPMENT VARIABLES IN DEVELOPING COUNTRIES:

MULTIVARIATE INTER-RELATIONSHIPS ANALYSIS AND STATISTICAL MODELLING USING BAYESIAN NETWORKS

Thesis submitted in accordance with the requirements of the

University of Liverpool for the degree of Doctor in Philosophy

2014

CELINE DONDEYNAZ

SCHOOL OF ENGINEERING

Supervisor: Dr César Carmona Moreno

THE AUTHOR

Celine Dondeynaz owns a Bachelor's degree in Civil and Environmental engineering from the Henri Poincaré University (Nancy, France), and a Master's degree in Environmental Sciences from the University of Bourgogne (France). She has been working for the past 6 years for the European Commission as scientific officer on cooperation and water resources management.

The following publications are related to this thesis.

International peer reviewed papers and book chapter

- Dondeynaz, C., López Puga, J., and Carmona Moreno, C.: Bayesian networks modelling in support to cross-cutting analysis of water supply and sanitation in developing countries, Hydrol. Earth Syst. Sci., 17, 3397-3419, doi:10.5194/hess-17-3397-2013, 2013. Available at http://www.hydrol-earth-syst-sci.net/17/3397/2013/hess-17-3397-2013.html
- Dondeynaz, C., Carmona Moreno, C., and Céspedes Lorente, J. J.: Analysing inter-relationships among water, governance, human development variables in developing countries, Hydrol. Earth Syst. Sci., 16, 3791-3816, doi:10.5194/hess-16-3791-2012, 2012. Available at http://www.hydrol-earth-syst-sci.net/16/3791/2012/hess-16-3791-2012.html
- Dondeynaz,C., Carmona-Moreno, C.,Vidal Legaz, B.: Cross Analysis and Key Indicators in the Water Sector, Brazil Profile and Scenarios. Chapter 2 in book: The BraSIS Project- EU-Brazil sector dialogue project on Sanitation, in press, 2013.

Conference papers

1. López Puga, J.,Dondeynaz, C., Carmona-Moreno,C.: Towards, dynamic Bayes Networks to model Times Series from the WatSan4Dev subset, Applied Stochastic Models and Data Analysis Conference (ASMDA2013), 25-28 June 2013, Editor C.H.Skiadas, p 133-134, 2013. The author wrote and presented the paper to the conference.
2. Dondeynaz, C., López Puga, J. Carmona-Moreno, C.: Modelling Inter-relationships among water, governance, human development variables in developing countries with

Bayesian networks. EGU General Assembly 2012, Geophysical Research Abstracts, Vol. 14, EGU2012-1675, 2012.

Available at http://meetingorganizer.copernicus.org/EGU2012/poster_programme/9165

3. Dondeynaz, C., Carmona-Moreno, C., Leone.A, and Chen,D.: Inter-relationships among water, governance, human development variables in developing countries: a database coherence analysis, EGU General Assembly 2011, Geophysical Research Abstracts, Vol. 13, EGU2011-7453, 2011. Available at

http://meetingorganizer.copernicus.org/EGU2012/EGU2012-1675.pdf

ACKNOWLEDGEMENTS

I would like first to deeply thank Prof Daoyi Chen from the University of Liverpool who makes this research possible through his active support. My profound gratitude goes to Dr César Carmona Moreno from the EC Joint Research Centre (Italy) for his patience, guidance, wisdom and permanent support during these years. I also express my deep thanks to Prof Jorge López Puga from the Catholic University of San Antonio (Spain) and Prof José Céspedes Lorente form the University of Almeria (Spain) who collaborated to this research sharing their ideas, skills and enthusiasm.

I extend my thanks to the Joint Research Centre of the European Commission for supporting and leaving me time to complete this thesis. In particular, I am grateful to Alan Belward of the Land Management Unit and Giovanni Bidoglio of the Water Resources Unit for their constant support. I also thank my colleagues at the JRC for the encouraging and friendly working environment so important to keep the research going. A special thank goes to Prof Jean-Pierre Prost who supported me to initiate this work. I finally thank Prof Timothy Short and Prof Diego Vazquez-Brust for the time they dedicated to review this thesis and their valuable comments.

A warm and special thank goes to my mother Françoise, and my family Guillaume, Mathilde, and Jean-Baptiste for their unconditional love, patience, understanding and support that gave me the necessary enthusiasm, motivation and strength to continue and complete this work.

This research resulted from a collaboration between the Joint Research Centre of the European Commission (Italy) and the University of Liverpool (UK). This research benefited from an Engineering and Physical Sciences Research Council grant and the financial support of European Commission through RALCEA and ACE-Water projects (EC contracts AA 2010/241-167 and AA 2009/220-992, respectively).

COPYRIGHT

DECLARATION

No portion of the work referred to in the dissertation has been submitted in support of the application for another degree or qualification of this or any other University or other institute of learning.

TABLE OF CONTENTS

LIST OF MOST USED ABBREVIATIONS AND ACRONYMS

Please refer to the Appendix E that provides the list of variables abbreviations and their signification while reading the paper version.

AP	Activity Pressure (agriculture, industries and municipal demand)
BNs	Bayesian Networks
CEC	Country Environmental Concern
DAG	Directed Acyclic Graph
DBNs	Dynamic Bayesian Networks
EU	European Union
FA	Factor Analysis
FAO	Food and Agriculture Organisation
HDP	Human Development and Poverty
IWRM	Integrated Water Resources Management approach
JMP	Joint Monitoring Programme from the United Nations
MAR	Missing at Random assumption
MDGs	Millennium Development Goals
OECD	The Organisation for Economic Co-operation and Development
ODA CI	Official Development Assistance Composite Indicator
ODA	Official Development Assistance
OLS	Ordinary Least Square regression
PCA	Principal Component Analysis
TI	Transparency International
UN	United Nations
WB	World Bank
WGI	Worldwide Governance Indexes
WHO	World Health Organisation
WSS	Water Supply and Sanitation

List of Tables

Ordered by Chapter

Chapter 2

Chapter 4

Chapter 4

CHAPTER 6

CHAPTER7

CHAPTER 8

APPENDIXES

List of Figures

Ordered by Chapter

Chapter 1

Chapter 3

Chapter 4

Chapter 5

CHAPTER 6

CHAPTER 7

CHAPTER 8

ABSTRACT

In the last decades, we have assisted to an important expansion of the number of indicators for measuring the development of a country− from the GDP per capita, households' consumption indicators, demographic and medical indicators, schooling rates to governance indexes. This has produced in a first time the development of composite indicators to explain and synthesise the spatial and temporal changes of these different indicators− the Human Development Index (HDI) and its adjusted versions, Multidimensional Poverty Index (MPI), or the Water Poverty Index (WPI), to provide policy makers simple figures to help them in their decisions. The main difficulty faced by the researchers was to explain complex behaviours through single indicators.

This research develops a framework to explain and contribute to the better understanding of the relationships between the existing single and complex indicators in the domain of Water Supply and Sanitation (WSS) in Developing Countries. This framework is based on the Bayesian Networks modelling method (Castelletti & Soncini-Sessa, 2007a), (Giné Garriga et al., 2009), (Dondeynaz et al., 2013).

In addition to building this analytical framework, this research also aims at measuring and analysing the distribution and the influence of Official Development Assistance (ODA) in recipient countries. The approach chosen is global, targeting cross-countries analysis and comparison to capture the principal key variables of water supply and sanitation coverage expansion and its benefits for the country development. Therefore, this research proposes a methodological framework using Bayesian models for analysing water supply and sanitation access levels together with governance, human development (education, health, and income), water resources, the uses of these resources and the ODA. The research outputs could support national decision making and/or donors' strategies, in particular the European Union.

Variables and data are collected at national country scale for 101 developing countries observations in a new database (WatSan4dev) for year 2004. Five country profiles are identified and ranged around five main thematic axes using multivariate and clustering analyses. The countries from profiles 4 and 5 were the least favoured in terms of development and access to WSS, therefore should benefit from ODA support. However, countries from profile 5 received

rather low ODA inputs in 2004, possibly as shown from the models because of their relative instability and poor governance. The modelling approach is led by the principles of robustness and replicability and took into account data availability and nature using Bayesian Networks. It is found that WSS access is strongly associated to country development (+35 % probability change) that is first sensitive, as expected, to the income level. The urbanisation level is the second strong factor associated to development with the limit of slums development. Health care and advanced governance complete these key factors. Lastly, WSS is sensitive to ODA CI where high-level ODA is estimate to benefit first to poor (45%) and middle (34%) development countries at 79% probability.

This modelling allowed, in addition, running probabilistic scenarios to test hypotheses and measure the probable changes on WSS and the development. The methodological process, the outputs of multivariate analysis, the five countries profiles, the Bayesian modelling as well as examples of scenarios are described and analysed. The reference date is first 2004. The analytical and modelling process is then applied to the 2000-2008 period.

CHAPTER 1: INTRODUCTION

1. INTRODUCTION

Access to safe drinking water and basic sanitation for all is still an issue in many countries. On the 6th March 2012, the UN announced that the Millennium target for safe drinking water had been reached, while the target for sanitation was still out of reach[1]. Only 63% of the world has improved sanitation access and this figure is only projected to increase to 67% by 2015, well below the 75% projected in the Millennium Development Goals (MDG's). At the same time, the UN estimates that by 2015, 92% of the global population will have access to improved drinking water.

Beyond these global figures, disparities appear across regions and urban/rural areas. "Only 61% of the people in sub-Saharan Africa have access to improved water supply sources compared with 90% or more in Latin America and the Caribbean, Northern Africa, and large parts of Asia" (UN, 2012). Access to basic sanitation still represents a challenge, given that of the "1.1 billion people who still practice open defecation, the vast majority (949 million) live in rural areas. This affects even regions with high levels of improved water access" (UN, 2012). In addition, relatively good rates of access to improved water supplies could be undermined by poor water quality. In fact, the quality of the water provided through the improved infrastructure is not systematically measured at global scale as mentioned for the case of Latin America by (Onestini, 2011) or more widely in (JMP, 2011) and not available up to now as one of the UN indicators.

Why are there such disparities and challenges? One reason could be that providing safe water or sanitation is more than a matter of infrastructure—it is affected by water resource availability, the existence of the capacities/skills/industry to manage and maintain infrastructure, funding, technology, governance general context and socio-cultural practices. Efforts for improvement should follow an integrative approach, trying to embrace the complexity and cross-cutting characteristics of water supply and sanitation issues. With this in mind, the question is how we are to capture, measure and consider all the elements influencing access to water supply and

[1] WHO press release of the 6th March 2012 is available online: http://www.who.int/mediacentre/news/releases/2012/drinking_water_20120306/en/, last access on the 30th April 2014.

sanitation. Having a clear picture of how to move ahead will help enhance and orient resources, policy making and, finally, actions.

This chapter first looks at the role of the WSS services in country development (section 2). The cross-cutting nature of WSS issues is then described in the context of the international movement toward a more holistic approach (Integrated Water Resources Management - IWRM) to organize the water sector (section 3.1). Developing countries are supported in their efforts towards development through international cooperation, with the action framework reflecting the vision of IWRM (section 3.2). The organisation of the thesis is presented together with the chapter's conclusions in section 4.

2. THE ROLE OF WATER SUPPLY AND SANITATION IN COUNTRY DEVELOPMENT

Having access to water supply and sanitation services (WSS) is one of the key contributors to the improvement of the well-being of the population, the environment **and** the economic development of a country. "The full magnitude of the benefits of water services is seldom considered for a number of reasons, including the difficulty in quantifying important non-economic benefits such as non-use values, dignity, social status, cleanliness and overall well-being" (OECD, 2011a).

However, some aspects of water supply have been economically evaluated, for instance the impacts of inadequate water supply and sanitation on health. (Hutton et al., 2007) estimate that WSS interventions would be cost beneficial in all the sub regions of the developing world if the time savings (reduced time fetching water) and the reduction in diseases that would result were taken into consideration. "In developing regions, the return on a US$1 investment was in the range US$5 to US$46, depending on the intervention. For the least developed regions, investing every US$1 to meet the combined water supply and sanitation MDG lead to a return of at least US$5 orUS$12"(Hutton et al. 2007). Improving the living conditions of the population provides direct economic benefits.

The observed WSS levels in any given country depend on several factors and conditions—it is important to note that an abundance of water resources is not necessarily synonymous with good access to safe drinking water. Algeria has over 75% water supply access despite its scarce resources, while the Democratic Republic of the Congo has less than 50% access, despite having abundant water resources (Fig 1.1 and 1.2).

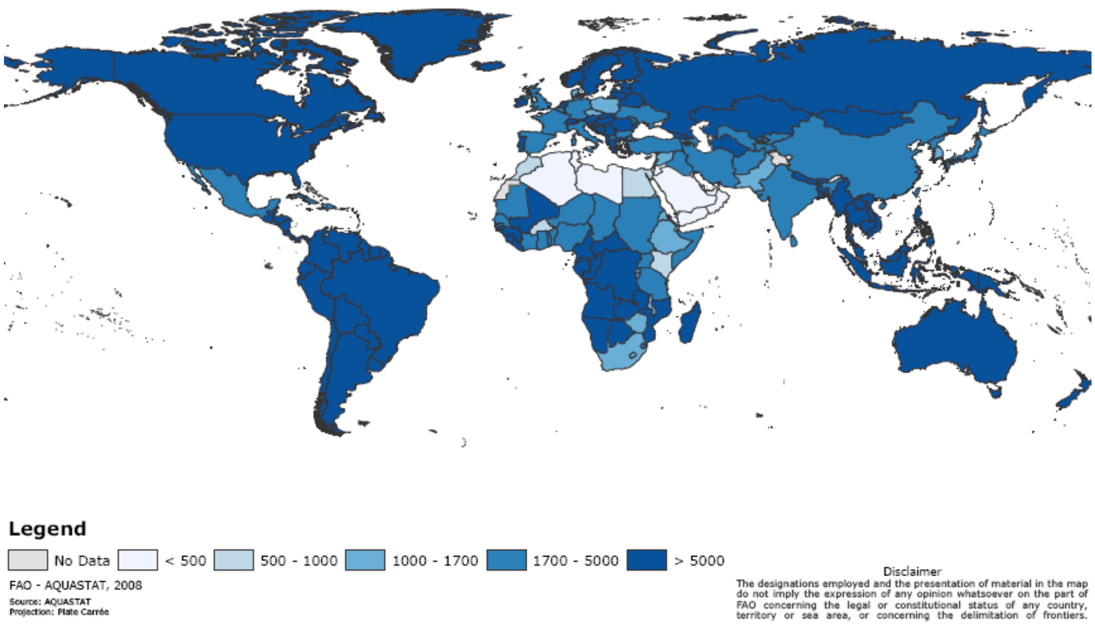


Figure 1.1: Total actual renewable water resources per inhabitant (m3/year) in 2005, (FAO, AQUASTAT)[2]

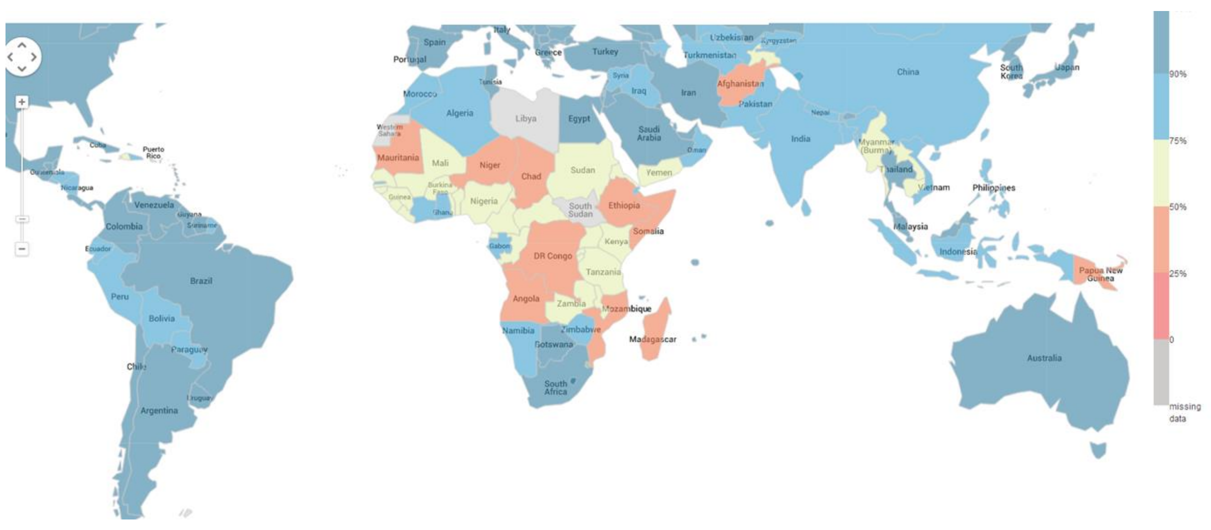


Figure 1.2: Access to an improved water source as percentage of the total population in 2005 (FAO, AQUASTAT) [3]

[2]Map available online and provided by AQUASTAT: http://www.fao.org/nr/water/art/2008/flash/aquastatmaps/gallery1.html, last access on the 30th April 2014.
[3]Map available online and provided by JMP: http://www.wssinfo.org/data-estimates/maps/, last access on the 30th April 2014

Access to a proper water supply and sanitation facilities is the "immediate visible part of the iceberg" and reflects infrastructure—the direct interface with the population. The whole water cycle, beyond the natural one—from water collection to its restitution in the environment—needs to be considered to ensure sustainable WSS services (Fig 1.3).

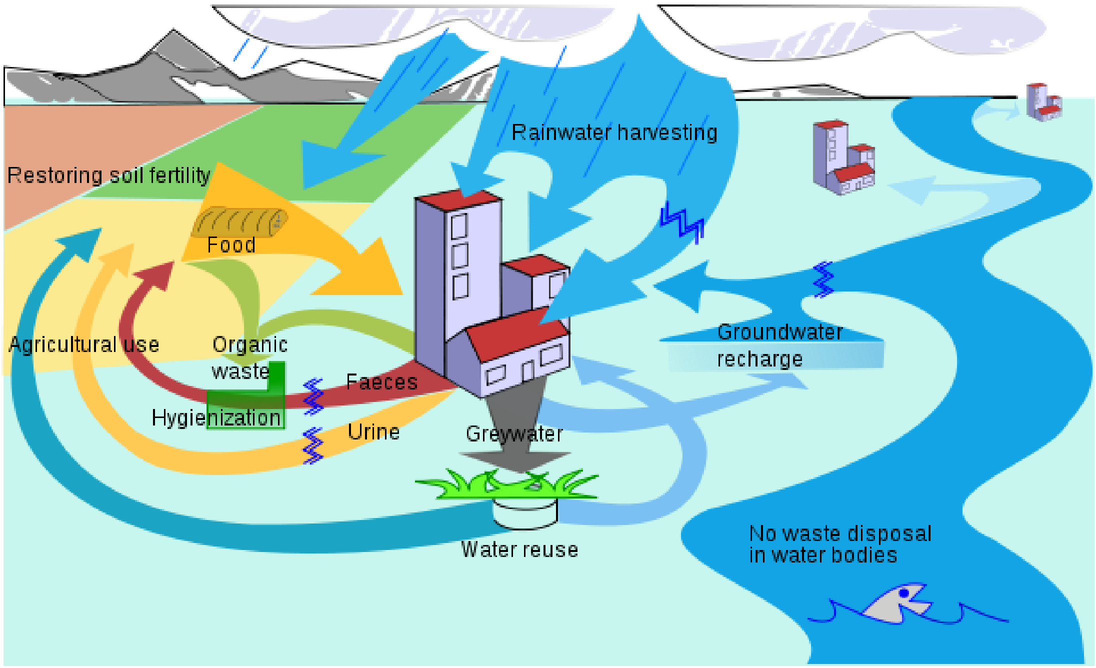


Figure 1.3: Ecological Sanitation Cycle. Source: GTZ.de[4], (UNESCO-IHP & GTZ, 2006)

Conceptually, interest should shift from the WSS infrastructure to the wider reality: the condition of the water sector. Urban demand for water and sanitation services puts pressure on the surrounding environment—the discharge of waste waters is an example of this pressure. This discharge can threaten the quality of the water supply if no preventive/treatment measures are taken. The quantity and quality of water resources are pressured in rural areas, as agriculture is a major consumer of freshwater globally (according to FAO's statistics on freshwater

[4]SWITCH water blog http://switchwatersummit.wordpress.com/2010/06/07/ecological-sanitation-cycle/, last access on the 30th April 2014.

withdrawals[5]). Rivers and ground waters are vectors that connect rural and urban areas, the various water consumers, water supply and sanitation. Ensuring sustainable WSS involves taking these interactions into consideration and aiming to find a balance, solutions or positive synergies between users to maintain the quantity and the quality of resources. Within this research, the water sector refers to all stakeholders, means, activities and resources involved in the delivery of water supply and sanitation services.

Sustainable WSS calls for water resource management to ensure the adequacy of resources to different uses and also to meet the needs of the environment/ecosystem. In fact, services provided by water ecosystems should also be preserved, as these benefit society in terms of “health, water treatment costs, fisheries and recreational values” (Tilman et al., 2002).

Institutional structure/organisation, and thus skills capacities, is required at all levels—from policy to operations—to organise the necessary processes and ensure they are functional. The complexity of the water sector is increased by the involvement of a broad range of stakeholders: public bodies—both at central and decentralised levels—private sector interests, civil society interests, etc. In the case of South Africa, the stakeholders participate at different level—from information to collaboration, at river basin and/or regional level (Du Toit & Pollard, 2008).Water supply and sanitation sector is related to a number of sectors, such as agriculture and food (i.e. performance or conflicts in uses), health (i.e. disease spreading), tourism (i.e. attractiveness), education (i.e. school attendance) and environment (i.e. pressures/pollution).

The WSS access expansion takes place in the context of each country and that context can either foster its expansion or limit it. For instance, the urbanisation process facilitates or complicates the implementation of WSS services depending on the rapidity, intensity or organisation of the population flows induced (Biswas, 2006). Governance and institutional strength and efficiency also appear essential to building and applying strategies to manage water resources and sanitation, to maintaining infrastructure and to negotiating between uses (Roger & Hall, 2003).

[5] FAO fact and figures on water withdrawals: http://www.unesco.org/new/en/natural-sciences/environment/water/wwap/facts-and-figures/all-facts-wwdr3/fact2-agricultural-use/, last access on the 30th April 2014.

Formalising this approach, a holistic management vision was officially adopted at the International Conference on Water and the Environment held in Dublin in January 1992[6]. The principles of the Dublin statement, adopted by the large majority of the international community, are as follows:

- "freshwater is a finite and vulnerable resource, essential to sustain life, development and the environment,
- water development and management should be based on a participatory approach, involving users, planners and policy-makers at all levels,
- women play a central part in the provision, management and safeguarding of water,
- water has an economic value in all its competing uses, and should be recognised as an economic good".

Based on these principles, Integrated Water Resource Management (IWRM) is defined as "a process which promotes the co-ordinated development and management of water, land and related resources, in order to maximise the resultant economic and social welfare in an equitable manner without compromising the sustainability of vital ecosystems (…)" (EuropeAid, 2009). This approach pushes for inclusive ways of thinking, planning and managing water resources, making coordination between all stakeholders and mechanisms of participation essential.

While the organisation of the sector should be handled by national governments in general as it develops, in a number of countries the international community also supports these efforts. The term "cooperation" or "international development cooperation[7]" in this research are defined as the capacity, operational, financial support received by a country to reinforce its capacity to alleviate poverty, whatever the specific domain of intervention (agriculture, energy, water resources, health etc.). States, international, public agencies and Non-Governmental Organisations are important actors in animating the international cooperation. "During the 1990s,

---

[6] Refer to Dublin Statement on water and sustainable development available online: https://www.wmo.int/pages/prog/hwrp/documents/english/icwedece.html, last access the 8th August 2014.

[7] OECD and European Commission use this terminology i.e. OECD glossary : http://stats.oecd.org/glossary/detail.asp?ID=608, last access on the 15th July 2014, or name of EuropeAid

the numbers and roles of international non-governmental organizations (NGOs) taking part in the foreign aid process grew" (Mavrotas & Mc Gillivray, 2009).

3. COOPERATION STRATEGY USED BY THE INTERNATIONAL COMMUNITY

3.1 The water supply and sanitation strategy

Since the Second World War, the international community has been interested in helping countries to fight poverty, to push for economic growth and, as a result, to increase their development. The varying motivations for such support are wide-ranging and often subject to polemics—propelled by anything from moral duties to economic strategies and/or interests. (Carbonnier, 2010) provides a short review of the donors' motivations and the different criticisms levelled at Official Development Assistance (ODA).

Experience shows that ODA was not always successful for a number of reasons, such as its short-link vision of how to develop a country. In brief, the development theory aimed to fulfil the financial gap using developed countries' support to reach the required investment level and thus achieve economic growth and overall development. (Easterly, 2003) argues that this association between aid and growth is not demonstrated while (Mavrotas & McGillivray, 2009) state that there are no consensus on the country-level impact of aid. Since the 1990's, the majority empirical studies have found that country growth would be "on average be lower in the absence of aid" (Mavrotas & McGillivray, 2009). Complementary there is a research interest to identify the conditional relationship between aid and growth and the factor of aid effectiveness (Neanidis & Varvarigos, 2009). However, there are claims that the aid does not contribute to country growth (Rajan, & Subramanian, 2008).

The cooperation related to the water sector often had a strong focus and investments on building infrastructure. The experience gained during this period shows that infrastructure is necessary, but far from sufficient on its own. One example that could be used to demonstrate this fact could be the "UN international Decade for Water and Sanitation" (1981-1990). Over this period, international organizations focused on promoting low-cost technology implementation, either for water supply or sanitation. Some of the goals set were not achieved as a result not only of insufficient investment, but also because there was too strong of a focus on building infrastructure and only weak incentives for sustainability and management (Cairncross, 1992). This resulted in the abandonment or under use of much of the infrastructure due—for instance—to the lack of maintenance, the recovery of costs not being foreseen and available (Evans, 1992) or inadequacy of the technology to local knowledge, capacities and needs (Cairncross, 1992). (Evans, 1992) already estimated in 1992 that 30-40% of the water supply systems in developing countries were not functional. Thanks to the lessons learned from this ten year experience, since the 1990's, the sustainability of infrastructure and the software component are stressed as crucial element to reach safe water and sanitation to all (Black, 1998). ""Promotion" has long been used in rural water supply, particularly in Latin America. It is related to a range of activities seeking to motivate, educate, and organize a community to contribute towards the construction, operation, and maintenance of a new water supply" (Cairncross, 1992).

International cooperation and national strategies have to go beyond the quantitative vision, i.e. the number of kilometres of pipe that needs to be built to benefit a certain number of people at the end of a project, and work on improving the sustainability of these services and efficient water resource management. This change began in the 90's with the international adoption of the IWRM principles in 1992 and the dissemination of this approach. South Africa is a good example of this movement (Karar, 2008), (Schreider & Hassan, 2011) while donors/international organisation have developed implementation guidance on the IWRM (EuropeAid, 2009), (FAO, 2004) or (UN ESCWA, 2004). The IWRM approach spreading is also supported by international networks for capacity building like Cap-Net[8] and Global Water Partnership (GWP[9]) that develop guidance, dialogue and training sessions on the IWRM.

On the donor side, the Paris Declaration on Aid Effectiveness (2005) moved integration forward by setting rules of conduct for donors—they were required to align their actions with the strategy/policy of the recipient countries. In this way, donors should coordinate among themselves, optimizing funds and allowing the implementation of more complex/cross-cutting projects. Among the rules adopted, the declaration pushes for accountability of actors,

[8]Refer to the Cap-net website: http://www.cap-net.org/, http://www.cap-net.org/, last access the 16th August 2014.
[9]Refer to GWP website: http://www.gwp.org/, last access the 16th August 2014.

monitoring of actions and capacities development within the country—elements key to the sustainability of infrastructure.

Since 2000, another international initiative—the Millennium Development Goals (MDGs)—has been orienting the development strategy and the sustainable development of WSS.

3.2 Fight against poverty: MDGs initiative

In 2000, the Millennium Development Goals (MDGs) were set by the international community with the aim of halving poverty by 2015[10]. The MDGs are designed to sustain and focus the commitment of all the stakeholders towards clear and common objectives. This global initiative covers several themes in its fight to tackle the various dimensions of poverty—namely hunger, education, maternal and child health, HIV/AIDS, gender equity, environmental sustainability and global partnership. This framework for the development activities of over 190 countries in ten regions has been articulated into over 20 targets and over 60 indicators. Water and sanitation is included in the environmental sustainability section as Target 7.C: "Halving, by 2015, the proportion of the population without sustainable access to safe drinking water and basic sanitation".

This research begins by looking at the two indicators designed to monitor progress toward this target: i) the proportion of the population with access to an improved water supply and ii) the proportion of the population with access to basic sanitation (JMP Portal[11]). According to the 2008 and 2012 MDGs assessments, significant efforts have been made towards achieving these two targets. In March 2012, the UN announced that the Water Supply target had been reached 5 years early. However, there are still some great disparities, with several regions not on track. Despite an average progression of 16 points of percentage between 1990 and 2010, the developing countries are still slightly under the set target (88% for WS and 75% for S). Sub-Saharan Africa and Oceania show very low coverage on average—61% and 54% in 2010, respectively (Fig 1.4) (UN, 2008). Sanitation is also still behind its target and was neglected in

[10] The full documentation on this MDG initiative is made available online, refer to http://www.un.org/millenniumgoals/, last access the 8th August 2014.
[11] WHO/UNICEF, JMP website http://www.wssinfo.org/, last access the 8th August 2014.

terms of financial commitment and consideration within national and international strategies and at a local level despite the specific focus in 2008 −year of the “international year for sanitation” (UN, 2012).

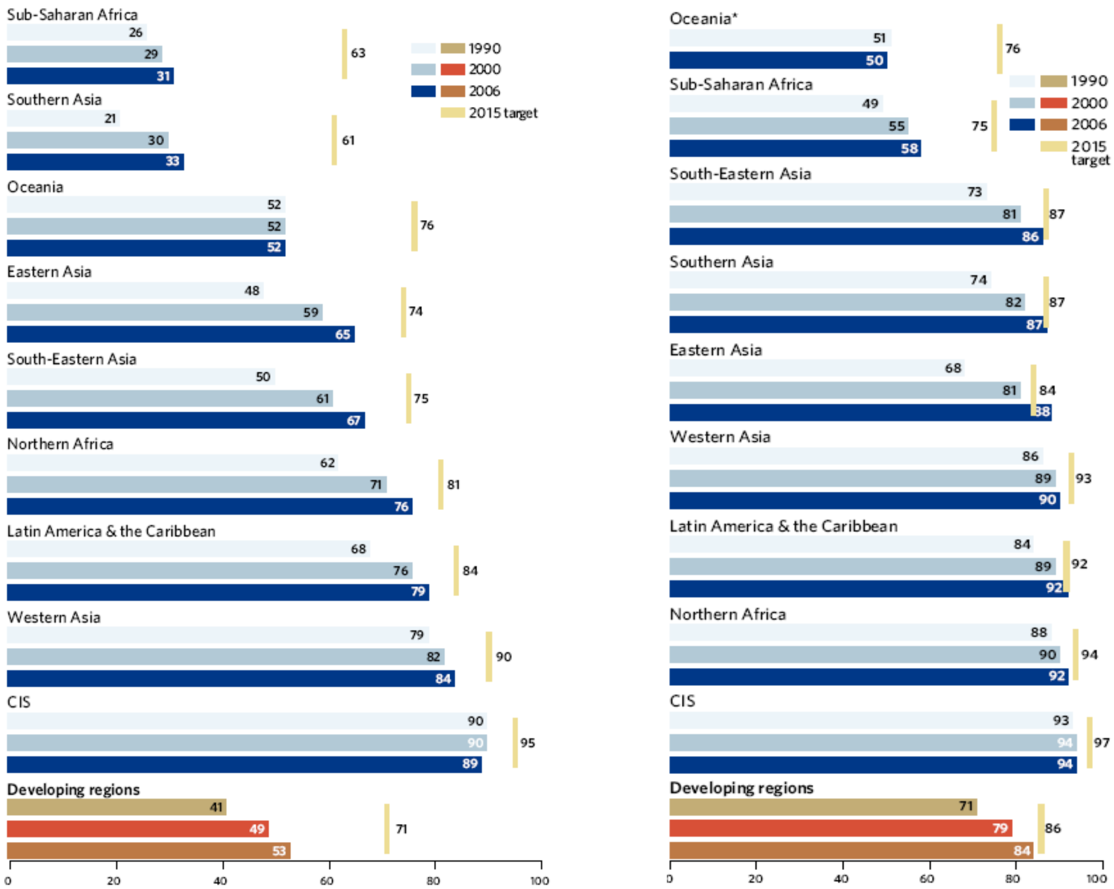
Efforts have been made in the area of sanitation although regions are off track (Fig 1.4). “Almost a quarter of the developing world’s population lives without any form of sanitation. An additional 15 per cent use sanitation facilities that do not ensure hygienic separation of human waste from human contact. Open defecation jeopardizes the entire community, not just those who practice it, because of the increased risk of diarrheal diseases, cholera, worm infestations, hepatitis and related diseases” (UN, 2008). In environmental terms, the lack of sanitation is an important source of surface and ground water resource contamination.

Improving this state of affairs requires both a financial commitment and the formulation of strategies that take into account the particularities of the water sector. The efforts of countries to reach MDGs targets may need to increase in intensity in the near future because of rising constraints, such as population growth pressure in some areas, informal urbanisation and the impacts of climate change.

According to the last (UN DESA, 2011) estimation “the world population is expected to hit 10.1 billion by 2100, reaching 9.3 billion by the middle of this century. Essentially all of the growth will take place in less developed countries and will be predominately among the poorest populations in urban areas.” In these circumstances, efforts towards improving WSS access need to be increased, with a focus on slums areas. Another concern is the way climate change modifies precipitation and temperature patterns, resulting in the recurrence of extreme events such as droughts or floods. The consequences of climate change can range from sizing issues with regards to supply and sanitation infrastructures to the modification of water quality due to water warming and various threats on water supplies (Bates et al., 2008).

a) Proportion of population using an improved sanitation facility, 1990, 2000 and 2006

b) Proportion of population using an improved drinking water source, 1990, 2000 and 2006



Source: the MDG assessment report (UN, 2008)
Figure 1.4: Proportion of population using improved WSS facilities in 1990, 2000, 2006

4. CONCLUSIONS

The implementation and maintenance of the WSS infrastructure are important elements for WSS delivery but is not sufficient (Evan, 1992). The improvement of the access to water supply and sanitation and its sustainability are complex issue. In this way, (Cairncross, 1992) shows that this is more intricate than only connecting a tap/water pump to the nearest water source as WSS involve other complex aspects. The whole water cycle, from the water withdrawal to its restitution to the environment, needs to be considered to ensure a sustainable management of the WSS. In addition, water resources are shared between the agricultural, industrial/commercial sectors but also households and ecosystems (Roger & Hall, 2003). Sustainable WSS services mean to ensure that all water user needs are met in a long term view (EuropeAid, 2009). For this purpose, the concept of Integrated Water Resources Management approach (IWRM) appeared in 1992 to formalize the support to the sustainable water resources management assuring the effective participation of all relevant stakeholders (Dublin statement, 1992). The IWRM is often applied at the river basin scale and, in some cases, like South Africa, at regional scale.

The lack of WSS services is a constraint in terms of health (disease spreading), education (attendance at school) (UNDP & UNICEF, 2006), labor productivity (due to time/resources necessary to fetch water) (Kiendrebeogo, 2012) that has a cost for the economy (Hutton et al, 2007) and the country development. The benefit of WSS services are multiple and directly linked to the population well-being.

Therefore, the international community is committed since 2000 to reduce poverty by defining the Millennium Development Goals (MDGs) [12]. The MDG target 7C aims at halving the population without an improved access to WSS by 2015. The last two reports on the MDGs observed some progress but noticed the persistence of disparities between urban and rural areas, sanitation and water supply and between regions (UN, 2008 and 2012). Mainly Sub Saharan Africa despite significant efforts and South Eastern are considered out of track for water supply access. All developing regions are out of track with regards to the sanitation's target (Fig 1.4).

Starting from these facts, the research proposed aims at identifying the key elements involved in the expansion and the sustainability of WSS services. This objective supposes to consider the developing countries situation and the multiple dimensions involved in WSS and water resources management. The organization of the thesis is in line with how the research was handled to respond to this main objective.

First, the problems and the background of the research are introduced as part of an overview of the research as it relates to water resource management (Chapter 1). The literature review in Chapter 2 covers development, water resources management, governance, financial aspects and highlights the main research gaps in the analysis and modelling of WSS issues. The subsequent

---

[12] The full documentation on this MDG initiative is made available online; refer to http://www.un.org/millenniumgoals/, last access the 8th August 2014.

research questions are also detailed in Chapter 2, section 4. The methodology, the results of the analyses performed and the modelling of the subset for year 2004 are the core work of this research (Chapters 3 to 6). The methodology used in this research work is described in Chapter 3. In Chapter 4, the data selection is made based on relevance to the objective and the processing methodologies used are detailed. Spatial and multivariate analyses of the variables and countries selected for 2004 are then described, interpreting the results obtained (Chapter 5). The description and the results of the modelling done in 2004 are presented, with emphasis on their relevance (Chapter 6). Finally, these methods and processes are extended to consider the period from 2000 to 2007. The integration of the three temporal points (2000, 2004 and 2007) and the modelling results are presented in Chapter 7.

To sum up, the structure of this research work is articulated around eight chapters briefly presented below:

- CHAPTER 1: Introduction

This chapter sets out the background of the various water resources management approaches and their main principles. It specifies the philosophy and strategies behind attempts at cooperation to alleviate poverty as it relates to the water sector.

- CHAPTER 2: Literature review

The literature review covers the various axes of research and concepts related to development and poverty, water ecosystems, resources and infrastructure, governance, financial flows. It also looks at the existing variables with their computation methods paying special attention to inclusive/cross-cutting analytical frameworks, and finally, reviews BNs methods to several applications (both static and dynamic BNs). At the light of the identified gaps, the research questions and main results are described.

- CHAPTER 3: Methodology

This chapter describes the methods used in this work but also the analytical framework built for the analysis of the WatSan4Dev dataset.

- CHAPTER 4: Description and coherency validation of the dataset

This chapter details the creation of the WatSan4dev dataset, with a description of the variables, pre-processing of data (cleaning, error check and imputation processes) and a validation of the data coherency using multivariate analysis.

- CHAPTER 5: Identifying keys variables and needy countries

This chapter presents the results of the analyses, multivariate analysis and linear regression performed on both the variables and the observations of the relevant WatSan4Dev subset. The identification of key variables and the five country profiles are described.

- CHAPTER 6: Modelling the WatSan4dev subset for 2004 using BNs

The BNs methods were applied to the subset for 2004 to validate the methodologies and verify their appropriateness and efficiency in modelling this dataset. The methodology, the different models and the interpretation of the results are detailed in this chapter.

- CHAPTER 7: Temporal analysis and modelling with DBNs

Building on the results of 2004, DBNs are applied to the three relevant years: 2000, 2004 and 2007. The retrospective models developed are analysed and tested to show their efficiency and appropriateness and the methodology is stressed.

- CHAPTER 8: Conclusions and further research

The main results and contributions of this research are summarized and several potential extensions of this research are detailed.

CHAPTER 2: LITERATURE REVIEW

1. INTRODUCTION

This chapter reviews the existing research related to the water sector as well as the different concepts and data sources available. This baseline allows developing and presenting the methodological approach and the objectives of this research. The literature on the water sector focuses first on the different approaches used that are classified by scale and scope. The literature review is organized by thematic areas in section 2, namely: development, governance, water resources and uses and external financial flows. The concepts and the main associated indicators are defined and discussed. In addition, a methodological review is made, summarizing the main advantages and limits of each analytical framework or method. A section is specifically dedicated to the Bayesian Networks method. Section 3 covers the methods used by the relevant data providers to address water-related issues. Section 4 gathers the research gaps identified in the literature review and details the methodological approach, the questions and contributions of this research.

2. THE WATER SECTOR

The existing research and empirical studies related to Water Supply and Sanitation (WSS) is vast and extensive because WSS is studied not only as such but also associated with other related domains. Therefore, the following literature review relates to development, water resources management and uses, governance and, finally to the financing aspects of the water sector.

2.1 The main approaches in studying multi-dimensional problems including Water Supply and Sanitation.

The literature can be classified according to two main approaches considering the scale and the scope (narrow or bi/multi-dimensional) of the study (Table 2.1).

The Integrated Water Resources Management (IWRM) way of thinking and acting within the water sector adheres to a systemic and inclusive vision (Chapter 1, section 3.1). This approach is reflected in this research, by passing the boundaries between sectors: WSS and more generally water resource management are analysed in the context of other issues such as health/education, agriculture/economic activities, governance, finance, etc. This has led to the development of cross-cutting analyses, models and tools to provide insight into the interactions between the various components of water resource management. This type of research can be classified based on two main criteria: i) the number of variables studied, and ii) the scale considered.

In the first type, studies have considered limited dimensions/or variables at national scale but are significant in that they are able to facilitate robust and understandable analyses of few relevant variables. The thematic scope is restricted but the scale is large enough to allow country comparison. For instance, water and sanitation variables were analysed together with one or two aspects, such as financial resources in (Adler et al., 2010a) and child health in (Botting et al., 2010). (Adler et al., 2010a) have built a framework for the analysis of human development index data, financial resources and several MDG targets. They evaluated a country's progress towards the MDGs, taking into consideration development measures and financial flow indicators. (Botting et al., 2010) analysed the impact of Official Development Assistance (ODA) on water and sanitation coverage and infant and child mortality.

Recently, emphasis has been put on governance, in particular the consequences of governance on the water sector and the mechanisms that could improve it. Indeed, it has been found that a large amount of WSS projects that have failed, were not sustainable or encountered difficulties during implementation because of a lack of capacity or governance (Cairncross, 1992) —with corruption proving to be a big issue (TI, 2008). To governance aspects, (Wolf, 2007) added ODA investments in Africa and their volatility to assess their effects on WSS delivery, expanding in this way the number of variables associated with WSS.

In another type of research, the local or basin scale is chosen, making a multi-dimensional analysis of the dynamics influencing water resource management feasible. The scope is large allowing the analysis/modelling of multiple aspects of water resources management, but the scale is restricted to a river basin/local areas. On this scale, the availability of data can be increased. Working on a small scale can also avoid comparability issues, as the number of providers can be limited. The scarcity and consistency of data are major constraints when focusing on developing countries. Therefore, the quantity, quality and overall dynamics of the water resources can be assessed together, taking into consideration withdrawals, constraints restricting their exploitation and specific uses. For instance, (Raskin et al.,1992) simulated the water supply and demand for the 1987-2020 period according to water use practices in the Aral Sea, assuming a “business as usual” scenario. (Mutie et al., 2006) combined Geographic Information System (GIS) tools and remote sensing together with a hydrological model to assess the pressure on the MARA river basin in Kenya. Following the same logic, the Soils Water Assessment Tool (SWAT) is a small watershed to river basin-scale model that simulates the quality and quantity of surface and ground water and predicts the environmental impact of land use, land management practices and climate change. (Braat & Lierop, 1987) and (Brouwer & Hofkes, 2008) explored the relationships between the water ecosystem and the economy and discussed how to integrate them as part of a decision support tool. (Cai et al., 2003), (Pulido-Velazquez et al., 2008) and (Lescot et al., 2013) have all offered modelling solutions that combine hydrological, agronomic and economic/cost-effectiveness features in an attempt to improve resource management choices on a river basin scale— based on the Syr Darya River basin in Central Asia, the Adra river system in Spain and the Gers river in France, respectively.

With more flexibility in terms of scale, the Water Poverty Index (WPI) is an analytical framework that directly associates four other components to the access to WSS services in order to assess “Water poverty”. The variables are distributed in five components: Resources (R), Access (A), Capacity (C), Environment (E) and Use (U) (Sullivan et al., 2003). The methodology has been often applied at community scale or river basin scale. The WPI has been computed once for 147 countries in 2002 (Lawrence et al, 2002). (Giné Garriga et al., (2009) have extended the WPI multi-dimension approach using Bayesian Networks. They have developed a probabilistic model that consists of several sub-networks that reflect the multiple constraints and variables affecting access to water resources in the Turkana district in Kenya.

The next step, then, would be to scale up the framework and methods used in (Sullivan et al, 2003) and (Gine Garriga et al., 2009). The gap resides in adapting a framework for a cross country analysis that considers the multiple factors involved in the WSS accesses (for instance, upscaling the work of (Gine Garriga et al., 2009). The challenge may be to develop a tool related to WSS services complexity within a wider country development trajectory. The extension of the geographic scope will allow worldwide comparisons and the identification of potential similar behaviours across countries.

Scale	Scope	Topics	Description	References
Local	Narrow	WSS infrastructure	Review of indicators and best practices related to performance of the WSS infrastructure	(Alegre et al., 2006), (Estache & Goicoechea, 2005)
	1 to 3 dimensions	Health and Urbanisation	Identification of determinants factors in Diarrhoea prevalence in slums –two cases studies	(El-Fadel et al., 2014)
		Education and WSS	Determinants of absence at school considering WSS access at school and at home in Kenya.	(Dreiblebis et al., 2012)
Local/river basin	Multiple dimensions	5 components Access, Use, Capacity, Environment, Resources.	Computation of the Water Poverty Index (WPI) to assess at community scale the conditions (easy or difficult) of access to WSS services in pilot sites.	(Sullivan et al., 2003)
Cross-country analysis	1 to 3 dimensions	Health, WSS access and ODA	Association of WSS coverage with children health and ODA-WSS assistance flow.	(Botting et al., 2010)
		Development and resources effectiveness	Performance of country towards 5 MDGs considering four financial inputs.	(Adler et al, 2009)
		WSS and rural development	Assessment of impact of water access improvement on rural labour productivity (27 countries)	(Kiendrebeogo, 2012)
	Multiple dimensions	WSS, governance, demographic Aid volatility	Assessment of the impact of Aid and its volatility on services delivery in Africa including demographic and governance indexes.	(Wolf, 2007)
River Basin	Narrow	Water resources and demand by uses	Development of analyses and models to assess the pressure on water resources (quantity/quality).	(Mutie et al., 2006), (Raskin et al., 1992)
	1 to 3 dimensions	Economic needs, hydrological and biological needs	Modelling for optimising the management of both economic and ecosystem needs.	(Braat & Van Lierop, 1987) and (Brouwer & Hofkes, 2008) among others
	1 to 3 dimensions	Water resources, agricultural water demand, climate change	Development of Decision Support tool for aquifer management considering climate scenarios (combination of models).	(Molina et al.,2010)
	1 to 3 dimensions	Water resources, agricultural pressure, cost effectiveness	Multi-dimensional modelling including land, agricultural, and hydrological models to support management of water resources in specific river basin.	(Cai et al., 2003), (Pulido-Velazquez et al., 2008), (Lescot et al., 2013) among others
	Multiple dimensions	Water resources, WSS accesses, governance, capacity, and environment	Computing and Modelling the WPI five dimensions in a Kenyan river basin (Sullivan, 2002).	(Giné Garriga et al., 2009)

Table 2.1: Scope, scale and description of example of studies related to Water Supply and Sanitation.

The next paragraphs review the concepts and indicators available for qualifying the broad dimensions following roughly the Water Poverty Index philosophy

2.2 Concepts and Indicators used to measure Development, Governance, Financial investment, Water resources and their uses.

The access to water supply and sanitation services is measured by the Joint Monitoring Programme in the framework of the MDGs. Table 2.2 summarizes the definition of improved and unimproved facilities for Water supply and Sanitation (WHO & UNICEF, 2006) on which data collection is carried out:

	Improved	Unimproved
Drinking-water	Piped water into dwelling, plot or yard Public tap/standpipe Tubewell/borehole Protected dug well Protected spring Rainwater collection	Unprotected dug well Unprotected spring Cart with small tank/drum Bottled water[a] Tanker-truck Surface water (river, dam, lake, pond, stream, canal, irrigation channels)
Sanitation[b]	Flush/pour flush to: - piped sewer system - septic tank - pit latrine - unknown place/not known where VIP latrine Pit latrine with slab Composting toilet	Flush/pour flush to: - elsewhere Pit latrine without slab/open pit Bucket Hanging toilet/hanging latrine No facilities or bush or field

[a]Bottled water is considered improved only when the household uses water from an improved source for cooking and personal hygiene.
[b]Shared or public facilities are not counted as improved.

Table 2.2 : Classification of improved and unimproved drinking-water and sanitation facilities (WHO & UNICEF, 2006)

These indicators are used in research at national scale and there is widespread consensus on these measures of WSS access. The main remarks have focused on the fact that "improved source" does not necessarily mean safe water resource (Onda et al., 2012), (Sorlini et al., 2013).

The multiplicative effect in terms of indicators occurs when assessing the efficiency of the services. There is an abundance of literature available providing indicators on infrastructure performance and the monitoring of WSS services. (Alegre et al., 2006) has detailed the methods, benchmarking indicators and best practices used to set up a monitoring system for water services. (Estache et al., 2005) reported several studies carried out on a national scale in the developing world on the performance of WSS (with special attention to private management). However, this type of research remains oriented towards operational management.

2.2.1. Qualifying country development and poverty

The literature on the definition and the measurement of country development and/or poverty is vast. Therefore, this section will only summarize the evolution of the conceptual approach used over the recent decades to estimate country development.

Alleviating poverty and fostering development have been first considered dependent to the economic growth (Glewwe & Van der Graag, 1988), (Easterly, 2002). The classification of country according to its development has been therefore made using income indexes such as the GDP/GNP per capita or setting poverty thresholds (Orshansky, 1969). Since the 80's, measuring the welfare or poverty has been extended conceptually as the income measure appears to "represent it inadequately" (Sullivan, 2002). To complement the measure, for instance, the household food consumption (Glewwe &Van der Graag, 1988), medical data in particular related to children (WHO, 1981) have been included.

The Human Development Index (HDI) is a good example of this will to enrich the measurement of the country development and/or poverty. Since 1980, the HDI is computed to measure an average achievement in three key dimensions of human development: "a long and healthy life, being knowledgeable and have a decent standard of living" (UNDP, 2013). The HDI measures health, education and income level but excludes the aspects related to inequalities, human security, empowerment, employment, gender issues, and etc.[13] (Alkire, 2007) provided an analysis of these "missing dimensions" where data are insufficient at national scale to evaluate "1) employment 2) empowerment 3) physical safety and,4) the ability to go about without shame." Pursuing this effort, several composite indexes have been developed; among others: i) in 2010, the Multidimensional Poverty Index (MPI) (Alrike & Santos, 2010), ii) Employment-

[13] Please refer to the UNDP page on Human Development Reports (http://hdr.undp.org/en/content/human-development-index-hdi) that gathers the extensive documentation on the HDI, last access on the 30th July 2014.

adjusted Human Development Index (Mihci et al., 2012) or iii) the Inequality-adjusted Human Development Index (IHDI) (Foster et al., 2005).

The latter introduces the concept of inequality that can balance/constrain the development at national level. The inequality first reflects an unequal wealth repartition and manifests itself in terms of access to education, health, employment, and WSS services. (Temkin, 1993) discussed the multiple dimensions and perceptions of inequality while the well-known GINI index was developed to evaluate the wealth distribution (Ceriani & Verme, 2012). To reduce the inequality in term of WSS, research and specific programs are looking on ways to combine the imperative of WSS services delivery to the poor with the sustainable financing of the WSS services, among others (Gillespie, 2005), (OECD, 2009). The gender is another factor of inequality that can be added to income (Meulder et al., 2004). Handling needs and empowering women in WSS management is clearly formalised in the IWRM declaration (Principle 3 of Dublin statement[14]) and the MDGs initiative set a specific goal to "promote gender equality and empower women" (Goal 3[15]). The woman's role in the provision of clean and sufficient water, and in the sanitation facilities maintenance, hygiene behaviour (O'Reilly, 2010) and general health at household scale is central; however, women's voices are less heard in the public sphere (Makoni et al., 2004) or were/are little considered for the WSS infrastructure development (Reed & Coates, 2003). A number of case studies has explored and demonstrated the benefit of the active women involvement to enable sustainable WSS services (Watts, 2004), (O'Reilly, 2010), (Oluyemo, 2012).

(Sullivan, 2002) highlighted that despite the enhancement of the concept of country development/poverty, the access to water and sanitation was poorly explicitly linked to poverty reduction. The Water Poverty Index therefore associates several development variables (on education, health, wealth) under the WPI component "Capacity" with the access to WSS services (Sullivan et al., 2003).

---

[14] Refer to the Dublin statement (https://www.wmo.int/pages/prog/hwrp/documents/english/icwedece.html), last access on the 10th August 2014.
[15] Refer to Goal 3 of MDG initiative (http://www.un.org/millenniumgoals/gender.shtml), last access on the 10th August 2014.

Evidences of the benefits of appropriate WSS on population well-being were and are provided by reporting/field studies and are: 1) the reduction of children mortality (UNDP & UNICEF, 2006) and diarrhoea prevalence[16] (UNICEF & WHO, 2009), 2) the time saved to fetch water (Hutton et al., 2007), 3) a better attendance at school (UNICEF, 2006). Academic literature reinforces and details these studies. For instance, (Botting et al., 2010) estimated that the countries with greatest gains in sanitation are 8-9 times more likely to have greater children mortality reduction. (Dreibelbis et al., 2013) found that the probability of school absence increases with the distance to the water supply source, in particular in households where children are in charge of fetching water. Rather than the availability, the cleanness of sanitation facilities at school was positively associated with a reduced probability of absence. (El-Fadel et al., 2013) related the diarrhoea prevalence observed into two slums areas to specific failures in the water supply and storage infrastructure, attic water tank, leaky plumbing and cleaning tank frequency.

The urbanisation component has also to be included into the characterisation of country development. Urbanisation process is associated with demographic changes (Notestein, 1945), (Dyson, 2011), (Canning, 2011). These population dynamics are very much relevant to WSS development services where the concentration of population in cities offers advantages and opportunities but also represent a challenge. On one side, "urban centres offer economies of scale in terms of productive enterprise and public investment," (Beall et al., 2010. Cities can be drivers for the economic dynamism, innovation, social changes etc. (Beall et al., 2010). The population agglomeration also facilitates the implementation of collective infrastructures. On the other side, "cities are also marked by social differentiation, poverty, conflict, and environmental degradation" (Beall et al., 2010). In fact, the urbanisation process affecting all developing countries is neither new nor homogeneous. This rapid urbanisation observed in the recent decades led to the increase of the population and the expansion of slums areas and is expected to continue in the two next decades (UN HABITAT, 2003, 2006, and 2011). The urbanisation organisation and management strongly influence the population access to basic services (waste management, WSS services, health basic care etc…). In fact, this phenomenon puts under

[16] See WHO webpage on water-related diseases: http://www.who.int/water_sanitation_health/diseases/en/, last access on the 11th October 2013

pressure the institutional/governmental capacities (from national to local level) that favour "urban inequalities" (Beall et al., 2010) and fragmented cities with heterogeneously served areas; for instance, in Africa (Wambui Kimani-Murage & Ngindu, 2007), (Foppen & Kansiime, 2009), or in Asia (Sharkar & Rahman, 2007).

The evaluation of the urbanisation process and its organisation have raised some methodological and definition issues (Bloom et al., 2010). The urban area definition has raised discussion on, for instance, the minimum number of agglomerated people considered as a city, the distance to which the population is considered agglomerated, etc. In addition, (Bloom et al., 2010) mentioned that countries "can manipulate statistics modifying the number and size of cities" through their definitions. The UN, which manage the data at worldwide scale, "has adjusted historical data ex post whenever possible; a proper reclassification of historical data can be an arduous, or even impossible, task" (Bloom et al., 2010). The UN measure the proportion of urban population living in slums where a slum household is defined as a group of individuals living under the same roof lacking one or more of the following conditions: the access to improved water, access to improved sanitation, sufficient-living area, durability of housing[17]. The essential point for this research is to obtain the estimation of urban population as it is perceived and computed as well as the amplitude of population living in slums for each country. These raw estimates allow to complement the analysis of country development and to detail the link between urbanisation, slums, WSS and development at the scale of developing countries.

This research adopts this multi-dimensional vision to measure the country development and poverty. However, synthetizing a country development in one single measure excludes the analysis of the organisation of and the relationships between sub-variables. Therefore, an alternative/complementary method is to be considered.

2.2.2. Estimating governance

The interest of measuring governance is relatively recent (1990's) in comparison with human development or poverty. The concept of governance as such has not one single definition (Hall,

[17] UN metadata on urban population living in slums online: http://mdgs.un.org/unsd/mdg/Metadata.aspx?IndicatorId=0&SeriesId=711, last access on the 8th August 2014.

2007) but generally expresses the way the power is exerted and evaluated (El-Mikawy & Oia, 2009). Governance can have multiple forms at different scales, thus the concept of multi-level governance. The multi-level governance implies that the state shares the power with regional, local authorities or non-state stakeholders like citizen communities or NGOs (Olowu, 2003). In addition, the governance concept can be declined by sectors, for instance, the health governance (Brabazza & Tello, 2014), the environmental governance (Gerdung, 2004) or the water governance (TI, 2008). Until the 1990's, the measurement of governance has been often more descriptive than empirical due to the difficulty to setup a robust measure of this wide concept (Saha, 2012). In addition to the large scope of governance, the evaluation of good or poor governance relies partially on a subjective, psychological perception. Therefore, it is easier to agree on what is good and what is bad but more difficult to agree on what is in between (Saha, 2012).

Specifically to the water sector and WSS services, a good national governance framework, political stability, regulatory capacities, and limited corruption, are assumed to facilitate the management and maintenance of the WSS services and foster the accountability to the customers as well as other services. (Wolf, 2007) and (Whitford et al., 2010) have found on two different datasets a small positive association of democracy while the regulatory capacities show a stronger positive association with good governance (Whitford et al.,2010). At local/municipal level, the weakness of WSS management and maintenance is often mentioned (i.e. Nigeria case in Ademiluyi & Odugbesan, 2008). At that level, poor governance manifests different effects depending on the organisation of WSS services. In fact, the governance modalities implementation are multiple‒institutional versus community/civil organisation, private/ public partnerships etc. This organisation vary from urban or rural areas where community – based management may be preferred and encouraged in rural areas (for instance by UNDP Community Water Initiative). However, (Ademiluyi & Odugbesan, 2008) stressed that establishing real partnerships with communities to ensure sustainable WSS programs and services require raising interest of the community and setting suitable conditions. Finally, (TI, 2008) highlighted the role and necessity to reduce corruption in the water sector while (Wolf, 2007) found a small positive association between control of corruption and WSS.

The first measures of governance at national level appeared around 1995-96[18]. Up to date, no specific indicator for water governance has been developed. The World Bank research department developed and made available, since 1996, a set of governance indicators: the Worldwide Governance Indexes (WGIs)[19]. These indicators catch and measure six dimensions that can constitute governance, namely: Voice and Accountability, Political Stability and Absence of Violence/Terrorism, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption (Kaufman et al, 2009a). This set of data has been used extensively in research studies such as (Li et al., 2000), (Kirmanoglu, 2003), (Morita & Zaelke, 2005), (Bayyurt & Yilmaz, 2012), (Sohaib Zubair & Khan, 2014) which explored the relationships between economic growth, sustainable development and governance.

The computation of the data set has also been discussed by the scientific community (Apaza, 2009).The main criticisms focus on the choice of the number of indexes and the validation of the construction approach (aggregation methodology) of the indicators. In short, (Thomas, 2009) argued that governance data are based on expert assessments, which are often subjective, that indicators are spotty, with standards error issues and, as a consequence, non-extendable and non-comparable between countries. (Langbein & Knack, 2010) questioned the number of indexes created to measure governance, and justify the number of distinct governance variables. In fact, "the WGIs indexes are designed to signify the relative absence or presence of some very closely related phenomena" (Langbein & Knack, 2010). They called for caution in making country evaluations using these indicators. (Thomas, 2009) also noted that "evidence must be provided to show that a purported measure of a theoretical construct is valid both in its conceptualization and its operationalization, by exploring predicted relationships with other observable variables". (Kaufman et al., 2009b) responded by stating that "construct validity is not a useful tool to assess the merits of the WGIs, and even if it were, Thomas provided no evidence of any practical consequences of failure to meet the criteria of construct validity".

[18] Refer to the WGIs website for updated data and documentation: http://info.worldbank.org/governance/wgi/index.aspx#doc, last access the 30th July 2014.

[19] Refer to the WGIs website for updated data and documentation: http://info.worldbank.org/governance/wgi/index.aspx#doc, last access the 30th July 2014.

An index specific to Sub Saharan Africa has been computed for the 2000-2006 period to measure essential political goods gathered under five categories: Safety and Security, Rule of Law, Transparency and Corruption; Participation and Human Rights; Sustainable Economic Opportunity; and Human Development (Rotberg & Gisselquist, 2008). This specific index would be more appropriate for a study on the African continent.

Since 1995, Transparency International (a Non-Governmental Organization) focuses on one symptom of poor governance, corruption, computing the Corruption Perception Index. This index aims at estimating the perception of individual experts in term of corruption practices and dissemination. This index combines individual surveys from multiple sources (minimum three). As for the WGIs, the use of individual sources that change over time and from country to country has been debated questioning its reliability (Galtung, 2005). (Lambsdorff, 2006) argued that the high correlation observed between CPI values from different sources ensure the overall reliability of the CPI. (Saha, 2012) also indicated the high correlation between CPI and other corruption measures such as the WGI’s Control of Corruption and International Country Risk Guide’s Corruption Index. As for WGIs, the CPI is well-know and used in research studies, such as (Canache & Allison, 2005) that compared and analysed CPI values in Latin American Democraties. (Fisman & Gatti, 2002) have analysed the relationship between decentralisation and corruption at national scale. The decentralisation was found to favour accountability to local users and therefore, the association with lower corruption. However, they mentioned that other theories argued for opposite processes and therefore, “effectiveness of decentralization in reducing corruption may vary significantly depending on the manner in which decentralization takes place”. (Pellegrini & Gerlagh, 2008) looked for cultural/historical and political organisation factors like WGI corruption index, protestants proportion, ethnolinguistic fractionalization, decentralization, fuels and minerals activities or instability that could favour corruption. They suggested a positive association between the proportion of protestants, income, and reduced corruption while political instability showed a negative association.

2.2.3. Assessing Water resources and their uses

This component relates to the water resources (including their quantitative and qualitative features) that can first be mobilized for water supply and sanitation purpose but also widely for country development. The water resources availability depends on biophysical conditions but is also subjected to variability and climate change. The water resources needs depend to the human uses: food production, industries and energy, domestic use and tourism purpose. The next paragraphs will focus on the approach and the related indicators available to assess both the resources and the burden.

Since the 70's, several attempts in estimating water resources availability globally and regionally have been made and are summarised in (Sullivan, 2002). Two main approaches are used, the national level estimates or the grid approach. Freshwater resources have been assessed at global scale and at country level (Shiklomanov et al., 1997). This evaluation is based on the mean annual river runoff. This includes "all the water coming in directly to the hydrological network during rainfall or snowmelt, plus groundwater from the upper aquifers feeding rivers more or less evenly throughout a year." (Shiklomanov et al., 1997). The country estimates also include water transfers from or to other countries. The comparison of resources to demand can only be done at country level (Sullivan, 2002).

Other water resources assessments have been performed on a gridded basis. (Arnell & King, 1998) estimated the local runoff for each grid. The Global Water Availability Assessment, GWAVA (Meigh et al., 1999) estimates the renewable water resources but also combines it with the estimation of the water demands (agriculture, population pressure and climate change elements), thus providing an estimation of water scarcity.

The current climate changes are modifying the water resources distribution over the globe, varying precipitation patterns and increasing the extreme events and disasters: "Pronounced long-term trends from 1900 to 2005 have been observed in precipitation amount in some places: significantly wetter in eastern North and South America, northern Europe and northern and central Asia, but drier in the Sahel, southern Africa, the Mediterranean and southern Asia. More precipitations now fall as rain rather than snow in northern regions. Widespread increases in heavy precipitation events have been observed, even in places where total amounts have decreased" (IPCC, 2008). The FAO computes a dryland index that estimates the proportion of national territory under desertification threat[20]. The climatologic variability affecting the quantity of water resources is an effect measured at long time scale. At least 30-40 years of climatic data are necessary for robust interpretations. Therefore, considering the time scale of this work (a decade), any climatologic interpretation will be avoided.

The water resources assessment should be put in parallel with the demand. The main sources of water demand come from the agricultural sector (livestock and crops), industries, and the population domestic needs. The FAO gathers data on water renewable resources[21] and also withdrawals, measured according to agricultural, industrial and municipal demand (FAO, 2011). The interaction between water resources availability and the various demands is widely studied and modelled, for example, with the WaterGAP model (Alcamo et al.,1997), the LISFLOOD model (De Roo et al, 1998) or the estimation of water scarcity/stress (such as in Meigh et al., 1999), (De Roo et al., 2012). The WPI integrates both components with its sub-indicators "resources" and "uses" (Sullivan et al., 2003).

Beyond the quantitative estimation of resources and demands, the quality of the water resources represents an issue (Onestini, 2011) together with the increase of pollution sources (of types "point" and "non-point") in particular in areas where the access to WSS, the environmental policy and its respect, are limited. Degraded water quality has multiple consequences −on population health, the agricultural/food production, the ecosystem health which has consequences on the economic/human activities. The water resources quality varies, spatially and in intensity, according to a) water resources physical characteristics, b) socio-economic context that extends or limits pollution sources and, c) the state of development that affect the state/stakeholders capacities to solve/mitigate the issue (Bos, 1999), (Ongley, 2001). An usual approach in developed countries to evaluate the water quality is to monitor key chemical/biological/physical elements present in surface water and groundwater measuring their concentration (Pearson, 1999). However, this monitoring approach is expensive both in financial

---

[20] Drylands are defined as the arid, semi-arid and dry sub-humid zones, or areas with lengths of growing periods of 1-179 days and hyper arid zones are excluded.

[21] Chapter 3 of Water Report n. 23 of the FAO available online: http://www.fao.org/docrep/005/y4473e/y4473e07.htm, last access the 1st August 2014.

and human resources therefore, partially available and variable in developing countries. (Ongley, 2001) already highlighted that this approach is not suitable for these countries. The monitoring of water resources is only the first obstacle to overcome, where (Ongley, 2001) suggested ways of improvement. The second step is to design and implement appropriate and efficient remediation processes where knowledge and capacity of the human resources are often important barriers (Ongley, 2001). Therefore, despite its importance, measuring water quality remains a challenge in many countries, for the following reasons: 1) the degradation of water quality can be due to complex factors and various sources (agriculture, industry, wastes, wastewater discharges, natural residues, etc.) to be identified; 2) because of the human and financial resources required for prevention and mitigation of the pollution. Therefore, the water quality inclusion in this work is expected to be difficult at developing country scale. The Joint Monitoring Programme conscious of its importance is looking at setting up a standard measurement protocol at national scale to obtain worldwide assessment (JMP, 2011).

Worldwide, the water resources are mainly consumed by the agricultural sector absorbing on average 72% of the withdrawn water. This percentage increases to an average of 90% when focusing on the developing world (Rosegrant & Ringler, 1998). A growing trend also applies to the water resources consumed by domestic and industrial demand (Meinzen-Dick & Appasamy, 2002). Generally, domestic and industrial demand is rather associated with urban areas while agricultural demand is located in rural areas. This distribution is globally true with some exceptions in case of agro-industrial activities in rural areas or urban agriculture. The challenge of water resources management is to provide sufficient quantity where it is needed to all users with the adequate water quality. Indeed, the domestic water supply requires high quality water while agriculture and industries accept lower quality level. For the last two sectors, the recycling and reuse of wastewater are possible (Meinzen-Dick & Appasamy, 2002).

The water demand of the environment to preserve a healthy ecosystem should be added to the human demand, as this supports human life in return. The concept of and the increasing research on ecosystem services explicate and value the role of the water ecosystem in the human well-being. This approach avoids opposing Human and Nature in the competition for water resources and demonstrates the interest to dedicate a share of the water resources to the environment. In fact, the ecosystem performs "valuable roles such as water purification and filtration, flood control, flora and fauna habitat provision and groundwater recharge" (Sullivan et al., 2003). The environmental flow[22] requirements vary from ecosystem to ecosystem and across seasons (Sullivan et al., 2003) Therefore, this estimation is made at the scale of a specific ecosystem or a river basin, through "a dialogue amongst scientists, policy-makers, water managers and users, and local populations, about sustainable water usage that balances priorities amongst competing demands" (Pahl-Wostl, 2012). As a consequence, no indicator on ecosystem water demand at national scale is available up to date.

2.2.4. Quantifying the external financial resources

The WSS sector can be financed, like other sector of the economy, by different sources: the country's own resources (estimated by the GDP for instance) or external flows. (Adler et al., 2010a) listed three main types of inputs: the foreign direct investment (FDI), the workers remittance (REM) and the Official Development Assistance (ODA). In term of importance, "remittance flows rank behind only foreign direct investment (FDI) as a source of external funding for developing countries. In 2004, workers' remittance receipts in developing countries exceeded "US$126 billion, much higher than total official development assistance and private non-FDI flows, and more than half of total FDI flows to developing countries" (WB, 2005). The estimation of remittances is made by the World Bank by recipient countries. The literature to demonstrate the benefit of remittances on the WSS sector is scarce (the case of Mexico in Abiba & Girod, 2010). However, the recipient households are expected to improve their living conditions which include WSS facilities (WB, 2005). The WPI includes the remittance as complementary household income; therefore, the variable is not specifically tested with WSS indicators.

The FDI represents the cross-border expenditure on production assets made by the private sector (Froot, 1994). This form of investment in the water sector has been supported by the privatization movement of the WSS services in the 1990's (Budds & McGranahan, 2003) and the reinforcement of multinationals specialised in environmental services. However, the share of

[22] Concept formalized in the Brisbane Declaration 2007. http://www.eflownet.org/download_documents/brisbane-declaration-english.pdf, last access the 1st August 2014.

FDI targeting WSS basic services may be limited compared to the FDI dedicated, to agricultural land buying that has significant impact on local water resources management, as for instance in Ethiopia (Bues, 2011), (Bossio et al., 2012). The World Bank provides global amount of the FDI at national scale without sectorial breakdown. The OECD provides the amount of FDI for Electricity, Gas and Water Sector but only relatively to the 34 OECD members. The data availability limits strongly the FDI inclusion at worldwide scale.

The official development aid is tracked through a common data system, the OECD Creditor Reporting System[23]. The system provides the ODA breakdown by sector and by type of donors. The scope of the ODA is wider than WSS services[24]. The ODA supports programs not strictly related to the WSS infrastructure but includes actions related to water resources management, participation, awareness and capacities of the water users' communities, to the organisation of the water sector at various levels. The literature on the reasons and the analysis of the ODA commitments is abundant at global level (Carbonnier, 2010), at donor/bilateral level (Tjonneland, 1998, Tuman & Ayoub, 2004, Larru & Tezanos-Vasquez, 2012) or focuses on specific associations of ODA (Knack, 1999), (Wolf, 2007), (Botting, et al., 2010). (Wolf, 2007) has shown that ODA volatility is positive and significant on WSS accesses in Africa while there is very positive association of the control of corruption index with the aid delivery to WSS. This suggests that limited corruption increases aid effectiveness. (Botting et al., 2010) found that countries receiving the most ODA-WSS are more likely than the countries in the lowest tertile of assistance to achieve greater gains in population access to improved water supply. This positive association was found not significant with improved access to sanitation. (Knack, 1999) highlighted some drawbacks of high dependency of a country to external assistance, where the administrative capacity of recipient governments can be degraded because of corruption practices are and reduced bureaucracy quality. The influence of assistance flow on governance quality is not simple where the amount, the country capacities and donors organisation can be favourable or not (Knack, 1999).

---

[23] Online OECD database: http://stats.oecd.org/qwids/, last access on the 12th August 2014.
[24]Refer to OECD DAC thematic classification used by all donors available online: http://www.oecd.org/dac/stats/purposecodessectorclassification.htm, last access on the 13th July 2014.

The OECD[25] makes regular syntheses at worldwide scale, by country and sector of the ODA commitments and presents the observed trends. However, these works focus mainly on the analysis of the source of ODA: the amount committed, its geographical dissemination, temporal evolution, efficiency, the targets and reasons of allocation etc. This can be also important to analyse the “needs” of ODA including all the potential countries in the attempt to rationalise the ODA commitments. It is clear that the geopolitical/economic strategies and spheres of influence will still continue to play a role in these commitments. The margins of improvement can be at better or larger identification on the same analytical base of the weaknesses of the recipient countries. The ODA prioritisation could be detailed beyond the wealth criteria (GDP per capita) available. In the framework of this research, the challenge resides in defining the features related to WSS sector, which could be measured and analysed for most recipient countries.

The vulnerability usually measures the potential damage in term of social-economic and environmental impacts of a hazard or disaster (Brikmann, 2006, Chapter 2). Following this general concept, this research could provide probabilities of a country to reach 100% access to WSS, and the estimated effect on the country development according to the most complete set of factors possible. This study will additionally include in analysis both the external financial support and governance, that are assumed to support the country capacities to extend WSS services.

2.3 Methodological frameworks to analyse WSS

This section summarizes the different methods used in the WSS related literature that can be classified in four categories: composite index framework, correlations and linear regression, Data Envelopment Analysis, Bayesian Networks. Table 2.3 recapitulates the main advantages and disadvantages of each method with an example. A common limitation described hereafter links the data availability with data quality.

---

[25] Refer to OCED cooperation development page where the reports related to aid effectiveness and statistics are quoted: http://www.oecd.org/dac/, last access on the 4th August 2014.

2.3.1. Composite indicator framework

As already mentioned, (Sullivan, 2002 and 2003) developed and discussed a new framework that synthetises several aspects affecting the provision of water into one composite index: the Water Poverty Index. It can be adapted to the data availability. It considers "1) the physical availability of both surface and groundwater, taking into account variability and quality as well as the total amount of water, 2) Access to water for human use, including distance to a safe source, time needed for collection per household. Access also includes water for irrigating crops or industrial uses. 3) Capacity Effectiveness of people's ability to manage water. Capacity is interpreted in the sense of income to allow purchase of improved water, and education and health, which interact with income and indicate a capacity to lobby for and manage a water supply. 4) Use Different uses of water, including domestic, agricultural and industrial.5) Environment Evaluation of the environmental integrity related to water and of ecosystem goods and services from aquatic habitats in the area" (Sullivan et al., 2003). This index has mainly been computed at local/river basin scale (Zhang et al., 2012), at scale of one country − (Heidecke, 2006) for Benin, (Pandey et al., 2012) for Nepal, (Jemmali & Matoussi, 2013) for Tunisia and in 2014 at scale of MENA region (Jemmali & Sullivan, 2014). The index was computed for 147 countries only in 2002 (Lawrence, 2002). Computing one composite index offers a synthetic and straightforward understanding but hides the underlying processes and the variable correlations.

2.3.2. Regressions and Data Envelopment Analysis methods for WSS analysis

A widely used approach to find determinants and/or impacts of a variable goes through the correlation computation (McGarvey et al., 2008), (Botting et al, 2010), (El-Fadel et al., 2013) and linear regression (Rodgers, 1979), (Nyong& Kanaroglou, 1999), (Wolf, 2007), (Kiendrebeogo, 2012), (Dreiblebis et al., 2012), (Tigabu et al., 2013). Linear regressions are robust methodologies that map and provide the strength of the associations between dependants and explicative variables (Wilcox, 2009). These analyses focus on the linear component of the relationships among variables. (Botting and al., 2010) used the Spearman rank correlations together with the performance of unconditional logistic regression to assess variables' association. The interpretation of the correlation can be done in both directions which can be seen as both an advantage and a limitation.

The Data envelopment analysis is a comparative approach that classifies observations according to a set of objectives corresponding to the most efficient set of observations (Ray, 2004), (Adler et al, 2010a). This approach is meaningful to classify the countries according to their efficiency to reach the MDG's targets. (Adler et al., 2010a), combined it with a PCA to mitigate the overestimation efficiency due to unsufficient number of observations compared to the number of variables (Adler & Yazhemsky, 2010b).

2.3.3. Literature review on Bayesian Networks (BNs)

Bayesian Networks (BNs) have been used extensively in the water sector for everything from ecological applications to their integration in decision support systems at local/river basin levels. Bayesian Networks are used to model the strains from which the water ecosystems suffer, such as climate change (Varis & Kuikka, 1997) and farming/irrigation practices (Borsuk et al., 2004); (Martín de Santa Olalla et al., 2007).(Baran & Jantunen, 2004), (Bromley et al., 2005), (Castelletti & Soncini-Sessa, 2007b), (Zorrilla et al., 2009) and (Carmona et al., 2011) highlighted the use of BNs as tools to facilitate stakeholders' participation and planning. BNs modelling has been used both alone and combined with other statistical techniques to support the management and decision-making processes at river basin level. (Giné Garriga et al., 2009) used BNs at river basin level to model several dimensions influencing the access to water (the physical availability of water resources, extended access to WS, people's ability and capacity for sustaining access, various uses of water resources and environmental factors that impact water resources) following the framework of the Water Poverty Index (Sullivan et al, 2003). The studies of (Castelletti & Soncini-Sessa, 2007a), (Molina et al., 2010), and (Susnik et al., 2012) combined a BN to model socio-economic components, included in a decision-making framework at local level, with a hydrological model. These multiple approaches go beyond the purpose of this research but demonstrate the flexibility of the BNs methodology. This research follows the applied multi-dimensional modelling approach of (Giné Garriga et al., 2009). It aims at scaling it up to national level to support regional or worldwide cross-country analysis, as JMP monitors WSS at this scale.

Dynamic Bayesian Networks (DBNs) are an extension of BNs and are used to model time series. Their application within water resource management has been minor compared to static BNs and specific literature is scarce. (Dawsey et al., 2007) use DBNs in the operational management of a water distribution system targeting the real time monitoring of the dissemination of contaminants within this network. (Shihab & Chalabi, 2007) and (Molina et al., 2013) addressed with DBNs the management of water resources at aquifer level. In particular, focusing on the quality of resources, (Shihab & Chalabi, 2007) explored the potential of DBNs in modelling water contamination (using salt water and various contaminants) in an aquifer using a network of monitoring wells and time measures. (Molina et al., 2013) have built a decision support system based on DBNs, combined with BNs and climate scenarios A1B and A2 from regional climatic models, to analyse and estimate the long terms impacts of climate change on aquifers. In this study, DBNs modelled the Serral-Salinas aquifer in Spain under double stress: land cover (irrigation) change and climate change impacts on rainfalls and ultimately on aquifer dynamics. The system estimates the evolution of the aquifer storage for each subsequent step (each step amounts to 5 years) under different climate change conditions using controlled water management interventions.

This work uses DBNs to create an "initial-building network" of variables and generate the probabilistic distribution of each node of this network for the observed period. This modelling integrates the knowledge of past periods into the simulations. The principles behind and descriptions of the BNs and DBNs are detailed in Chapter 3.

Indicators	Main methods	Description	Main advantages	Main disadvantages	References
Indicators available under: Resources (R), Access (A), Capacity (C), Use (U), Environment (E) components	WPI Computation is the weighted average of the 5 components	Each sub indicator (R, A, U, E, C) and the WPI are computed	• Simple and synthetic measure. • Multivariate assessment of items • Adaptable to several scales	• Not including external financing • hiding relationships between variables	(Sullivan et al., 2003)
Infant mortality, child mortality rate, ODA dedicated to WSS from 2000 to 2006, % change of improved water supply and sanitation.	Spearman rank correlation coefficients, unadjusted and adjusted odds ratios with 95% confidence intervals	Spearman Rank correlation coefficients are computed to identify statistically significant relationships between variables. To assess the associations between variables of interest, unadjusted and adjusted odds ratios and 95% confidence intervals were estimated by unconditional logistic regression.	• robust computation • comparison and qualification of the gap between highest and lowest tertiles of the distribution.	Two-direction association	(Botting et al, 2010)
Variables defining to WSS, Health and education services (2 variables each), variables qualifying the institutions involved (federalism, freedom of press, WGI VA, WGI CofC) and socio-economic variables.	OLS regression for each public service (WSS, health and Education)	Water supply and Sanitation indexes are set as dependent variables where explicative variables are the aid share and its volatility, institutional variables, population density and Africa dummy variables. AFRICA	• robust computation and model • qualify the strength of the association between variable as in (Botting et al.,2010)	• Linear method • Two-direction association	(Wolf, 2007)
10 indexes related to the first 4 MDG, and 4 variables related to financial resources (GDP, FDI, REM, AID) for three periods 1990-1994/1995-99/2000-2004,	Data Envelopment Analysis combined or not with PCA:	Non parametric method that compares each country with the most appropriate performer based on multiple high correlated objectives (here MDGs). Basic DEA, restricted DEA and combined PCA-DEA are performed and compared	• non parametric method • classify decision making units (based on inputs/outputs) according to user defined efficient set of observations	• overestimation according of the ratio variables/observations PCA-DEA reduces this bias	(Adler et al 2010a)
5 composite indicators following WPI index for a total of 81 single variables collected and computed for a river basin in Kenya	Object Oriented Bayesian networks	Structured for the 5 dimensions a conventional BNs where instance node linked the different networks	• Combined/ordered numerous and various dataset types • Uncertainty is dealt with transparency • Straightforward simulation according to parameters set		(Gine Garriga et al, 2009)

Table 2.3: Examples of the main methodologies to analyse WSS access: description, main advantages, and limitations

3. DATA PROVIDERS

International data providers, having the international mandate of collecting and pre-processing this data, are selected according to two main criteria:

- The institution is recognised by the international and scientific community in the domain and publishes data that serves for multiple scientific and policy studies/analyses being used by the stakeholders. The extensive use of these data increases the detection and correction of errors, and the discussion on the indicators themselves (Thomas, 2009) (Dondeynaz et al., 2012), (Saha, 2012).

- The collection and computation methods are described, and validation and harmonisation processes carried out (Esty et al., 2005), (UNICEF & WHO, 2012)

The Food and Agriculture Organisation (FAO), the Joint Monitoring Programme (JMP), the World Bank (WB) and the Organisation for Economic Co-operation and Development (OECD) are the main providers that establish standard definitions and common data collection and/or processing methods and collect data from the national states accordingly. These international institutions compile the majority of the data available worldwide. Non-Governmental Organisations (NGOs) and universities compute relevant indicators at the same scale. Regional and national institutions provide significant amounts of data that should be reviewed in case of local/regional studies.

3.1 International data institutions

The international institutions collect data either using surveys completed by states and institutions or directly using geospatial tools. The next paragraphs will detail the general methodology, data quality control and harmonisation procedures of each provider.

First, the Bank for International Settlements (BIS), the European Central Bank (ECB), the European Union Statistical Office (EUROSTAT), the International Monetary Fund (IMF), OECD, the United Nations (UN) and the World Bank (WB) have joined together to focus on exchange practices in the field of statistical information with the setup of SDMX (Statistical Data and Metadata Exchange[26]). Its goal is to explore e-standards that promote improved efficiency and avoid duplicating efforts. SDMX consists of common technical and statistical standards and guidelines, together with IT architecture and IT tools, all of which are designed to be used for the efficient exchange and sharing of statistical data and metadata.

FAO statistics provides national statistical offices with internationally recognized definitions, concepts and classifications to promote the consistency and comparability of information at global level. AQUASTAT is FAO's global water information system and was developed by the Land and Water Division. AQUASTAT[27] gathers information on water resources, water uses and agricultural water management at country level.

To harmonize data, the FAO performs cross-comparisons with similar countries and with historic data for the country in question and mathematically checks for consistency and correctness in addition to validation through human interaction. The harmonization processes are specifically adapted to the nature of the indicators: from surveys (i.e. a country survey on water use in agriculture and rural development) or from geospatial data (Joseph, 2005). The data series are released with their metadata; concepts and definitions are gathered in the glossary of AQUASTAT[28].

The JMP is the institution in charge of monitoring the MDG targets related to WSS appointed by the United Nations. For each country, survey and census data are plotted on a timescale from 1980 to the present according to improved facilities definition (WHO & UNICEF, 2006). The total estimates are population-weighted averages of the urban and rural numbers. For imputing potential missing data, a linear trend line, based on the least-squares method, is drawn through these data points to provide estimates for all the years between 1990 and 2011 (wherever possible)[29].

---

[26] Statistical Data and Metadata Exchange website where standards are detailed: http://sdmx.org/, last access 14th August 2014.
[27] AQUASTAT homepage: http://www.fao.org/nr/water/aquastat/main/index.stm, last access 14th August 2014
[28] AQUASTAT glossary http://www.fao.org/nr/water/aquastat/data/glossary/search.html?lang=en, last access 14th August 2014
[29] JMP definitions and methods: http://www.wssinfo.org/definitions-methods/method/, last access 14th August 2014

Since the 1990s[30], the following aspects have improved the quantity and quality of the data available (UNICEF & WHO, 2012):

- A shift from provider-based to user-based data: initially, the JMP relied almost exclusively on government data, which were largely drawn from water-utility companies and line ministries (based on the number of facilities constructed). These figures did not include facilities that had fallen into disrepair or were constructed outside of government-supported programmes. A key improvement to the accuracy of these figures was a shift to user-based data—collected through household surveys and population censuses—which more accurately reflects the actual use of water and sanitation facilities by individual households.
- More standardization of data: in response to the lack of comparability of data on drinking water sources and sanitation facilities, WHO and UNICEF helped incorporate harmonized questions into major household surveys, and in 2006 they published 'Core Questions on Drinking Water and Sanitation for Household Surveys' to encourage their more widespread use. This increased standardization has greatly enhanced the comparability of data.
- Increased availability of data: the 90's saw an unprecedented increase in the availability of household survey data, largely due to the implementation of the UNICEF-supported Multiple Indicator Cluster Survey (MICS) and the Demographic and Health Survey (DHS) initiated by the United States Agency for International Development (USAID)[31].
- Expanded sources in database: in 2000, some 220 sources of data could be found in the JMP database; the current issue reflects more than 1,400 sources.

Lastly, JMP forecasts using linear regression. Wherever necessary, JMP extrapolates the linear regression line up to two years before or after the earliest or the latest data point. Outside of these time limits, the extrapolated regression line is flat for up to four years. If the extrapolated regression line reaches 100% coverage or beyond, or sits at 0%, a flat line is drawn from the year

[30]Box 1 p5 of the Progress report for drinking water and sanitation 2012, http://www.wssinfo.org/fileadmin/user_upload/resources/JMP-report-2012-en.pdf, last access 14th August 2014
[31] JMP data sources: http://www.wssinfo.org/definitions-methods/data-sources/

prior to the year where estimates would reach 100% (or 0%). Where insufficient data exists for linear regression, the slope of the regression is assumed to be zero, i.e. no progress is made.

The WB maintains several databases: among these, the World Development indicators that cover a wide range of topics, focusing on developing countries: Agriculture and Rural Development, Aid Effectiveness, Climate Change, Economic Policy and External Debt, Education, Energy and Mining, Environment, Financial Sector, Gender, Health, Infrastructure, Labour and Social Protection, Poverty, Private Sector, Public Sector, Science and Technology, Social Development, and Urban Development. The WB has made an effort on the development of standardization, with—for instance—the use of survey-based harmonized indicators in Africa to improve the monitoring of country development. The WB assesses collected data using a Data Quality Assessment Framework (DQAF) that has been developed by the IMF in collaboration with the World Bank (IMF, 2010). This methodology for assessing data quality brings together best practices and internationally accepted concepts and definitions in statistics, including those of the United Nations Fundamental Principles of Official Statistics and the General Data Dissemination System[32] (GDDS). This data quality assessment is carried out through "intensive, iterative consultation with experts from national statistical offices, international organisations, IMF staff and data users outside the Fund" (Carson, 2001). It facilitates a comprehensive view of data quality, one that recognizes interrelations, including trade-offs, among elements of quality and allows emphases to vary across data categories and uses/users. The IMF also evaluates the compliance of countries to data dissemination international standards (DQAF) through national assessments[33].

As the previous international data providers, the OECD[34] collects data (from households or state/institutional surveys) and the necessary metadata using international and common standards to limit the potential errors and inconsistencies at this step. Data verification is then carried out,

---

[32] IMF full documentation on GDDS: http://dsbb.imf.org/pages/gdds/home.aspx, Last access on the 14Th August 2014

[33] IMF Reports on the Observance of Standards and Codes (ROSCs) includes an assessment data dissemination and quality. The list of reports is available on the IMF website: http://www.imf.org/external/NP/rosc/rosc.aspx Last access on the 14Th August 2014

[34] Refer to the full documentation of OECD quality framework for statistics. http://www.oecd.org/std/qualityframeworkforoecdstatisticalactivities.htm Last access on the 14Th August 2014

with the aim of tidying up data and metadata, providing information to users about the accuracy of the data and metadata disseminated and providing the basis for future improvements to the statistical activity process. Data verification involves data item edit checks to ensure values are within valid ranges and respect accounting identities. It also includes comparisons over time, across data items, across datasets and across countries (OECD, 2011b).

This research uses variables included in these different international databases, as they are the references used by the international community.

3.2 Other data providers

3.2.1. Transparency International

NGOs often support their analyses compiling their own data. Related to the five dimensions considered in this work, Transparency international (TI) provides three main measures related to corruption aspects: Corruption Perception Index (CPI), the Bribe Payers Index (BPI) and the Global Corruption Barometer (GCB). Since 1999, the BPI evaluates "the supply side of corruption - the likelihood of firms from the world's industrialised countries to bribe abroad" (TI portal)[35]. Launched in 2010, the GCB is a survey that assesses the general public attitude towards, and experience of, corruption in dozens of countries around the world.

The CPI is the first indicator developed by TI that catches the perceived levels of corruption determined by expert assessments. The CPI's computation is made under the supervision of a group of experts, the Index Advisory Committee (IAC). This index combines individual surveys from multiple sources (three minimum per country). Selected experts fill the individual survey for a specific country. The sources by country as well as their assessment are published together with CPI in a yearly basis.

Since the 2012 methodological update, the CPI is comparable over time. Previously, the CPI was based on perceptions of corruption in each country/territory, relative to the other countries scored and ranked on this index. The 2012 CPI uses the raw scores given to any country/territory and

---

[35] Refer to TI pages to BPI: http://archive.transparency.org/policy_research/surveys_indices/bpi, last access on the 11th August 2014.

then converts these raw scores to fit the CPI scale. The scale on which the CPI is presented has also been updated, from [-3, 3] to [0-100] (TI, 2012).

Related to data uncertainty, The CPI score is reported alongside a standard error and a confidence interval which reflects the variance in the values of the source data that are comprised in the CPI score (TI, 2012). The standard error term is calculated as the standard deviation of the rescaled source data, divided by the square root of the number of sources (90% confidence interval is calculated assuming a normal distribution.).

3.2.2. Universities and Research consortia

Several worldwide issues, such as climate change and, the environmental assessment, have favoured scientific collaboration to carry out the necessary data collection and analytical work. Close to this research interest, the Yale Centre for Environmental Law and Policy of Yale University and the Centre for International Earth Science Information Network of Columbia University in collaboration with the World Economic Forum and the Joint Research Centre of European Commission, have developed the Environmental Sustainability Index (ESI) and the Environmental Performance index (EPI).

The Environmental Sustainability Index (ESI) framework has been developed in 2000 to characterize the environmental sustainability at the national scale. The approach aims at combining the characterization of the environment with the human pressure and the capacity available to support its sustainable management at national scale. The ESI is composed of 76 datasets and 21 indicators that cover the following five broad categories: Environmental Systems, Reducing Environmental Stresses, Reducing Human Vulnerability to Environmental Stresses, Societal and Institutional Capacity to Respond to Environmental Challenges, and Global Stewardship (Esty et al., 2005). In practical terms, “the higher a country’s ESI score, the better positioned it is to maintain favourable environmental conditions into the future” (Esty et al., 2005). The number of aspects has been first conditioned by the availability of data to assess the initial selected dimensions[36]. The data went initially through a review by country to reach a

[36] Appendix G of the report on ESI 2005 available online: http://www.yale.edu/esi/g_idealset.pdf, last access the 7th August 2014.

60% minimum data coverage before a posteriori verification by multiple researchers (Esty et al., 2005). Markov Chain Monte Carlo simulation was used to impute possible missing data. The ESI allows cross-country comparison and has been published from 2000 to 2005.

The same consortium has computed from 2006 to date the Environmental Performance index (EPI) since 2006. The EPI focuses on the evaluation of characterization of the environment and the vitality of the ecosystems (Esty et al., 2006). The EPI considers 16 variables that are included in 6 sub-indicators: the environmental health qualify, water resources, air quality, biodiversity and habitat, productive natural resources, and sustainable energy. These are themselves aggregated into two components, environmental health and ecosystems vitality, to finally compute the global EPI value. To make the 16 indicators comparable, each was converted to a proximity-to-target measure with a theoretical range of zero to 100. The excessive indicator values were adjusted while no indicator values above the long-term target were used. A PCA has been used to weight the sub-indicators.

Time series analyses of ESI and EPI are not possible because of non-consistency of methods. In the same way, the inability to truly compare rankings between versions of the ESI and EPI are due to changes in data sources, imputations, methodology, framework, target setting, weighting, and aggregation (Esty et al., 2005) and (Esty et al., 2006).

This review of data sources has focused on providers who work at global scale. However, more data providers may be found in particular at regional level: for example, the economic commissions or regions and the regional development banks in Latin America (CEPAL and BID[37]), in Asia (ADB[38]) and Africa (i.e. SADC, ECOWAS and COMESA[39]), national agencies

[37] For the Comisión Económica para América Latina (CEPAL) database, refer to http://www.cepal.org/ and for the Bank inter-Americana de Desarollo (BID) to: http://www.iadb.org/en/research-and-data/statistics-and-databases,3161.html, last access on the 11th August 2014.

[38] Asian Development Bank database is online, http://www.adb.org/data/statistics, last access on the 11th August 2014.

[39] Regional Economic Communities are regional organisations recognised by the African Union, for instance, ECOWAS for Western Africa, SADC for Southern Africa and COMESA for Eastern Africa with a total number of eight. http://www.un.org/africa/osaa/reports/new-reports/Background%20Note%20to%20the%20RECS%20briefings%20to%20Member%20States.pdf, last access on the 11th August 2014.

for research and cooperation such as the French Institute of Research for Development (IRD), the Spanish Centro de Estudios y Experimentación de Obras Públicas (CEDEX) .All of them can hold data and analyses at national/local scale depending on their interest and intervention areas. Studies realized at continental scale are delivered to international organisations.

4. RESEARCH QUESTIONS AND CONTRIBUTION

Building on previously established scientific and field knowledge at national level**,** this research offers an analytical framework and proposes a model for water supply and sanitation access levels involving governance, human development (education, health, urbanisation and income), water resources, the uses of those resources and financial assistance.

4.1 Main gaps

From the literature review, the following gaps have been identified:

First, there is a need for cross-country and cross-cutting analysis of the WSS variables. The cross-cutting approach (in both analysis and modelling) has often been applied at small or river basin scale (Chapter 2, section 2.1), but this research will attempt to scale it up to national country level. In fact, the WSS general strategy and orientations are often considered at national level by governments and international donors. This scale also allows making country comparisons and drawing conclusions/results at a global level (i.e. in MDGs targets are established at country scale). This work will extend the inclusive principles of the Water Poverty Index (Sullivan et al., 2003) and of the Integrated Water Resources Management approach to find and measure the interconnection of Water supply and Sanitation with the different above reviewed dimensions (Chapter 2, section 2).

Then, the Official Development Aid (ODA) has not been associated and analysed with the multiple dimensions (country development, governance, water resources availability and uses) influencing the WSS services delivery. Building this comprehensive view can support decision-making strategies. Therefore, this research will integrate for the first time Official Development Assistance together with environmental, well-being, water demand and governance indicators, extending the thematic scope of existing multi-country analyses (Adler et al., 2010a), (Botting, et al., 2010) and (Wolf, 2007).

Lastly, the ODA was often studied considering the donor's side with the analysis of the motivations and characteristics of the ODA allocation. ODA was rarely studied together with the country situation and needs. Only (Adler et al., 2010a) adopted such a view and built a framework to assess the country efficiency in progressing toward the MDGs. The analysis of the "demand" could orient ODA allocation. This research will attempt to analysis developing countries according to a common framework and identify weaknesses and similarities.

The study of the last two gaps related to ODA will provide donors, in particular the European Commission as main beneficiary of this research, an extended picture of the recipient country status to better target their financial support.

4.2 Research questions

In terms of methodology, this work is organised around the following two questions:

1. *Are the different indicators collected by the international institutions coherent enough to establish relationships between WSS access and that can influence the development of these basic services? Do the temporal data allow analysis of trends over time?*

The task mainly lies in gathering representative and consistent variables for the main five dimensions, namely country development, governance, water resources and water demand, Official Development Aid on a national scale. An important number of indicators provided by international institutions, universities and NGOs are reviewed. Special attention is paid to the consistency of the computation and collection methods to allow national comparisons. Missing data is expected to be the main constraint, particularly when looking at temporal points.

An iterative approach is taken: the highest possible number of variables that can relate to WSS access is gathered. The database is then progressively restricted by an iterative multivariate analysis process and, driven by field knowledge combined with statistical criteria, to the more relevant and significant variables with regards to the research questions. The variables with strong correlation (>0.8, Appendix D, section 1) or showing relationships not supported by literature of field knowledge are excluded from the study.

A combination of methods is set up in this work using multivariate analyses, clustering and regression methods to observe the organisation of the dataset. These results become the knowledge basis used in the development of the various necessary models.

2. *Can data be integrated into a model to better understanding the processes involved in WSS development considering its cross cutting nature?*

The modelling reinforces and deepens the results from the first analysis step. A quantitative modelling approach would require accurate and high-quality data that are not available at the scale considered. The quality of the data collection varies across countries because of the heterogeneity of the data collection. A qualitative approach is taken considering the nature and heterogeneity of available data. Therefore, exploring trends and relationships among these dimensions on a qualitative scale is suitable.

This step involves mapping the influence of the selected variables on the WSS level. This constitutes a new step forward because of modelling such a number of relevant variables at national scale. This provides a global picture of developing countries according and the relationships between the different selected dimensions. Bayesian Networks (BNs) is the modelling methodology chosen in this work because adapted to the particular characteristics of the data (detailed hereafter) and its capacity to generate probabilistic scenarios. These models can support the identification and simulation of the efforts needed to increase WSS access and how these interact with the development indicators at country level.

This analytical and modelling framework is designed to respond to the following key thematic questions:

- *Can key variables that foster WSS access be identified at national level, and can their influence be measured and modelled? An associated question is: can the effects of the WSS access improvement be evaluated in terms of country development?*

The improvement of WSS services access is one of the worldwide targets set by the international community in terms of poverty alleviation (MDGs, target 7.c). WSS is also a crucial factor in the population well-being, therefore its interaction with other development variables is to be explored.

* *Is the ODA supporting the WSS coverage? How ODA is allocated across recipient countries?*

The efficiency of ODA dedicated to WSS was found not significant using OLS regression (Wolf, 2007). (Botting et al., 2010) indicated a small positive association with sanitation access using spearman rank coefficient. The use of Bayesian Networks because modelling linear and non-linear behaviours may bring additional results on the effect and the distribution of ODA to these parametric methods and extend the non-parametric spearman correlation method.

* *May different groups of countries be defined according to a large and multi-dimensional basis to support the identification of countries' strengths and weaknesses?*

The analysis of selected countries according to the same multiple variables can provide comparative inputs on each country status. This geographic analysis of the variables is also included to complement the cross-country assessments of the MDGs targets made by international institutions such as the Joint Monitoring Program. The aim is also to support the prioritisation of ODA invested by donors (in particular the European Union Institutions) with as much synthetic and clear information as possible.

4.3 Research results

As for the research questions, the main results and contributions of this research are related to data and analytical methodology and the WSS theme being:

1. A **new extended database**: the WatSan4Dev database covering several aspects of the development, is created and validated focusing on 101 developing countries and including several temporal points (2000, 2004, 2007). The database is adapted and, extends the WPI framework in particular with ODA variables (Sullivan et al., 2003) (Chapter 4).
2. An **adapted analysis framework** is developed to qualify the variables influenced or influencing WSS access levels in developing countries (Chapter 3). It includes multivariate analyses and probabilistic modelling where the multivariate analysis outcomes are used to model the data with Bayesian Networks. This is the added value of this framework compared to others where the BNs scenario allow running on demand

simulations to estimate potential interventions (several scenarios will be described in this thesis).

This analytical and modelling framework applied first on 2004, then on 2000-2007 period, provides several thematic results:

- The multivariate analysis and BNs modelling of relationships between 25 variables highlight key elements contributing to WSS coverage and, therefore on country development (Chapters 5 and 6). The outputs confirm but also probabilistically measure the empirical knowledge and literature with regard to the importance of GDP per capita, and the strong association with health variables. They provide additional elements on governance role and highlight the positive effect of irrigation and slums reduction on WSS services access.
- A **new classification** of countries according to these variables with the description of 5 country profiles (Chapter 5, section 4). This classification allows comparing the country status towards the development with ODA support received. This highlights in a visual and synthetic way where the countries in most need of help among developing countries are, namely country profiles 4 and 5. This helps prioritising the actions on one/several sectors.
- The **analysis of the 2000-2007 period** of the WatSan4dev and the probabilistic estimation of efforts would have been necessary to reach 75% Sanitation and 88% Water supply access as set by the MDGs for 2015 (Chapter 7).

5. CONCLUSIONS

The literature review described the different concepts, the methodological frameworks and indicators available to characterize the main dimensions influencing the WSS services (Chapter 2, section 2). Indeed, WSS is recognised to be a transversal/cross-cutting issue that interact with country development, governance, water resources availability and their uses (IWRM approach). The literature related to external financial flows that can support WSS and Development was also reviewed.

The concept and measurement of the country development have evolved and been enriched over time. Development has been extensively studied starting from an economic view to the integration of social conditions/well-being variables (health, education, wealth distribution…etc.). This enriched concept is therefore supported by a significant amount of data and indicators.

Governance is a rising field in research since the 1990's. This concept can have multiple forms and is involved in WSS at different scales (from national to local level). Therefore, controversies and debates on what to measure (with one, two or several indexes…) and how to catch governance complex nature are numerous (Apaza, 2009).

The estimation of the water resources, their characterisation and their uses has implied an important amount of environmental research. Several methods/models were developed to estimate the water resources (Shiklomanov et al., 1997) (Meigh et al., 1999). The main water uses estimated are restricted to agriculture, industries and households consumptions but represent a good estimation of the water resources withdrawn (Meinzen-Dick & Appasamy, 2002). Only ecosystems water needs are rarely estimated in particular in developing countries. The main gap remains the estimation of water resources quality at national scale.

Finally, the three main external financial flows that can be committed to WSS can be classified by their main sources: a) from the companies and investors (FDI); b) from the individuals (remittances); and, c) from donors (ODA). The latter has induced an abundant research on the motivations/determinants of ODA, its implementation, distribution and impacts.

The review on data providers highlighted the efforts and protocols set up to ensure the robustness and consistency of datasets (Chapter 2, section 3). The review followed a multi-dimensional approach expanding the principles of the Water Poverty index (Sullivan et al., 2003). The research questions were therefore developed (Chapter 2, section 4.2) and the work organized to address the identified gaps (Chapter 2, section 4.1): 1) scaling up existing cross-cutting analytical and modelling frameworks to country scale (Sullivan et al., 2003) (Giné Garriga et al., 2009). This allows identifying key elements that foster WSS and mapping the relationships between WSS, development, ODA, governance, water resources and water uses; 2) analysing the influence of ODA on WSS and widely, on the country development; 3) identifying the needs and strengths of countries in the attempt to provide insights for donors to orient their financial support.

Next chapter will present the methodologies used and the analytical framework built to address these gaps.

CHAPTER 3: METHODOLOGY

1. INTRODUCTION

Based on the literature review, a tailored analytical framework need to be developed to match data heterogeneity and to provide insights on the research questions described in the previous chapter in section 3.2. In fact, several methodological steps are necessary: 1) collect data and build suitable dataset; 2) analyse both variable and observations to map relationships between variables and identify countries with similar characteristics; 3) model the data to build a tool to support decision making. The review of methods is presented following these main steps.

First, the normalisation and the multivariate imputation methods are selected to process and considerate the WatSan4dev dataset (Chapter 3, section 2). The combination of methods necessary for analysing both variables and observations include multivariate analyses, clustering and linear regression methods (Chapter 3, section 3). Lastly, section 4 of this chapter describes and provides the advantages and inconveniences of the BNs in both static and dynamic versions.

2. NORMALIZATION AND DATA IMPUTATION METHODS

One of the major difficulties in implementing and analysing the WatSan4Dev dataset was managing missing data (Table 4.3). This is a recurrent issue when dealing with indicators from developing countries most probably because of the lack of financial and human resources. Standard normalization processes are used in accordance with the data distribution (square root, logarithmic, Box Cox, standard deviation transformations and Ordinary Least Square (OLS) regression normalization (Wilcox, 2009).

Several methods of imputation are available depending on the dataset and the missing data pattern. Data imputation methods use one of the three main assumptions on the missing data patterns: i) missing values are completely at random (MCAR), ii) are missing at random (MAR) or iii) values are missing not at random (MNAR). (Clarke & Hardy, 2007) mention the stratified MCAR assumption where the missing data pattern is conditioned by "variables which are fully observed and the remaining variables are MCAR". The autocorrelation table of missing values (Table 7.2) shows that missing values are random (special case is WGIs and WSS indicators as collected at the same time by the same teams and following same methodologies) and therefore the dataset can be considered as MAR.

2.1 Overview, advantages and disadvantages of data imputation methods

The following section summarizes the different imputation methods according to the classification proposed by (Clarke & Hardy, 2007).

2.1.1. Simple methods

As proposed by (Greenland & Finkle, 1995), (Crawford et al, 1995), simple methods as indicator, linear regression and mean show serious bias in variances and covariances computation even when the data follows the MCAR assumption. For this reason, these methods are therefore not considered in this research.

The complete cases analysis (CC) "fits the analyst's model to the subsample of complete cases". The CC is attractive "for univariate patterns with fully observed covariate data, provided that MCAR or Stratified MCAR assumptions are valid" (Clarke & Hardy, 2007). However, when the response ratio is low, then the efficiency is greatly reduced. This is not one of the assumptions for our dataset and therefore it is not suitable to this work (multivariate missing data pattern).

The Hot deck imputation (HD) is a non-parametric method that imputes the value according to the "best donor" (corresponding to the completed cases). The limit is that if the number of variables is large, the best donor could be not the closest to the incomplete case. The latter being the case in this research, HD is not suitable.

The regression imputation (RI) is not used when the covariates are missing (Little & Rubin, 2002), as in this work. In addition, (Clarke & Hardy, 2007) mentioned the difficulty to implement this method from scratch because standards error estimations are extremely complex to compute. Since the RI method presents the advantage of its implementation simplicity, the naïve application of this method is lost because of non-trivial situations with regards to MCAR assumptions (which is not the case of the dataset in this research) and the standards errors estimation.

The HD and RI methods are possible alternatives to the CC method when the number of complete cases is small but both will underestimate the standards errors.

2.1.2. Proper methods

The Maximum likelihood (ML), Multiple imputation (MI), inverse probability weighting and Bayesian imputation are included in this methods group.

The maximum likelihood method can use the Expectation–Maximization (EM) Algorithm described as follow (Dempter et al., 1977), (Little & Rubin, 2002), (Clarke & Hardy, 2007):

Providing an analyst model p(Y,E,X| θ), a starting value $\hat{\theta}^{(0)}$, and setting p=1, the Expectation-step computes:

$$Q\,(q|\,\hat{\theta}^{\,(p-1)}) = E\,\{\log L_{\text{obs}}(\theta)|\text{observed data},\,\hat{\theta}^{\,(p-1)}\} \quad (1)$$

and the Maximisation-step maximizes $Q\,(q|\,\hat{\theta}^{\,(p-1)})$ to obtain $\hat{\theta}^{\,(p)}$.

The EM algorithm uses the "information from the complete and incomplete case, no matter the assumption on missing value pattern."

The Bayesian full probability model use Monte-Carlo Markov chain methods where missing values are treated as parameters, meaning quantities where uncertainty is recognized (Ibrahim et al., 2005). This method includes and estimates more the data uncertainty than computing missing values.

The MI methods may be described as a "multiple imputation as a generalization of simple imputation methods to allow for this uncertainty by simulating multiple missing values from a given imputation model, and also allowing for uncertainty in estimation of the imputation model itself. The total variation in these imputed values thus accounts for the uncertainty." (Rubin, 1987).

The Inverse Probability Weighting (IPW) method allows making adjustment for missing data imputation using weights. The missing data weights are commonly calculated from the predicted probability of response from a logistic model in which all variables predictors of non-response are included. This approach is simple and flexible to use, effective in large samples but only make use of the complete data cases. This is not appropriate to the WatSan4Dev dataset as most data cases are incomplete. Therefore, the likelihood and multiple imputation methods are the only two possible options. The advantages and disadvantages of each method are detailed by (Clarke & Hardy, 2007) as follow:

- The ML method depends on the analysis model. This implies that when the covariates (or exogeneous variables) are fully observed, the estimation of θ can be used through the maximum likelihood without having to make any more modelling assumptions regarding the missing values. When the covariates are incompletely observed, the likelihood method must incorporate the modelling assumption about covariate distributions.
- The MI is more general than the ML method because of the distinction made between the analysis model $p(Y|E,X;\theta)$ and the model for the complete multivariate dataset $p(Y,E,X|\varphi)$. The real advantage comes when the data does not follow a multivariate normal distribution, as in this work, as standards error estimates are easily calculated.

(Rubin, 1987) indicates that "in case of MCAR or MAR assumptions the MI method is a better option than ML method". The solution adopted for the WatSan4Dev data combines the EM algorithm and MI method.

2.2 EMB multiple imputation method

The solution adopted was to process the missing data using multiple imputation methods (Horton & Lipsitz, 2001). The multiple imputation methods used compare national observations based on several indicators and compute missing data without modifying the statistical nature of the variables. Missing data (**m**) computations are performed to obtain values issued from the data distribution (realistic value – the hypothesis is that any imputed indicator value already exists in the data distribution in a development country with a similar profile) rather than accurate quantitative values computed from a model. This ensures the coherency of the data base and significance of the analysis performed, while avoiding quantitative interpretation.

In this work, the Expectation-Maximization Bootstrap Algorithm (EMB) [40] (Honaker & King, 2010) was chosen as a multiple imputation method. This method involves imputing **m** values for each missing cell in the data matrix and creating **m** completed data sets. Across these completed data sets the observed values are the same, but the missing values are filled in with a distribution of imputations that reflect the uncertainty introduced by the missing data.

In details, this algorithm combines the classic EM algorithm with a bootstrap approach:

- For each draw, the algorithm bootstraps the data to simulate estimation uncertainty and then runs the EM algorithm to find the mode of the posterior for the bootstrapped data.
- Several completed sets are created and then combined under the analyst's supervision. The assumptions are 1) that the imputation model assumes that the complete data (that is, both observed and unobserved) are multivariate normal and, 2) that the data are randomly missing (MAR). This algorithm is fully described in (Honaker et al., 2012).
- The dataset is incrementally completed by imputing missing data for variables with less missing data before processing the more incomplete ones.

The imputation process can only be applied to data following a normal distribution. Therefore, prior to this imputation step, the normalization of the distribution is performed on the 41 variables to allow better missing values imputation.

The threshold of missing data, from which the imputation induces too high error, is not easy to define. (Brown et al., 2012) argue that this error "is likely to be a combination of factors including, but potentially not limited to, the size and shape of the data set, the distribution of the missing data, and the choice of estimation method." (Kyureghian et al., 2011) has indicated that MI methods show the best chance to estimate the true value within the 95 % interval in case of both 20% and 50% missing rates. (Lorez & Boroni, 2011) has found that EMB (Expectation-Maximization to Bootstrapped data) algorithm had a limited bias at 60 % missing data. (Honaker et al., 2012) do not advise any missing data rate which forbids the use of EMB algorithm and

[40] Imputation have been made using Amelia II software (Honaker,J.,King,G., and Blackwell,M., Amelia II: a program for missing data, version 1.2-17,2011. http://r.iq.harvard.edu/docs/amelia/amelia.pdf)

Amelia II software. A 50% missing data threshold is set to select variables, following (Kyureghian et al., 2011).

3. ANALYSIS METHODS

The data analysis methods used in this research are based on the Principal Component Analysis (PCA) and Factor Analysis (FA). Both are variable reduction methods allow the analysis of important number of variables (indicators) and observations (countries). Effectively, by analysing the correlations among variables in a dataset, the variables can often be reduced to a smaller group of Factors or Principal Components. Both methods provide a set of loadings (correlations between original variables and extracted factors/components) and a set of scores (values given to each data item/observation by the extracted factors/components after the variable reduction). The multivariate analyses performed here use the Pearson correlation coefficient.

Principal Component Analysis (PCA) is the main multivariate analysis method used in this research work to identify the preliminary structure and analyse the data organisation. The choice of a multivariate method depends mainly on the objectives of the analysis. FA is appropriate when defining and describing the latent factors is the main goal, while PCA is preferred to analyse the relationships between variables and reach a smaller number of variables (Rencher & Christensen, 2012).

The main objective of the multivariate analysis in this first step is to track potential errors in the pre-processing data. The study of the covariance matrix provides information about coherence noises associated to the items/measures/observations. The PCA is found appropriate to measure the linear associations between the variables as it requires no specific assumptions on the relationships between variables. This allows tracking high dependent variables. This approach was applied by (Kass & Li, 2012) who used PCA for "the removal of random uncorrelated noise as well as for the separation of coherence undesired signals from those used to UXO and UXO-like anomalies". In the same way, (Duffy et al., 1995) proposed an objective approach using unrestricted PCA to the reduction of large datasets to produce multichannel spectral coherence data. The Cygnus Research International (CRI) also used PCA applied to a group of times series (called Empirical Orthogonal Function) to extract coherent variations that are dominant among a group of time series for further analysis. Using this approach, PCA is used to extract a coherent data (data cross-correlation showing relationships among the different variables) among the indicators collected by the international institutions.

In a second step, PCA and FA are then used to analyse the relationships among the selected variables and to define the number of main components, maximizing the variance on the first components (Rencher & Christensen, 2012). PCA and FA are powerful tools for such analyses, since only covariance matrix of high dimension data could become difficult to interpret. In addition, the FA analyses are also performed here to compare the found underlying factors with the variable reduction proposed by the PCA. Lastly, this knowledge gained on the variable relationships will drive the Bayesian Networks modelling phase.

The following paragraphs summarise the theory principles of the PCA and FA methods as described by (Cureton & D'Agostino, 1983), (Harman, 1976), (Jolliffe, 2002), (Stevens, 2002), (Hair et al., 2009), (Rencher & Christensen, 2012).

3.1 Principal Component Analysis (PCA)

Summarizing the history of the development of PCA, (Jolliffe, 2002) highlights the importance of (Pearson, 1901) and (Hotelling, 1933) considered as the earliest descriptions of the PCA. PCA computation details can be found in these early publications.

The Principal Component Analysis (PCA) is a statistic description technique allowing an optimized graphic representation of multi-dimensional data. This representation allows for a simultaneous description of the relationships (correlation matrix) between the variables (**N**) and the similitude (coordinates of the observations in the space of the Principal Components) among the observations (**M**). One of the main advantages of this method is that it reduces the initial N-dimensional space into a low dimensional map (the optimal view for a variability criterion) and builds a set of P uncorrelated factors (called Principal Components). The PCA technique is widely used in the multivariate analyses domain to derive dominant patterns of variability (Jolliffe, 2002).

From a formal point of view, the PCA uses an orthogonal transformation to convert a set of observations (**M**) of possibly correlated variables (**N**) into a set of values of uncorrelated variables, referred to as principal components (Rencher & Christensen, 2012). The number of principal components is less than or equal to the number of original variables. However, this transformation is defined in such a way that the first principal component has as a higher variance as possible (accounting for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (uncorrelated with) the preceding components. The interpretation is facilitated by the reduction of the number of dimensions as the graph is directly readable. The independence of the principal components is ensured because the dataset is jointly normally distributed.

Several main contributions since (Hotelling, 1933) are mentioned in (Jolliffe, 2002):

1) (Anderson, 1963) discussed "the asymptotic sampling distributions of the coefficients and variances of the sample principal components".

2) (Rao, 1964) brought "new ideas concerning uses, interpretations and extensions of PCA",

3) (Gower, 1966) focused on the distance between individuals referring to the principals axes considering different multivariate analyses methods, and finally,

4) (Jeffers, 1967) "gave an impetus to the really practical side of the subject by discussing two case studies in which the uses of PCA go beyond that of a simple dimension-reducing tool." Insights are provided, for instance, on the data transformation and the selection of factors discussing the rule of thumb for the eigenvalue threshold of 1.00. The literature mentions two main rules of thumb for selecting the number of factors for interpretation. Factors with an eigenvalue above 1 are selected (Kaiser, 1960) where the threshold effect questions the meaningfulness of this rule (Fabrigar et al., 1999). The other considers the "substantial drop in magnitude" of the eigenvalues (Kaiser, 1970) where the "substantial drop" is subject to certain subjectivity (Fabrigar et al., 1999), (Jolliffe, 2002). (Fabrigar et al., 1999) call for relying on multiple criteria when deciding the appropriate number of factors.

3.2 Factor Analysis (FA)

Factor analysis (FA) is a statistical method used to describe variability among observed and correlated variables in terms of a potentially lower number of unobserved and uncorrelated variables, referred to as factors (Salkind, 2014). While PCA analyses all the observed variance in order to find optimal ways of combining variables into a small number of subsets, FA uses regression modelling techniques to test hypotheses and produce error estimates by analysing only the shared variances (Bartholomew et al., 2008). The latter method facilitates the identification of the structure underlying such variables and the estimation of scores used to measure the latent factors themselves. FA computation details can be found in (Thurstone, 1947).

Literature in the domain suggests that a minimum number of cases is required to reduce the number of variables. There are two categories of general recommendations in terms of minimum sample size for using FA and PCA. The first type of recommendations uses the absolute number of cases (**N**) as the main criteria, while the other recommendations argue that the subject-to-variable ratio (p) is important. (Arrindell & van de Ende, 1985), (Velicer & Fava, 1998) and (MacCallum et al., 1999) have reviewed many of these recommendations on the sample size.

In the literature, authors do not reach a consensus on the minimum sample size (**N**) (Rouquette & Falissard, 2011). Several rules are suggested going from N=100 (Gorsuch, 1983), (Kline, 1979, p. 40), (Gorsuch, 1974, p. 333); (Arrindell & van der Ende, 1985, p. 166); (Garson, 2008) proposing N=150 together with (Hutcheson & Sofroniou, 1999); the rule of 200 supported by (Guilford, 1954, p. 533), (MacCallum et al., 1999, p84) and (Arrindell & van der Ende, 1985; p. 166); the rule of 250 given by (Cattell, 1978) and discussed in (MacCallum et al., 1999, p84); the rule of 300 proposed by (Norusis, 2005, p.400), (Garson, 2008); to several hundreds. Significance rule (Lawley & Maxwell, 1971) suggests 51 more cases than the number of variables to support chi-square testing (Garson, 2008) to more than 500; and, even more than 1000 in (Comrey & Lee, 1992).

Regarding the “subjects-to-variables (**STV**)” ratio (p), again, the literature does not show any consensus. The suggested optimum ratio varies from 20:1 (Hair et al., 2009) and (Hogarty et al., 2005) to a ratio of 2 (Kline, 1979, p. 40) passing through the rule of 10 (Garson, 2008), (Everitt, 1975), (Nunnally, 1978, p. 276), (Arrindell & van der Ende, 1985, p. 166), (Kunce et al., 1975), (Marascuilor & Levin, 1983), (Velicer & Fava, 1998, p. 232); the rule of 5 (Bryant & Yarnold, 1995), (Garson, 2008), (Gorsuch, 1983), (MacCallum et al., 1999), (Everitt, 1975), (Gorsuch, 1974), (Arrindell & van der Ende, 1985, p. 166); and, lastly, a ratio from 3 to 6 is acceptable if the lower limit of variables-to-factors ratio varies from 3 to 6 (Cattell, 1978, p. 508), (Arrindell & van der Ende, 1985, p. 166).

As stated by (Zhao, 2009) for what regards the number of factors versus the sample size, research has demonstrated that the general rules of thumb for the minimum sample size are not valid and useful (MacCallum et al., 2001), (Preacher & MacCallum, 2002). The minimum sample size (**N**) depends on other aspects of design, such as:

- *the communality of the variables* (defined as percentage of variance in a given variable explained by all the factors jointly) that can be interpreted as the reliability of the indicator (Garson, 2008). (MacCallum et al., 1999, p96) suggested that communalities should be all greater than 0.6, or the mean level of communality to be at least 0.7.
- *the degree of overdetermination* of the factor (number of variables versus number of factors). Six or seven variables per factor and a rather small number of factors is considered as high overdetermination of the factors if many or all communalities are under 0.5 (MacCallum et al., 1999). They should be at least four measured variables for each common factor and perhaps as many as six (Fabrigar et al., 1999, p. 282). A factor with fewer than three variables is generally weak and unstable (Costello & Osborne, 2005, p. 5).
- *the loading size*. 5 or more strong loading items (0.50 or better) are desirable and indicate a solid factor (Costello & Osborne, 2005, p. 5) (Warne & Larsen, 2014, p.110). The sample-to-population pattern fit is very good in case of high (0.80) loading condition, moderate in case of middle (0.6) loading condition, and very poor for low (0.40) loading condition (Velicer & Fava, 1998, p. 243).

In conclusion, after analysing the disparity of the thresholds, rules, etc., the general rules of the thumb for the minimum sample size cannot be considered as valid and useful if these are not associated with other statistical indicators, thematic and data inherent considerations.

3.3 Comparison between PCA and FA methods

The most important differences between PCA and FA are explained in (Rencher & Christensen, 2012, section 13.8) and (Jolliffe, 2002, chapter 7):

1) PCA analyses the total variance present in the dataset. PCA is used to find optimal ways of combining variables by establishing potential relationships among the dataset variables so as to obtain an empirical summary of the data.

2) FA analyses only the common variance that means uncontaminated by unique and error variability (Cureton and D'Agostino, 1983); (Harman, 1976); (Stevens, 2002). FA is thus less sensitive to noise in the dataset, as it is based on assumed underlying/latent factors. A thematic knowledge (a set of hypotheses that form the conceptual basis of the study) of the dataset is required when applying the FA.

When performing the multivariate analyses, several tests are undertaken:

- the Kaiser-Meyer-Olkin (KMO) test. KMO is an assessment of sampling adequacy (Warne & Larsen, 2014) and,
- the Cronbach's alpha. Alpha was developed by (Cronbach, 1951) to provide a measure of the internal consistency of a test/survey/scale. The Cronbach's alpha is the most widely used objective measure of reliability because of its simplicity. It is expressed as a number between 0 and 1. Internal consistency describes the extent to which all the items in a test measure express the same concept or construct and hence it is connected to the inter-relatedness of the items-observations within the test. However, in spite of the widespread use of alpha in the literature, the meaning, proper use and interpretation of alpha is often not clearly understood (Schmitt, 1996), (Cortina, 1993). If the items in a test are correlated with each other, the value alpha is increased. However, a high coefficient alpha does not always mean a high degree of internal consistency. This is also affected by the length of the test. If the test length is too short, the value of alpha is reduced. Thus, to increase alpha, more related items testing the same concept should be added to the rest. It is also important to note that alpha is a property of the scores on a test from a specific sample of testees. Therefore, investigators should not rely on published alpha estimates

and should measure alpha each time the test is administered. In consequence, it is important to understand the associated concepts of internal consistency (it is concerned with the interrelatedness of a sample of test items) and homogeneity (it refers to unidimensionality) (Tavakol & Dennick, 2011). Having these different concepts in mind, in the concrete case of this research work, the Cronbach's alpha measure will be low by definition because of the specificity of the WatSan4Dev dataset but this will not attempt the reliability of the observations of this work. This because the variables collected by the international institutions are measuring very heterogeneous concepts such as access to water supply, access to sanitation, health/education, agriculture/economic activities, governance, finance, … for a restrictive number of countries (101 developing countries prioritized by the European Commission). The dataset coherence, the conclusions of this work that are confirmed by the international scientific literature in the domain and knowledge reported from the field, cross-validate the assumptions of this work.

Summarising the use of multivariate analyses in this work:

1) PCA was first used to verify the coherency and robustness of the WatSan4Dev dataset, tracking data errors and incoherencies among the variables and observations (Chapter 4, section 3);
2) After, PCA was used in an iterative process to select the relevant variables out of the 41 variables included WatSan4Dev dataset (Chapter 5, section 2). PCA was suitable for assessing multi-dimensional objects and the behaviour of 25 variables for 101 observations (countries); and,
3) Lastly, PCA provided a global view of the correlations among variables and countries, thus indicating areas for further analysis and to drive the Bayesian Network modelling during the second phase of this work (Chapters 6 and 7).

3.4 Clustering methods

K-mean clustering is used to identify countries with similar characteristics (country profiles) and key variables involved in a good WSS access.

Several methods are available to cluster multivariate data (Omran et al., 2007), (Madhulatha, 2012).The k-mean and the agglomerative hierarchical clustering methods were considered for the Watsan4dev subset.

The agglomerative hierarchical clustering (AHC), as defined in (Johnson, 1967), generates a cluster tree from singleton clusters to one cluster, merging most similar clusters. This is the bottom-up approach versus the less used top-down approach named divisive hierarchical clustering. Several merging methods are available (Murtagh, 1983). This agglomerative process is visualized through a dendrogram, an informative tool to understand the dataset. This is the main advantage of this method, as discovery tool of data. The main disadvantage is that depending of the distance metric considered, the clustering results could vary. No provision can be made for a relocation of objects that may have been 'incorrectly' grouped at an early stage. Therefore, the AHC should be run several times with different metrics (Murtagh, 1983).

The k-mean clustering algorithm was developed by (MacQueen, 1967). K-mean clustering is an algorithm used to classify or to group objects based on attributes/features into a k number of groups. The grouping is done by minimizing the sum of squares of distances between the data and the corresponding cluster centroid. Considering m_i the mean of the vectors in cluster i (the centroid), k the number of clusters, a value x belongs to the cluster i if $|x - m_i|$ is the minimum distance of all k-distances $|x - m_k|$ [41]. The coordinates of each country as well as its Euclidian distance to the centroid, the computed representative of the cluster (also named profile in this work), is presented in the recapitulative table per cluster (Chapter 5, section 4.2). The farther from the centroid, the bigger the divergence with the profile. The advantage of this method is that it defines tight clusters with no overlap. The main disadvantage is to predict the appropriate number of clusters and what k would be. The initial partition is made randomly; therefore the final clusters can be slightly different that implies running several times the k-mean analysis. Globular data favours a good performance of the clustering.

[41] Matteucci.M., Online tutorial on clustering Algorithms. Retrieved from http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html, last access on the 14th August 2014

The AHC method is used as an exploratory analysis of the appropriate number of clusters structuring the data and the k-mean method is used to analyse countries' observations in an attempt to define different country profiles**.**

3.5 Ordinary Least Square (OLS) Linear Regression analysis

The classical Ordinary Least Square (OLS) regression analysis is used to obtain a preliminary exploratory identification of the key explanatory variables influencing the water supply and sanitation access variables (Stone & Brooks, 1990), (Wilcox, 2009), (Slakind, 2014). The OLS method provides insights on the main key variables but excludes some other variables that may influence the WSS and the human development because not considering the whole variability of the dataset. The full variability of the dataset (including both linear and non-linear components) is handled by Bayesian Networks (Chapters 6 and 7).

The OLS regression analysis is a method for predicting the value of dependent variables Yi, based on the values of independent variables Xi. The equation can be written as follows:

$$Y_i = \beta_{0,i} + \sum_{j=1}^{j=p} \beta_{j,i} X_{j,i} + \varepsilon_i \qquad (2)$$

where Y_i is the dependent variables, $\beta_{0,i}$ is the intercepts of the model, $X_{j,i}$ corresponds to the *j*th explanatory variable of the model (j= 1 to p), and ε_i is the random error with expectation 0 and variance σ^2. The $\beta_{0,i}$ and $\beta_{j,i}$ parameters and ε_i errors are estimated from the observations (Wilcox, 2009).

The results of the OLS analysis are validated by the goodness of fit coefficients of the model (the coefficient of determination, R^2), the variability explained by the model and the analysis of the variance. The Ficher's F test is also used to estimate the risk of assuming a null hypothesis. Finally, the relative influence of the explanatory variables is considered significant if they comply with field experience, classical cases studies or the relevant scientific literature.

4. Bayesian Networks Methods

Thomas Bayes (1702-1761) is the mathematician who first used probability inductively and established a mathematical basis for probability inference. The latter is a mean of calculating the probability that an event will occur in future trials based on the number of times it has not occurred (Heckerman, 1995). Based on this rule, BNs are statistical tools that were first generated in the field of artificial intelligence as models to manage uncertainty. A BN (also known as belief network, causal network, influence diagram or probabilistic expert system) is a statistical graphical model pertaining to highly structured stochastic systems (Cowell et al., 1999) conceived to represent probabilistic conditional relationships between variables. According to (Aguilera et al., 2011), the BNs can be formally specified within two different dimensions. On one hand, a BN is a graphical representation of a problem containing a set of related variables, and on the other hand the network is defined by a set of probability distributions. The following section technically defines each of these dimensions. The section 4.2 of this chapter details the advantages and limitations of such methods. A literature review of the applications of BNs in the context of the water sector is available in the initial literature review (Chapter 2, section 2.3.3).

4.1 BNs principles

4.1.1. Graphical representation

A BN is a graphical representation of a problem. In the context of BN modelling, the graph referred to as a *directed acyclic graph* or *DAG*, is defined a as a *pair* $G=(V,E)$ *where* V *refers to a finite set of vertices, nodes or variables and* E *is a subset of ordered pairs in the Cartesian product* $V \times V$ *called links or edges* (Jensen, 2001), (Aguilera et al., 2011).

The term *directed* refers to a directionality concept that implies that the edges or links between variables are directed. For instance, if $(A,B) \in E$ but $(B,A) \notin E$, this means that there is a direct link between A and B, which is represented by $A \rightarrow B$.

The term *acyclic* denotes that loops are forbidden in the network. A directed connection between two variables in a BN is interpreted as a statistical or relevance dependency. In the case of the example above, B depends on A. Talking in causal terms, A causes B or B is the effect of A.

The *principle of conditional independence* is the concept used to spread evidence within the model. In the case of three variables or set of variables, x, y and z, x and y are (conditionally) independent given z if $p(x|z)=p(x|yz)$. In other words, the two variables x and y are independent given a third variable z if, and only if, $p(xy|z) = p(x|z) \times p(y|z)$ (Pearl, 1988). This principle is important because it makes probability updates of the three basic types of sub-structures (converging, diverging and serial) existing in a BN possible (Fig 3.1).

a) Serial structure b) Converging structure c) Diverging structure

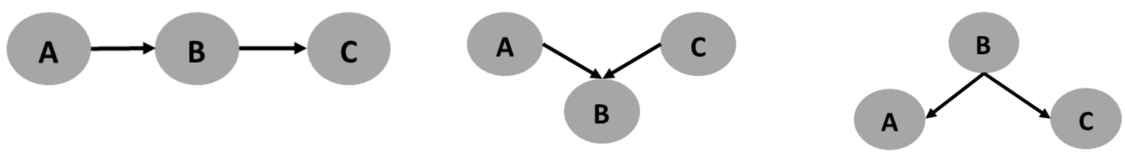


Figure 3.1: Basic structure used in a BN with A, B, C as nodes or variables (Jensen, 2001).

4.1.2. Set of probabilities: Qualitative dimension

There are three basic quantitative elements that characterize a BN: 1) a subjective conception of probability; 2) a set of conditional probability functions; and 3) the Bayesian theorem as a rule essential to updating probabilities.

1) The BNs are based on a vision of probability as a subjective issue (so-known as *Bayesian*) (Cox, 1946), (Cowell et al., 1999), (Dixon, 1964), (Heckerman, 1995). This takes into consideration the fact that probability is just a *degree of belief* as regards to the occurrence of events. Probability is understood to be a subjective evaluation based on our state of knowledge about nature or reality (Dixon, 1964). Far from being a single approach, there are at least three views on Bayesian probability (Neapolitan & Morris, 2004): i) a radical vein within which probability does not exist and is just a mental construction; ii) another school that deliberately rejects its objective essence; iii) a third group of theoreticians use its frequentist principles without making any philosophical justification. Probability is a measure to quantify uncertainty and BNs use it in subjective terms.
2) A Bayesian network is parametrically determined by a set of conditional probability functions. Generally, those parameters are specified in a Conditional Probability Table (CPT). The CPT contains a set of probability values that correspond to all the possible combinations of node states and parent node states. These probabilities or parameters are the materials with which the Bayesian theorem works.

3) The Bayesian theorem is a useful rule derived from the concept of conditional probability as applied to the intersection of events. Derived from (Bayes, 1763), it is a valuable tool when updating knowledge about an event based on evidence related to another linked event. In its simplest form, the Bayesian theorem can be formulated as follows:

$$p(B \mid A) = \frac{p(A \mid B) \times p(B)}{p(A)} \quad (3)$$

The *principle of conditional independence* allows probabilistic factorizing of the whole model meaning—or in practical terms, the assessment of probabilities given the evidences. In a BN, the factorization processes follow the below equation (2):

$$p(x) = \prod_{v \in V} p\left(x_v \mid x_{pa(v)}\right) \quad (4)$$

where x_v refers to the variables in the model and stands $x_{pa(v)}$ for x_v´s parents or ancestors (Hoeting et al., 1999), (Nadkarni & Shenoy, 2001) (Aguilera et al., 2011).

4.2 Advantages and limitations of BNs

There are several reasons one might use BNs modelling in the environmental sciences. Firstly, and most significantly, BNs are able to differentiate and efficiently manage the quantitative and qualitative dimensions of a problem. Concretely, this means that discrete and continuous variables can be modelled together (Edwards, 1998), (Heckerman, 1995). BNs also manage the presence of missing data in input data (Nadkarni & Shenoy, 2004) (Jansen et al., 2003) and “perform the proper predictions with the model built from them” (Aguilera, 2011). This is especially important in datasets related to developing countries, where a high proportion of missing data is a common issue. (Uusitalo, 2007) highlights that BNs have technically no minimum sample sizes required to perform the analysis and can show a better prediction accuracy even with rather small sample sizes by the fact that it is a probabilistic model (Kontkanen et al., 1997). In addition, BNs can combine prior knowledge with empirical data in order to develop models (Nadkarni & Shenoy, 2004) and assess interaction effects. Non-linear relationships are modelled efficiently (Lee et al., 2005). BNs carry out *local computations* in order to avoid computing the whole joint distribution in one model, which makes the inference task a relatively undemanding process (Pearl, 2001), (Xiang, 2002).

However, (Uusitalo, 2007) and (Aguilera et al., 2011) note that there are several drawbacks associated with the use of BNs in the field of environmental sciences: 1) “the building process and the parameters require more data as the number of variable increases” (Aguilera et al., 2011). Note that no advised ratio is found in the literature as this is close-linked to each data distribution (Pradhan et al., 1996), (Ordoñez Galán et al., 2009), (Chen & Pollino, 2012); 2) although BNs can handle hybrid databases (containing continuous and discrete variables), solutions are not yet available in commercial software; 3) the complexity increases rapidly when modelling times series (DBNs). This makes medium size models usually intractable due to the number of variables.

The Bayesian Networks are flexible methods that are also able to handle temporal datasets in their dynamic version.

4.3 Dynamic Bayesian Networks (DBNs)

Dynamic Bayesian Networks (DBNs) or Temporal Bayesian Networks are probabilistic graphical models oriented to model time series data structures. Although the theoretical origins of those statistical techniques could be tracked to the late 1980s (Dean & Kanazawa, 1988 and 1989), (Cooper et al., 1989), (Murphy, 2002) these directed acyclic models are relatively less common than their static counterparts.

Considering three variables Y_n, Y_i, Y_j, the intra-slice edges i.e. $Y_i^t \rightarrow Y_n^t$ and inter-slice edges i.e. $Y_i^t \rightarrow Y_i^{t+1}$ or $Y_i^t \rightarrow Y_n^{t+1}$, the DBNs proceed as follow:

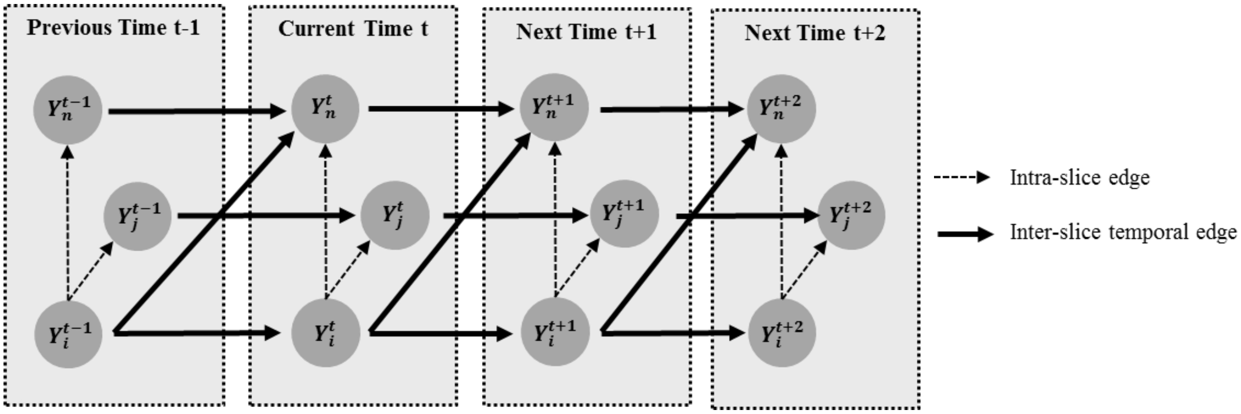


Figure 3.2: Dynamic Bayesian networks structure and organisation (adapted from Nicholoson & Krob, 2004)

DBNs can be described as a probability function for the sequence of T hidden-state variables $X = \{x_0 \dots . x_{T-1}\}$ and the sequence of T observables variables $Y = \{y_0 \dots . y_{T-1}\}$, where T is the time boundary for the event being investigated (Mihajlovic & Petkovic, 2001). It can be expressed as

$$Pr(X,Y) = \prod_{t=1}^{T-1} \Pr(x_t|x_{t-1}) \prod_{t=0}^{T-1} \Pr(y_t|x_t) \Pr(x_0) \quad (5)$$

where three sets of parameters need to be defined (Mihajlovic & Petkovic, 2001):

- Transition state of the probability distribution functions $\Pr(x_t|x_{t-1})$ that specifies time dependencies between states or times slices.
- Observation of the probability distribution functions $\Pr(y_t|x_t)$ that specifies dependencies of observation nodes as they relate to other nodes at time slice *t*.
- Initial state distribution $\Pr(x_0)$ that outlines the initial probability distribution at the beginning of the process.

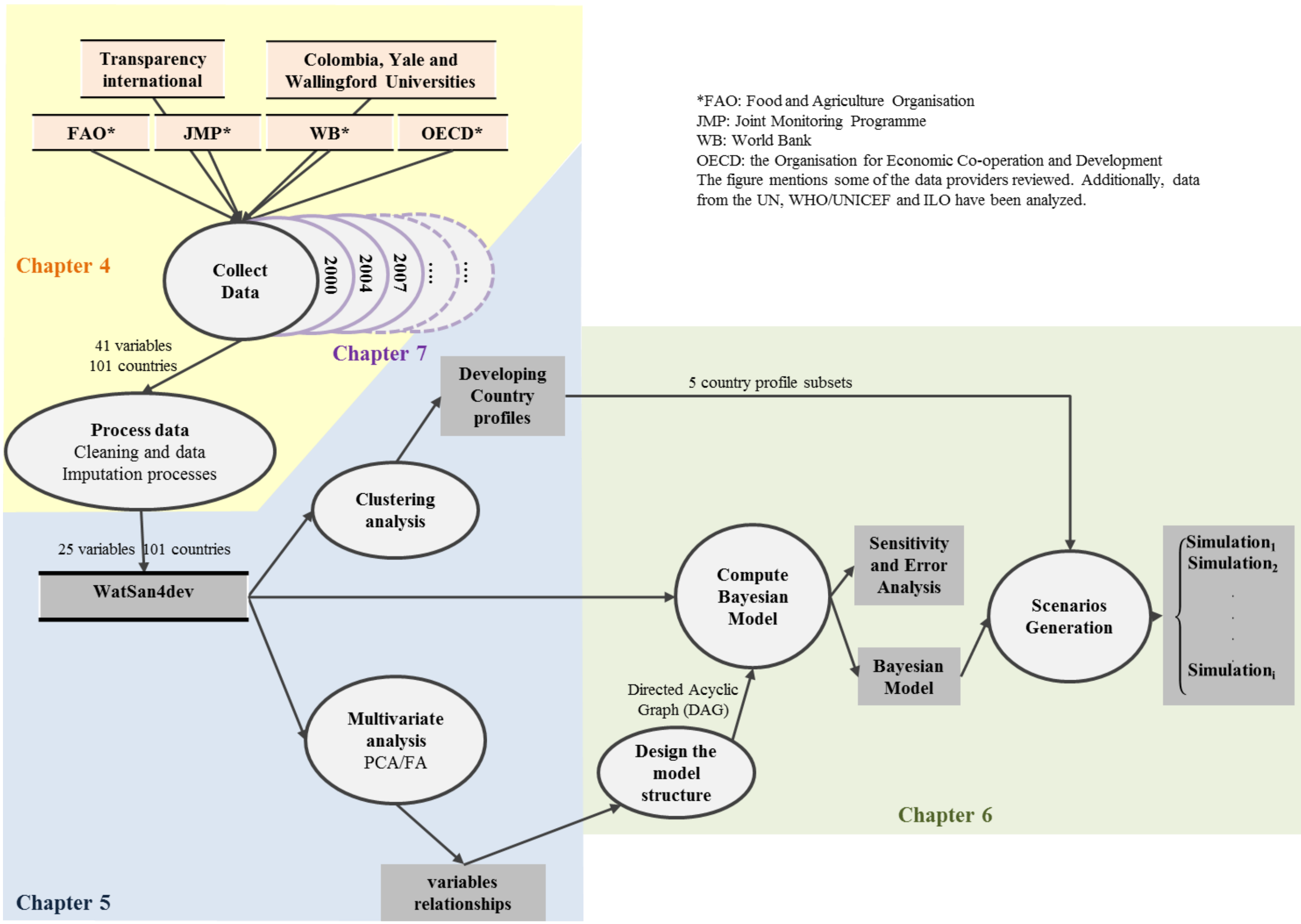
DBNs are useful when modelling time-dependent and dynamic problems, like those found in meteorological forecasts and population growth forecasts and the areas of robotics and speech recognition (Zweig, 1998). (Zweig, 1998) summarises the advantages of using DBNs in modelling temporal processes as follows: 1) the DBNs handle non-linear behaviour through the use of a “tabular representation of conditional probabilities”; 2) the interpretation process is facilitated by the fact that each variable represents a specific concept; 3) the maximum factorisation of the joint distribution favours both statistical and computational efficiency; 4) the DBNs can handle a large number of variables that could then be included in the graph structure; and 5) the DBNs have "precise and well-understood probabilistic semantics". A literature review on DBNs applications for water management is available in Chapter 2, section 2.3.3. The Bayesian Networks are found appropriate to model the WatSan4Dev dataset (both linear and non-linear components).

5. CONCLUSIONS

Based on literature and methods review, this study is organised around the following methodological steps (Fig 3.3):

- WSS Data collection description (Chapter 4): The data used in this work will be described.
- Preliminary analysis of the data from a first selection of the 41 variables (Chapter 4). PCA and FA will be used for a preliminary analysis of the dataset to track data errors and incoherent variables.
- Iterative preliminary multivariate analyses to identify variables that lead to the consolidation of a 25-strong variable subset (Chapter 5). Here the PCA will be used for variable reduction.
- Analysis of the consolidated dataset (Chapter 6). In this chapter, both PCA and FA are used to analyse the relationships among the different dimensions to drive the BNs design. K-mean clustering is used to classify countries in several profiles.
- Modelling of the dataset using BNs (Chapter 7). In this chapter, BNs are used to explain the variability of the dataset, build and run the simulations.

This methodological process is first performed for the year 2004 to verify and demonstrate its validity and use. Once validated, the whole process is extended to other temporal points—2000 and 2007—in an attempt to model the last decade (Chapter 7). The research on times series is more oriented towards the methodology rather than thematic contributions, as BNs are used less in their dynamic terms (DBNs), rather than the interpretation of results.



*FAO: Food and Agriculture Organisation
JMP: Joint Monitoring Programme
WB: World Bank
OECD: the Organisation for Economic Co-operation and Development
The figure mentions some of the data providers reviewed. Additionally, data from the UN, WHO/UNICEF and ILO have been analyzed.

Figure 3.3: Recapitulative Functional Data flow Diagram where a rectangle with two black borders represents a database, a circle indicates a process, and a grey rectangle a result.

CHAPTER 4: DESCRIPTION AND COHERENCY VALIDATION OF THE WATSAN4DEV DATASET

1. *INTRODUCTION*

The development cooperation experience in the water sector has highlighted that the major issue is that WSS infrastructure alone operates inefficiently in bringing sustainable water supply and sanitation services to the population. WSS is a complex issue that influence and is influenced by other sectors in society—health and environment most heavily, but also institutional capacities and economic sectors such as agriculture and industry.

Taking such a comprehensive overview requires gathering various indicators on economic and social status, governance and environment. Starting with the standard MDG indicators (the percentages of the population with access to improved water supply and sanitation), the search for available data covered most international providers, such as the World Bank, as well as specific NGOs and internationally recognised universities (Fig 4.1).

This chapter presents the data sources and variable selection criteria included in the WaterSan4Dev database and the methods applied to establish the coherence of the data subset (section 2). Finally, it presents the results of the coherency verification process (section 3). The year 2004 is taken as the reference point because it was the latest data available on the MDG variables regarding water and sanitation when this study began.

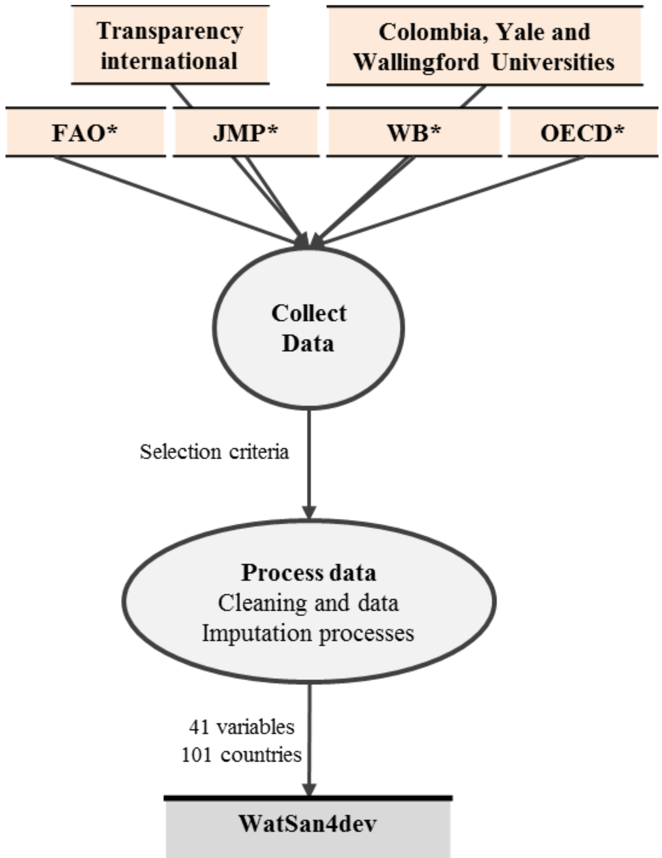


Figure 4.1: Functional Data flow Diagram related to the WatSan4Dev database creation, where a rectangle with two black borders represents a database, a circle indicates a process, and a grey rectangle a result

*FAO: Food and Agriculture Organisation
JMP: Joint Monitoring Programme
WB: World Bank
OECD: the Organisation for Economic Co-operation and Development
The figure mentions some of the data providers reviewed. Additionally, data from the UN, WHO/UNICEF and ILO have been analysed.

2. *Database Description*

2.1 WSS related Data collection and Selection

International institutions collect and provide data with detailed methodologies to support and orient their actions (Chapter 2, section 3). The indicators that monitor progresses towards the MDGs are under the responsibility of the United Nations (UN). These data are freely available and accessible via online databases and can be considered reference data.

The indicators used in this work were collected from the World Bank (WB), the Organisation for Economic Co-operation and Development (OECD), the Food and Agriculture Organization (FAO), the World Health Organization (WHO), the United Nations Department of Economic and Social Affairs (UN DESA), the United Nations Development Programme (UNDP), the United Nations Statistics Division (UNSD), the United Nations Human Settlements Programme (UN

HABITAT) and the Joint Monitoring Programme (JMP). Some indicators come from research institutions such as universities (Yale, and Columbia), NGOs (Transparency International) or institutes (Wallingford, Centre for Ecology and Hydrology) that benefit from international recognition within this domain (more details in Chapter 2, section 3).

The compatibility and consistency of this dataset, in terms of geographical and temporal scales, is a major constraint in the analysis process. The national scale was first chosen because most of the data were supplied at this level. For instance, the ODA is only collected and computed at country scale because of the difficulties (usually associated to high survey costs) to downscale the indicator at river basin level. The indicators computed from household surveys related to socio-economic indicators (i.e. Income per capita, education, demographic indicators) are often released at such national scale.

This scale is appropriate to the one of the objectives of this research: analysing the relationships between of the ODA and the other development indicators and, in particular, with the WSS coverage at country level. This is of particular interest for international donors and cooperation agencies. Without underestimating the river basin level, the international donors prefer in a first approach prioritising the support at national level versus the regional or local scales as activities and key actors at that level are considerably more complicated (high costs are associated) to identify. Once country priorities (activities) are established, the local and regional activities are identified (European Commission, 2011). The national scale allows this comparative analysis of countries and avoid considering local/distinctive features. This provides insights on common and different behaviours in term of development, water resource management and official development assistance' distribution. Such results can contribute to establishing strategy/policymaking (European Commission, 2011).

The river basin scale would be more suitable when looking for a coherent management solution of water resources for a specific river basin. The analysis outcomes are valid and adapted to this basin. At this scale, additional data collection can be directly done to fulfil the existing gaps (i.e. through households survey) depending of the size of the river basin. However, extending or scaling up the methods, outcomes and lessons learnt may be difficult. The literature review details and orders the available research according to the approaches considered− national or river basin scale (Chapter 2, section 2.1).

Data sets for 2004 are used because the latest Joint Monitoring Programme (JMP) report on WSS access levels dealt with this time period as at the beginning of this research. Moreover, the majority of the other indicators present more completed datasets for 2002 to 2004. In the Chapter 6, the dataset will be extended to two other temporal points, 2000 and 2007, available at the time of the data collection. The data collection covers countries all over the world. Around 130 indicators were analysed based on the following main criteria (Appendix A): i) **Relevance**: the variable plays a potential role in the water supply or sanitation level; ii) **Data availability**: the dataset has enough observations(less than 50% of country observations for a variable; iii) **Reliability**: the variable is produced by trustworthy providers using clearly outlined methods.

Missing data is an important constraint for an analysis of developing countries. Several variables were excluded from this global analysis because of a significant amount of missing values (greater than 50% missing data), namely Gini Index (80%). Biological Oxygen Demand (BOD) (50%), diarrhoea prevalence in children (81%) and the literacy rate for youth aged 15-24 (51%). The threshold of missing data is set to 50% of the sample to limit poor imputation (Chapter 3, section 2.1). However, variables with missing data rate around 50%, like BOD, were first tested in the multivariate analyse to analyse their behaviour. In case of no significant added value or high dependency with other variables, the variable is excluded from the WatSan4Dev dataset. The section 4 of this chapter discusses more in depth the variables selection.

The analysis of the indicator’s relevance is in line with the framework of the Water Poverty Index (WPI) in (Sullivan, 2002) but not only. As developing countries are the main focus of this research, Official Development Assistance (ODA) delivery is included in the database (Dondeynaz et al., 2012). The WPI (Sullivan, 2002) considered five dimensions −Resources, Access, Capacity, Use, Environment− to which the ODA is added in this work (Chapter 2, section 2.3.1). Scaling up and expanding this framework, the indicators are analysed and selected to characterize these dimensions namely: **water resources** (water resources and environmental variables), **human pressure on water resources** (human activity pressure variables), **governance** (variables measuring various aspects of governance, including environmental aspects), **human development** (social, education and health variables) and **official development assistance** (globally disbursed official assistance provided to developing countries by donors).

After applying the above mentioned criteria, Relevance, data availability and Reliability, to the whole set of found variables, 41 variables were selected for this study (Table 4.1) and included in the WatSan4Dev database. A description of all the variables is given in Appendix B.

Variables to be explained			Data provider
Access to Water Supply and Sanitation services		Access to Improved Water source (WS) Access to Improved Sanitation (S)	Joint Monitoring Programme Joint Monitoring Programme
Dimension	**Sub-dimension**	**Indicator**	**Data provider**
Human development	Income	Human Development Index (HDI) Gross Domestic Product – Purchasing Power Parity (GDP per cap) Poverty Rate (%Poverty) Female economic activity rate (Femal eco)	UNDP World Bank World Bank UNDP ILO
	Health	Malaria cases (Malaria) Fertility Rate (Fertility) Mortality Rate for children under 5 (Child Mortal-5) Life expectancy at birth (Life Expect Birth) Health expenditure (Health exp)	WHO/UNICEF WHO/UNICEF UNICEF UN DESA Population WHO
	Education	Ratio girls to boys in primary school (School G/B) Gross enrolment at primary, secondary and tertiary school (School enrol)	World Bank UNESCO
Environment state	Environmental general characteristics	Environmental Sustainability Index (ESI) Water Poverty Index (WPI) Percentage of land under risk of desertification (Desert risk) Surface of Water bodies (WB) Precipitations (Precipit) National Biodiversity Index (NBI)	Yale and Columbia Universities Wallingford, Centre for Ecology and Hydrology. FAO FAO FAO Convention on Biological Diversity.
	Water resources availability	Total renewable water resources (TWRR)	FAO
Human pressure	Water demand	Total Withdrawals (Total withdrawals) Withdrawals by activity sectors: agriculture, industries and municipal sector* (Withdrawal municipal, Withdrawal industrial)	FAO FAO

Dimension	Sub-dimension	Indicator	Data provider
Human pressure	Agricultural pressure	Agriculture area (%Agri area)	FAO
		Total surface of irrigation (% irrigation)	FAO
		Water use intensity in Agriculture (Water Use Int Agri)	FAO
	Demographic pressure	Urban population (Urban Pop)	UN- HABITAT
		Urban Slum population (% slums)	UN- HABITAT
		Urban population growth (Urban growth)	UN- HABITAT
		Rural population growth (Rural growth)	UN- HABITAT
Governance	Governance efficiency	Voice and accountability (WGI VA)	World Bank
		Political Stability and Absence of violence (WGI PS AV)	World Bank
		Government effectiveness (WGI GE)	World Bank
		Regulatory Quality (WGI RQ)	World Bank
		Rule of Law (WGI RofL)	World Bank
		Control of corruption (WGI CofC)	World Bank
		Corruption Perception Index (CPI)	Transparency international
	Environmental concern	Environmental governance (Env Gov)	World Economic Forum
		Participation in International Environmental agreements (Particip)	International conventions and agreements.
Financial flow	Assistance flow	Official Development Assistance (ODA)	OECD
		Disbursement for the water sector (ODA-WSS)	OECD

*The variables of sector withdrawals are mathematically dependents; therefore agricultural withdrawal is excluded for next multivariate analysis to avoid to high collinearity.

Table 4.1: Indicators-Variables implemented in the WatSan4Dev dataset

The number of countries was also reduced from 237 to 101 through the application of four filters.

- 1st Filter. The first filter was applied to focus the data on developing countries targeted by this research work (Table 4.2). Since the World Bank set the threshold for developed country at more than 10,065 current $ per capita in 2004[42], Bahrain, Kuwait, Qatar, and the United Arab Emirates are included in this research because considered emerging and developing countries by the International Monetary Fund (IMF,2004) that year.

[42] WB historical thresholds and country classification file is available online https://datahelpdesk.worldbank.org/knowledgebase/articles/378833-how-are-the-income-group-thresholds-determined, last access the 2nd September 2014.

- 2nd Filter. The second filter was applied to low populated countries (less than 150000 inhabitants) like Saint Kitts and Nevis, ….as well as non-independent state such as French Guyana, Guadeloupe, Puerto Rico, etc. (Appendix C) because of their specific status regarding international cooperation.
- 3rd Filter. An additional filter concerns the countries with more than 35% missing values that are excluded to limit analysis perturbations and, in particular, to avoid bias in the imputation process. The threshold of 35 % results from an analysis of the observations affected by a missing data rate above 30% of the variables. The observation is excluded when the missing values are concentrated on specific dimension(s) as for Afghanistan (7 variables of the human development dimension out of 11 are missing).
- 4th Filter. Finally, Eastern European countries are excluded from this research as non-benefiting from the European Aid policy. The EC aid efforts focuses first on the African – Caribbean – Pacific (ACP) countries; then, on Latin American, Southern Asia regional and national cooperation. Central Asia is excluded because not being part of the main priority area of the EU development cooperation[43]. Table 4.2 lists the countries considered in the WatSan4Dev dataset and the appendix C details the countries excluded.

WatSan4Dev countries	GDP per cap 2004	WatSan4Dev countries	GDP per cap 2004
Burundi	94	Egypt, Arab Rep.	1082
Congo, Dem. Rep.	118	Philippines	1089
Ethiopia	139	Bhutan	1094
Liberia	149	Indonesia	1143
Malawi	210	Paraguay	1201
Sierra Leone	221	Angola	1239
Rwanda	232	Honduras	1316
Niger	243	Congo, Rep.	1349
Madagascar	251	Syrian Arab Rep	1389
Eritrea	257	Morocco	1867
Nepal	272	Guatemala	1932
Gambia, The	274	Cape Verde	1981
Mozambique	281	Swaziland	2083

[43] Priority areas for the 80% EuropeAid' expenditure http://ec.europa.eu/europeaid/how/ensure-aid-effectiveness/documents/2006-list-of-46-ec-priority-countries_en.pdf, last access on 11th August 2014 and in Appendix F.

WatSan4Dev countries	GDP per cap 2004	WatSan4Dev countries	GDP per cap 2004
Uganda	289	Jordan	2157
Central African Rep	321	Iran, Islamic Rep.	2369
Tanzania	348	Dominican Rep	2414
Burkina Faso	370	Thailand	2442
Mali	382	Ecuador	2471
Togo	390	Peru	2559
Guinea-Bissau	390	El Salvador	2620
Haiti	397	Algeria	2624
Cambodia	405	Colombia	2765
Bangladesh	408	Tunisia	2832
Guinea	412	Suriname	3009
Ghana	420	Namibia	3233
Lao PDR	442	Cuba	3400
Zimbabwe	450	Brazil	3610
Kenya	464	Belize	3738
Chad	466	Jamaica	3842
Zambia	486	Argentina	3994
Mauritania	522	Uruguay	4145
Benin	547	Venezuela, RB	4304
Vietnam	558	Costa Rica	4390
Sudan	578	Panama	4456
Comoros	579	South Africa	4695
Lesotho	603	Mauritius	4840
South Asia	628	Malaysia	4875
Pakistan	629	Gabon	5340
Nigeria	644	Botswana	5425
Papua New Guinea	660	Lebanon	5453
India	668	Libya	5906
Yemen, Rep.	693	Chile	5929
Sao Tome & Principe	710	Mexico	7224
Senegal	760	Equatorial Guinea	8886
Nicaragua	834	Oman	10374
Djibouti	839	Saudi Arabia	10784
Cote d'Ivoire	873	Bahrain*	16726
Cameroon	919	Kuwait*	27148
Bolivia	977	United Arab Emirates*	28371
Sri Lanka	1054	Qatar*	44292
Guyana	1057		

*Countries considered as emerging and developing countries by the IMF (IMF, 2004)
Table 4.2: Countries included in the WatSan4Dev database (Gdp per cap in current US$)

2.2 Imputation of the WatSan4dev Dataset

Since the list of the countries being analysed in this work is 101, the first imputation process is performed using the 156 observations to increase the ratio subject-to-variable (156/41, 3.80:1) (Chapter 3, section 2.2). This improves the missing data imputation because increasing and populating the range of the whole data distribution in WatSan4Dev. However, the imputation was manually verified and specific case-by-case corrections made according to the diagnostic provided by Amelia II software. The rest of the analysis will involve the 101 countries (STV = 101/41 = 2.46:1) (Table 4.2).

Variables-2004	% missing values	List of observations with missing data
WS and S	4	Bahrain, Kuwait, Oman, Saudi Arabia
ODA, Fertility, Child Mort-5, Life Expect Birth, Agri Area, Urban Pop, WGI WA, WGI PS-AV, WGI GE, WGI RQ, WGI-CofC, WGI RofL 2004, Precipit, TWRR, Rural Growth, Urban Growth	0	
GDP per cap	6	Liberia, Bhutan, Myanmar, Libya, Cuba, Qatar
%Poverty	34	Saudi Arabia, Kuwait, United Arab Emirates, Libya, Argentina, Myanmar, Iran, Qatar, Syria, Uruguay, Cuba, Ivory Coast, South Africa, Oman, Botswana, Mauritius, Central Africa, Belize, Sudan, Gabon, Congo, Dem Rep Congo, Comoros, Suriname, Equatorial Guinea, Liberia , Lebanon, Angola, Bahrain, Djibouti, Namibia, Bhutan, Sao Tome and Principe, Cape Verde
Malaria	22	Liberia, Lesotho, Dem Rep Congo, Cameroon, Equatorial Guinea, Congo, Gabon, Gambia, Lebanon, Algeria, Jordan, Mauritius, Jamaica, Libya, Tunisia, Qatar, Bahrain, Uruguay, Kuwait, Cuba, United Arab Emirates, Chile.
HDI	1	Liberia
School G/B	12	Sierra Leone, Guinea-Bissau, Dem Rep Congo, Angola, Haiti, Pakistan, Bhutan, Sri Lanka, Guyana, Belize, Chile, Liberia
Withdrawal municipal and Withdrawal industrial	1	Sao Tome and Principe
Total withdrawals	4	Guyana, Saudi Arabia, Jordan , United Arab Emirates, Sao Tome and Principe
%irrigation	3	Comoros, Papua New Guinea, Equatorial Guinea
WaterUse IntAgri	2	Belize, Sao Tome and Principe
CPI	20	Burundi, Lesotho, Burkina Faso, Rwanda, Cambodia, Laos, Swaziland, Guinea- Bissau, Guyana,, Guinea, Central African Republic, Comoros, Bhutan, Togo, Equatorial Guinea, Mauritania, Liberia, Cape Verde, Sao Tome and Principe, Djibouti
WPI	4	Libya, Cuba, Liberia, Sao Tome and Principe
Desert Risk	7	Cape Verde, Central African Republic, Comoros, Bahrain, Mauritius, Venezuela, Sao Tome and Principe

Variables-2004	% missing values	List of observations with missing data
ESI	13	Djibouti, Qatar Belize, Equatorial Guinea, Suriname, Lesotho, Eritrea, Swaziland, Cape Verde, Comoros, Bahrain, Mauritius, Sao Tome Principe
NBI	13	Djibouti, Qatar, Belize, Equatorial Guinea, Suriname, Lesotho, Eritrea, Swaziland, Cape Verde, Comoros, Bahrain, Mauritius, Sao Tome and Principe
Env gov	47	Sudan, Kuwait, Yemen, Saudi Arabia, Iran, Burundi, Lebanon, Libya, Mauritania, Sierra Leone, Liberia, Syria, Dem Rep Congo, Togo, United Arab Emirates, Rwanda, Niger, Burkina Faso, Ivory Coast, Oman, Guinea, Guinea-Bissau, Cambodia, Cuba, Laos, Myanmar, Bhutan, Congo, Papua New Guinea, Central Africa Republic, Gabon, Guyana, Djibouti, Qatar, Belize, Equatorial Guinea, Suriname, Lesotho, Eritrea, Suriname, Lesotho, Eritrea, Swaziland, Cape Verde, Comoros, Bahrain, Mauritius, Sao tome and Principe
Particip	13	Djibouti, Qatar, Belize, Equatorial Guinea, Suriname, Lesotho, Eritrea, Swaziland, Cape Verde, Comoros, Bahrain, Mauritius, Sao Tome and Principe
School enrol	4	Liberia, Bhutan, Haiti, Ecuador
Health exp and Femal eco	1	Liberia
ODA WSS	6	Syria, Libya, Belize, United Arab Emirates, Kuwait, Qatar
%slums	22	Oman, Algeria, Liberia, Botswana, Eritrea, Uruguay, Cuba, Papua New Guinea, Sao Tome and Principe, Sri Lanka, Malaysia, Mauritania, Swaziland, Tunisia, Djibouti, Bhutan, Mauritius, cape Verde, Libya, United Arab Emirates, Kuwait and Qatar

Table 4.3: Missing data rate per variable and list of missing observations.

The WatSan4Dev dataset should only be considered for qualitative analysis by the research community and not for quantitative interpretation purposes because of the data collection and computation methods and heterogeneity of the data sources. The term qualitative approach implies that data should be used rather for analysis of trends and statistical behaviours rather than accurate/quantitative analyses. Little is known about data uncertainty, particularly regarding socio-economic indicators and water resources demand. Standard errors are available for Worldwide Governance Indicators (4-7%, in average) and Corruption Perception Index (7%, in average) (from WGIs and CPI 2004 datasets[44]).

[44] Retrieved from WGIs database http://info.worldbank.org/governance/wgi/index.aspx#home, and TI database http://www.transparency.org/research/cpi/overview, last access on the 25th August 2014.

The data uncertainty comes from various sources related to the human and financial capacities committed to data collection that vary from country to country. In a general way, remote areas, the size of the country, the skills/structuration of institutions in charge of the data collection, the amount of money committed affects the accuracy, the extension of the data collection and the quality of data. (Alkema et al., 2012) highlight the example of the fertility rate estimation in West Africa comparing the different data sources and UN estimates (Demographic and Health Survey - DHS, and non DHS), with sometimes limited observations.

However, the variables are collected from official international providers that do substantial works towards the verification and consolidation of datasets. For instance, JMP uses "data reconciliation processes" to reduce discrepancies within data national sources and the international estimates generated by the JMP (UNICEF & WHO, 2012). These processes aim to ensure the quality and significance of the data being collected in a country; more details are provided on computation methods and quality control in Chapter 2, section 3.

One of the purposes of this research work is to show that despite of the data collection heterogeneity and methods, there are cross-correlations among the different indicators and countries with similar development profiles. These results would contribute to prioritize indicators and identify where most efforts should be done by the international community devoted to development and also ranking countries by priority.

3. *COHERENCY VERIFICATION OF THE WATSAN4DEV DATASET.*

An important number of indicators (41 variables) and observations (101 countries target of this work) have been collected by the different international institutions to build the WatSan4Dev dataset (Chapter 2, section 3 and Appendix A). The verification process first goes through principal component analysis (PCA) to examine the correlation matrix, the relationships between variables and their compliance with known behaviours described in the scientific literature and/or attested to through field experience. FA is used for comparison with the PCA results in order to identify potential divergences and errors, not to confirm an underlying structure of the latent factors (Kass. & Li, 2012). The Multivariate analyses were performed with SPSS and XLSTAT[45].

Two essential steps are to be checked prior to initiate any analysis process:

1) Each indicator and its associated meaning are reciprocally consistent, coherent and keep the same significance at inter-countries level. The first two are verified in the following sections and the latter was already performed by the different international institutions (Chapter 2, section 3).

2) The different indicators (governance development, human development, economic development, water resources availability, concern of the countries regarding environment) and observations (countries) are cross-correlated. This will allow in a second phase to cluster the observations-countries and to model relationships among the different variables with Bayesian networks. This point is analysed and discussed in the following sections through the use of multivariate analysis.

3.1 Principal Component Analysis (PCA)

This section details the results of the PCA performed using 41 variables and 101 country of the WatSan4Dev dataset. The appendix D, section 1 shows the correlation matrix, the eigenvalues for components over 1. The Kaiser-Meyer-Olkin (KMO) indicator measuring the global value reaches 0.865 (average) for the 41 variables, the Cronbach's Alpha is equal to 0.73 and the mean value of commonalities is 0.772 (min value = 0.512 and max value = 0.943). These different measures indicate the validity of the dataset for representing the different concepts presented in this research work, but also, the first results obtained show a coherency (cross-correlation) between the different variables and observations. The following paragraphs present a first interpretation of the results cross-validated with the scientific literature and the field experience in the domain.

[45] The statistical software used are XLSTAT pro (http://www.xlstat.com/) and SPSS standard (http://www-03.ibm.com/software/products/en/spss-stats-standard/).

The first six components (Table 4.4) explain up to 74% of the variability of the dataset. These components are selected because of gathering a significant total variation and the drop in magnitude observed between F6 and F7 (scree plot in Appendix D, section 1). The composite indexes (WPI, ESI, HDI) are projected in the space of the principal components as supplementary variables (Fig 4.2). These variables are projected to avoid the bias of overlapping between these composite indicators (produced from independent variables used in the PCA) and the independent indicators themselves. This is also the case of the Worldwide Governance Indexes (WGIs) which were also projected as supplementary variables as the literature highlights the high-correlation among these six indexes (Thomas, 2009), (Langbein & Knack, 2010) (Chapter 2, section 2.2.2). The correlation matrix in Appendix D, section 1 clearly shows this latter point from the literature.

From the list of variables associated to each principal component (Table 4.4), the first component (F1) represents the variables linked to socio-economic conditions. The second component (F2) represents the environmental conditions and the third (F3) the human activity development (industrial and intensive agricultural pressure). The fourth factor (F4) mainly represents official globally disbursed assistance. The fifth component (F5) shows high low for the urban population growth and surface water bodies. Finally, the sixth component (F6) represents the country commitment to environment international agreements associated with the degree of civil freedom and government accountability. This will be further developed and justified hereafter.

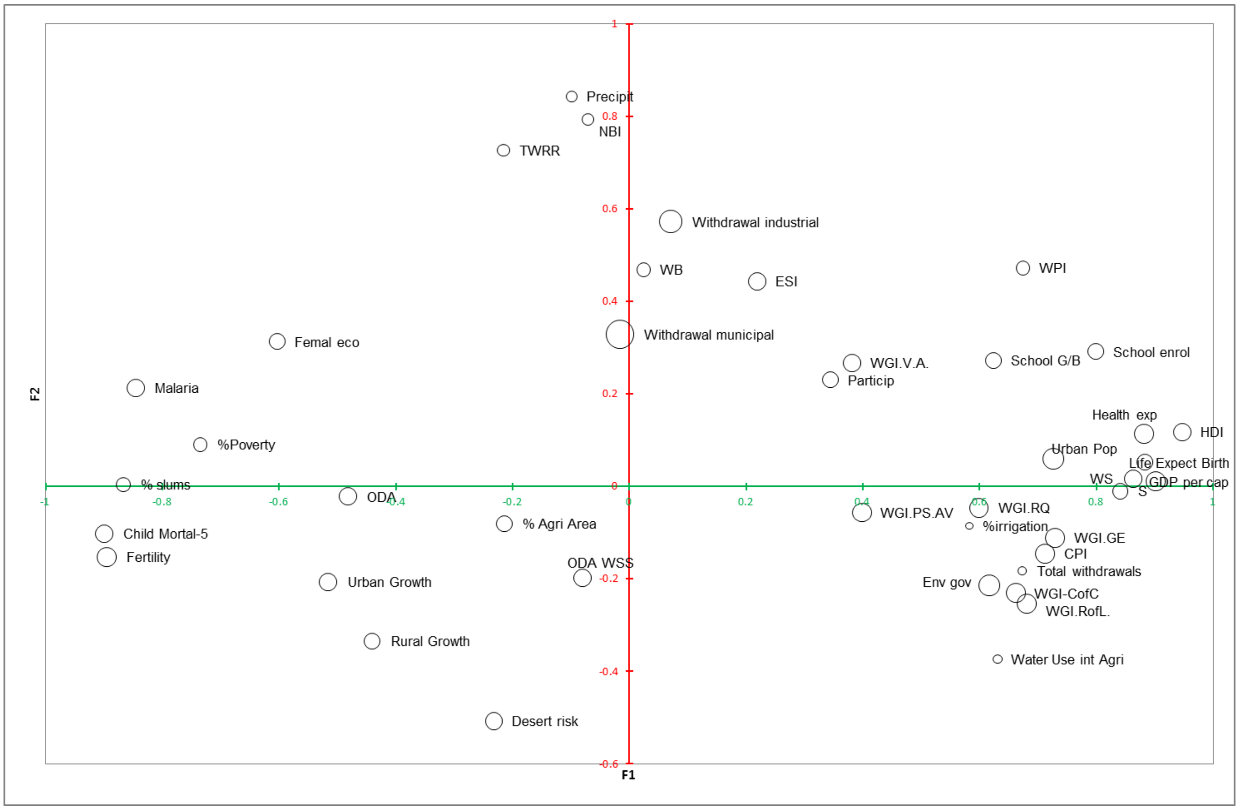


Figure 4.2: Three first components of the PCA performed on the WatSan4Dev dataset (Cumulative variability explained up to 58.84%). In green the first component F1, in red the second component F2 and the size of the point represents the factor loading in the third component F3.

Percentage of variance						
Components	F1	F2	F3	F4	F5	F6
Eigenvalue	12.419	3.647	2.762	1.852	1.451	1.436
Variability (%)	38.808	11.396	8.632	5.788	4.534	4.488
Cumulative %	38.808	50.205	58.836	64.624	69.158	73.647
PCA - Factor loadings						
Variables *	F1	F2	F3	F4	F5	F6
WS	**0.863**	0.017	0.102	-0.063	0.170	0.036
S	**0.841**	-0.010	-0.002	0.049	0.116	-0.041
% poverty	**-0.735**	0.091	-0.153	-0.063	-0.019	0.141
% agri area	-0.214	-0.081	-0.019	**-0.541**	0.276	0.325
Urban Pop	**0.727**	0.060	0.345	-0.102	-0.041	-0.108
Env gov	**0.616**	-0.214	0.391	0.136	0.116	0.255
School enrol	**0.799**	0.291	0.000	-0.015	-0.034	0.075
Femal eco	**-0.603**	0.313	-0.026	-0.062	0.241	0.212
% slums	**-0.867**	0.004	-0.144	0.059	0.054	-0.104
Fertility	**-0.896**	-0.152	0.162	0.023	0.075	0.010
Withdrawal municipal	-0.016	0.329	**0.833**	-0.152	0.118	-0.042

Variables *	F1	F2	F3	F4	F5	F6
Withdrawal industrial	0.070	**0.572**	**0.486**	0.082	-0.030	-0.157
CPI	**0.713**	-0.146	0.263	0.255	-0.015	0.256
TWRR	-0.215	**0.727**	-0.254	0.166	-0.398	0.050
Health exp	**0.882**	0.114	0.207	0.017	-0.082	0.086
Life Expect Birth	**0.883**	0.052	-0.057	0.041	0.198	-0.100
Child Mortal-5	**-0.899**	-0.102	0.083	-0.048	-0.157	0.115
School G/B	**0.624**	0.273	-0.025	0.119	0.010	0.236
Rural Growth	**-0.441**	-0.335	-0.058	0.367	0.397	-0.181
Urban Growth	**-0.516**	-0.207	0.046	0.109	**0.564**	-0.131
NBI	-0.071	**0.794**	-0.232	-0.156	0.106	0.114
Particip	0.344	0.230	-0.023	-0.343	0.134	**0.561**
GDP per cap	**0.902**	0.011	0.213	0.103	-0.084	-0.097
% irrigation	**0.582**	-0.085	**-0.635**	0.075	0.097	0.028
ODA WSS	-0.080	-0.198	0.141	**0.753**	0.023	0.380
ODA	-0.482	-0.022	0.135	**0.524**	-0.162	**0.446**
Malaria	**-0.845**	0.213	0.074	0.075	-0.156	0.149
Water Use int Agri	**0.631**	-0.373	**-0.535**	0.042	0.077	0.103
WB	0.025	**0.469**	-0.186	0.111	**0.577**	0.216
Precipit	-0.098	**0.843**	-0.334	0.195	0.036	-0.075
Desert risk	-0.232	**-0.507**	0.039	-0.408	-0.213	**0.415**
Total withdrawals	**0.673**	-0.182	**-0.553**	0.038	-0.217	0.031
Supplementary variables						
WPI	**0.674**	0.472	-0.147	0.150	-0.267	-0.059
ESI	0.219	**0.443**	0.043	0.179	-0.259	0.205
WGI VA.	0.382	0.267	0.143	0.059	0.034	**0.474**
WGI PS AV	**0.399**	-0.056	0.263	0.372	-0.019	0.314
WGI GE	**0.730**	-0.111	0.224	0.169	0.075	0.333
WGI RQ	**0.599**	-0.046	0.268	0.165	0.156	0.345
WGI RofL	**0.681**	-0.254	0.215	0.223	0.130	0.330
HDI	**0.947**	0.118	0.042	0.051	0.085	-0.098
WGI CofC	**0.662**	-0.231	0.269	0.225	0.099	0.288

*The main factor loads for the variable are in bold

Table 4.4: Factor loadings for the first six components (73.6%). The HDI, WPI and ESI are projected in the PCA space as supplementary variables because of the collinearity with the independent variables. The WGIs are also projected as supplementary variables in the space of the PCA.

The variables related to poverty (female economic activity rate – Femal eco, the percentage of slums - % slums, child mortality rate - child Mortal-5, the fertility rate - fertility, poverty rate - % poverty) are negatively correlated with development indicators (the income per capita – GDP per cap, health expenditure – Health exp, sanitation and water supply level - WS and S, Human development Index – HDI, life expectancy at birth – Life Expect Birth), as was expected. These poverty indicators are also negatively correlated with governance variables (CPI, Env gov, WGIs) and with less strong correlation with education (School enrolment – School enrol, ratio Girls to boys at school - School G/B).

The evolution of rural/urban populations (Urban Growth and Rural Growth, Urban pop) and the interaction between these variables is rather complex, implying demographic, economic and urban transition processes explain their position on the figure between poverty and development parameters. Indeed, developing countries are evolving in a multi-dimensional transition that combines a demographic transition and an accelerated urbanisation process (called Urbanisation second wave" in UNFSA, 2007). As it is a long and complex process, countries can sit at diverse stages within it; that said, they should be moving in the direction of urbanization, with a lower level of growth across their total population due to the drop in the fertility rate and increases in income (UNFSA, 2007). A deeper analysis of these transitions is available in Chapter 5, section 2. It describes the selection of the suitable data subset to answer the questions set in this research work.

From the second component (F2), it can be deduced that the variables related to water availability, precipitations (Precipit), water bodies (WB) and total renewable water resources (TWRR) are negatively correlated (-0.531,-0.308 and -0.241 correlation, respectively) with the risk of desertification (Desert risk). The Environmental Sustainability Index (ESI) and Water Poverty Index (WPI) include development or governance factors in their calculations, justifying their loads either in F2 or F1. WS access and S are lowly correlated with the water resources availability (TWRR), with -0.232, and -0.199 correlation respectively. Sanitation is negatively correlated with desert risk (-0.208 correlation). The negative direction of the correlation may be induced by the selected country sample, where low access to WSS services are observed in either rainy or dry countries as mentioned on the introduction (Chapter 1, section 2).

The third component, F3, presents more complex but coherent correlations: Water use intensity levels in agriculture (Water Use Int Agri), the proportion of irrigated area (% irrigation) and the total water withdrawals are strongly linked to each other (0.785 and 0.687 correlation with Total withdrawals respectively). Water Use Int Agri and % irrigation are negatively correlated with municipal (-0.535 and -0.5 correlation respectively) and industrial withdrawal levels (-0.443 and -0.223 correlation, respectively). This reveals the competition between the various uses of water:

industrial and municipal withdrawals often associated with urbanisation and industrialisation processes (Meinzen-Dick & Appasamy, 2002) while agriculture remains generally the first consumer in developing countries (Hinrichsen et al., 1997). Intensive agriculture is associated with the development variables being on positive side in F1.

The fourth component F4 represents significantly the Official Development Assistance. The global Official Development Assistance (ODA) is most strongly correlated with health (0.460 correlation with Child Mortal-5), access to sanitation and water facilities (-0.450 and -0.412 correlation respectively) and poverty variables (0.408 correlation with % poverty, 0.405 correlation with % slums). The loads of ODA poverty on the negative side of 1 is expected as ODA target less favoured countries a priori (Wolf, 2007). The ODA-WSS is significantly represented on the fourth one (0.791 factor load). It presents a correlation with the governance indexes (0.363 correlation with WGI PS AV, 0.2 correlation with CPI, 0.223 correlation with WGI CofC, and 0.227 correlation with WGI RofL), and with the growth of rural population (0.358 correlation with Rural Growth). Note that no significant correlation between WSS access and the specific ODA dedicated to WSS sector is observed (-0.098 and 0.004 correlation respectively).

The fifth component F5 show significant loadings for the urban population growth (0.564 load) and the surface of water bodies (0.577 load). The literature highlights some cases studies where expansion of urban areas have led to the reduction of water bodies size, in addition to the deterioration of water quality like in China (Du, 2010), or in India (Chandra Prasad and al., 2009).

The last component F6 presents significant loadings for the environmental formal commitment of countries (0.561 load - Particip), the voice and accountability governance index (0.474 load - WGI VA), ODA (0.446 load – ODA) and the desertification risk (0.415 load – Desert risk). Voice and Accountability index (WGI-VA) measures the ability of citizens to participate in selecting their government and examining its actions. The indicator also reflects the level of freedom of expression/media and freedom of association that are present, as these constitute the necessary conditions for a civil society to be in control of its government. (Barrett & Graddy, 2000) and (Neumayer, 2002) suggest that an increase in civil and political freedoms improves the national commitment to the environment, foster the reduction pollution and hence improves the quality of the environment. Desertification risk may stimulate the country to environmental commitments. International donors, through ODA, support these environmental efforts of the countries.

Finally, the proportion of agricultural area in a given country (%agri area) is loading F4 (- 0.541 load on F4) and associated to ODA-WSS and ODA (0.753 and 0.524 load, respectively). This is not coherent with the levels of correlation with those variables (-0.154 and -0.014 correlation respectively). The percentage of agricultural area also shows significant but negative correlations with WPI (-0.366 correlation), ESI (-0.248 correlation) and, finally, a positive even if low correlation with desert risk (0.243 correlation). Considering these incoherencies, this variable is to be taken with caution.

3.2 Comparison with the Factor Analysis (FA)

FA is performed to compare the factor loading configuration outlined by the PCA. The first sixth factors explain 74% of the variability and refine the PCA results (Table 4.5). At this stage, FA is used to explore the organisation of the variables comparing with PCA rather than to confirm an underlying structure of the latent factors. As for the PCA and from the scree plot (Appendix D, section 2), the first six factors are selected for interpretation.

The main output of this comparison is the consequences of the weight of the governance (D3) variable on the FA results. The third component (D3) represents the governance in its various aspects including the six Worldwide Governance indexes (WGIs), Corruption Perception Index (CPI) and the environmental governance index (Env gov). These variables show high correlations between themselves− often above 0.800 (correlation matrix in Appendix D). The existence of a governance component demonstrates the coherence of governance indicators whatever the data source (Saha, 2012). A reduction of their number could limit the governance weight within the FA and reveal additional relationships within the dataset.

The first factor (D1) represents the socio-economic development of the country based on all the variables related to health, education, poverty and urbanization following the PCA results.

The second factor (D2) negatively associates urban population growth (-0.542 load - Urban growth), environmental variables−the amount of water resources (0.840 load- TWRR) the environmental sustainability (0.736 load -ESI), the amount of precipitation (0.513 load - Precipit), and the Water Poverty Index (0.607 load - WPI). The correlations between urban growth and environmental variables are non-significant (-0.188 correlation with TWRR, -0.041 with NBI, 0.053 with WB, -0.077 with Precipit, and 0.037 with Desert risk). Urban growth also presents a significant loading on F1 (-0.470). Therefore, this association is subject to caution, and may relate to the perturbation due the governance factor.

The fourth factor (D4) expresses the pressure put on water resources according to its various uses that includes the industrial, the municipal withdrawals (-0.652 and -0.901m load respectively) and the global amount of water withdrawn (0.685 load - total withdrawal). The intensive agriculture, represented by the %irrigation variable (0.687 load) and the Water Use intensity (0.731 load - Water Use Int Agri) competes with the others water consumers. This component, representing a balance between water needs, is more complex to interpret. It can be an indicator of intensive agriculture or on the contrary.

The fifth factor D5 mainly associates ODA WSS (-0.494 load) with the proportion of agricultural area (0.621 load- %agri area) and the commitment to international environmental agreements (0.630 load- Particip). As mentioned in the previous paragraph, this first association is to be taken with caution as the %agri area variable has shown loadings not supported by literature and correlation matrix within the PCA. The financial flows into developing countries (ODA and ODA WSS) show different loads as observed in the PCA results: the global ODA remains mainly linked to health and social factors (-0.629 load on D1) but is also represented on D5 (-0.312 load). This was also observed in the PCA through the ODA significant loading on F1 (-0.482 load). The ODA WSS is also associated to governance factor (0.489 load on D3), in line with the correlations observed in the PCA.

Percentage of variance						
	D1	D2	D3	D4	D5	D6
Variability (%)	30.646	8.047	16.865	7.564	4.649	6.561
Cumulative %	30.646	38.692	55.557	63.121	67.770	74.331
Rotated Component Matrix[a]						
	D1	D2	D3	D4	D5	D6
WS	**0.798**	-0.034	0.352	0.030	0.123	0.062
S	**0.792**	0.014	0.250	0.127	-0.040	0.039
% poverty	**-0.685**	0.017	-0.299	-0.006	0.099	0.109
% agri area	-0.210	-0.253	-0.050	-0.009	**0.621**	0.035
Urban Pop	**0.746**	0.078	0.205	-0.253	0.003	-0.143
Env gov	0.423	-0.116	**0.695**	-0.089	0.018	-0.162
School enrol	**0.756**	0.269	0.223	0.013	0.078	0.164
Femal eco	**-0.609**	0.088	-0.080	-0.158	0.188	0.341
%slums	**-0.775**	-0.131	-0.370	-0.026	-0.100	0.116
Fertility	**-0.831**	-0.234	-0.209	-0.220	-0.050	-0.075
Withdrawal municipal	0.103	-0.068	0.081	**-0.901**	0.096	0.018
Withdrawal industrial	0.153	0.233	0.045	**-0.652**	-0.098	0.245
CPI	0.462	0.070	**0.784**	0.038	-0.068	-0.130
TWRR	-0.229	**0.840**	-0.172	-0.035	-0.120	**0.326**
Health exp	**0.786**	0.205	0.415	-0.068	0.040	-0.045
Life Expect Birth	**0.835**	-0.026	0.292	0.145	-0.017	0.168
Child Mortal-5	**-0.848**	-0.013	-0.291	-0.153	0.023	-0.190
School G/B	**0.530**	0.225	0.294	0.060	0.020	0.220
Rural Growth	**-0.453**	-0.438	0.037	0.123	-0.351	0.118
Urban Growth	-0.470	**-0.542**	-0.036	-0.064	-0.109	0.222
NBI	-0.030	0.453	-0.173	-0.109	0.290	**0.633**
Particip	0.189	0.213	0.336	0.050	**0.637**	0.175
GDP per cap	**0.834**	0.107	0.404	-0.049	-0.130	-0.101
%irrigation	0.504	-0.015	0.036	**0.687**	0.020	0.211
ODA WSS	-0.294	-0.042	0.489	0.094	**-0.494**	0.017
ODA	**-0.629**	0.095	0.246	-0.047	-0.312	-0.006
Malaria	**-0.826**	0.210	-0.186	-0.229	-0.036	0.054
Water Use int Agri	0.504	-0.136	0.191	**0.731**	0.036	-0.033
WB	0.016	-0.059	0.047	-0.002	0.114	**0.769**
Precipit	-0.053	0.513	-0.199	-0.038	-0.104	**0.734**
Desert risk	-0.339	-0.069	0.081	0.172	0.454	**-0.578**
Total withdrawals	0.565	0.218	0.088	**0.685**	-0.028	-0.110
WPI	**0.644**	0.607	0.102	0.075	-0.145	0.198
ESI	0.068	**0.736**	0.302	-0.084	-0.063	0.102
WGIVA	0.120	0.329	**0.705**	-0.054	0.301	0.216
WGI PS AV	0.137	0.129	**0.766**	-0.026	-0.197	-0.041
WGI GE	0.465	0.036	**0.824**	0.063	0.085	-0.041
WGI RQ	0.348	-0.003	**0.818**	-0.027	0.126	0.060
WGI RofL	0.406	-0.102	**0.852**	0.122	0.015	-0.084
HDI	**0.904**	0.084	0.316	0.047	-0.051	0.112
WGI CofC	0.401	-0.071	**0.836**	0.056	-0.016	-0.113

a. Extraction Method: Principal Component Analysis with varimax rotation, general cronbach's alpha: 0.73
Values in bold correspond for each variable to the factor for which the squared cosine is the largest
Table 4.5: factor matrix: the six components representing 74% of the variability

The sixth factor (D6) represents water resource availability and environmental quality, showing high loadings for Precipitations (Precipit, 0.734 factor load), Total Water Renewable Resources (TWRR,0.326 factor load), Water Bodies surface (WB, 0.769 factor load), the Desertification risk (desert risk,-0.578 factor load), and the National Biodiversity Index (NBI, 0.633 factor load). The organisation of this factor is the same as the component F2 of the PCA.

4. DISCUSSION

The review of the variables available has been extensive (Appendix A). However, gaps and variables choices are discussed in next paragraphs as they influence the outcomes of the analyses made. The main gap relates to the characterization of water resources at country level. The environmental variables—namely TWRR, Desert Risk, Precipit, NBI, WB, and ESI are coherent and correlated with each other, but they show little correlation with the percentage of the population that has access to WSS (Appendix D, Section 1). The ESI is a cross-cutting and very complex index that aims to evaluate the capacity of a country to manage its environment in a sustainable way (Appendix B). It is highly correlated with TWRR (0.635 correlation) and WGI VA (0.421 correlation), but is only correlated with sanitation access (0.215 correlation). Desert risk is also negatively correlated with sanitation access (-208 correlation). The WB and NBI are correlated with the amount of precipitation (0.439 and 0.732 correlation respectively) but have a non-significant correlation with WSS (0.098 correlation with WS; 0.063 correlation with S; and, NBI has a -0.033 correlation with WS and -0.058 correlation with S). Since the TWRR variable has a low correlation with WSS variables, at the level of this analysis these correlations can be considered as coherent (-0.253 correlation with WS and -0.239 correlation with S). Therefore, although this research demonstrates that there are few significant correlations (Appendix D, section 1) between the environmental variables and WSS services, it is clear that WSS has a relationship with the environment (Pearson, 1999). For example, wastewater treatment or sanitation excreta have a direct impact on surface or groundwater quality when not adequately treated and vice versa (Addo et al., 2013).

This suggests that the quantity of water resources is not sufficient to establish links with WSS, but rather needs to be combined with qualitative water indicators. The main issue is that on a national scale, indicators of the quality, accessibility and management capacities of water resources are not available in most of developing countries. Several indicators have been examined in the search of water quality indicators (Chapter 2, section 2.2.3). The Biological Oxygen Demand (BOD) was the only variable found at national scale that could characterize the quality of the surface waters. It estimates the emissions of organic water pollutants as measured by biochemical oxygen demand, which refers to the amount of oxygen that bacteria in water will consume in breaking down waste. This is a standard water-treatment test for the presence of organic pollutants (Hettige et al., 1997). High Biological Oxygen Demand (BOD) can be due to wastewater discharge in urban and industrial areas or by the agriculture.

This BOD index shows two main limitations to characterize the qualitative aspects of water resources in developing countries : i) the index computes only measures of industrial water pollution; and, ii) the missing data rate reaches 50% of the dataset (Table 4.3). The index considers, indeed, the share of manufacturing in total output, the sector composition of manufacturing and the intensity (per unit of output) of industrial pollution at the end-of-pipe (Hettige et al., 1997). At the global scale of developing countries, industrial sector often consumes a minor part of the water withdrawn, agriculture remaining the main customer. However, the BOD was included in multivariate analysis to analysis the correlations with other WatSan4Dev variables. The BOD is highly correlated with development indicators, such as with the informal urbanisation (-0.603 correlation - % slums), Gross Domestic product (0.505 correlation - GDP per cap), the access to sanitation (0.530 correlation - S) and to water supply (0.454 correlation - WS), the mortality rate of children under 5 (-0.545 correlation - Fertility) and at a lower level with other development indicators. Few significant correlations are found with the water quantity related indicators (-0.044 correlation with WB, 0.083 correlation with Precipit, -0.136 correlation with Desert risk, -0.089 with NBI, 0.413 with TWRR) as well as with the total and municipal water demands (0.117 with Total withdrawals, 0.196 with Withdrawal municipal). Therefore, BOD, being an environmental indicator, can be considered as a proxy of the degree of country development. In addition to the BOD variable, the water quality sub-indicator included in the ESI computation is excluded of the analysis of the dataset because showing around 85% missing values.

Indicators on agriculture water pollution (organic or chemical pollution) and on urban pollution discharge, if available for the developing world, could also be included as agriculture and cities are important source of pollution (Ongley, 1996). Scaling down analyses may permit the availability of such indicators at, for instance, a basin level. (Giné Garriga et al., 2009) obtained information on water quality (water point protection, pollutant agents from livestock or waste, water salinity) through qualitative questionnaires when modelling data in the Turkana district in Kenya.

The country economic general development is accompanied by the population's well-being according to the strength of wealth distribution mechanisms. The inequality of wealth distribution maintains a large proportion of poor people who have generally no access to WSS services like other basic services. This is an important aspect that enriches the characterization of the country social-economic development (Chapter 2, section 2.2.1). The GINI index has been developed to capture and measure the inequality through the analysis of the income distribution (Ceriani & Verme, 2011). The World Bank handles the computation of the GINI at national scale from governmental statistics. Unfortunately, the missing data rate of GINI dataset is around 80% leading to the index exclusion (World Bank database). Looking in details at the GINI dataset, a specific multivariate analysis on Latin America may be considered for further research (Chapter 8, section 2.1).

The remittances received from the population emigrated could have been included as an external financial input (Chapter 2, section 2.2.4). Workers' remittances are a financial flow that supports directly the households' income which was higher than the ODA in 2008, except for the Sub Saharan African region (Driffield & Jones, 2013). This inclusion of remittances data may be explored in further research as these financial flows grow as the ODA's disbursements do particularly since 2000-2003 period (Driffield & Jones, 2013, Fig 2 and 3). This analysis can allow analysing the impact of multiple individual investments on the WSS coverage at national level.

This research has chosen to focus on the institutional financial flow (ODA) in the attempt of measuring the relationship with the WSS and the country development. Aid effectiveness has become imperative since the Paris declaration in 2005 (Chapter 1, section 3.1). The current context of economic stagnation since the 2008 crisis pressures donor countries and, particularly, European institutions to ensure the efficiency and effectiveness of public expenses.

As mentioned in the literature review (Chapter 2, section 2.2.2), the development of Worldwide Governance Indicators (WGIs) has been discussed and mainly focused on: 1) the construction methods of the different indexes (Thomas, 2009, and the specific answer of Kaufman et al., 2009b) and, 2) the relevance and significance of having six different indicators (Langbein & Knack, 2010). Despite these limitations and criticisms, this research considers that governance measures are relatively new and have been subject to caution since the 90's. This study uses the WGIs but makes sure to verify their coherency using other indicators provided by other institutions—namely environmental governance (sub-indicator included in the Environmental Sustainability Index 2005 provided by the World Economic Forum) and corruption perception indexes (CPI provided by Transparency International). Their coherency was confirmed throughout the preliminary analysis performed (FA and PCA) in Chapter 4 section 3.

5. CONCLUSIONS

The WatSan4Dev dataset presented in this chapter is a new database for development (socio-economic, environmental and governance indicators), collecting the indices widely dispersed across the different international institutions in charge of their management (Chapter 4, section 2.1). The raw data, 41 indicators, have been processed, normalized with standard methods and imputed with EMB multivariate imputation method to fill in missing data (Chapter 4, section 2.2). Several criteria data were applied for selecting variables: i) the relevance to WSS issue; ii) the data availability with a missing data threshold set at 50%; iii) the reliability of the variable implying the trustful-official providers and the fact that the method is fully described. 101 countries have been selected for further analysis.

As a result of the statistical processing, the coherence and robustness of the different indicators has been increased. The multivariate analyses and their compliance with the current knowledge and field experience in the water sector proved and cross-validated the coherency of the WatSan4Dev dataset (Chapter 4, section 3). The data availability was a major constraint that has led to the non-inclusion of several aspects: water quality and inequality indexes were relevant to complete the characterisation of the water resources and country development dimensions respectively. The WatSan4Dev indicators can be used for further, deeper studies; however the WatSan4Dev dataset should only be used for qualitative estimations and analysis.

CHAPTER 5: IDENTIFYING KEY VARIABLES AND NEEDY COUNTRIES

1. INTRODUCTION

The Watsan4Dev is the base from which this research wants to identify the key relationships that influence or are influenced by WSS. A complementary objective is also to identify the neediest countries according to selected variables. In fact, this information may support strategy making within the international community. The analysis of the different variables are performed at two levels (Fig 5.1): (a) the analysis of the variables themselves, which provides insight into the relationships between the different indicators and their relevance to the water sector; and (b) the analysis of the observations (countries), which offers a classification of the countries based on their water sector and overall level of development. This geographical analysis is also used to assess country status based on the selected variables and to highlight countries need to be made a priority. Several countries' profiles are built as a tool to identify weaknesses restricting country development and population well-being.

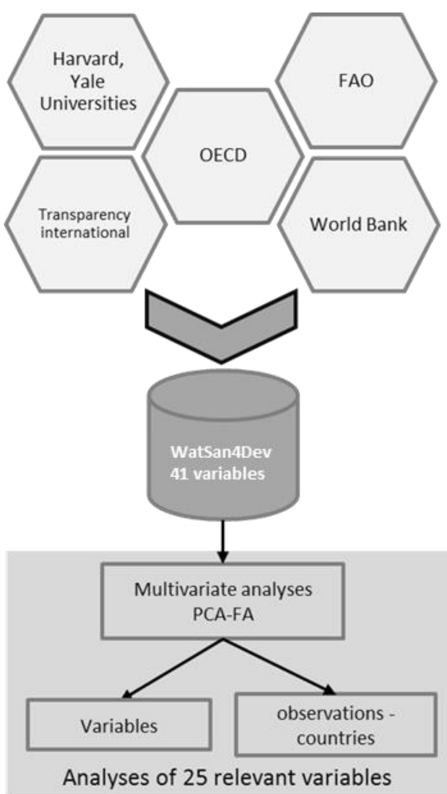


Figure 5.1: Approach followed that provides two main results: (a) the identification of key variables impacting the level of WSS services and (b) geographical analysis identifying several country profiles.

The figure mentions some of the data providers reviewed. Additionally, data from the UN, WHO/UNICEF and ILO have been analysed

This chapter presents first the dataset used, which was reduced to include only the relevant and necessary variables (section 2). Section 3 describes the main outputs of the multivariate analyses performed on the Watsan4dev subset and describes its organisation. An analysis of countries' behaviours is provided with the definition of five profiles (section 4). The key variables interacting with WSS and identified using linear regression and are detailed in the section 5 of this chapter.

2. A FIRST REDUCTION OF THE NUMBER OF VARIABLES OF THE WATSAN4DEV DATASET

In this research work, the first constrain to analyse the WatSan4dev covariance proceed from the fact that indicators come from different international data providers and themselves collect data from regional and local authorities in developing countries. The data providers verify and clean data to ensure the coherence and homogeneity of each indicator collected (Chapter 2, section 3). However, no cross-validation among indicators was performed. This research work shows the coherence and reliability of the WatSan4Dev dataset through statistical tests such as KMO (0.865), Cronbach's Alpha (0.73), Communalities (average = 0.772, min = 0.512, max = 0.93) and Subject-to-Variable (STV) ratio (101 countries-observations/41 indicators-variables, STV = 2.46:1) (Chapter 4, section 3).

At this stage, the number of variables needs to be reduced in order to represent the different dimensions considered in this work. This reduced number of variable facilitates establishing clusters of observations-countries with similar profiles and the thematic interpretation. Because of the low STV (2.46:1) combined with the high thematic heterogeneity of the indicators in WaterSan4Dev, low correlations may thus be obtained at the end of this analysis process (Appendix D, section 3); and the relationships among variables and countries may appear independent despite a linear relation between them. The first consequence is that the statistical indicators associated to the dataset and the subset of variables selected will be degraded. Despite of that, the WatSan4Dev dataset is coherent regarding the scientific literature and the experience reported from the field. Therefore, to obtain a reliable reduction of the variables, we will follow this procedure:

1. Remove variables until KMO overall is over 0.6

2. Check the communality of each variable. To drop those variables that have the smallest communality, until the communalities of all variables are above 0.60. Since the general conclusions and interpretations are coherent with the literature, key variables considered thematically important for the interpretation will be kept despite of their low communality (in any case the variable selected will be greater than 0.5). This concerns the proportion of agriculture area variable (0.567 communalities -%agri area).

3. Check the mean value of all communalities to ensure that the mean value is over 0.7. If not, repeat step 2.

4. User Kaiser Strategy (dropping all components with eigenvalues under 1.0) and Scree plot to determine the number of factors.

5. Set the loading size cut-off value at 0.35 (usually, this value should be greater than 0.6, but taking into account the heterogeneity of the dataset and the fact that this analysis is only oriented to drive the BNs modelling phase, this value is set here to 0.35) and drop the factors that has less than 3 variables.

This procedure is based on the analysis of the relevant literature described in Chapter 3, section 3. The selection of variables also follows thematic criteria to ensure and enrich the final interpretation and conclusions. These thematic criteria are described hereafter.

Governance variables, namely the Worldwide Governance Indicators (WGIs), Corruption Perception Index (CPI) and Environmental Governance (Env Gov), are highly correlated (correlation values range from 0.645 to 0.852), reflecting the coherence of the data despite the different gathering methods and data sources used. Their number is reduced to avoid multi-collinearity. The variable kept is the one showing the highest correlations with all other dependant variables and the highest communalities (correlation matrix and communalities in Appendix D, sections 1 and 2). Rule of Law (WGI RofL) is taken as representative of several other governance indicators that have high correlations with the Corruption Perception Index (0.874 correlation – CPI), Governance Effectiveness (0.906 correlation – WGI GE), Regulatory Quality (0.826 correlation – WGI RQ) and Environmental Governance (0.792 – Env Gov) and, has 0.924 communality value. The Rule Of Law index (WGI RofL) as representing these above cited governance indexes is named "advanced governance" later on in this thesis. The minimal political stability (WGI PS AV) and democratic/civil freedom context (WGI VA) appear as basic conditions to build an effective organisation able to make and abide a country by rules and to respond to the population needs. Therefore, the other WGIs are named "advanced governance", because "a good government" is rather probable and able to develop if these basic governance conditions are minimally satisfied.

In the same way, the Child Mortality rate (Child Mort-5) represents life expectancy at birth (-0.889 correlation-life expect) and the fertility rate (0.827 correlation - Fertility) and has the highest communality value (0.848). A high child mortality rate means a low life expectancy at birth and therefore is correlated with a high fertility rate in developing countries. Health expenditure (Health exp) has a high correlation (0.9 correlation) with income per capita (GDP per capita) (Appendix D, section 1).

The water bodies surface (WB) and National Biologic Index (NBI) are correlated with few variables—mainly with the amount of precipitation (0.439 and 0.732 correlations, respectively). The NBI is also correlated with the amount of water resources available per capita (0.589 correlation value between NBI and TWRR). These two variables do not add specific information regarding the water resources availability. They also show no significant correlations to the WS (0.098 correlation with WB and -0.033 correlation with NBI) or S (0.063 correlation with WB and -0.058 correlation respectively with NBI) variables (Appendix D, section1); therefore are excluded from the analysis.

Rural and urban population growth (Urban Growth and Rural Growth) variables represent complex processes that are not targeted in this research. Effectively, the evolution of rural/urban populations requires considering both the repartition of the population (Urban Pop) and their own dynamics (Urban Growth and Rural Growth). The interaction between these variables involves demographic, economic and urbanisation transition processes[46] (Chapter 2, section 2.2.1). Developing countries can be at different stages of this multi-dimensional transition. The

[46] A complete analysis can be consulted in the UN Department of Economic and Social Affairs, Population Division, 2006 and in UNFPA, 2007.

demographic transition mechanism is described among others by (Notestein, 1945) and discussed in (Canning, 2011). This interconnection between urbanisation and demographic transition is discussed in (UNFPA, 2007), (Dyson, 2011). Referring to demographic-urban transition observed in developed countries:

- the first stage corresponds to a demographic behaviour where the mortality, birth and fertility rates are high, thus, a limited naturel growth. The population mainly lives in rural areas.

- the transition stage sees the progressive reduction of the mortality and the fertility rate in a context of industrial development, accompanied by a population migration from rural to urban areas. Therefore, the population growth of both rural and urban population first increases before slowing down as the transition is progressing. The possible cases regarding both dynamics and repartition of population in rural and urban areas in a country are multiple during this transition. For instance, the rural population growth is high while percentage of the rural population remains high; both urban and rural population grow in a context of urban migration etc.

- the last stage corresponds to the achievement of this demographic and urban transition where the population live mainly in urban areas, and has a low naturel growth (low mortality and birth rates).

This European and historic example may diverge from what has occurred in recent decades, for instance in developing countries. Urbanisation process and the decline of mortality and fertility occur "in poor settings that are neither very urban nor very industrial" (Dyson, 2011). The combination of these transitions is not simple. Indicators to describe the state of the population (urban/rural distribution and its demographic behaviour) and indicators on the dynamics are useful to fully describe these mechanisms. However, this makes complex the interpretation of the multivariate analysis. It is decided to keep the variables describing the state of population repartition in the country–the percentage of the urban population, the proportion of urban population living in urban slums. The demographic behaviour of the country is expressed by the under-five children mortality rate. The state of the urbanisation as well as the demographic context dropping the dynamic side of these phenomena are found sufficient to situate the country in these transitions. The two indicators on population growth rate are therefore excluded.

The Environmental Sustainability Index (ESI), the Water Poverty Index (WPI) and the Human Development Index (HDI) are also excluded being composite indicators over-lapping active variables. They were only used as reference indicators projected into the space of the PCA (as supplementary variable) to support the interpretation of the relevant factors.

3. UNDERSTANDING RELATIONSHIPS BETWEEN VARIABLES

In the end of this selection procedure, the principal component analysis is performed and provides five factors with 25 variables, representing a STV ratio of 4.04:1 (101/25) (Table 5.1).

Variable	Minimum	Maximum	Mean	Std. deviation
WS	52.335	109.107	76.712	15.484
S	16.333	105.018	57.428	26.247
%Poverty	3.742	71.300	37.136	19.135
Urban Pop	10.000	98.000	49.248	23.819
School enrol	22.700	94.100	61.916	17.233
Femal eco	17.600	91.800	52.049	16.571
% Slums	0.000	97.000	44.639	24.805
Withdrawal domestic	0.000	44.308	25.674	10.329
Withdrawal industrial use	0.000	43.041	17.299	9.604
TWRR	20.337	126.950	85.225	20.433
Child Mortal-5	-14.706	156.637	80.497	48.417
School G/B	89.354	102.732	96.043	3.920
Particip	30.000	90.000	65.033	15.488
GDP per cap	0.000	10.000	4.337	2.540
%irrigation	0.000	100.000	47.197	27.137
ODA WSS	0.000	100.000	47.215	27.935
ODA	11.710	52.100	31.073	11.737
Malaria	0.000	100.000	47.019	30.269
Water Use int Agri	-6.931	102.518	65.064	20.651
Precipit	56.778	81.476	69.127	7.237
Desert risk	0.000	100.000	50.000	29.300
Total withdrawal	31.532	75.017	53.274	12.741
WGI.V.A.	0.000	100.000	48.796	22.114
WGI.PS.AV	0.000	100.000	55.920	22.792
WGI.RofL.	0.000	100.000	42.637	22.401

Table 5.1: Descriptive statistics of the 25 variables selected and processed.

Percentage of variance					
	F1	F2	F3	F4	F5
Eigenvalue	9.567	3.042	2.352	2.081	1.432
Variability (%)	38.269	12.169	9.409	8.324	5.727
Cumulative %	38.269	50.438	59.846	68.170	73.898
PCA-factor loadings					
Variables	HDP	AP	WR	ODA CI	CEC
WS	**0.855**	0.071	-0.045	-0.052	0.206
S	**0.825**	0.170	0.010	-0.044	0.068
%Poverty	**-0.745**	-0.037	0.115	-0.026	-0.055
Urban Pop	**0.786**	-0.205	-0.077	-0.069	0.054
School enrol	**0.769**	0.065	0.240	-0.042	0.239
Femal eco	**-0.639**	-0.196	0.206	-0.080	0.265
%slums	**-0.842**	-0.064	0.070	-0.005	-0.239
Withdrawal municipal	0.161	**-0.894**	-0.050	-0.044	0.060
Withdrawal industrial	0.206	**-0.643**	0.380	-0.007	0.066
TWRR	-0.276	-0.027	**0.800**	0.006	0.191
Child Mortal-5	**-0.875**	-0.202	-0.133	0.052	-0.129
School G/B	**0.589**	0.102	0.289	0.122	0.232
Particip	0.195	0.038	0.001	-0.151	**0.827**
GDP per cap	**0.921**	0.004	0.008	0.105	0.066
%irrigation	0.438	**0.726**	0.123	-0.097	0.055
ODA WSS	-0.063	0.038	-0.024	**0.823**	-0.076
ODA	-0.470	-0.097	0.068	**0.695**	-0.080
Malaria	-0.822	-0.271	0.173	0.115	-0.003
Water Use int Agri	0.494	**0.759**	-0.154	0.013	0.046
Precipit	-0.113	-0.009	**0.922**	-0.058	0.074
Desert risk	-0.345	0.112	**-0.677**	-0.044	0.341
Total withdrawals	0.515	**0.721**	0.044	-0.069	0.088
WGI.V.A.	0.298	-0.081	0.153	0.325	**0.743**
WGI.PS.AV	0.398	-0.055	-0.029	**0.691**	0.295
WGI.RofL.	**0.639**	0.111	-0.256	0.458	0.380

*The highest factor loading is in bold

Table 5.2: Rotated PCA factors matrix after varimax rotation and kaiser normalization. The five components explain 73.9% of the variability. The highest factor loading is in bold.

The overall KMO is 0.847, the KMO minimum is 0.515 (ODA-WSS and Precipit), the KMO maximum is 0.970 (Water Supply); the average value of the communalities is 0.739, the maximum value of communalities is 0.879 (Precipitat) and the minimum value is 0.504 (School G/B). The Variable-to-Factor ratio is 5 (25 variables/5 Factors). It is important to remind that "as long as communalities are high, the number of expected factors is relatively small, and model error is low (a condition which often goes hand-in-hand with high communalities), researchers and reviewers should not be overly concerned about small sample size" (Preacher & MacCallum, 2002, p.160). This can be considered as a moderate to high degree of overdetermination. The first five components explain more than 73% of the total variability (Table 5.2 and Fig 5.2). These components are used to analyse the relationships between variables. The PCA parameters and the correlation matrix are available in the Appendix D, section 3. The interpretation is provided and detailed per component in next sections.

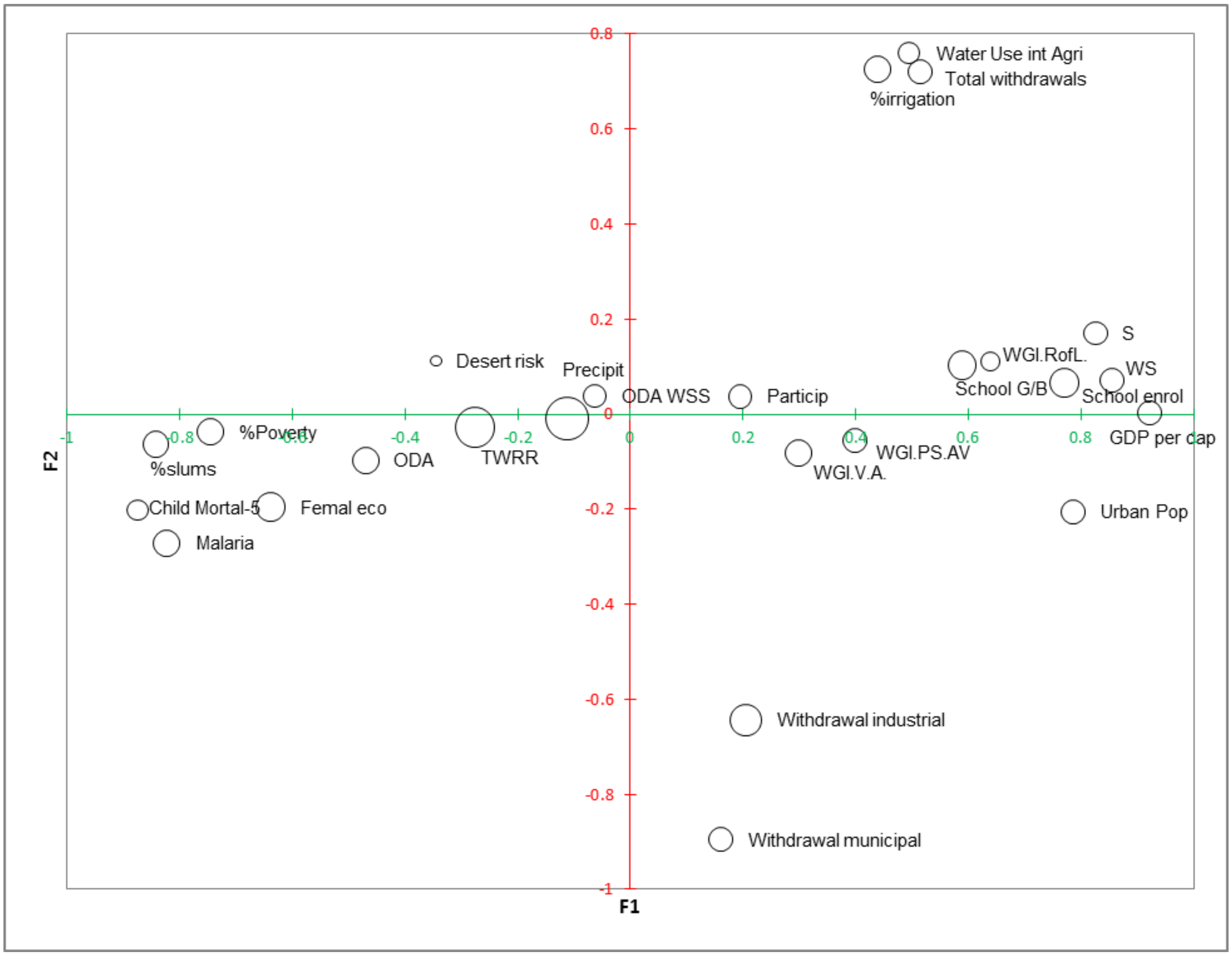


Figure 5.2: Three first components of the PCA (60%). The first component HDP is in green, the second component AP is in red and the size of the point represents the factor loading in the third component WR.

3.1 Factor 1: Human Development and Poverty variables

Factor 1 (HDP) represents the human development in the country, taking into account health (Child Mortal-5, Malaria), education (School enrol, School G/B), living conditions (Urban Pop, % Slums, WS, S), income (GDP per cap, femal eco, %poverty) and advanced governance (WGI

RofL). Unlike the HDI index, the HDP indicator computed in this study also factors in living conditions and governance.

As expected, figure 5.2 shows that variables related to poverty (female economic activity rate - femal eco, the percentage of slums - % slums, child mortality rate - Child Mortal-5, poverty rate - %poverty) are negatively correlated with development indicators (income per capita - GDP per cap, sanitation and water supply - WS and S, school enrolment- School enrol). The literature, the number of case studies demonstrate the negative consequences of poverty on health and education like in (Botting et al., 2010), (Tacke & Waldmann, 2013) and in Chapter 2, section 2.2.1. The Human Development Report (HDR)[47], published each year, analyses specifically the progress toward poverty reduction, health and education level improvement. The same applies to the positive relationship between the access to basic social services and the population well-being, like in case of upgrading slums (Turley et al, 2013).

The level of urbanization (Urban Pop) is negatively correlated (-0.650 correlation) with urban slums (% slums) accordingly to the dynamics of the rural-urban population transition detailed in Chapter 2, section 2.2.1. Generally speaking, urbanization represents an opportunity for the development (0.786 load on HDP) and for WSS services (0,643 and 0,624 correlation respectively). The agglomeration of the population provides the necessary critical mass for collective infrastructures and improves cost effectiveness. The organization and development/rehabilitation planning of urban areas is crucial to limiting the proportion of the population living in slums which cancel these positive effects.

The participation of women in economic activities is correlated with poverty variables − malaria (0.589 correlation), urban slums proportion (0.511 correlation), water supply access (-0.498 correlation), the mortality of children under 5 years old (0.477 correlation-Child Mortal-5). (Mammen & Paxson, 2000) explain that female participation in the labour force is represented by a u-shaped pattern when cross-referenced with the income level per capita for the 1970-1985 period (Fig 5.3). Female economic participation rates are high for low income countries. For the period considered, this rate decreased with the growth of average income to around $2,550 per

[47] HDR website: http://hdr.undp.org/, last access the 02nd September 2014.

capita (the minimum of the u-shape). Beyond this threshold, the female participation in the economy progressed again together with the growth of income. It can be assumed that the threshold ($2,550 per cap) has increased in absolute terms since 1985, but that the u-shape is still valid. The hypothesis is that the decreasing trend observed in our dataset corresponds to the first part of the u-shape (Fig 5.4). A number of observations, like in sub-Saharan Africa (Burundi, Mozambique Niger, Uganda….) show a low income per capita and a very high female economic participation rate (above 50%). This agrees with the characteristics of female economic participation described by (Boserup, 1989), (Mammen & Paxson, 2000), and (Beguy, 2009). A high female economic participation is therefore an expression of poverty in this work, where women contribute to the household's income. This supports the positive correlation observed between the female economic participation and poverty variables. However, the economic participation of women is based not only on income but also on several other factors, such as rural-urban context, fertility, social and cultural parameters that make the explanation of this phenomenon more complex (Boserup,1989), (Ahn & Mira, 1998), (Beguy, 2009). This may concern for instance, the middle-East countries where women employment is limited by legal and/or social constraints (Martin, 2006).

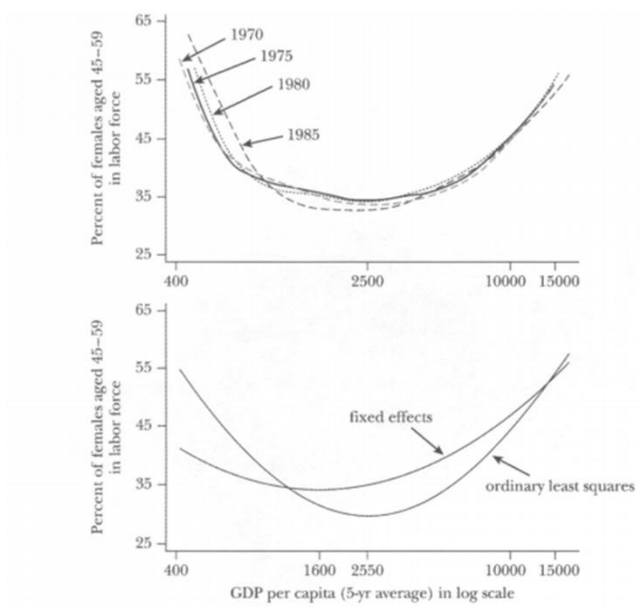


Figure 5.3: Distribution shape of female participation in labour force (for females aged 45-59 – although also holds true for other ages) (Mammen & Paxson, 2000, p 148).

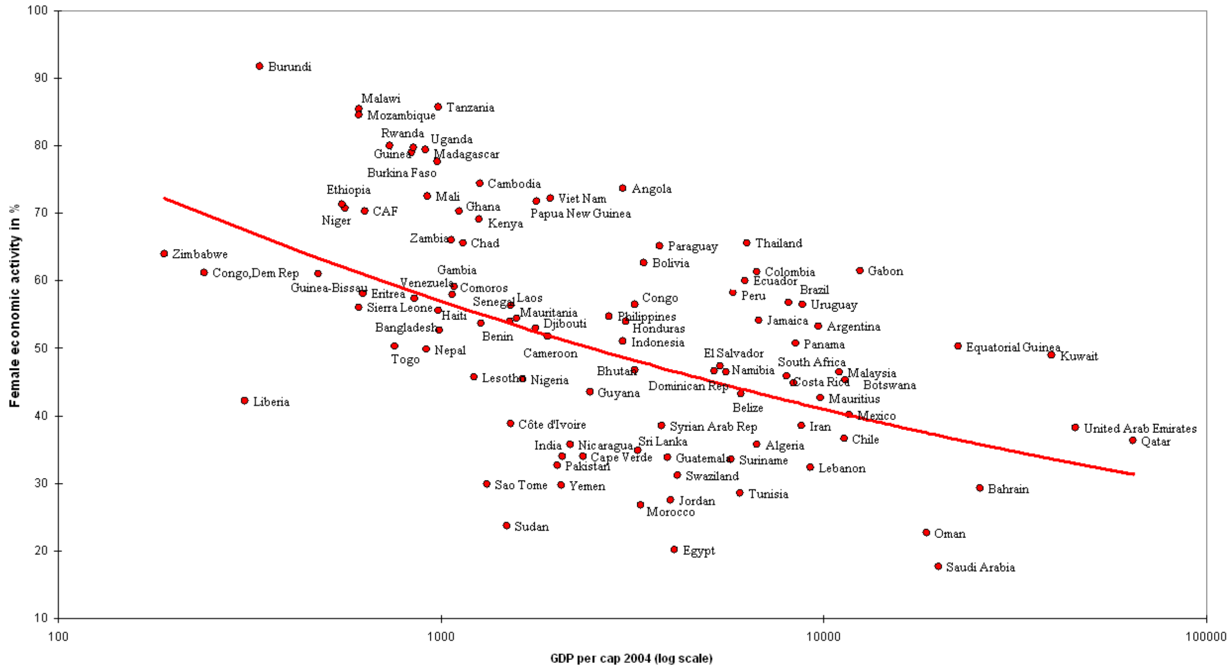


Figure 5.4: Female economic participation versus income per capita, distribution and trendline for the WatSan4dev dataset in 2004

Governance indexes have a significant load on F1 where WGI RofL is the stronger (0.639 load). Advanced governance (WGI RofL) is mainly correlated with the development of the country: a 0.674 correlation with Gross Domestic Product (GDP per cap), a 0.621 correlation with water supply access (WS), a 0.513 with basic sanitation access (S), and a 0.518 correlation with school enrolment (School enrol). Generally, the governance is assumed to foster the delivery of social-basic services and widely favourable economic framework (Wolf, 2007), (Whitford et al., 2010), (Chapter 2, section 2.2.2). Poor governance practices can have different forms and scales that range from the direct infrastructure management at municipal scale to sector policy making, planning and implementation at national scale. (TI, 2008) points out and describes the consequences of corruption on the water sector.

3.2 Factor 2: Water demand according to uses

Factor 2 (AP) shows high factor loadings for variables related to water consumption according to its use (Total Withdrawals, Withdrawal municipal and Withdrawal industrial), together with indicators related to irrigation (%Irrigation) and intensity of agriculture (Water Use Int Agri). It can be considered a measure of how economic activities put pressure on water resources in a country. It is well documented that the majority of water demand in developing countries is for agricultural purposes (Hinrichsen et al, 1997), (Rosegrant & Ringler, 1998). 82% of the developing countries considered (83 out of 101) dedicate more than 50% of their water resources to agriculture. Therefore, total water withdrawals are negatively correlated with municipal (-0.553 correlation) and industrial (-0.251 correlation) withdrawals (Meinzen-Dick & Appasamy, 2002). The more positive the factor loading is, the more predominant agriculture is. On the contrary, the more the factor tends to be near 0 or negative, the greater the share of municipal and industrial activities in water consumption is. The total withdrawals variable is mainly driven by agricultural demand, the main competitors with which are municipal and industrial demands.

Apart from high correlations with withdrawals and agriculture intensity (Water Use Int Agri), the level of agricultural land equipped with irrigation systems (% irrigation) is more correlated with the level of development of the country than with the water resources component (TWRR and Precipit). We advance the hypothesis that the irrigated perimeters captured by this variable are medium or large scale schemes; very small irrigations parcels are not included because of the data collection scale and data gathering methodologies. Irrigation development is furthered by investments, technological capacities and a strong political will for agricultural development. The cultural traditions in agriculture also play an important role in the development and maintenance of irrigation scheme, in particular in case of peasant community irrigation (like in the Andean area, India and South East Asia). For instance, (Clark et al., 1986) describes this type of irrigation associated with rice crop through several examples mainly in South East Asia. Sub Saharan Africa is less concerned by irrigation practises (except Maghreb areas) as the rain fed agriculture is predominant (Xie et al., 2014). Water resource availability appears to account for little in the choice of irrigation development in this case (Kweku Baah Inkoom & Zomanaa Nanguo, 2011).

3.3 Factor 3: Water resources estimates

Factor 3 (WR) gathers high factor loadings for variables expressing the amount of water resources available through the amount of precipitation (0.929 load - Precipit), the estimated amount of renewable water resources available per capita (0.770 load - TWRR) and the proportion of land under desertification risk (-0.667 load - desert risk). The variable desert risk is logically negatively association with the amount of water resources (-0.204 correlation) and precipitations (-0.531 correlation). The amount of water resources available and the desertification threat represent constrains to the satisfaction of the human water needs. (Albuquerque Sardinha, 2008) highlights the management methods that allow sustaining agriculture production in dry land areas facing desertification (with examples in Sahel area). A range of technologies, from rainwater harvesting (Helmreich & Horn, 2009) to sea desalination (IFAD, 2009), are available to fit the amount of water resources available to some extent. This gives a certain margin to meet human water needs provided that sufficient financial and human resources are available. This way to cope with limited water resources may explain the relatively high WSS access in countries, like in the Maghreb, with scarce water resources.

3.4 Factor 4: Official Development Assistance flows

Factor 4 (ODA CI) represents Official Development Assistance—both the overall flow of assistance (0.695 load) and the flow of assistance specifically dedicated to WSS (0.823 load). The governance indicator related to political stability and the absence of violence (WGI PS AV) is significantly also included (0.691 load). This raises attention on the criteria on which Official Development Assistance (ODA) is allocated to recipient countries. Several studies have explored the link between the amount of ODA and the quality of governance however they do not provided a unique conclusion (Alesina & Dollar, 2000), (Neumayer, 2003a and 2003b), (Bandyopadbyay & Wall, 2006). In fact, each donor acts differently according to ODA allocation criteria that most probably change over times. (Neumayer, 2003a) reminds that since 1990's international donors have given to "good governance "a central place in the aid agenda". Evidences of divergences between multilateral donors in affecting aid are found: the political freedom/democratic context appears to be rewarded by high assistance while the respect of personal integrity is little considered (Neumayer, 2003a). (Bandyopadbyay & Wall, 2006) found that Aid is responsive positively to governance effectiveness (WGI GE) and civil freedom/rights (WGI VA), the other WGI indexes not being considered. (Alesina & Dollars, 2000) suggest that the aid delivery is responsive to democratic conditions of recipient countries but varies importantly from donor to donor. It is also stressed the important role of the former colonisation link (i.e. for France), the strategic foreign policy (i.e. for US) and the UN voting pattern (i.e. for Japan).

In this work, the ODA committed to WSS (ODA-WSS) appears more sensitive to the political stability of the recipient country (0.363 correlation) than the ODA general flow (no significant correlations with WGIs, Appendix D, section 3). The latter is correlated with health (0.460 correlation with Child Mortal-5), access to sanitation and water services (-0.450 correlation with S and -0.412 correlation with WS) and poverty variables (0.408 correlation with% poverty or 0.405 correlation with % slums). This is coherent with the nature of ODA that concerned social aspects (health, education, basic services) for around 30% of the total amount in 2004 (OECD statistics, QWIDS) and with (Bandyopadbyay & Wall, 2006) findings. ODA are funds dedicated to development rather than given for emergency purposes, therefore a minimum level of political stability needs to be established for donors to commit to transferring ODA to beneficiary countries. This relationship is stronger when considering the assistance provided to the water sector (ODA-WSS).

3.5 Factor 5: Environmental concern and democracy.

Factor 5 (CEC) expresses the degree of concern a country demonstrates for the environment (0.827 load – Particip), empowered by democratic and government accountability (0.743 load - WGI VA). Signing international environmental agreements formally engages countries in global environment protection through the United Nations Framework Convention on Climate Change (UNFCCC), the Vienna Convention on the Protection of the Ozone Layer, the Convention on the Trade in Endangered Species (CITES) and the Basel Convention on the Transboundary Movement of Hazardous Waste. Therefore, a government committed internationally to the environment is assumed to implement internal environmental management strategies/policy to fulfil its engagements. The advanced governance (WGI RofL) accompanies the environment commitment with a 0.380 factor load.

Deeping Chapter 4, section 3.1, the civil freedom and widely, a democratic context where the government is accountable to citizens seem favourable to the environmental commitment and to improve of the quality of the environment (Barrett & Graddy, 2000), (Neumayer, 2002). The latter reviews the statistical evidences of the positive relationship between democracy and environmental commitment, and describes its environmental outcomes (i.e. the quality of environment is improved). This is argued that democracies show a stronger commitment to environment, even the translation in term of environmental outcomes is not systematic or immediate. The environmental outcomes can rather be correction actions than the prevention of environmental degradation (Gleditsch & Sverdrup, 1996).

The role of the civil society in pushing global environmental policy appears important in areas such as climate change mitigation, the protection of biodiversity and hazardous waste management. Citizens, organised in non-governmental, community organisations or groups of interest, can be key actors because, aside from economic-industrial interests, they are expected to push more for environmental policy/strategy making than other stakeholders (for instance, at international level, Gerdung, 2008). In the water sector, the civil society participation is favoured and pushed because of the development of integrated water resource management (IWRM); stakeholder participation is the principle n.2 of the Dublin Statement[48] (1992). A good context of governance (WGI RofL) is expected to favour the setup of such process and the implementation of the environmental commitment.

3.6 FA results

FA confirms the PCA results by showing the same factor loading matrix (Table 5.3) and therefore should be interpreted in the same way. The correlation matrix and parameters of the FA are in Appendix D, sections 3 and 4.

[48] The Dublin statement is available online: https://www.wmo.int/pages/prog/hwrp/documents/english/icwedece.html, last access the 3rd September 2014.

Percentage of variance					
	D1	D2	D3	D4	D5
Variability %	35.296	12.529	9.951	8.233	7.888
Cumulative %	35.296	47.825	57.777	66.009	73.898
FA-factor loadings					
Variable	HDP	AP	WR	ODA CI	CEC
WS	**0.855**	0.071	-0.045	-0.052	0.206
S	**0.825**	0.170	0.010	-0.044	0.068
% poverty	**-0.745**	-0.037	0.115	-0.026	-0.055
Urban Pop	**0.786**	-0.205	-0.077	-0.069	0.054
School enrol	**0.769**	0.065	0.240	-0.042	0.239
Femal eco	**-0.639**	-0.196	0.206	-0.080	0.265
% slums	**-0.842**	-0.064	0.070	-0.005	-0.239
Withdrawal municipal	0.161	**-0.894**	-0.050	-0.044	0.060
Withdrawal industrial	0.206	**-0.643**	0.380	-0.007	0.066
TWRR	-0.276	-0.027	**0.800**	0.006	0.191
Child Mortal-5	**-0.875**	-0.202	-0.133	0.052	-0.129
School G/B	**0.589**	0.102	0.289	0.122	0.232
Particip	0.195	0.038	0.001	-0.151	**0.827**
GDP per cap	**0.921**	0.004	0.008	0.105	0.066
% irrigation	0.438	**0.726**	0.123	-0.097	0.055
ODA WSS	-0.063	0.038	-0.024	**0.823**	-0.076
ODA	-0.470	-0.097	0.068	**0.695**	-0.080
Malaria	**-0.822**	-0.271	0.173	0.115	-0.003
Water Use Int Agri	0.494	**0.759**	-0.154	0.013	0.046
Precipit	-0.113	-0.009	**0.922**	-0.058	0.074
Desert risk	-0.345	0.112	**-0.677**	-0.044	0.341
Total withdrawals	0.515	**0.721**	0.044	-0.069	0.088
WGI VA.	0.298	-0.081	0.153	0.325	**0.743**
WGI PS AV	0.398	-0.055	-0.029	**0.691**	0.295
WGI RofL	**0.639**	0.111	-0.256	0.458	0.380

Table 5.3: Rotated FA factors matrix after varimax rotation and Kaiser Normalization. The first five components explain 73.9% of the variability. The highest factor loading is in bold.

4. ANALYSING COUNTRY OBSERVATIONS

4.1 Clustering countries with similar profiles

Hierarchical Agglomerative Clustering was used to explore the dataset through the analysis of the dendrogram. The AHC analysis (using dissimilarities and unweight group average aggregation method) showed a structure that allows different truncations: 3, 4/5 and 9 clusters. As recommended by (Murtagh, 1983), the AHC was repeated with different distance measures: Euclidian with Ward's aggregation method, Spearman dissimilarities and Pearson correlation coefficient. The structure of the dataset was similar with main three levels of possible truncation, at 3, 4/5 and greater than 8 clusters). Five clusters option is taken as a suitable trade-off as three clusters present to wide group of countries without presenting a clear profile while eight/nine clusters atomizes the dataset in small group of countries. The k-mean clustering method is used to finalize the five groups of countries with similar profiles among the 101 country observations (Fig 5.5) as the k-mean methods provided more relevant clusters (Chapter 3, section 3.4).

The *k* mean procedure was run 10 times[49] providing stable results. The *k* mean procedure was run 10 times providing stable results. Only a maximum of 9 peripheral countries (Bolivia, Belize, Jamaica, Lesotho, Pakistan, Sao Tome & Principe, Sudan, Yemen and Zimbabwe) in the clusters out of 101 countries shifted during the different runs. This can be considered as a minor impact in the computation and interpretation.

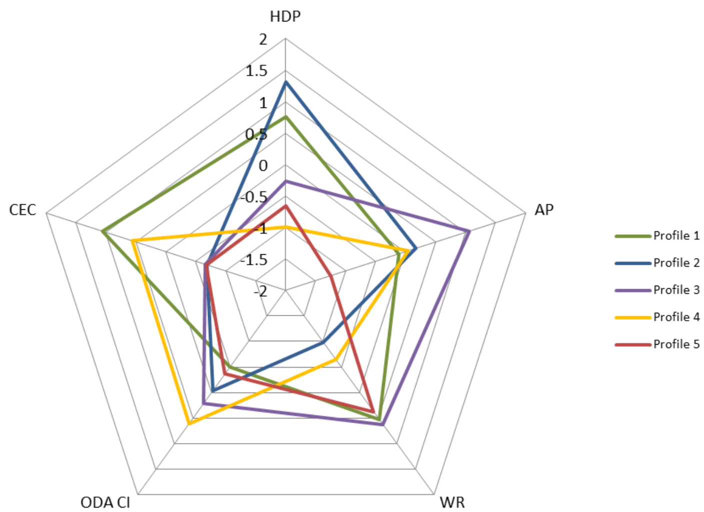


Figure 5.5: Characteristics of the five identified clusters. The reference value is the centroid of the cluster HDP (D1) stands for Human Development against Poverty, AP (D2) for Activities Pressure, WR (D3) for Water Resources, ODA CI (D4) for Official Development Assistance and CEC (D5) for Country Environmental Concern.

Cluster	HDP	AP	WR	ODA CI	CEC
1	0,762	-0,116	0,530	-0,500	1,057
2	1,310	0,169	-0,979	-0,036	-0,693
3	-0,239	1,200	0,684	0,177	-0,707
4	-0,981	-0,024	-0,638	0,635	0,545
5	-0,649	-1,246	0,376	-0,360	-0,676

Table 5.4: Coordinates of the centroid of each cluster. The random seed is 123456789

[49] The seeds used for the ten runs of the k-means clustering were: 4292015, 1683579780, 1866235603, 8032576,522649800, 850943609, 465352614, 732015342, 706934192.

Country profiles are ordered from most (profile 1) to the least (profile 5) advantaged country status when considering these five axes (referring to cluster centroid coordinates, Table 5.4). Profile 1 presents high WR, HDP and CEC values relatively to the others profiles (ranking 1st in CEC, 2nd in HDP and WR), implying that there is little need for external support. The water demand (AP) reflects a relative balance between municipal/industrial activities on one side and agriculture on the other side. Profile 2 shows a low commitment to the environment, associated with weaknesses in terms of accountability and civil freedom (ranking 4th in CEC). Profile 3 indicates that the economy is mainly driven by agricultural activities, facilitated by abundant resources (ranking 1st in AP and WR). Profiles 4 and 5 are the less favourable profiles in the context of human development (ranking respectively 5th and 4th in HDP). One difference between these two is that profile 4 benefits from a high level of ODA commitment (ranking 1st in ODA CI), associated in this analysis with a relative political stability and absence of violence in these countries. It’s reminded here that ODA allocation by donors depends on multiple factors (Chapter 5, section 3.4). Profile 5 reflects an economy mainly oriented towards the exploitation of primary resources (negative value in AP), while profile 4 reflects a more balanced economy (value around 0 in AP).

In figure 5.6, the African continent reflects the highest diversity, home to countries of all five profiles. Latin America is largely home to the “best profiles” with the exception of Bolivia, while South East Asia presents a mixed distribution, with countries ranging between advanced development (profile 1) and agriculturally-oriented (profile 3).

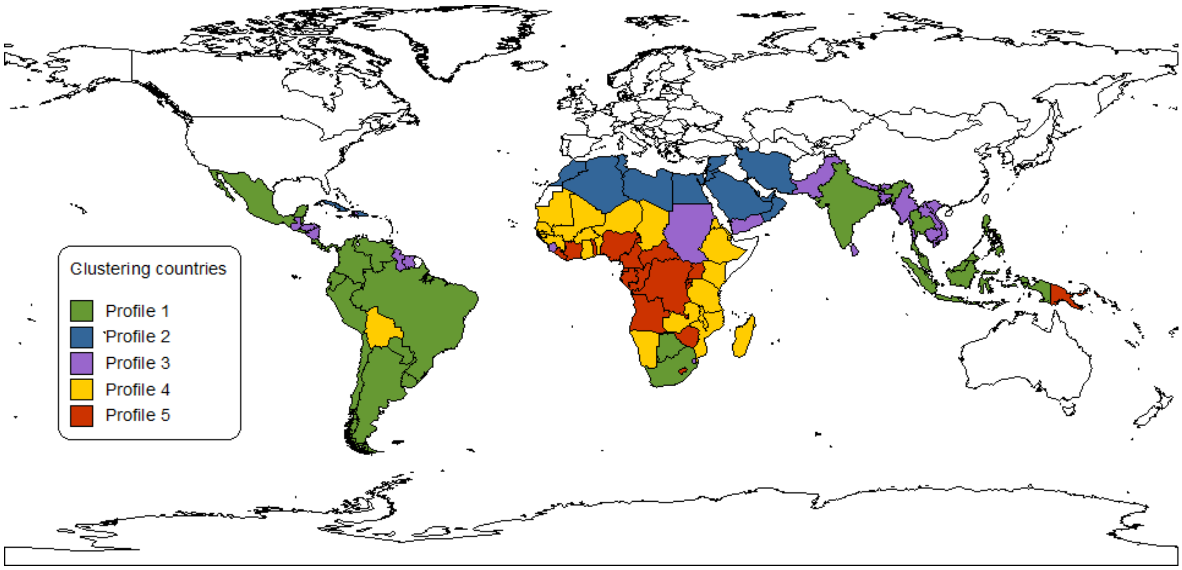


Figure 5.6: Geographical distribution of the country profiles.

4.2 Description of the five profiles

The main behaviour is first described by profile according to the cluster centroid values (Table 5.4). The divergences to the main behaviour are measured by the distance to the centroid, (Chapter 3, section 3.4) and are pointed out for each profile.

Profile 1: On the way to well-being (Table 5.5, Fig 5.7)

Profile 1 is characterized by good development (0.762 centroid value in HDP), a high environmental engagement and level of freedom (1.057 centroid value in CEC). Although the water resources in these countries are still mainly dedicated to agriculture, their resources are also devoted to municipal, commercial and/or industrial uses, as shown by the average value ranging around 0 (-0.116 centroid value in AP). These countries benefit from high water resource availability (0.530 centroid value in WR) and have the lowest level of external assistance (-0.500 centroid value in ODA CI). This profile is considered the most advantageous, reflecting relatively high levels of human development, environmental concern and expression freedom and a certain balance between water uses (agriculture versus industry/municipal uses), indicating diversified economic development. The context implies relatively high access levels to WSS (around 80% for WS), although basic sanitation always falls behind (around 70%).

The Philippines is the most representative country in this cluster (the closest to the centroid, Table 5.5). Belize shows divergences in its ODA CI (1.380 versus -0.500 centroid value), WR values but the amplitude of HDP, AP and CEC values justifies its inclusion in the profile 1 group. In Botswana a major share of their water withdrawals are for municipal and industrial activities (59%)—the agriculture is not dominant. While its economy is based on natural resource exploitation, it has succeeded in maintaining an income in other sectors such as tourism, financial services, subsistence farming and cattle breeding[50]. This tends to the "ideal" economic diversification for this profile. Jamaica (0.535), Malaysia (0.438) and Mauritius (1.165) show above-average values for ODA CI (-0.5 centroid value). Thailand and India show a maximum in AP (1.573 and 0.921 respectively), indicating the dominance of agricultural activities in these countries (consuming 96% and 87% of their water resources, respectively). Thailand has a close-to-average profile, with the exception of its AP value (1.573). India also shows a very low human development index (-0.183). Being at the periphery of this cluster, India could be seen as an intermediary country, sitting between profiles 1 and 3. The clustering algorithm has included it in this cluster because three components (CEC, ODA CI, WR) out of the five match the average values of this profile.

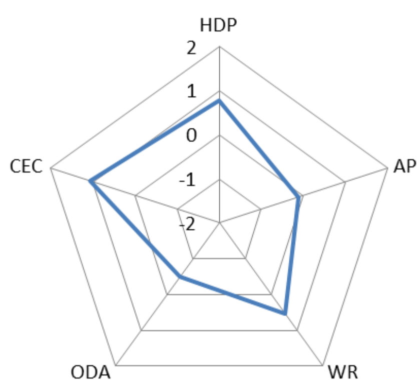


Figure 5.7: Profile 1 – this profile corresponds to the most advanced countries having a relatively good human development (HDP) based on a diversified economy (AP) and political/democratic organization (CEC). Their WSS level is, therefore, relatively high compared to that of other developing countries; consequently, these countries receive low ODA from donors.

[50] CIA factbook Botswana https://www.cia.gov/library/publications/the-world-factbook/geos/bc.html, last access on the 11th October 2013

Observation	HDP	AP	WR	ODA CI	CEC	Distance
Philippines	0.613	0.095	1.142	-1.166	0.586	1.052
Uruguay	1.318	0.714	0.630	-0.091	1.097	1.085
Brazil	0.774	-0.787	0.710	-1.277	1.379	1.091
Peru	0.381	0.524	1.263	-0.368	0.773	1.091
Colombia	0.563	-0.479	0.610	-1.511	0.728	1.143
Panama	0.732	-0.808	1.192	-0.993	1.482	1.158
Costa Rica	1.333	-0.342	1.536	-0.318	1.302	1.218
Ecuador	0.406	0.825	1.062	-1.025	1.231	1.266
Argentina	1.142	-0.370	-0.214	-1.352	1.679	1.369
Mexico	0.963	0.144	-0.684	-0.743	1.611	1.396
El Salvador	0.743	-0.561	0.731	0.311	-0.031	1.442
Paraguay	0.343	-0.915	0.117	-1.551	1.186	1.451
Malaysia	1.404	-0.376	1.376	0.438	0.643	1.498
Venezuela	0.586	-0.563	0.746	-1.998	0.602	1.651
Indonesia	0.119	0.829	0.866	-1.440	0.358	1.670
Chile	1.471	0.022	-0.304	0.324	2.026	1.685
South Africa	0.591	-0.603	-0.994	-0.185	1.595	1.726
Thailand	0.733	1.573	0.771	-0.879	1.037	1.748
Jamaica	0.781	-1.084	0.841	0.535	0.053	1.764
Mauritius	1.271	0.446	0.408	1.165	1.236	1.842
India	-0.183	0.921	-0.809	-1.303	1.416	2.129
Botswana	0.727	-1.568	-0.741	0.545	1.645	2.272
Belize	0.705	-0.304	1.937	1.380	0.667	2.388

Table 5.5: Factor loadings and Euclidian distances from the centroid of profile 1.

Profile 2: Freedom/democracy black spots (Table 5.6, Fig 5.8)

This profile has a high HDP value—above to that of profile 1 with 1.310 centroid value—but a low-to-medium level of ODA CI with -0.036 centroid value. Agriculture consumes the majority of the drawn water resources, but enough remain to satisfy municipal needs; industrial needs are marginal (0.169 centroid value in AP). The main differences between this profile and profile 1 lie in water scarcity (-0.979 centroid value in WR) and levels of CEC (-0.693 centroid value). They are relatively scarcely exposed to global environmental matters (e.g. climate change, protected species trade, hazardous waste transport) when compared to other countries. The amount of freshwater resources is very limited and under pressure, as these countries have part of their territory under desertification threat or in dry/desert areas (-0.979 centroid value in WR). Efficient irrigation system and unconventional water supplies (i.e. Desalinisation of sea water) are examples of measures implemented to mitigate water scarcity impacts (IFAD, 2009).The majority of these countries leave little space for citizens to express their views, many being monarchies or authoritarian regimes, resulting in the low level of CEC. In terms of WSS, these countries present the best levels of access (close to 100%). This fact confirms that water supply coverage is not systematically dependant of water quantity resources.

Lebanon is the reference in this cluster (Table 5.6). Cuba and the Dominican Republic diverge from the average because they benefit from better climate conditions and more water resource availability (0.676 and 0.494 respectively in WR). Saudi Arabia and Libya have the poorest environmental involvement at international level, together with little accountability to their citizens (-1.996 and -1.849 respectively in CEC versus -0.693 centroid value), implying a high divergence on CEC.

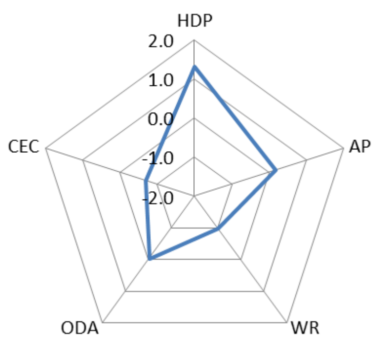


Figure 5.8: Profile 2 – Despite of a good human development (HDP), this profile presents very low values in terms of levels of accountability/democracy and low interest in environmental matters (CEC). Their relative wealth allows them to circumvent their scarce water resources (WR) with high levels of WSS services.

Observation	HDP	AP	WR	ODA CI	CEC	Distance
Lebanon	0.815	0.475	-0.917	0.234	-0.340	0.734
Oman	1.412	0.591	-0.770	0.318	-1.178	0.770
Egypt	0.987	1.132	-0.782	0.017	-0.600	1.040
Bahrain	1.861	0.018	-1.089	-0.037	0.183	1.052
Tunisia	1.195	-0.137	-1.166	0.796	0.001	1.147
Libya	1.540	0.126	-1.055	-0.070	-1.849	1.183
Kuwait	2.093	-0.372	-1.575	-0.416	-0.564	1.193
Morocco	0.511	0.386	-1.439	0.168	0.241	1.345
Jordan	1.269	0.140	-1.342	1.267	-0.057	1.495
Algeria	0.997	-0.812	-1.131	-1.193	-0.753	1.557
Dominican Rep	1.004	0.000	0.494	0.425	-0.609	1.585
United Arab Emirates	2.170	-0.393	-1.370	0.948	-1.522	1.692
Syrian Arab Rep	0.484	1.347	-1.092	-0.889	-0.356	1.711
Saudi Arabia	1.597	0.149	-1.362	-1.172	-1.996	1.794
Cuba	1.489	-0.003	0.676	-0.769	-1.269	1.916
Qatar	2.197	-0.697	-1.565	1.406	-1.323	2.087
Iran, Islamic Rep	0.641	0.920	-1.160	-1.639	0.210	2.104

Table 5.6: Factor loadings and Euclidian distances to average for profile 2. Iran is considered in periphery of this cluster, diverging from the average values of the profile.

Profile 3: Agricultural economy (Table 5.7, Fig 5.9)

These countries demonstrate a medium level of human development (-0.239 centroid value in HDP) sustained by important levels of external assistance (0.177 centroid value in ODA CI). The economies of these countries are largely oriented—in some cases exclusively—towards agricultural activities (1.200 centroid value in AP). This agricultural predominance is often accompanied by practices of irrigated agriculture (like in South East Asia). This agricultural intensity is facilitated by an abundance of water resources (0.684 centroid value in WR). Environmental concerns and government accountability come last in importance (-0.707centroid value in CEC). These countries are also defined by an important gap in term of access to basic sanitation (S) when compared to their access to an improved water supply (WS). Sanitation access is clearly neglected, sitting at least 25-30% less than their WS access figures.

Nepal is a good example with its WS access at around 90% and its sanitation access level at 35%. The fact that 81% of the population living in rural areas[51] facilitates the permanence of cheaper, easier and socially-accepted alternative sanitation options such as open defecation. Vietnam is the country closest to the centroid of this profile, with agriculture counting for 68% of its Water withdrawals and 31% of its agricultural land being equipped with irrigation. Table 5.7 indicates the distance different countries sit from the centroid.

Several countries diverge from the average profile: i) Honduras, Nicaragua and Guatemala have developed little irrigated/intensive agriculture compared to the South East Asian countries, with less than 3% irrigated areas; however, in these countries, agriculture accounts for around 80% of their water consumption; ii) Yemen's water resources are low (-1.213 value in WR), however agricultural represents almost all of its water consumption (at 95%); iii) Bhutan receives more ODA than the average (>125$ per cap, 5th position); iv) Bangladesh has a higher level of concern for environmental issues (0.400 value in CEC) than the average (+1.107 than centroid value). One hypothesis is that Bangladesh's concern is motivated by the consequences of environmental issues like floods or sea level rising. In addition, Bangladesh frequently relies on grass-root participation for basic services implementation and management (WSSCC-B, 2009). For instance, has been pioneer in implementing community-led total sanitation initiative which promotes community management of the sanitation facilities (Chambers, 2009).

Although from the 80's and 90's East Asian countries has seen an important development of manufacturing activities (starting from Japan spreading to South Korea, Taiwan, Hong Kong and Singapore and, more recently to Malaysia, Thailand and China (Razeen, 2010). Vietnam and Cambodia are the last recent countries initiating this economic evolution), the water withdrawal indicators in 2004 shows that the agriculture activity (AP) continues to be the most important water consumer in those countries. This has also spread in some Southern Asian countries such as Bangladesh attracting in particular textile manufacturing.

---

[51] CIA factbook NEPAL https://www.cia.gov/library/publications/the-world-factbook/geos/np.html

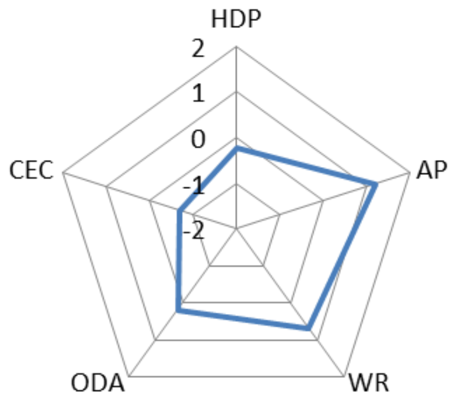


Figure 5.9: Profile 3 – mainly organized around agriculture (AP), and often irrigation, the countries in this profile show limited human development (HDP), with low accountability and low concern for the environment (CEC). An important gap in basic sanitation access is observed, generally falling far behind water supply access.

Observation	HDP	AP	WR	ODA CI	CEC	Distance
Viet Nam	0.196	0.866	0.983	-0.208	-0.270	0.854
Sri Lanka	0.383	1.818	0.450	0.195	-0.351	0.975
Cambodia	-1.045	1.685	1.224	0.488	-0.513	1.145
Sierra Leone	-1.034	0.484	1.209	0.316	-0.703	1.199
Laos	-0.526	1.107	1.443	0.611	-1.752	1.396
Haiti	-0.704	0.811	0.377	-0.967	-1.366	1.485
Guatemala	0.371	0.071	1.237	-0.259	-0.397	1.496
Swaziland	-0.388	1.724	-0.470	0.840	-1.162	1.509
Nepal	-0.640	1.643	-0.051	-1.049	-0.248	1.616
Nicaragua	0.168	0.122	1.323	1.392	-0.167	1.872
Myanmar	-0.444	1.929	1.176	-1.415	-1.386	1.952
Bangladesh	-0.522	1.876	1.468	-1.102	0.400	2.003
Honduras	0.430	-0.081	1.502	1.382	-1.044	2.079
Bhutan	-0.100	1.337	0.730	2.328	-0.564	2.165
Guyana	-0.003	1.798	1.871	1.852	-0.352	2.180
Yemen	-0.486	0.933	-1.213	-0.299	-1.660	2.206
Suriname	0.938	1.450	1.672	1.679	-0.210	2.220
Sudan	-0.930	1.341	-0.955	-1.055	-1.370	2.267
Pakistan	-0.212	1.881	-0.980	-1.367	-0.318	2.402

Table 5.7: Factor loadings and Euclidian distances to average for profile 3. Pakistan is considered in periphery of the cluster, diverging from the average values of the profile.

Profile 4: Essential external support (Table 5.8, Fig 5.10)

This profile is mainly defined by high levels of ODA CI (0.365 centroid value) that could be justified by low levels of human development (-0.981 centroid value in HDP). Despite this, these countries show a significant level of commitment towards environmental international issues and offer a significant level of freedom to their citizens (0.545 centroid value in CEC). Agriculture monopolizes a majority of the drawn water resources (-0.024 centroid value in AP), a small part remains for the development of other activities (domestic, commercial and/or industrial) in the context of water resource stress (-0.636 centroid value in WR). Indeed, most of the countries with this profile are facing stress on their water resources as they are mainly located in Sahel area and East Africa where climate change is pushing towards desertification (Nicholson, 2011). According to FAO data, around 75 % of the profile 4 countries face desertification on their territory[52] and/or are partly covered by desert (FAO, 2000). Ghana, Guinea, Benin, Mozambique and Bolivia show lower values between 11 to 28% of dry lands (AQUASTAT Database). (Reich, 1998) and (De Roo et al., 2012) have mapped, respectively, the areas vulnerable to desertification (worldwide scale) and the areas estimated to face water scarcity (for Africa). Most of these countries are located in vulnerable zones and/or under water stress. The access to water services is poor in these countries and has even worse access to sanitation services. Some very low levels were observed, such as those in Ethiopia—22% for water supply and 13% for sanitation.

Table 5.8 shows the distances the different countries sit from the centroid. Zambia is the most representative of the profile, having dominant agricultural activity (76%) but with 86 % of its land at desertification risk. Zambia also benefits from almost $100 ODA per cap.

Bolivia is the only country from Latin America in this profile. Bolivia is second in line as a target country for ODA (after Honduras) in Latin America, having one of the lowest human development indices (0.695 HDI value in 2005) in the region. Its economy is dominated by

[52] Drylands are defined as the arid, semi-arid and dry sub-humid zones, or areas with lengths of growing periods of 1-179 days and hyper arid zones are excluded (FAO, 2000).

agriculture, but it also exploits its natural resources[53], as indicated by the average value in AP (-0.298). This country is also concerned by environmental issues, as it is facing, among other things, risk of desertification (Reich, 1998). Having had a democratic regime since 1982, the country has the high CEC value (1.292), characteristic of this profile (Hudson & Hanratty, 1989). Madagascar almost exclusively dedicates their drawn water resources to agriculture (95%, like in profile 3). This is higher than the average value in this profile (1.627 value versus 1.200 centroid value in AP), however the other four components correspond to this profile. Eritrea diverges with a low value in CEC (-1.102). This country is under a strong authoritarian regime and ranks last for instance with regards with Press freedom index of Reporters without borders[54]. Cape Verde has superior development levels to the other countries but fits this profile for the other factors. Djibouti distinguishes itself with its scarce water resources (- 1.767 value in WR). As an existing desert area, this country is not considered to be "under desertification risk" by definition.

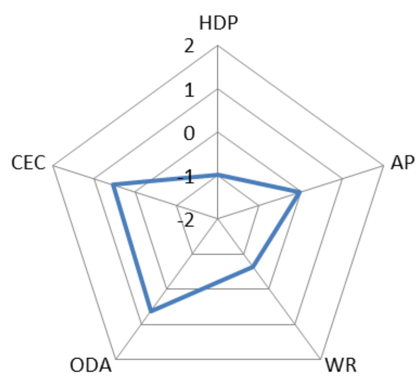


Figure 5.10: Profile 4 – The mean feature of this profile is a relatively high level of ODA, facilitated by relatively good political accountability and stability (CEC) despite low levels of human development (HDP). In this general context, access to WSS is still below the MDG's targets.

[53] Refer to CIA Factbook https://www.cia.gov/library/publications/the-world-factbook/geos/bl.html, last access on the 20th August 2014.
[54] Reporter without borders, ranking 2014: http://rsf.org/index2014/en-index2014.php last, last access on the 20th August 2014.

Observation	HDP	AP	WR	ODA CI	CEC	Distance
Zambia	-1.111	-0.237	-0.331	1.195	0.526	0.685
Mozambique	-1.455	-0.312	-0.296	0.755	0.797	0.708
Malawi	-1.168	-0.077	-0.323	0.082	0.820	0.719
Burkina Faso	-1.568	-0.109	-1.473	0.784	0.918	1.100
Guinea-Bissau	-1.499	0.132	-0.072	0.366	-0.196	1.111
Tanzania	-1.307	0.405	-0.606	0.100	1.404	1.147
Gambia	-0.588	-1.066	-0.824	0.198	0.759	1.230
Guinea	-1.107	0.337	0.178	0.119	-0.165	1.258
Ghana	-0.572	-1.107	-0.238	0.917	0.702	1.267
Senegal	-0.846	0.595	-1.065	1.367	1.258	1.276
Mali	-1.492	0.702	-1.002	0.793	1.522	1.378
Niger	-1.900	0.670	-1.366	0.478	0.987	1.440
Mauritania	-0.631	0.341	-0.875	1.294	-0.673	1.493
Bolivia	-0.384	-0.298	0.502	0.601	1.292	1.513
Benin	-0.694	-1.464	-0.503	1.079	0.175	1.584
Kenya	-0.989	-0.580	-0.892	-0.624	1.371	1.625
Sao Tome & Principe	-0.290	0.326	0.479	1.437	0.110	1.638
Ethiopia	-1.527	0.645	-1.069	-0.556	-0.207	1.707
Namibia	-0.069	-0.764	-0.626	1.841	1.086	1.769
Chad	-1.747	-0.239	-1.054	-0.906	0.048	1.851
Eritrea	-1.200	0.638	-1.255	0.028	-1.102	1.987
Djibouti	-0.155	-0.835	-1.767	0.994	-0.582	2.004
Cape Verde	0.325	0.099	-1.057	2.260	0.441	2.133
Madagascar	-1.581	1.627	0.227	0.629	1.794	2.321

Table 5.8: Factor loadings and distances to average for profile 4. Madagascar is considered in periphery of the cluster, showing divergence from the average values of the profile.

Profile 5: Instability and Primary material consumption (Table 5.9, Fig 5.11)

This profile is characterized by a rather abundance of water resources (0.376 centroid value in WR) that shapes the profile. However, its main characteristic is the low values observed on the other axes. First, in a context of limited water exploitation, the withdrawn water is mainly dedicated to industrial and municipal uses, as the agriculture sector is often underdeveloped in these countries (-1.246 centroid value in AP). In fact, countries with this profile generally have an economy based on mining or primary material exportation, and this is the main consumer of water resources. This is the case in Gabon, Congo, Angola Liberia, Papua New Guinea, Togo and Nigeria, all of which exploit sub-sols resources such as oil, gold, copper, phosphates or diamonds, or other natural resources such as timber.

This profile includes both non-favourable conditions for a sustainable development (-0.649 centroid value in HDP) and an economy focused on primary resource exploitation, the processes of which consume water resources in a non-sustainable way. These countries also have poor human development and political instability (-0.676 centroid value in CEC). The possible consequences of this are that they receive only limited external financial commitments from international donors. Despite having low levels of human development, these countries receive rather low levels of ODA CI (-0.360 centroid value). One hypothesis is that this is because of the political instability and violence prevalent in these countries. As already mentioned, ODA is effectively dedicated to development and not to humanitarian emergencies. International donors may consider a minimum level of security and stability necessary to engage and implement a medium-long term vision (Chapter 5, section 3.4). In terms of WSS, the situation is similar to that of profile 4, with populations having low levels of access to water and sanitation. Comoros acts as the reference country for profile 5. Although Gabon has a high HDP value (0.427), its other components fit the profile well. Zimbabwe and Burundi diverge from the average profile in that agriculture is a central economic activity (-0.010 and -0.402 respectively versus -1.246 centroid value in AP). These two countries can be seen as intermediate cases between profiles 3 and 5. They belong to cluster 5 because three components out of the five (HDP, ODA CI, WR) fit with the average behaviour of this profile. Equatorial Guinea has had an economy based on oil exploitation since 1996 and is in the group of least developing countries[55] (low HDP). Its main divergence is its very negative CEC value (-2.283), which is a result of the country being governed by an authoritarian regime[56] since 1979, therefore low value in WGI VA index (-1.670) and lack of commitment to the environment (Neumayer, 2002). However, its political stability since 1979 allows the delivery of ODA ($63 per capita). Lesotho enjoys political stability and benefits from significant external financial flows ($48 per capita). It also has a relatively high level of CEC (0.286), indicating a high level of freedom (similar to profile 4).

[55] Least developed countries list http://www.unohrlls.org/en/ldc/25/, last access on the 11th October 2013
[56] CIA World fact book: Since 1979, Teodoro OBIANG is the president of the republic elected with 95.8% of votes.

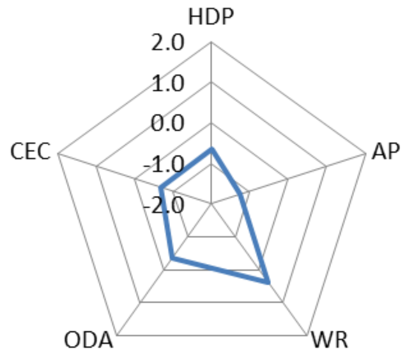


Figure 5.11: Profile 5 - This profile shows low values in human development (HDP) and/or accountability towards the population and the environment (CEC). Because of the political instability and/or violence prevalent within some countries, their external assistance remains marginal. Environmental concerns may be undermined by country's type of economy, many of which are based on raw material or natural resources exploitation.

Observation	HDP	AP	WR	ODA CI	CEC	Distance
Comoros	-0.701	-1.170	0.594	-0.487	-0.324	0.443
Cameroon	-0.490	-0.828	0.215	-0.273	-0.334	0.592
Congo, Dem Rep	-1.388	-1.362	0.753	-0.568	-1.070	0.949
Rwanda	-0.845	-0.799	0.321	0.485	-0.533	0.988
Uganda	-0.793	-1.519	0.363	0.089	0.167	1.004
Congo	-0.403	-2.164	0.501	-0.517	-1.029	1.034
Togo	-0.686	-1.069	-0.570	-0.692	-0.161	1.142
Liberia	-0.758	-1.157	1.322	-0.601	-1.324	1.180
Angola	-1.112	-1.267	-0.331	0.203	-1.323	1.204
Burundi	-1.412	-0.402	0.399	-1.012	-0.408	1.338
Ivory Coast	-0.501	-0.949	0.194	-1.685	-1.028	1.422
Central African Rep	-1.045	-2.216	-0.176	-1.004	-1.159	1.432
Gabon	0.427	-1.531	1.289	-0.228	-0.564	1.450
Nigeria	-0.648	-0.757	-0.639	-1.374	-0.550	1.521
Papua New Guinea	-0.690	-1.818	1.992	0.050	-0.187	1.829
Zimbabwe	-0.350	-0.010	-0.812	-1.373	-0.348	2.039
Lesotho	-0.193	-1.611	0.087	1.331	0.286	2.052
Equatorial Guinea	-0.090	-1.797	1.264	1.173	-2.283	2.517

Table 5.9: Factor loadings and Euclidian distances to the centroid of profile 5. Equatorial Guinea is considered in periphery of the cluster, showing divergence with the average values of the profile.

5. *IDENTIFYING KEY VARIABLES INTERACTING WITH WSS*

A multiple regression is concerned with the relationship between a dependent variable− the WSS variables and a series of m independent variables− 23 variables. This is one of the methods used to identify determinants (Chapter 2, section 2.3.2) and is adapted to our normalized and quantitative variables (Chapter 3, section 3.5). Therefore, an OLS linear regression analysis was performed with both water supply (WS) and sanitation (S) as dependent variables with the objective of identifying key explanatory indicators, rather than to obtain a model (Tables 5.10 and 5.11). For this reason, rather low values of R^2 are accepted. The first linear regression (method stepwise and adjusted R^2 of 0.711 for analysis with WS and 0.680 for analysis with S) highlighted that, in order, the proportion of urban slums, the mortality rate of children under five years and the prevalence of malaria are key variables explaining both WS and S. The child mortality rate and malaria prevalence variables are removed from the 2nd regression so as to find additional variables and to look for potential differences between WS and S variables. The results are presented in the next two sections.

5.1 Water Supply

OLS Regression 1st run						
Variable	Standardized Coefficient		t*	Sig.*	95% confidence Interval for β	
	β*	Std error*			Lower Bound	Upper Bound
Intercept	101,432	1,780	56,975	< 0.0001	97,899	104,966
% Slums	-0.420	0.078	-5.403	< 0.0001	-0.574	-0.265
Child Mortal-5	-0,301	0,102	-2,966	0,004	-0,503	-0,100
Malaria	-0,219	0,092	-2,374	0,020	-0,403	-0,036
OLS Regression 2nd run						
Variable	Standardized Coefficient		t*	Sig.*	95% confidence Interval for β	
	β*	Std error*			Lower Bound	Upper Bound
Intercept	66.894	5.906	11.327	< 0.0001	55.171	78.616
GDP per cap	0.305	0.106	2.881	0.005	0.095	0.515
% irrigation	0.196	0.063	3.094	0.003	0.070	0.322
Urban Pop	0.179	0.085	2.108	0.038	0.010	0.348
% slums	-0.333	0.099	-3.364	0.001	-0.529	-0.136

*Where β is the coefficient of the explanatory variable, Std error is the standard error of β, t is β divided by the std error and Sig. is the p value.

Table 5.10:Models parameters with WS as dependent variable - method stepwise, adjusted $R^2 = 0.711$ for regression 1st run and adjusted $R^2 = 0.686$ for regression 2nd run.

This analysis indicates that the keys to WS interpretation are urbanization (% slums), health care represented here by the child mortality rate (Child mortal-%) and malaria prevalence (malaria), the level of income (GDP per capita) and the proportion of land equipped for irrigation.

The urban areas benefit globally from a higher access to WSS services for multiple reasons such as the concentration of the population, economic, accessibility or critical mass advantages (JMP statistics). Cities provide opportunities for investment in WSS because the critical mass to create new infrastructure and management capacities is easily reachable (Chapter 2, section 2.2.1). The JMP continues to highlight this rural-urban gap with a specific report on equality and disparities published in 2011 (JMP, 2011). Therefore, efforts should be oriented particularly to rural areas.

The analysis highlights that informal urbanization developments (% slums) contribute in a negative way (negative coefficient) to WS. Over the last 20 years, developing countries are going through a rapid, important and thus, often informal urbanization process (Cohen, 2006). Municipal authorities often encounter difficulties in facing and the structuring the population flows from rural areas and/or urban growth, which explains the negative coefficient β (Table 5.10). Therefore, the city expansion is not accompanied by WSS services development. This non-simultaneity of the two adds another constraint due to the implementation of WSS facilities a posteriori in very high density areas. In addition, slums often develop in rough land configuration that make complex this infrastructure implementation afterwards in such areas (Banes, 2000). Slum development is expected to increase, with urban dwellers accounting for 66% of populations in less developed regions by 2050 (UN DESA, 2011). Therefore, extending percentage of access to WSS in these conditions can be difficult.

The positive impact of access to improved WSS services on human health is recognised (Cairncross, 1992), (Botting et al., 2010) (Chapter 2, section 2.2.1). Related to malaria, improved water supply but also good maintenance of the water infrastructure – in particular storage or irrigation schemes, is important to avoid the stagnant water bodies necessary for malaria vector breeding (Matsuno et al.,1999) in particular in case of irrigation (Konradsen et al., 2004), (Lewis, 2011).

The average income per capita (GDP per cap) is understood to be an indicator of living standards and is the key for households’ access to an improved WSS infrastructure (Sbrana, 2009).

Lastly, the development of irrigation is a positive factor. This positive relationship may be due to the multipurpose nature of the irrigation infrastructure: water supply sources used for crops are often used for water delivery for livestock and people (Van der Hoek et al., 1999).

5.2 Sanitation

OLS Regression 1st run						
Variable	Standardized Coefficient		t*	Sig.*	95% confidence Interval for β	
	β*	Std error*			Lower Bound	Upper Bound
Intercept	98,187	3,173	30,949	< 0.0001	91,890	104,483
% Slums	-0,356	0,082	-4,363	< 0.0001	-0,518	-0,194
Child Mortal-5	-0,302	0,107	-2,826	0,006	-0,514	-0,090
Malaria	-0,264	0,097	-2,721	0,008	-0,457	-0,072
OLS Regression 2nd run						
Variable	Standardized Coefficient		t*	Sig.*	95% confidence Interval for β	
	β*	Std error*			Lower Bound	Upper Bound
Intercept	32.269	12.061	2.675	0.009	8.330	56.207
Total withdrawals	0.287	0.069	4.142	< 0.0001	0.149	0.425
Urban Pop	0.259	0.080	3.255	0.002	0.101	0.417
% Slums	-0.432	0.087	-4.954	< 0.0001	-0.605	-0.259

*Where β is the coefficient of the explanatory variable, Std error is the standard error of β, t is βdivided by the std error and Sig. is the p value.

Table 5.11: Model parameters with sanitation as the dependent variable, stepwise method with Adjusted R^2= 0.680 for regression 1st run and Adjusted R^2= 0.635 for regression 2nd run.

As for WS, the type of urban development, the child mortality rate and malaria prevalence that exists in a country are key to explain the level of access to sanitation for the reasons mentioned in Chapter 5, section 5.1. The total water withdrawals represent the capacity of the society to mobilize their available water resources as a whole. It is generally accepted that the total water consumption per inhabitant (total water withdrawal) increases with the development of a country because of the increase of household consumption (Gleick, 1996) and/or increases in the water volume needed due to economic activity development (industrial, commercial and agricultural). The link with sanitation in this context may be indirect. With economic development and the resulting growth of household consumption capacities, the demand for sanitation services increases. Sanitation is indeed an external/social sign of well-being. At the same time, the economic dynamic allows for development to satisfy this demand. As the total withdrawal variable is driven by the agriculture water demand (Chapter 5, section 3.2), the analysis may also indicate that sanitation benefits from the development of the agricultural sector. The necessary development of water supply capacities (technological aspects, operators' capacities, financial abilities, etc.) to respond to domestic or agricultural water demand may also overflow into sanitation development. Stakeholders involved in water provision start investing at the end of the water cycle—meaning wastewater treatment and sanitation—to address pollution threats.

6. Discussion and Conclusion

This chapter details the analyses of both the variables and the observations gathered in the WatSan4dev subset. A preliminary analysis (Chapter 5, section 2) demonstrated that 25 of the variables were actually independent and suitable to this research's goal. The variables organize themselves around five factors: Factor 1 (HDP) represents human development, including the governance of the country. Factor 2 (AP) expresses how economic activities pressure water resources in a country. Factor 3 (WR) gathers high factor loadings for variables expressing the amount of water resources available. Factor 4 (ODA CI) represents the external financial flows (ODA and ODA WSS) conditioned by political stability and the absence of violence (WGI PS-AV). Factor 5 (CEC) links the involvement of a country in International Environmental Agreements with the strength of its civil society (WGI-VA).

The HDP factor expresses the development of a country not only as a function of income, life expectancy and education like HDI, but also taking into account the living conditions of the population (Urban Pop, % Slums, WS, S) and advanced governance (WGI RofL).

Governance is a cross-cutting factor; therefore WGIs are associated with several components. The reduction of governance variables to three allows seeing better their respective loadings on different components (Chapter 5, section 3). This discrimination gives additional insights on the role of governance by detailing the governance association with Official Development Assistance (F4) and Commitment to the Environment (F5). In fact, the political stability index (WGI PS AV) is associated to ODA CI (0.657 loading on F4). In the same way, the democracy/civil freedom index (WGI VA) is associated the international environmental commitment (0.762 factor loading on F5) (Neumayer, 2002). The first association completes the prior analysis of the role of "good governance" in the allocation of Official Development

Assistance. The ODA committed to WSS appears more sensible to the political stability of the recipient country (0.363 correlation with WGI PSAV).

Extending the discussion of chapter 4, the water resource composite indicator still requires reinforcement of data collection, which involves finding proxies or designing appropriate new indicators on a national scale. Water quality can be assessed on two levels: the quality of surface-ground waters and the drinking quality of the water supply. The FAO and UNEP/GEMS-Water[57] are mainly engaged in efforts to provide data on biochemical and nitrogen concentrations in surface waters on a national or basin scale. The JMP performed pilot country projects (Ethiopia, Jordan, Nicaragua, Nigeria, and Tajikistan) in 2004-2005 to investigate new methods of monitoring the quality of drinking water supplies. The conclusions of these pilots have been available since 2010 and estimated that 30% of the water supplies monitored do not comply with WHO water quality standards. The generalized data collection methods have not been initiated globally yet[58]. The JMP also set up a task force to study and suggest ways to monitor drinking water quality (JMP, 2010). To summarize, further research on country indicators defining quality (accuracy and uncertainty), accessibility, the management capacities of resources and the state of the environment are needed to refine the analysis of the relationships between the environment, human development and WSS.

Country profiles have been established around these five factors and are ordered from the most favourable (profile 1) to the least advanced (profile 5) country status. This new classification helps identify countries in difficulty and the particular weaknesses restricting country development and well-being. The method synthetizes the large amount of information to five composites and, therefore, provides an understandable status of a country/or group of country. The external assistance flow is for the first time included with the 5 keys dimensions: socio-economic development, water resources, water demand, and environmental commitment including governance.

[57] UNEP program for the development of a global freshwater quality information system with a series of national and international partners, http://www.unep.org/gemswater/, last access on the 11th October 2013.
[58] Refer to Water Quality documentation on the JMP website: http://www.wssinfo.org/water-quality/introduction/, last access on the 11th October 2013.

Countries in profiles 4 and 5 are less advanced. They need to put efforts in different domains in addition to give the priority to social and economic development. Profile 4 countries are facing water scarcity in the context of low development. Therefore, efficient resource management, mitigation measures and adapted solutions are crucial to improving WSS. In the case of profile 5, efforts should be concentrated on ensuring political stability and the reduction of violence, which is the basis of any development.

In attempting to identify key variables to improved access to WSS, it was found that health improvement and urbanization and more precisely, the type of urbanization plays an essential role in access to improved water supply and sanitation facilities. The relationship between child mortality under 5 years (Child Mortal-5) and access to water supply and sanitation (WS and S) is clearly depicted in this analysis. Indeed, WSS is clearly correlated more generally with health, which is essential to improve living conditions (Butala, 2010), (Turkey et al., 2013).

Cities are clearly a favourable context for WSS infrastructure maintenance and expansion thanks to the urban population agglomeration. However, informal urbanization developments (slums) correlates negatively with the water supply and sanitation conditions because of the difficulty local authorities in the cities to face structuring massive population flows from rural areas (Cohen, 2006). This risk of slums is expected to increase by 2050 (UN DESA, 2011) and urbanization is expected to continue over the next 30 years, mainly in medium and small cities. (Cohen, 2006) mention that these latter city categories are facing greater health issues and lacks of capacity than large cities. Fighting and finding solutions to organize urbanization processes—in medium and small cities in particular—should therefore be a priority. The extension of access to sanitation and water supply services at country level may be slowed down due to these urban dynamics.

Lastly, irrigation facilitates the improvement of the water supply thanks to a multi-purpose infrastructure. Sanitation facilities are associated widely to the level of water demand. The access of Sanitation is expected to follow an increase of water needs that manifest an increase of living standards (Gleick, 1996). Sanitation services appear to benefit from the strengthening of the water supply sector. Note that environmental conditions and water resources availability (total water renewable resources, the precipitations, desert risk) are secondary factors explaining the WSS level (Chapter 4, section 4).

The OLS method provides insights on the main key variables but excludes some other variables that may influence the WSS and the human development. These methods consider the linear component to explain around 68-70% of the variance of the dataset variables (Chapter 3, section 3.5). Next chapter will deal with the modelling using Bayesian Networks to explain both the linearity and non-linearity components of the variables.

CHAPTER 6: MODELLING WATSAN4DEV SUBSET FOR 2004 USING BAYESIAN NETWORKS

1. *INTRODUCTION*

Linear models in the precedent Chapter 5 were built to identify the key variables that foster sustainable WSS development with good statistical performance (Chapter 5, section 5). However, these models explain 68-70% of the variance of the WSS variables (with adjusted R^2 between 63% and 68%, respectively). Therefore, further modelling methods were examined and the Bayesian Networks (BNs) methods were chosen because of their performance in managing missing data among other characteristics as described in Chapter 3, section 4.

This modelling of the 25 variables subset using BNs will measure the effects of WSS variations on country development and will estimate the relationship with the external assistance flows invested by donors at national level. Having an understanding of these relationships may help orient efforts and provide incentives to expand access to WSS. Various probabilistic models were built and several scenarios were run in this attempt to describe these relationships and "map" the processes behind WSS behaviours (Fig 6.1).

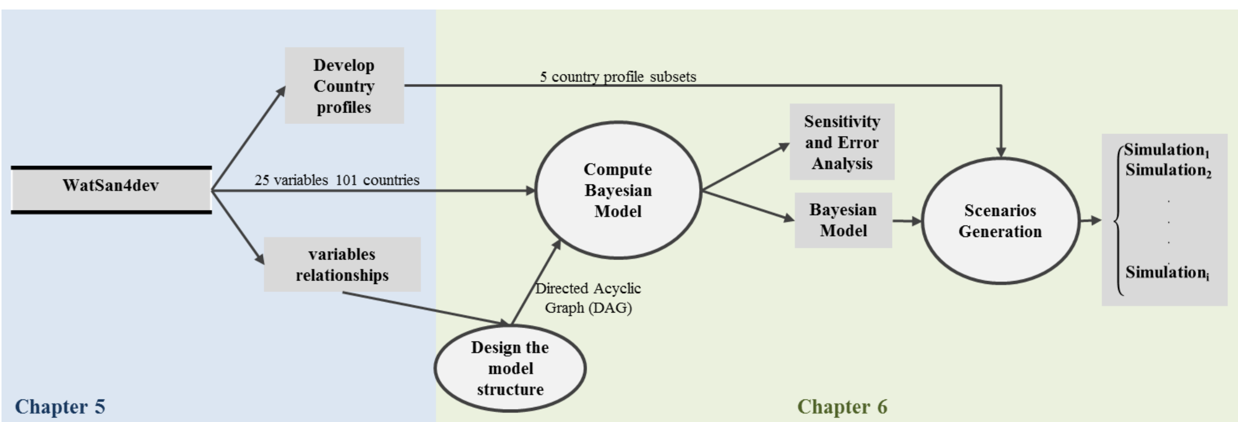


Figure 6.1: Functional Data flow Diagram related to 2004 modelling where a rectangle with two black borders represents a database, a circle indicates a process, and a grey rectangle a result.

This chapter first details how BNs are applied concretely to the WatSan4Dev subset. The different general and thematic models built are presented in sections 3 and 4, respectively.

Simulations and interpretations based on the model results are also included in these sections. The modelling methodology is then applied to the best profile (profile 1) and the two worst country profiles (profiles 4 and 5) to deepen the analyses of these key groups of countries (section 5). Finally, the results are discussed in section 6 of this chapter.

2. BNs applied to WatSan4Dev subset

The dataset modelling is done in two mains steps: 1) integrating composite indices for each dimension in a global model; 2) creating a model for each dimension. Separated models are built for WSS. The global models offer a synthesis of the mechanisms behind WSS, while the modelling by dimensions examines WSS behaviour in detail. This section describes the four processing steps necessary to model our subset using the BNs methods:

1) **Variables categorization:** The variables are categorized using the k-mean method (McQueen, 1967) (Chapter 3, section 3.4). Although the categorization process produces a loss of information (Aguilera et al., 2011), it is performed to make the problem tractable and to homogenize the dataset. The objective is to create three categories or levels (low, medium, high) that match the qualitative nature of the selected variables. Table 6.1 indicates the limits of each class for each variable used. WSS and WGIs variables show only two levels (high and low) because of the shape of their distribution, which increases the classification error rate. (Uusitalo, 2007) mentions that categorization can be done through an automatic process or according to expert knowledge where the second may be preferred depending of the problem. At this stage, an automatic method (k-mean clustering) is selected as only one year is considered and therefore the consistency of scale over years is not an issue.

2) **Building thematic composite indicators:** Composite indicators are built for each thematic dimension, namely HDP, CEC, AP, WR and ODA CI. The first component of the specific PCA, which includes the dimension sub-variables, is categorized in three levels using the k-means clustering method, as were the single variables. Regarding the composite HDP, the PCA is performed excluding the two WS and S variables to avoid their overestimation in the models.

3) **DAG design**: Two-steps modelling strategy is used to develop the general and specific models. The global or general models are built using the PCA/FA results and using expert knowledge (literature and knowledge reported from the field) to specify link directions. The knowledge reported from the field corresponds to reports, working papers describing direct observations and/or analyses from organisations working on the ground such as NGOs and UN local agencies. This expert knowledge is combined with the multivariate analysis's results presented in Chapter 5 to design the structure of the models.

   This expert and statistically-guided method for developing graphs allows for the testing of different structures in order to assess the goodness of fit of each. This method is used as a non-parametric alternative to regression models trying to generate more parsimonious (in terms of the number of parameters to estimate) and usable models from a decision-making point of view.

4) **Beliefs computation**: the Bayesian theorem is applied to and probabilities are computed for each variable and each level (high, mid, low) according to either the whole input dataset (101 developing countries in sections 3 and 4 of this chapter) or any subset of observations-countries (Chapter 6, section 5).

5) **Probabilistic scenarios**: these are then run to describe the interrelationships between variables or the frame of a specific country profile.

Variables	Categories				Unit or intervals
	LOW		MEDIUM	HIGH	
	Minimum	Upper limit	Upper limit	Maximum	
WS	22	82	na	100	%
S	9	58.5	na	100	%
Poverty	7.6	25.	49	71.3	%
Urban Pop	10	39	69	98	%
School enrol	22.7	46.4	69	94.1	%
Femal eco activity	17.60	40.2	63.1	91.8	%
% Slums	0.00	29	58	97	%
Withdrawal Municipal	1.00	19.34	52.88	84.00	%
Withdrawal industrial	0.00	25.99	54.75	73.00	%
TWRR	7.64	459	9465	326116	m^3/y/per cap
Child Mortal-5	7.00	28	95	265.00	$^0/_{00}$
School G/B	0.63	0.91	0.98	1.02	[0.63,1.02]
Particip to IEA	0.30	0.40	0.70	0.90	0-1
GDP per cap	0.58	2.19	6.75	24.03	thousand $per cap
% Irrigation	0.00	0.7	8	99	%
ODA WSS	0.00	0.35	1.85	16.51	$ per cap
ODA	0.60	16	45	289.27	$ per cap
Malaria	0.00	1.81	60	428.14	$^0/_{00}$
Water Use int Agri	0.50	37	1070	28333	m^3/ha/yr
Precipitat	51	716	1649	3141	mm/y
Desert risk	0	0	26	100	%
Total withdrawal	6.4	127.8	500	2161	m^3/cap/yr
WGI.V.A.	-2.14	-0.53	na	1.17	[-2.5,+2.5]
WGI.PS.AV	-2.46	-0.65	na	1.10	[-2.5,+2.5]
WGI.RofL	-1.82	-0.32	na	1.20	[-2.5,+2.5]
CEC	-2.70	-0.69	0.8	2.78	na
HDP	-4.56	-0.803	2.14	5.04	na
ODA CI	-2.35	-0.809	0.843	3.01	na
AP	-4.39	-0.959	1.904	3.23	na
WR	-2.43	-0.584	0.880	2.87	na

Table 6.1: Limits of the variables categorized in 3 categories: high, medium and low and are expressed in the original scale except for Composite indicators HDP, AP, WR, CEC,ODA CI.

2.1 Designing the DAGs

Global models

The global model is based on composite indicators compiled from the previously defined dimensions of variables: Human Development-Poverty (HDP), Water Resources (WR), Human activity pressure on WR (AP), Official Development Assistance (ODA CI) and Country Environmental Concern (CEC). These different nodes are linked with directed edges by order according to correlations observed previously (Appendix D, section 1) and according to our purpose. In fact, the composite indexes computed qualify complex phenomena and the relationships between them can be in both directions. They can reflect some kind of chicken and egg situations. The WSS access fosters for instance, the attendance of youth people at school (Dreibelbis et al., 2012) that will permit their better access to employment. The economic growth favours the rise of households' income that then allows the population to pay for WSS services (provided that the wealth distribution is relatively fair). The WSS access increases the rural productivity (Kiendrebeogo, 2012) while the implementation of irrigation scheme supports water supply access for the rural population. This is clear that both-direction dynamics are valid and can exist depending on the country specific context. Therefore, the hedge directions are set from WSS to HDP according to the research goals (Fig 6.2a).

Thematic models

The links between CI and their sub-variables are modelled according to two options (Fig 6.2b):

- a simple structure where hedges are set from the CI to each Sub-variable
- a non-naive structure where sub-variable is linked via an intermediate variable to the CI

The WR is set as a parent node to AP composite index. The reason for this structure is the very low and non-significant correlations between WR variables and WSS variables, hence, AP set as intermediate variable (Chapter 5, section 3.2).

Financial assistance flow (ODA CI integrating variables ODA, ODA WSS and WGI PS AV) is included in the models according to the information given in the previous PCA analysis: i) ODA shows higher positive correlations with poverty variables than with development variables. It also shows a negative correlation with WS (-0.450) and S (-0.412); ii) ODA WSS is correlated with the political stability and absence of violence index (0.363 correlation - WGI PSAV,) and iii) WGI PS AV is positively correlated with other governance variables such as the indicator measuring voice and accountability within a country (0.524 correlation -WGI VA). Therefore, ODA CI is linked with HDP, CEC and WSS (Fig 6.2). The relationships between ODA, ODA WSS, WGI VA and WGI PS AV variables is explained when presenting the multivariate analyses performed (Chapter 5, section 3.4).

ODA WSS shows non-significant correlations with WSS (-0.098 and 0.004 correlation respectively). (Botting et al., 2010) observed the positive association between sanitation and children mortality but obtained no significant correlation between the percentage change of sanitation access and ODA-WSS input. They suggested that “this may be due to ineffectiveness in investments, a weak capacity of the mandated national institutions, or perhaps due to success on behalf of local, non-internationally funded efforts”. WSS disbursements were 8% of the total ODA in 2004-2005 (OECD, 2007). The low level of communality and low KMO may also explain the low linearity of the ODA WSS (Appendix D, section 4).

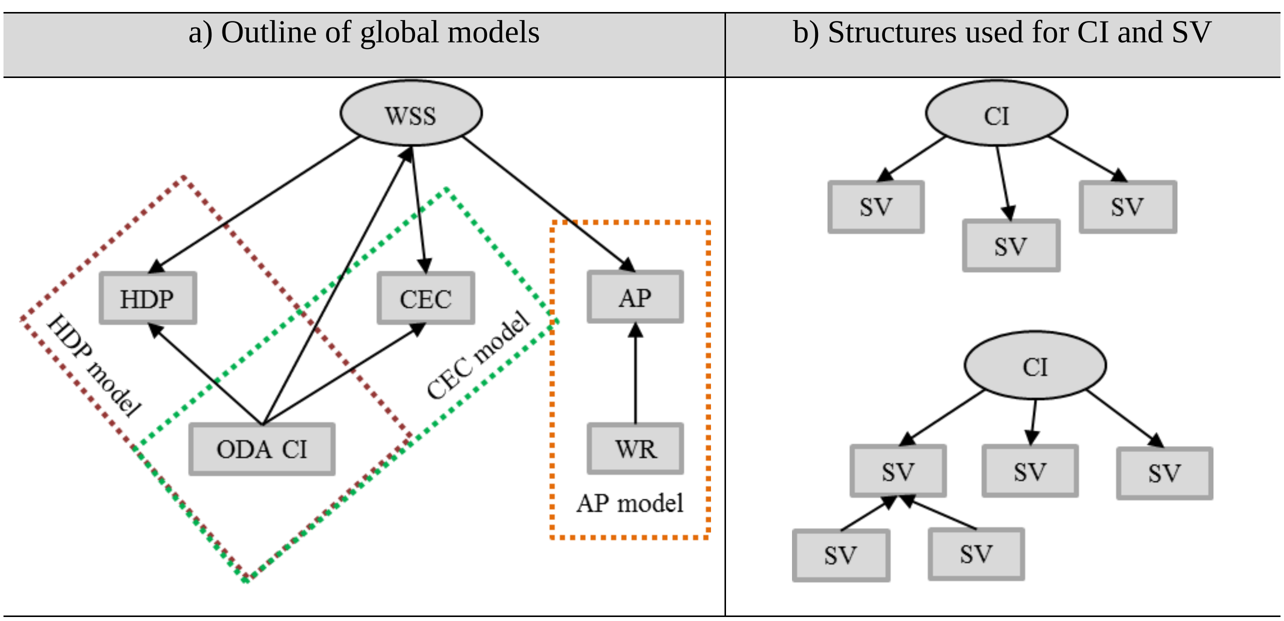


Figure 6.2: Outline of the models structure. In figure a), the DAG global model is in grey and the coloured rectangles delimit the thematic sub-models. CI stands for Composite Indictors and SV for Sub-Variable. In figure b) presents the two structures used to model the links between CI and SV.

2.2 Models performance and sensitivity analyses

The results of the statistical tests in Table 6.2 summarize the performance of each model detailed in sections 3 and 4 of this chapter. Statistical validation is carried out[59] and measured using the Error Rate, Logarithmic Loss (LL), Quadratic Loss (QL) and Spherical Payoff (SP) coefficients (Pearl, 1978). The error rate represents the percentage of cases incorrectly classified—the lower the value, the better the predictive validity of the model. Logarithmic loss (LL) ranges between zero and infinity (where zero is the best fit). The quadratic loss (QL) ranges between zero and two (zero is the best fit). The spherical payoff (SP) varies between zero and one (one is the best fit).

Model	Error rate	LL*	QL*	SP*
a) WS general model	11.88%	0.2814	0.1673	0.9083
b) S general model	9.901%	0.2921	0.1622	0.9143
c) HDP model	4.95%	0.168	0.09021	0.9517
d) CEC model	19.80%	0.4931	0.2967	0.8353
e) AP model	11.88%	0.2688	0.157	0.9129

*LL = Logarithmic Loss, and QL = Quadratic Loss where best value is 0, SP = Spherical Payoff where best value is 1.

Table 6.2: Performance and Error Analysis for WS, S, HDP, CEC and AP models

A sensitivity analysis is carried out to evaluate the relative impact of each variable in the model. Entropy reduction (ER) (or mutual information) refers to the expected reduction in the query variable (WSS) due to a finding for any other variable of the model (Pearl, 1991). Its value varies between zero (meaning complete independence between the query and the instantiated variable) and the entropy value of the query without any evidence in the model. The variance of node belief (BV) and the Root Mean Square (RMS) change of belief are also computed (Neapolitan, 1990). Both statistics range from zero to one (the closer the value to zero, the stronger the independence between the query and the instantiation variable). Table 6.3 summarizes the results of the sensitivity analyses for each model described and analysed in the following sections. Note that only directly linked variables with composite indicators or WSS variables are included in

[59] To build and test the models, the Netica Application (Norsys Software Corp.) version 4.16 is used (www.norsys.com)

Table 6.3. WR, Withdrawal municipal and industrial are not included in the analysis because indirectly linked to WSS and parent nodes of AP and Total withdrawals, respectively.

The sensitivity analysis measures the influence (probability change) of ad dependent/child node. Therefore, it provides the key variables for WS and S improvement according to set parent nodes.

Models	Variables	ER	BV	RMS	rank
a)WS global model	HDP	0.36475	0.1136945	0.3372	1
	CEC	0.06549	0.0221944	0.149	2
	AP	0.05969	0.0202661	0.1424	3
	ODA CI	0.01701	0.005855	0.07652	4
	WR	na	Na	Na	5
b)S global model	HDP	0.3201	0.099637	0.3157	1
	AP	0.12310	0.0394685	0.1987	2
	CEC	0.01982	0.006738	0.08208	3
	ODA CI	0.01258	0.004285	0.06546	4
	WR	na	Na	Na	na
c)HDP model	GDP_per_cap	0.75752	0.197325	1	
	Child Mortal-5	0.52090	0.110159	2	
	% Slums	0.51033	0.098481	3	
	School enrol	0.44786	0.059585	4	
	Malaria	0.43765	0.077565	5	
	% Poverty	0.38850	0.050855	6	
	Urban Pop	0.34604	0.067429	7	
	School G/B	0.28987	0.068078	8	
	WGI RofL	0.26849	0.035107	9	
	Femal eco	0.18374	0.038455	10	
	ODA CI	0.00195	0.000416	11	
d)CEC model	Particip	0.41790	0.083801	1	
	WGI VA	0.37631	0.048051	2	
	ODA CI	0.03400	0.004356	3	
e)AP model	% Irrigation	0.68640	0.161407	1	
	Water Use Int Agri	0.52215	0.090917	2	
	Total withdrawals	0.07318	0.011538	3	
	WR	0.03537	0.005556	4	

*ER = Entropy reduction, BV = Belief Variance, RMS = Root Mean Square change

Table 6.3: Sensitivity analysis results by model.

The sensitivity analysis main outcomes extend the linear regression analysis (Chapter 5, section 5). As expected, the human development variables are ranking first: by order GDP per capita, children mortality rate, the percentage of slums, for both WS and S variables. The AP variable switches ranks with CEC, ranking 2nd in Sanitation analysis and 3rd in Water Supply's one. The governance also plays a role that is more transversal or/ multi-dimensional influencing WSS access in several ways, by order, through advanced governance (WGI RofL), democracy/civil freedom (WGI VA) and political stability (WGI PSAV). It is worth noting that ODA CI is little correlated with the WSS variables coming last in the sensitivity analysis.

The description of the developed models and the following simulations allow measuring the potential underlying processes. After developing the BNs, several simulations were run to measure the probabilistic variations of key variables. The interpretation of these results constitutes a proposition on the potential underlying processes introduced in the following paragraphs. The suggestions are mainly formulated based on the simulation, but also takes into account the relevant scientific literature and knowledge reported from the field.

3. GENERAL MODELS: STATISTICAL VALIDATION AND SCENARIOS

3.1 Water Supply (WS) general model

The general model classifies 86% of the WatSan4Dev sub-set correctly (Table 6.4). Figure 6.3 presents a DAG of this model.

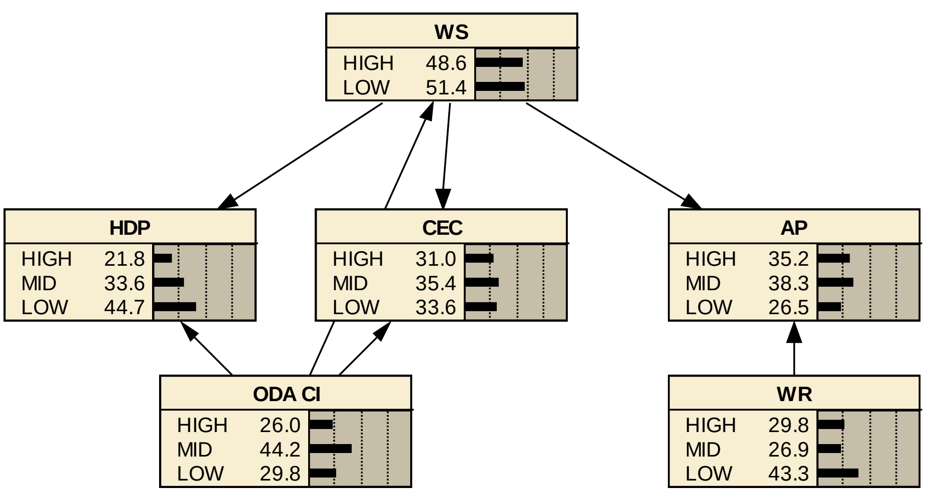


Figure 6.3: WS general model; DAG shows the relationships between the WS and the dimensions and initial probabilities computed.

Table 6.4 presents the initial probabilities (IP) of each variable of the model for a specific level. For each simulation, a specific variable is set at 100% probability for one category. The model then computes the new probability distributions of the other variables based on this change. The results of these simulations are reported in Table 6.4. The positive increments of variables (recomputed probability minus initial probability) are represented in green and the negative increments in red. Simulations are ordered according to WS variation (HDP implies the highest change, ODA-CI the lowest).

Simulation 1 (HDP set to 100%) estimates the level of improvement to WS needed to support poverty reduction efforts. Indeed, in keeping with the Millennium Goals initiative, efforts should be halving poverty worldwide through improvements to WS access (halving the population without access to improved Water Supply and basic Sanitation-target 7C). Simulation 2 (CEC set to 100%) estimates the influence of WS access and WS management on the environmental concern indicator. Simulation 3 examines the type and level of pressure of human activity (water demand according to sectors) on resources in light of WSS development. Simulation 4 (ODA CI set to 100%) looks at the influence of high ODA CI mobilization on the national WSS access.

Variables	Levels	Initial Probabilities	Simulation 1 HDP	Simulation 2 CEC	Simulation 3 AP	Simulation 4 ODA CI
HDP	HIGH	21.8%	100	6.9	5.5	0.7
ODA CI	LOW	29.8%	1	2.1	2.3	100
CEC	HIGH	31.0%	10.5	100	5.1	3.1
AP	HIGH	34.4%	8.8	5.4	100	2.6
WR	HIGH	29.8%	(=)	(=)	(=)	(=)
WS	HIGH	48.6%	84.5%	70.3%	66.3%	59.4 %

Table 6.4: General model – Water Supply WS simulations: Positive variations of probabilities are in green.

Simulation WS1 (HDP set to 100%): pushing poverty reduction to foster development

In this simulation, increasing HDP calls for major efforts in poverty reduction. High access to WS needs to reach almost 85%, thus significant improvements to WS access are needed when compared to IP. Two main explanations on this relationship can be suggested:

- Safe access to a source of drinking water close to consumers is a primary basic service that enables health and the social and ultimately economic development of a country. Indeed, the benefice of WS on health is well known in terms of water born disease or children mortality (El-Fadel, 2013), but its positive effect goes far beyond this in contributing to the creation of a favourable context for development (see Chapter 6, section 4.1 for details).

- Governments engage in WSS programs as an important component of their poverty reduction strategies (to increase their HDP). In conjunction with state efforts, international donors finance programs to fight against poverty and mobilize resources to improve WSS services. In 2004, DAC donors committed 3 billion and disbursed around 2.1 billion in ODA for WSS improvement (OECD, 2007).

The current model estimates the contribution of WS to HDP, but there are multiple relationships between HDP and WS. In the context of economic development, households tend to dedicate financial resources to improving their WS access when this is possible. A better education is a long term effect that contributes to increase in income (Houthakker, 1959), making living conditions (including WS) a priority. Note that in cases of unequal distribution of wealth, these positive impacts can a priori be undermined. An indicator of inequality could not be integrated into these models (Chapter 4, section 4). Indeed, the Gini index that measures inequality of income at a national level had too many missing values when considering developing countries. The advanced governance included in HDP (control of corruption, government efficiency, policy making capacities) is definitely favourable to WSS sustainability: “The strength and quality of government institutions and a strong policy environment are just as important to achieving success in the environment sector as they are to other sectors” (Burtaine & Parks, 2013).

Simulation WS2 (CEC set to 100%): WS management fostering environmental concern and the role of civil society

In this simulation, the increment of CEC implies an increment of WS up to 70.3%. Complementary, high WS (100% probability) is estimated to push high CEC to 45% probability (+14%). CEC expresses the country’s commitment to major environmental issues (Particip). The degree of civil society's freedom of expression and the level of accountability of government (WGI VA) are also included. The bigger the increment of CEC, the more the country is committed to its environment and its citizens as observed in multivariate analyses (Chapter 5, section 3.5). Based on this model, the expansion of the access to water supply reinforces the environmental commitment and is associated with a strong civil society. Participative management of the water services and of water resources is spreading since the formalisation of the IWRM[60] that directly reinforces CEC. This approach aims at organising the management of WSS services through the involvement of all stakeholders, for instance, with the development of water user associations and/or using consultation processes (Schouten & Moriarty, 2003). This multi-stakeholders approach can be implemented at a local scale, at river basin and at sector scale. South Africa has engaged its whole water sector in the development of stakeholder platforms for collective/negotiated water management as part of its strategy to expand WSS coverage (Du Toit & Pollard, 2008). The Water sector can act as a starting point calling citizens to organize themselves to participate in decision making and management processes beyond their warning roles regarding pollution or WSS services interruption.

Simulation WS3 (AP set to 100%): WS related to the development of human activities

According to the model, the increment of activity pressure on WR (AP) resulted in an increment of WS access of up to 66%. The AP composite index should be interpreted as follows: the bigger the increment of the AP factor, the more agriculture monopolizes the available water resources, often through irrigation practices. It is important to remember that AP behaviour occurs in countries where agriculture often represents an important economic activity. The municipal water supply network (referred to as municipal demand by FAO) is still used to draw a small amount of water in the areas considered. Agriculture is the main consumer of drawn water resources (consuming over 50% of water resources) in 80% of the selected countries. Low values indicate that a significant part—although still not the majority expect in very few cases—of

[60] Principle n.2 of IWRM, formalised in DUBLIN Statement 1992 : http://www.wmo.int/pages/prog/hwrp/documents/english/icwedece.html#principles, last access the 11th October 2013

absolute limited water consumption is linked to industrial activities including raw material extraction (implying a specific water supply) and/or domestic/municipal uses (served by network supply).

This behaviour indicates that agriculture is positively associated with water supply coverage, creating a positive synergy between these sectors: 1) in cases of irrigation agriculture, water supply infrastructure can serve a multi-purpose logic, providing water to cover population, agriculture and often energy needs (Chapter 5, section 3.2); 2) a strong water supply sector ensures services to the population but the existence of such sector (capacities, market organisation…) can be beneficial to agriculture (Kiendrebeogo, 2012) and in particular to irrigation plants. However, irrigation practices (over 20% of irrigated areas) were the exclusive concern of countries in South East Asia, the Middle East and Egypt. Latin America and Africa are largely unconcerned with irrigation.

Simulation WS4 (ODA CI set to 100%): the multiple relationships between ODA CI, and HDP, WS, CEC.

The first group: ODA CI, HDP and WS

In this simulation, high donor investments (ODA CI) aimed at supporting partner countries indicate a low-medium HDP situation (78% probability) and, therefore, there is a 61% probability that access to WS is poor. This complies with the purpose of ODA: providing financial support to less advanced countries (low HDP) with poor access to WS. In fact, the model shows that countries having higher levels of WSS need less ODA for development.

In 2004, investments in social infrastructures and services (government and civil society support, education, and health issues) represented 30% of the total ODA provided by all DAC donors and the EU (OECD statistics in 2004[61]). Humanitarian emergency aid is excluded from ODA. The model shows that poor countries (with low HDP) had a 45% probability of having high levels of ODA. In general, the lower the HDP of the country is, the higher their ODA CI is. However, the

[61] OECD: http://stats.oecd.org/qwids/, last access 15th February 2013.

model estimates that 21% of the countries that benefit from relatively high ODA CI still show high HDP. The Chapter 5, section 3.4 highlights the main motivations attributed to ODA allocation and controversies over donors' visions and strategies of cooperation (Carbonnier, 2010). In addition, excessive dependence on external financial support can have negative effects on the quality of governance, creating "corruption in government, bureaucratic quality, and the rule of law" (Knack, 1999)—all of which are elements included in the HDP variable. The ways in which external assistance is delivered may further explain the observed behaviour of ODA CI, HDP and WS. For instance, donor fragmentation combined with a low capacity to fund implementation can favour corrupt practices.

<u>The second group: ODA CI, WS and CEC</u>

The model estimates that the increment of ODA CI (also including political stability, WGI PS AV indicator) induces an increment of high CEC (also including civil society voice, WGI VA indicator) from 31% to 34.3%, with the majority of countries (43.9%) showing a medium CEC level.

Several hypotheses can be formulated to explain this link:

- Programs/projects supported by external assistance promote citizen involvement in the management of WSS services because the IWRM participatory approach drives in that direction.
- Minimum political stability (WGI PS AV) and civil society voice (WGI VA) facilitate efficient use of the funds and sustainable infrastructure implementation. In fact, cooperation has shown that external assistance is facilitated with political stability, democratization and, more generally, with "good governance" (Santiso, 2001) (Neumayer, 2003b). A good governance context facilitates the mobilization of funds, while poor governance creates risks.

These two hypotheses can coexist depending on the governance context and the donors' strategy in a specific country. Donors may deliver assistance despite the risks involved in some countries based on a long term or geo-political strategy. The need to support less advanced countries can also prevail over high assistance effectiveness expectations.

3.2 Sanitation (S) general model

The S model follows the same structure as WS (Fig 6.4). The model performance is good with 10% error rate (Table 6.2b).

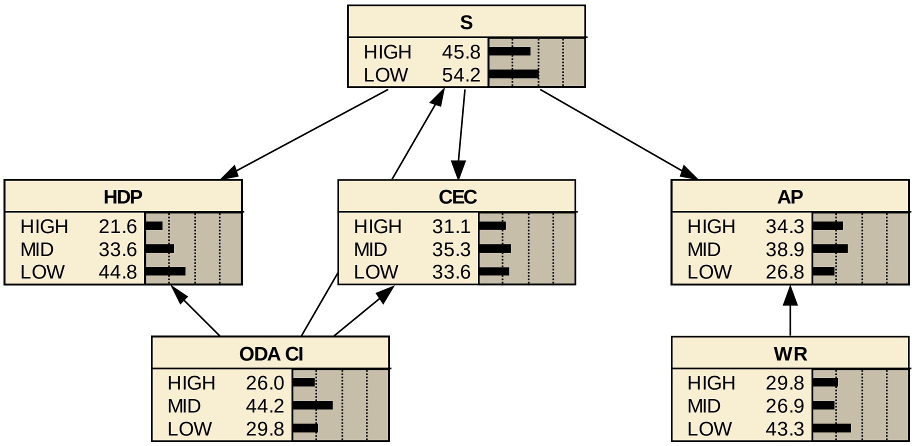


Figure 6.4: S general model, DAG shows the relationships between the S and dimensions and initial probabilities computed.

Table 6.5 includes the computed probabilities of the different simulations made.

Variables	Levels	Initial Probabilities	Simulation 1 HDP	Simulation 2 AP	Simulation 3 CEC	Simulation 4 ODA CI
HDP	HIGH	21.6%	100	6.3	1.4	0.7
ODA CI	LOW	29.8%	1	1.8	4.9	100
CEC	HIGH	31.1%	1.9	2.1	100	5.1
AP	HIGH	34.2%	9.9	100	2.3	2.1
WR	HIGH	29.8%	(=)	(=)	(=)	(=)
S	HIGH	45.8%	80.3%	66.7%	53.9%	53.1%

Table 6.5: General model - Sanitation S simulations. The positive variations of probabilities are in green.

S is mainly sensitive to HDP (Table 6.3b). The development of sanitation services helps improving living conditions and is a key element to the formalization of slums and the improvement of health and education (Chapter 5, section 5). ODA CI is mainly (44.5%) delivered in countries with low HDP and low S.

The main difference with WS's model lies in the relatively lower sensitivity of Sanitation to CEC which drops of one rank for the benefit of AP (Table 6.3). In simulation 2 and Table 6.5, the probability of high S that reaches 66.7% probability in case of high AP, is similar to the high WS probability (66.3% probability - Table 6.4). Therefore, the commitment to the environment and the civil society organisation has less impact on the expansion of sanitation services. High S (set to 100% probability) is estimated to push high CEC to 36% probability (+7%) while this probability increases up to 45% in WS model. Sanitation is a rather neglected topic associated to the private and family sphere. Therefore, ensuring "sanitation for all" has not been considered as a higher public priority than the access to drinking water (Chapter 6, section 6 details some of the cultural barriers linked to sanitation). The commitment to sanitation is recent and has followed international awareness campaigns. For instance, the eThekwini Declaration on sanitation was signed in 2008 by 19 African states (African Union, 2008). Bangladesh and India are two leader countries in developing pedagogic/cultural approach around sanitation and community-based management[62] (WSSCC-B, 2009), (Chambers, 2009). The latter approach aims at making sanitation a collective issue that improves the environment and health of the whole community. The paradox is that sanitation has a significant effect on the environment, in particular in impacting water resources. The water contamination by sanitation source is multiple: leakages from pipe or non-sealed latrine, direct open defecation, incomplete wastewater treatment or at household level poor hygiene behaviours (Addo et al., 2013).

High AP (100% probability) supposes a similar probabilities of high S and WS access (66%). This indicates that there is a synergy between agricultural sector and WSS sector which foster sanitation services. This may be due to the organisation of the suppliers, the market and the rural community necessary to ensure the production that benefit to the sanitation sector. In addition, Agriculture and Sanitation can also develop a direct synergy like, for instance, the reuse of human excreta for soil fertilization. This opportunity to add value to sanitation/wastewater products via their use in agriculture is available both in rural and urban areas. It is easier in rural areas, where sanitation products can be collected and transformed into fertilizer at an affordable

[62] Community-Led Total Sanitation Program http://www.communityledtotalsanitation.org/, last access on the 24th August 2014.

price using composting toilets[63] or solar drying processes. In urban areas, wastewater sludge can, if treated correctly, be reused for agriculture. However, it is difficult to estimate how effective these practices could be because there is very little data available regarding the reuse of sanitation or wastewater products in developing countries. (Raschid-Sally & Jayakody, 2008) studied wastewater reuse in agriculture in 53 cities across the world.

4. SPECIFIC MODELS FOR THEMATIC DIMENSIONS

General models provided the core idea about the relationships between the thematic dimensions and WSS and specific models were used to refine this analysis. In this section, the analysis of relationships is scaled down and sub-models for each dimension are described.

The specific models' DAGs are set with the composite indicator (HDP, CEC or ODA CI) acting as a parent node to its sub-variable nodes. As a consequence, the variation change in the sub variables is estimated according to the CI behaviour within the global model. This allows the association between sub variables and the corresponding CI to be measured. Therefore, the link between CI and its sub variables is a matter of the strength of the relationship rather than a conditional link between two variables. An indirect link is created between WSS variables and sub variables with the CI as intermediate variable. The ODA CI is included in the thematic models to respect the structure of the global model. This also gives insights on ODA CI relationships with HDP, CEC and AP sub-variables. The following sections summarize the sensitivity analysis results for the thematic models (described in Table 6.3) and details the simulations run.

Note that the order of variation can differ between the sensitivity analyses and the simulation tables. The sensitivity analyses report the cumulated variations across all the categories (High, Mid, Low). The simulations table reports the probability variations for a specific category.

[63] See composting toilet fiche: http://www.sswm.info/category/implementation-tools/wastewater-treatment/hardware/site-storage-and-treatments/composting-t, last access 04th July 2013.

4.1. Human Development and Poverty (HDP) model

The statistical analyses establish a model with direct links between HDP, the variables in the HDP dimension and the ODA CI dimension. The suggested model explains up to 95% of the variability (Table 6.2c). Figure 6.5 shows the structure of the HDP model and the initial probabilities calculated for all variables.

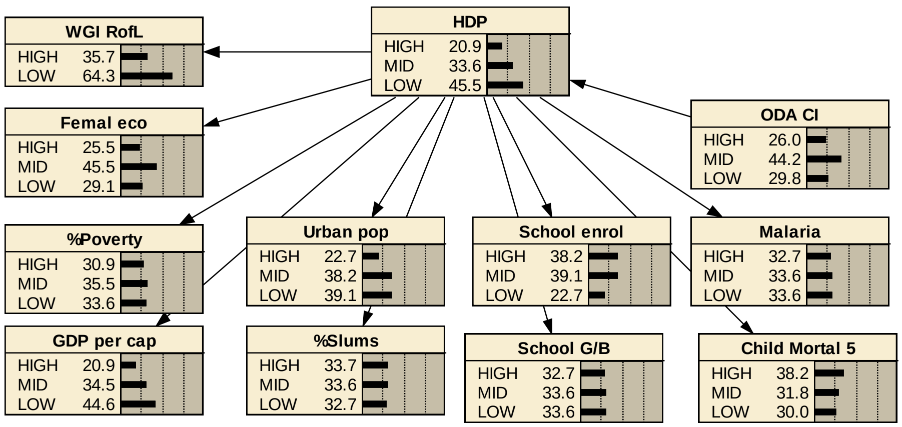


Figure 6.5: Probabilities in % for the variables involved in the HDP Dimension discretized in High (HIGH), Medium (MID) and Low (LOW) levels.

The three main variables influencing HDP are income per capita, child mortality under 5 years (fighting efficiently against mortality) and the proportion of slums (pushing the organization of urbanization) (Table 6.3c). These are the variables that authorities and/or donors should focus their efforts on in order to foster development. Two other important aspects are the support given to school enrolment and the control of malaria.

Table 6.6 sums up the probabilities computed based on several scenarios presented in this research. Simulation 1 estimates the HDP level when the level of child mortality under 5 years has been significantly reduced. Simulation 2 estimates the improvement of HDP necessary to support the reduction of informal urbanization (% slums). Simulation 3 observes how the improvement in human development is translated in terms of malaria control. Simulation 4 observes the benefits of good development on governance conditions. Simulation 5 examines the association between human development and education (from primary level to university).

Finally, simulation 6 examines the role of gender through girls' primary education (School G/B) and female economic participation (Femal eco) variables.

Variables	Category	Initial Probabilities	Simulation 1 Child Mortal -5	Simulation 2 % slums	Simulation 3 Malaria	Simulation 4 WGI RofL	Simulation 5 School enrol	Simulation 6 School G/B
School enrol	HIGH	38.2%	30.7	30.6	29.7	7.4	100	17.5
School G/B	HIGH	32.7%	16.4	16.1	15.2	12.6	15.1	100
Malaria	LOW	33.6%	27.5	27.2	100	21.3	26.1	15.5
%slums	LOW	32.7%	28.6	100	26.4	21.9	26.1	16
%Poverty	LOW	33.6%	30.1	29.5	27.7	23	27.2	16.8
Child Mortal-5	LOW	30%	100	26.2	24.5	20.4	23.8	14.9
Femal eco	LOW	29.1%	15.1	15	14.4	11.7	14.8	8.6
Urban pop	HIGH	22.7%	22.8	22	20	17.1	18.7	12.4
GDP per cap	HIGH	20.9%	23.8	22.7	20.1	17.7	18.2	12.8
WGI RofL	HIGH	35.7%	27	26.5	25	100	24.7	12.4
ODA CI	LOW	29.8%	1.1	1.1	1.2	0.9	1.4	0.7
HDP	HIGH	20.9%	54.6%	52.8%	48.7%	45.2%	45.3%	38.9%

T**able 6.6**: Thematic model - HDP Simulations. The positive variations of probabilities are in green.

Simulation HDP1 (low Child Mortal-5% set to 100%): WSS is essential to reduce children mortality

If child mortality (Child Mortal-5) is reduced through a focus on basic health care, then according to the model HDP would be high, with a 54.6% probability. Fighting child mortality requires better access to safe drinking water together with basic health care and quality nutrition. Improved sanitation considerably reduces human/water contamination by dissemination of pathogenic material. WSS access considerably reduces the incidence of waterborne diseases such as diarrhoea and consequently the risk of mortality, particularly in young children. Pneumonia and diarrhoea are the two greatest causes of child mortality in developing countries—diarrhoeal disease kills 1.5 million children every year (UNICEF & WHO, 2009). South East Asia and Africa are the regions most affected by diarrhoea and malaria.

Simulation HDP2 (low% slums set to 100%): the role of urbanization in development

The type of urbanization taking place in a country is important and depends on the demographic and migration dynamics, as already mentioned in Chapter 5, sections 2 and 5. A reduction of slums (low% slums set to 100% probability) implies major efforts to control and organize urbanization processes; in areas where this is taking place, high HDP probability is estimated to rise up to almost 53%. Organizing and implementing basic services in these suburbs/districts is an obvious priority. The model suggested in this work estimates that in 2004 almost 40% of developing countries presented with low urban population rates while the level of the population living in slums was high in 34% of cases. The availability of WSS services significantly improves living conditions—the model estimates a reduction of poor HDP by 34 points of percentage in case of 100% high WS access and, by 30 points of percentage in case of 100% high S access. The consequences of a lack of sanitation are amplified in terms of disease spreading and water contamination by population density. The limitations of slums (from high to medium level) are translated in the model by the significant reduction of the probability that countries will have a low level of HDP from 91% to 38%. This influences essential variables such as school enrolment (Dreibelbis et al., 2013), household income (GDP per cap), the child mortality rate and malaria prevalence.

Simulation HDP3 (low Malaria set to 100%): WSS improvement limits malaria prevalence

The reduction of malaria prevalence (Malaria) increases HDP by up to 48.7%. According to the CDC[64], the majority of the countries included in the WatSan4Dev database are affected by malaria in either all or part of their territory. Among other socio-economic factors, WSS services availability also contributes to a nation’s ability to reduce malaria prevalence (HDP as connecting variable). The lack of a water supply infrastructure leads to outdoor water storage, which encourages mosquitoes and thus the dissemination of malaria (Lewis, 2011). Limiting inappropriate storage is essential, and can be avoided, for instance, through the use of piped water and a rainwater harvesting infrastructure.

[64]CDC malaria map: http://cdc-malaria.ncsa.uiuc.edu/, last access 04[Th] July 2013

Malaria prevalence may also be reduced as a proxy of good, efficient sanitation. The existence of sanitation facilities reduces the development of mosquitoes by reducing the conditions favourable to their growth. Improving sanitation, and by extension wastewater collection, reduces the prevalence of stagnant and potentially pathogenic water areas. Indeed, in the case of pipe infrastructure, "when population growth outpaces the existing infrastructure, wastewater treatment systems are unable to cope with the influx, garbage and sanitation facilities cannot contain the increased refuse, and access to clean, treated drinking water may not be available. All of these conditions contribute to create the perfect environment for disease causing mosquitoes to breed" (Lewis, 2011).

Simulation HDP4 (WGI RofL set to 100%): Advanced governance and HDP relationships

In the highly advanced governance (WGI RofL) simulation HDP increased to 45.2%, more than doubling the initial probability (20.9%). In fact, the ability of a nation to define rules and laws, to organize public services, to be committed to the implementation of suitable policies and to fight corruption (as represented by the WGI RofL) is positively associated with its development. This positive association is a chicken-egg situation. Corrupt practices together with weak institutional authorities are common in a poverty context. Corrupt practices discourage investment that might be used to extend services to the poor, divert finance from the maintenance of deteriorating infrastructure, and take cash from the pockets of the poor to pay escalating costs and bribes for drinking water (TI, 2008); thus undermine WSS services delivery (Wolf, 2007) Conversely, country development is boosted by good governance.

This tells us that the development and maintenance of WSS services helps build up a nation's capacities through the setup of participative/efficient management rules—at least at a local level. Good governance supports sustainable WSS delivery (TI, 2008), (Witford et al., 2010), while the interest and participation of users encourages better governance and transparency.

Simulation HDP5 and HDP 6: WSS role in education and gender issue (school enrolment, ratio girls to boys at school and female economic activity analysis).

In the HDP 5 simulation, school enrolment (School enrol) is fostered, and HDP is enhanced to the point of reaching up to 45.3% probability. In the HDP 6 simulation, the improvement of the girls' schooling rate (School G/B) shows a positive correlation with HDP, which reaches up to 38.9%. A medium/high HDP (high HDP with 37.6% probability and Mid HDP with 43.7% probability) indicates a limited female economic participation (low Femal eco) (Boserup, 1989). Beyond HDP improvement, several explanations on the relationships of these factors with WSS are possible:

- Water availability in households is an important factor in the enrolment, attendance and drop-out rates of children (Dreibelbis et al., 2013). When water access was improved in communities, school enrolment rates increased in Tanzania, India and Bangladesh by up to 15% (UNDP & UNICEF, 2006). Back-to-school campaigns in countries in transition are especially dependent on water supplies: "Teachers are hesitant to relocate to communities without a reliable and safe source of water" (UNDP & UNICEF, 2006).
- Girls in the developing world are generally in charge of the daily water fetching, setting time barriers to school attendance. Separate sanitation facilitates the attendance of girls, particularly after they reach puberty, thanks to privacy and security advantages. Despite these advantages, "in some countries in Africa and Asia as few as 10 per cent of schools have adequate and separate sanitation facilities, while student-to-latrine ratios can be as high as 150:1" (UNDP & UNICEF, 2006).
- In poverty context, women are massively active in the economy so as to sustain their families—translated into statistics, this reads as a high female economic activity rate (Chapter 4, section 3.1). Women in most developing countries handle both domestic work, including water fetching, and an economic activity: "Providing water sources frees up time for women and leads to both direct and indirect opportunities to engage in activities which provide an income" (WSSCC, 2006). Therefore, the level of female activity may remain more or less the same with the access to WSS services, but women's daily lives are qualitatively improved.

4.2 Country Environmental Concern (CEC) model

The performance of the CEC model is higher than 80% (Table 6.2d). Figure 6.6 shows the initial probabilities of the variables in the CEC model following the same methodology as for HDP.

The World Governance Voice and Accountability index (WGI VA) only presents two levels (High and Low), following the structure of the variable; its range is [-2.5, 2.5], where 0 can be considered the limit between poor and good governance.

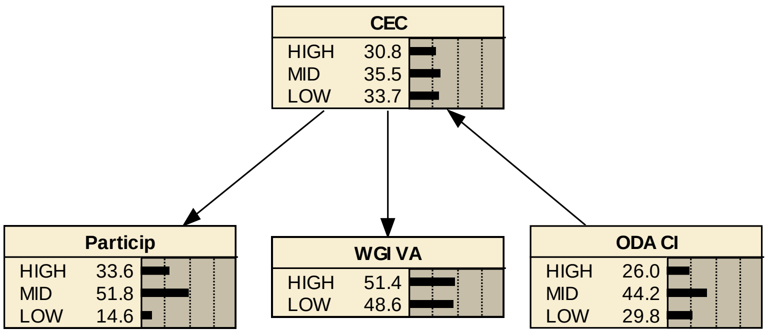


Figure 6.6: Probabilities in % for the variables involved in the CEC Dimension categorized in High (HIGH), Medium (MID) and Low (LOW) levels.

Table 6.3d illustrates CEC sensitivity to sub-variables in decreasing order: participation in international environmental agreement (Particip), government accountability and citizen freedom of expression (WGI VA) and finally the ODA CI (including ODA, ODA WSS and WGI PS AV). Simulation 1 analyses the country's commitment to international environmental concerns. Simulation 2 estimates the influence of citizens' level of freedom of expression on environmental issues and, more widely, on democratic conditions and government accountability. Simulation 3 considers the relationship of ODA CI with environmental and democratic commitment (Table 6.7).

Variables	Level	Initial probabilities	Simulation 1 Particip	Simulation 2 WGI VA	Simulation 3 ODA CI
Particip	HIGH	33.6%	100	18.47	1.37
WGI VA	HIGH	51.4%	28.3	100	-1.1
ODA CI	LOW	29.8%	1.3	-0.6	100
CEC	HIGH	30.8%	70.1 %	54.4 %	36.2 %

Table 6.7: Thematic model, CEC simulations. Positive variations are in green and negative variations of probabilities are in red.

Simulation CEC1 (Particip set to 100%): commitment towards environment includes WSS management

In this simulation, high international environmental commitment of countries (Particip) induces an increase in CEC of up to 70.1%. This indicates national commitment toward environment and environmental services such as water sanitation, energy, and waste. Water and sanitation delivery starts with the implementation of infrastructure, but rapidly calls for integrated and environmental management to ensure its sustainability. The global models estimate that a high formal country commitment (at 100% probability) is associated with high WSS (at 70% for WS and 54% for S).

This positive association between environmental concern and WSS may differ according to the priority given to and perceptions of WS and S. A safe drinking water supply is directly related to human survival; hence it is a priority, while sanitation is often perceived as very much secondary. Sanitation is generally neglected by governments/public institutions, despite it being good leverage for the improvement of living standards, health and the environment (Chapter 6, section 6). In 2012, the sanitation target had still not been reached. This difference is supported by the global models: CEC is more important in the sensitivity analysis of WS than of S (Table 6.3).

Simulation CEC2 (WGI-VA set to 100%): the role of freedom of expression and accountability to citizens in national commitment to the environment

The model estimates that an increase in civil society freedom (WGI VA) translates into an increase of CEC from 30.8% to 54.4%. Citizens being able to express their criticisms and put pressure on their governments according to their interests is key to efficient environmental management.

As mentioned in Chapter 5, section 3.5, the voice of the citizens can both directly and indirectly push the government and other economic actors to sustainably manage the environment. The population is primarily concerned with a lack of water resources, poor waste management, ecosystem degradation and the pollution impacts. Civil society pushes for and plays a major role in the control of sustainable management of the environment if allowed to within their institutional and political context (Barrett & Graddy, 2000), (Neumayer, 2002).

Simulation CEC3 (ODA CI set to 100%): external assistance influences a country's environmental concerns

The relationship between ODA CI and the CEC is not linear. The main trend is that high ODA CI indicates medium (45%) or good (34%) CEC. The model estimates that high financial external flows are less common in contexts where environment concern is poor (present in only 21% of cases). This illustrates international donors' preference to intervene in relatively stable political contexts, supporting local population involvement and governance aspects more generally (Chapter 6, section 3.1, simulation WS4).

4.3 Activity Pressure on Water Resources (AP) model

The structure of the AP model is set differently according to the correlations revealed in the PCA (Chapter 5, section 3.2). Municipal and industrial withdrawal variables are correlated with Total Withdrawal (Fig 6.7). The model reflects a classification performance of higher than 88% (Table 6.2e).

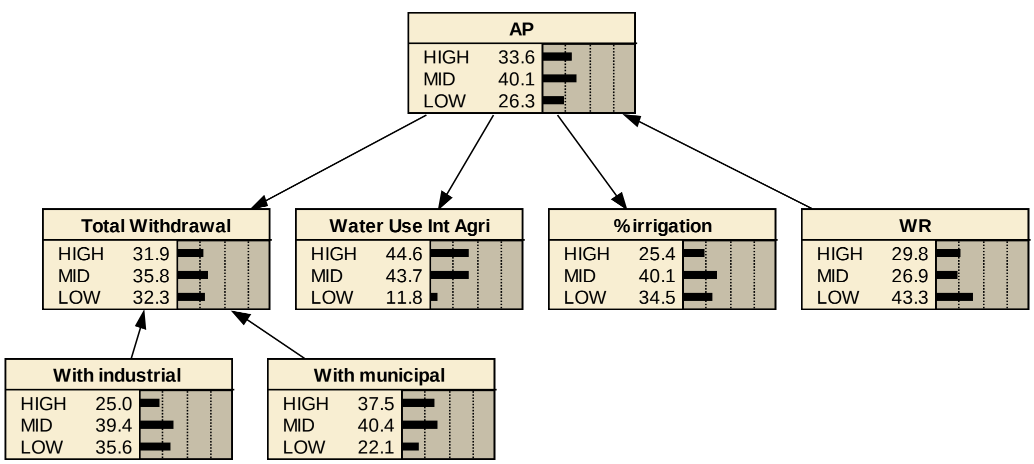


Figure 6.7: DAG and the initial probabilities for the variables involved in the AP Dimension categorized in High (HIGH), Medium (MID) and Low (LOW) levels.

The sensitivity analysis is computed for the four directly linked variables: Water Use Intensity in Agriculture (Water Use Int Agri), proportion of irrigated areas (% Irrigation), the total amount of water withdrawn (Total Withdrawal) and water resources composite index (WR). In keeping with the results in Chapter 5, section 3, the weight of the WR variable is the lowest in the model and therefore, accounts for little in AP (Table 6.3e).

Simulation 1 assesses the weight of intensive agriculture water usage in AP, combining the intensity of Water Usage in Agriculture (Water Use Int Agri) and the proportion of irrigated areas (% Irrigation). Simulation 2 examines the competition between different water uses through their respective withdrawal amounts (total withdrawal, industrial withdrawal, municipal withdrawal). Simulation 3 estimates the effects of a potential reduction of available WR under the pressure of climate change on economic activity (Table 6.8).

Variable	LEVEL	Initial probabilities	Simulation 1	Simulation 2	Simulation3
% irrigation	HIGH	25.5%	100	9.8	-1.3
Water Use Int Agri	HIGH	44.7%	100	11.1	0.9
Total Withdrawal	HIGH	31.9%	13.7	100	0.2
Withdrawal industrial	HIGH	25%	(=)	1.5	(=)
Withdrawal municipal	HIGH	37.5%	(=)	-12.7	(=)
WR	LOW	29.8%	(=)	(=)	100
AP	HIGH	33.7%	98.5%	48.3%	31.9%

Table 6.8: Thematic model - AP simulations, negative variations are in red, the positive variations of probabilities are in green.

Simulation AP1 (% irrigation and Water Use Int Agri set to 100%): WSS sector and the water intensity in agriculture

Intensive agriculture practices are represented in this model by the percentage of irrigated areas (% irrigation) and the amount of water provided per hectare (Water Use Int Agri). In the simulation, an increase in intensive agriculture practices means an increase in AP from 33, 6 % up to 98.5%. The positive association of WSS with AP is mainly responsive to intensive agriculture practices (Chapter 5, section 3.2). The model estimates that High AP with 100% probability means high irrigation practices at 70.3% and high level of water use intensity at 90%.

Looking at the WatSan4Dev subset, only 23% of the countries considered present an irrigation rate of higher than 10%—and these are mainly in South East Asia and the Middle East. These countries show high WS access (but not necessarily high S) associated with irrigation schemes. Even small irrigation plants structure rural areas and thus facilitate the organization of WSS services. For the rest of the developing countries, the exclusive use of water resources for agriculture (high AP) can indicate an economy based on rainfall agriculture with low and weather-driven harvests. In this context, the population is poor and lives in rural areas where WSS access is generally low (low WSS is estimated at around a 33% probability despite high AP). The dissemination of a population throughout rural areas can reinforce barriers to WSS delivery.

Simulation AP2 (Total Withdrawal set to 100%): the analysis of water withdrawals by usage

The AP model estimates that an increase in the total water withdrawal increases AP by up to 48.3%; the municipal withdrawal variable decreases to 24.6% and industrial withdrawal goes up slightly to 26.5%. This means that the industrial variable behaves slightly differently to municipal withdrawal with respect to Total Withdrawal: a certain level of industrial water demand can co-exist with dominant agricultural pressure.

Indeed, the behaviour of the total amount of withdrawals largely depends on agricultural water needs, as confirmed by the AP model. Therefore, the relationship between municipal and industrial withdrawal (as a percentage of the total withdrawal) and the total amount of water withdrawals (amount of water drawn per capita) are negatively correlated (-0.553 and -0.251 respectively). Based on the WatSan4Dev subset, the selected developing countries dedicate the majority of their available water to agricultural purposes according to three main profiles: i) 35% of the countries dedicate water almost exclusively to agriculture (above 85%); ii) 40% of the countries currently dedicate the majority of their water resources to agriculture but this proportion is slowly being taken over by household consumption (corresponding medium level AP); iii) 25% of countries have substantial industrial/mining/forestry activity, i.e. higher than 10%, and this is often combined with high municipal consumption, significantly reducing the share dedicated to agriculture. This is the case in Lesotho, where 40%— 42% in Papua New Guinea— of water withdrawals are dedicated to mining activities.

Simulation AP3 (WR set to 100%): the influence of the availability of water resources on the activity development

According to the AP model, the reduction of water resources due to, for instance, climate change, slightly decreases the probability of high activity pressure (AP) from 33.6% to 31.9% (-5%). At the same time, the probability of low AP still decreases from 26% to 19.1% (-26.5%) that preserves the medium AP category. This suggests that countries can find solutions adapted to mitigate the consequences of declining resources. This can be confirmed by WSS data: arid countries such as those in the Middle East have high WSS access thanks to their economic activities being less dependent on water resources. These countries have funds available to implement infrastructure adapted to water scarcity, such as desalination plants (IFAD, 2009).

Irrigation can be a solution to increasing the efficiency of water use in agriculture. Factors limiting such development are political commitments, policies, investments and agronomic and management skills. The FAO published recommendations encouraging the reinforcement of policies for planning and infrastructure development, taking into account social factors and capacity building for staff and farmers (FAO, 1987). (Kweku Baah Inkoom & Zomanaa Nanguo, 2011) analysed the reasons for the under-utilization of irrigation infrastructures in the upper west region of Ghana. They highlighted the importance of non-environmental factors such as lack of ownership, management and skills. The development of irrigation and the intensive use of water resources are the results of the interaction of very complex processes and factors, even when water scarcity or desertification is the starting point for intensive agriculture.

To summarize the analyses of the AP model, the WSS sector supports agriculture, particularly when it comes to intensive practices. Agriculture participates in rural development and the structuring of these areas can favour the development of WSS. However, this positive effect can be undermined when agriculture monopolizes water resources and/or results in poor economic development (e.g. agriculture does not generate an income beyond that needed to survive).

5. MODELLING COUNTRY PROFILES

In Chapter 5, section 4.2, five country profiles were built around 5 axes that corresponded with the first five PCA components. Profile 1 (on the way to well-being) is considered the most favourable profile, with high values for the WR, HDP and CEC dimensions. Profile 2 (environment commitment and freedom/democracy black spot) has weaknesses in terms of commitment to the environment, which are associated with a low level of accountability and civil society freedoms (CEC). Profile 3 (agricultural economy) presents an economy mainly driven by agricultural activities in a context of abundant natural resources. Profile 4 (essential external support) and profile 5 (primary material consumption) are the less favourable profiles with regards to HDP. However, profile 4 benefits from higher levels of freedom, CEC, political stability and high ODA CI. Profile 5 shows an economy mainly based on natural resource exploitation, often in a context of political instability.

The general models described in section 3 of this chapter are used to model each country profile. The advantage of these models is that they refine the analysis of each country profile, which improves the accuracy of the interpretation of WSS status. They also complete the characterization of each group of countries by providing the probabilistic distribution of each axis. Finally, they allow us to measure the behaviour specific to a profile and to run different scenarios. The statistical performances of country profile models are greater than 79% (Table 6.9).

Model name	Nb. of observations	error rate	LL*	QL*	SP*
WS-profile1	23	4.35%	0.084	0.033	0.98
S-profile1	23	8.7%	0.31	0.18	0.90
WS-profile2	17	5.88%	0.16	0.079	0.96
S-profile2	17	0%	0.054	0.0071	0.998
WS-profile3	19	21.05%	0.42	0.27	0.85
S-profile3	19	10.53%	0.32	0.19	0.91
WS-profile4	24	0%	0.096	0.036	0.96
S-profile4	24	8.33%	0.24	0.14	0.93
WS-profile5	18	16.67%	0.37	0.22	0.88
S-profile5	18	0%	0.034	0.0052	0.998

* LL = Logarithmic Loss, QL= Quadratic Loss where best value is 0, SP = Spherical Payoff where best value is 1.
Table 6.9: Statistical performance of country profile models for WS and S

5.1 Analysing WSS behaviours across profiles

In a general sense, as shown in figure 6.8, WSS show a coarse linear behaviour in WatSan4Dev database that is expressed in the following equation:

$$S = 1.0285WS - 21.82 \quad (3)$$

with adjusted R^2=0.5265.

This linear model only explains 52.65% (R^2) of the data variability, however it does highlight the general gap that exists between S and WS (-21.82 points of percentage) and is often observed in data and in the field (UNICEF & WHO, 2008). Some countries where there is an increased gap between WS and S are: Botswana (where there is a 53% gap between WS and S), Nepal (55%) and Namibia (62%). Conversely, Libya experienced a negative gap between WS and S in 2004 (72% versus 97%, respectively). These latter examples are among those that cannot be explained by the linear model introduced in this paper.

In the top right-hand corner of figure 6.8, countries belonging to profile 1 (91% for WS and 75% for S) and profile 2 (93% for WS and 86% for S) show high average WSS levels. In the bottom left corner, countries from profiles 4 (62% for WS and 35% for S) and 5 (65% for WS and 38% for S) show low average WSS levels.

Countries from profile 3 follow a different WS and S statistical profile (refer to the purple dots on Fig 6.8). WSS values in this case are less homogeneous, showing higher standard deviations: the WS average value is 73% with a 15.2% standard deviation, while S reaches 55% with a 22% standard deviation. Profile 3 is built around the AP dimension, with a predominant role played by agriculture. WSS services have higher or lower levels of development according to the agriculture sector and its organization/dynamics. In this group, irrigation practices are linked to relatively high WS—all countries with a high irrigation level show a high level of access to WS (above 75%). However, several countries without irrigation like Guatemala, Honduras, Nicaragua and Suriname show good access to WS as well. This can be explained by their on-going urbanization, which favours WS access in a context where agriculture is still dominant. In fact, these countries show an urban population rate of around 45-50% and their informal slums are relatively limited (below 45%). This relationship is less clear with regards to sanitation.

The best and the two worth profiles are analysed to a more in-depth level and then modelled in the next two sections. Profile 1 is interesting to analyse as a reference given that it is the most favourable state. Profiles 4 and 5 are also explored and modelled because these countries experience many difficulties on which efforts should be concentrated.

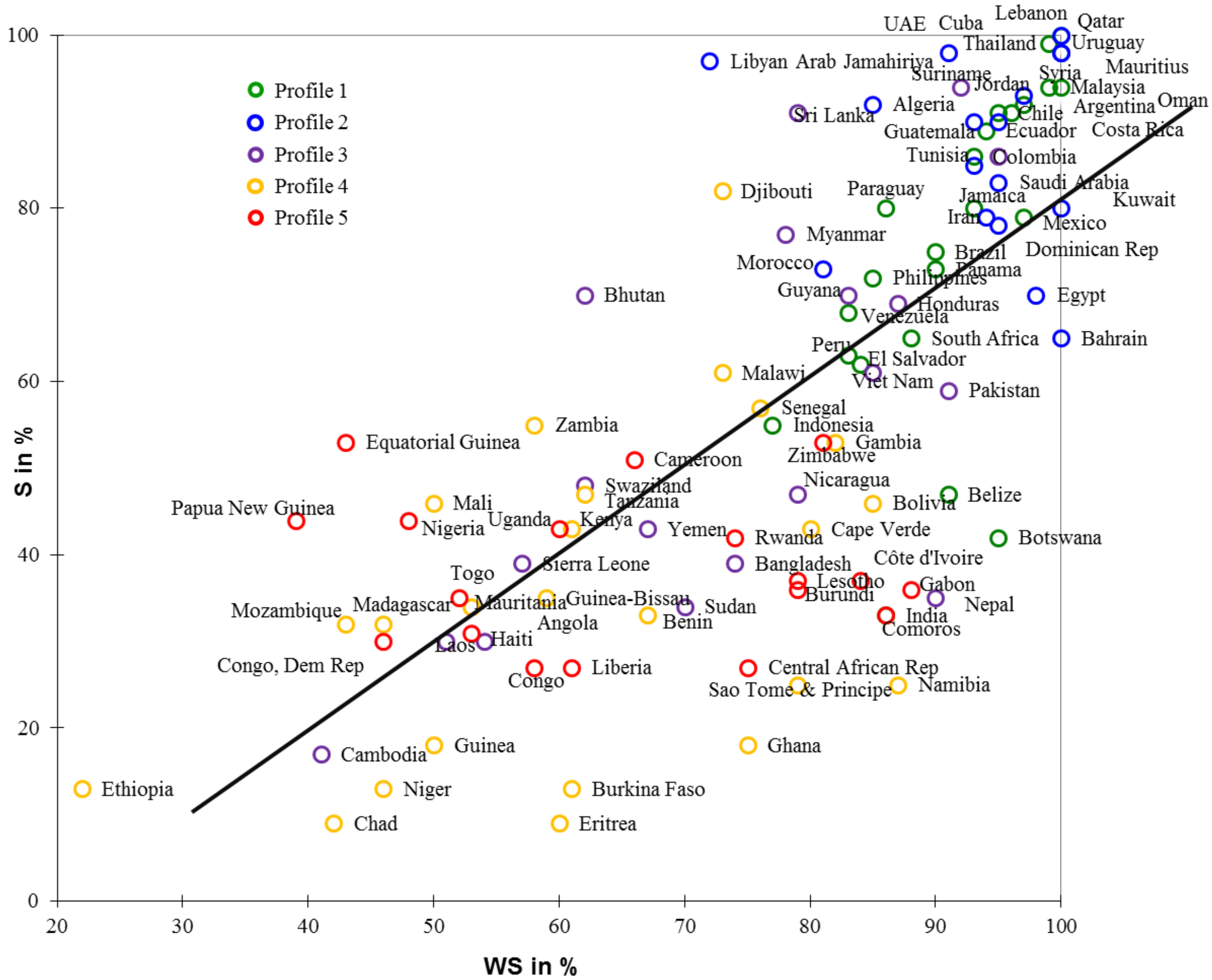


Figure 6.8: WS versus S according to 5 country profiles with linear equation as $S = 1.0285WS - 21.82$ and R2=0.5265

5.2 Profile 1: The favourable profile

The countries belonging to profile 1 are more advanced in terms of development, freedom and democracy, as well as demonstrating relatively good management of their water resources according to their uses. Many are “emerging countries” with an increasing role on the world scene due to their economic weight, such as Brazil, Argentina and India. They are the most balanced across the 5 axes, despite there being a section of the population without improved

WSS access (Chapter 5, section 4.2). The clustering tool facilitates a cross-cutting analysis of specific countries. Brazil and its neighbouring countries are used to demonstrate this.

5.2.1. Comparison with other Latin American countries

Compared to the other countries in the profile (Fig 6.9), Brazil demonstrates good human development (0.774 value in HDP, similar to the centroid that is the reference value for the group), a very good value in terms of environmental concern and citizens' voice and accountability (1.379 value in CEC, above the reference) and strong water demand for the industrial sector, services and domestic purposes (-0.787 value in AP, lower than the reference). This indicates that the agriculture sector in Brazil is still important, consuming 62% of the total of water withdrawn. However, other users—domestic, commercial and industrial consumers—also monopolize a significant amount of the withdrawn water resources. A balance is found between these various water users in a general context of abundance (0.710 value in WR, slightly above the reference). The level of ODA CI is therefore very limited (-1.277); the ODA CI value falls significantly below the reference value (-0.500, Table 5.4). This was accompanied by a slightly negative value (-0.0817) for WGI PS AV in 2004, which indicates that the country does have an issue with violence and/or political instability.

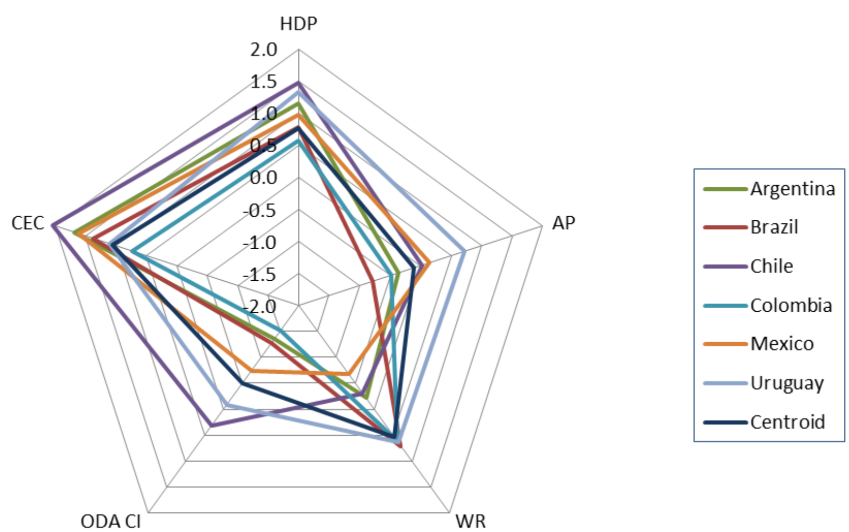


Figure 6.9: Profile of Brazil, Argentina, Uruguay, Mexico, Colombia and Chile.

Mexico, Argentina, Chile and Uruguay outdo Brazil in terms of HDP, but when comes to the national commitment to the environment and civil society strength (CEC), Brazil has slightly higher value than the reference (1.091 for Brazil versus 1.057 for the centroid).

Looking at the data in detail, Brazil has higher values for poverty variables when compared with others in this group. The black spot signifies the percentage of the urban population living in slums, where Brazil shows the highest value of all profile 1's country in 2004 (29%). This makes the percentage of the population living under the national poverty line important (21% - the 2nd highest value). This translates into child mortality under 5 years and female economic participation rates (with Brazil ranking second for both variables). In terms of access to WSS, Brazil showed high levels of both (with WS at 90% and S at 75%) in 2004 but still has the lowest values out of the 5 countries in the profile. Basic sanitation access in particular sits well outside the average of the selected countries (89%). Brazil does also face malaria like Colombia, which does not exist in the other countries. Advance governance and income per capita have medium values. Therefore, Brazil ranks 4th out of the 5 countries in terms of HDP in this regional comparative analysis.

Brazil has the lowest AP value because it combines a demanding agricultural sector (62% of withdrawals), with significant industries (18 % of withdrawals) and services, commercial and population demand (20% of withdrawals). Chile combines intensive agriculture (using irrigation) industrial sector consumption (25% of withdrawals). Argentina and Mexico show a similar configuration, with a significant level of agricultural consumption (73% of the withdrawals); irrigation practices in Mexico give it a higher AP value. Agriculture is the central economic activity in Uruguay, consuming almost all of the water withdrawn, but the country's low levels of irrigation limit its AP value.

Only Chile and Uruguay are beyond the reference value for ODA CI—all the others fall below. Brazil is considered to be an emerging country, having supported its own development for a number of years, and thus external flows are at a minimum.

Finally, with Mexico, Argentina and Chile all experiencing water stress in parts of their territories, their WR values are below the reference (0.530 centroid value in WR). Brazil also has territories suffering from water stress (i.e. the north east region), but the abundance of resources in the Amazon masks these regional disparities.

The model (Fig 6.10) confirms that profile 1 countries benefit from a high level of access to an improved water supply (87%). Sanitation access is lower (with 76.5% high probability), but basic services are ensured for a majority of the population. 88.3 % of countries belonging to this profile benefit from medium or high levels of HDP together with high CEC (for 69% of countries). The pressure on water resources is well-distributed among the various uses ranging from industries, agriculture (cattle and/or crops) and municipal activities (this includes services and commercial and domestic uses). Note that while agriculture is the main consumer of water drawn (over 50%), other significant usages exist. Therefore, the AP value mainly sits at a medium level (almost 45%) or at a high level (35%) in countries where agriculture creates more pressure on water resources. Financial support is secondary in these countries (with 50% benefiting from low ODA CI), a fact that reflects their increasing capacity to support their own development. Most of the countries (53%) enjoy an abundance of resources.

a) Water Supply (WS) b) Sanitation (S)

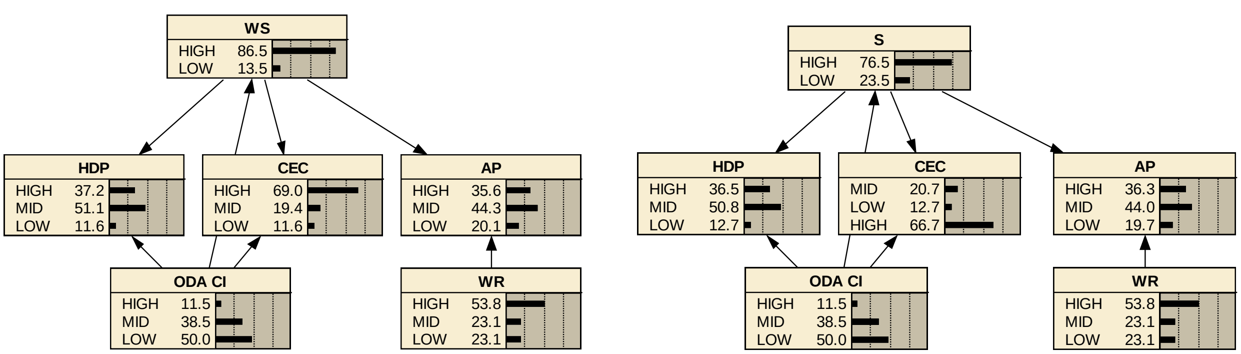


Figure 6.10: DAG and initial probabilities computed by the models - 5a for Water Supply and 5b for Sanitation - related to countries belonging to profile 1.The percentages calculated for each variable and category describe the probabilistic distribution of the profile 1 countries. This gives the main characteristics of the profile; i.e that 51.1% of the countries have medium level HDP, while only a small minority (11.6%) is estimated to have a low level.

5.2.2. Scenarios for Profile 1 (Table 6.10 and Fig 6.11)

WS and S levels are often already high for this group of countries and therefore observable variation changes are limited. The trends for WS and S are similar, although sanitation results are lower. It is important to remember that the scenarios are based on data related to profile 1 countries as of 2004.

As examples, two scenarios are run: i) using a general model, high S is set to 100% probability to visualize what would imply reaching this target (Simulation 1 – Table 6.10); ii) using a sub-model, high HDP is set to 100% probability (Simulation 2 – Fig 6.11).

Simulation 1 aims to estimate the development effects of access of above 58% to basic sanitation services for all profile 1 countries. Simulation 2 estimates the influence of good, advanced governance (WGI RofL) on the HDP indicator, as well as the different sub-variables included such as education, school enrolment and living conditions. Indeed, governance factors are often crucial to supporting a country' socio-economic development and vice versa (Chapter 6, section 4.1).

Simulation S1: 100% high sanitation—what are the benefits?

Within profile 1, the expansion of access to basic sanitation (at a high level for all countries) is thought to lead to the improvement of living conditions: +5% for HDP. The probability of a country continuing to have poor HDP sits at 9%.

Variable*	Level	Initial Probability (IP)	Re computed Probability	Δ Simulation1**
HDP	HIGH	36.5%	38.5%	5%
ODA CI	HIGH	11.5%	7.5%	-35%
CEC	HIGH	66.7%	73.2%	10%
AP	HIGH	36.3%	35.6%	-2%
S	HIGH	76.5%	100%	31%

*WR is excluded as water resources are a constant in 2004.
**Δ Simulation1 corresponds to the variation rate between S1 and IP.

Table 6.10: Simulation 1 for Sanitation (general model). The error rate of classification compared to reality is low (8.69%). Only probabilities related to relevant categories are reported in the table and used for interpretation.

Indeed, access to sanitation has a direct impact on health, with reductions in water pollution and in the dissemination of diseases such as malaria and diarrhoea. It also improves school attendance and living conditions, particularly in slums. Hutton et al. (2007) estimate that WSS interventions are cost beneficial when considering the consequent time savings (no longer any need to fetch water) and reduction in disease. "For developing countries, the return on a US$1 investment was in the range US$5 to US$46, depending on the intervention." This improvement logically influences the ODA CI, with an estimated reduction of 35%.

For this group of countries, the extension of sanitation services improves CEC by 10%, while slightly decreasing AP (-2%). The slight change in AP is assumed to be due to the improvement of municipal activities that is facilitated by access to basic services. The positive association between S and CEC can be illustrated by the systematic involvement of civil society in the management of infrastructure; an example of this the community-led total sanitation program[65].

Simulation S2: the positive influence of advanced governance

When simulating good governance (WGI RofL) for all profile 1'countries, the HDP becomes high for half (52% probability) of this group of countries (+12% if compared with IP). The probability of their being good governance in a poor country remains, in this case, very low (8% probability).

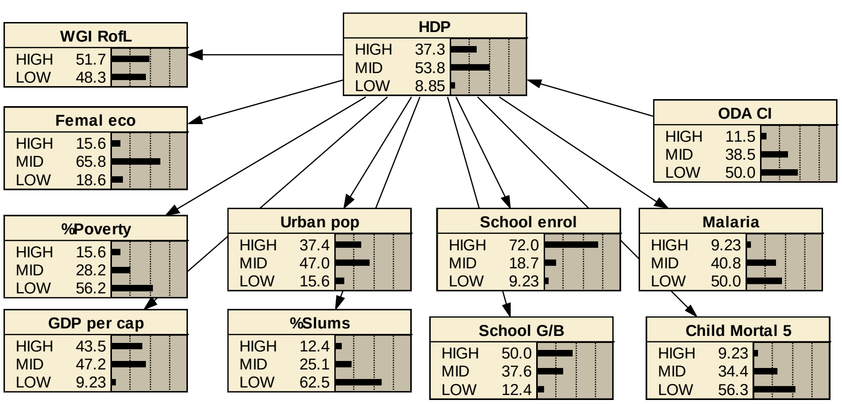


Figure 6.11:DAG for HDP sub-model

The error rate of classification compared with the reality is limited, at 4.35%.

[65] http://www.communityledtotalsanitation.org/, last access 13th October 2013

In the context of the profile 1' countries, good governance (WGI RofL set to 100%) is therefore mainly associated with:

- Increased income (+10%) and a reduction in the poverty rate (+6.4% in low category),
- Reduced urban slums (+5% of the low category), malaria (+5% in low category) and child mortality (+2% in low category) as a result of better health care.
- The expansion of education (+2.1%), in particular for girls (+4%)

The external flow is expected to remain stable. Improved governance is not the most important factor in increasing development but does contributes significantly by enabling growth and encouraging public authorities to be efficient.

5.3 Profiles 4 and 5: less advanced in WSS

Countries from profiles 4 (essential external support) and 5 (primary resources consumption) each have a less favourable WSS status (Fig 5.5). The analyses were extended to refine their profile characteristics and several scenarios were tested in order to identify a potential "path" towards WSS improvement. This modelling was performed on all five profiles, but in this work the analyses for these two groups of countries are presented.

5.3.1. Model running for Profile 4 (Table 6.11)

For these countries, the sensitivity analyses rank HDP in first place, followed by CEC and ODA CI and finally, AP. CEC ranks third in the S model, switching with ODA CI. Countries belonging to this profile are in a less advantageous situation when it comes to WSS access than countries from profile 5 (Fig 6.8). However, their HDP, CEC and ODA CI values are better. Indeed, they benefit from a stable political context that attracts external financial flows: 56% of countries show a high ODA CI. Their environmental concern is a favourable base on which to make efforts in the WSS sector: 73% of these countries show a medium-high level CEC.

These countries' primary issue is poverty, which can be limited by the organization of urbanization, improving health, elevating education from basic to college-level and by taking measures to improve general governance. Simulation 1 estimates that high WS should increase significantly up to 35.8% and S up to 37% if such efforts are made (high HDP set to 100 % probability).

Water resource availability is also limited, with around 96% of countries having low-medium WR levels. In fact, 23 out of the 24 countries are in Africa and have areas under desertification threat or being a desert. 75 % of countries of this profile have part of national territories being potentially affected by desertification in 2004 (FAO AQUASTAT). Sustaining adaption or mitigation measures are essential to fighting against this constraint and ensuring long term WSS services.

Variables* - WS model	Initial Probabilities	Simulation 1 HDP	Simulation 2 AP	Simulation 3 CEC	Simulation 4 ODA CI
HDP	11.5%	100	1.1	0.1	-2.33
ODA CI	55.6%	-11.4	0.5	11.6	100
CEC	36.2%	0.3	1.1	100	7.6
AP	16.4%	2	100	0.7	0.2
WR	3.7%	(=)	(=)	(=)	(=)
WS	16.5%	35.8%	23%	22.9%	18.8%
Variables* - S model	**Initial Probabilities**	**Simulation 1 HDP**	**Simulation2 AP**	**Simulation 3 CEC**	**Simulation 4 ODA CI**
HDP	11.7%	100	1.7	-1.75	-3.11
ODA CI	55.6%	-14.8	-1.5	12.4	100
CEC	38.3%	-5.7	-1.3	100	8.6
AP	15.6%	2.2	100	-0.4	-0.4
WR	3.7%	(=)	(=)	(=)	(=)
S	16.4%	37%	25.7 %	11.3%	12.5%

*Values concern the High category of each variable. Negative variations are in red and positive variations of probabilities are in green.

Table 6.11: Profile 4 - WS and S simulations

In simulations 3 and 4, ODA CI highlights some differences from the general model (Chapter 6, section 3) regarding S and CEC. With the increment of ODA CI, high S drops down to 12.5% and CEC increases up to 46.9% (note that Table 6.11 is computed for High ODA CI category). In contrast with WS model (Chapter 6, section 3.1), High S still increases up to 20%, while CEC is concentrated on the medium-low category (34 % probability) when incrementing ODA CI at a medium level (100% probability). In addition, an increment of ODA CI pushes the CEC at a higher level (46.9% high probability) compared to the S global model (34.9% probability).

This suggests that ODA CI has a better efficiency on CEC than in a general behaviour. More investigation in these specific countries may shed light on this specific behaviour for this profile and explain the reason in particular S is concerned.

5.3.2. Model running for Profile 5 (Table 6.12)

According to both sensitivity analyses, WSS variables are influenced in order by HDP, AP, CEC and finally ODA CI. Table 25 indicates that major efforts should be put into the three main dimensions (AP, HDP, and CEC) to improve the substantially the level of WSS. Indeed, these 18 countries present with low access to WS in 75% of cases and low S in 88% of cases.

Variables* WS model	Initial Probabilities	Simulation 1 AP	Simulation 2 HDP	Simulation 3 CEC	Simulation 4 ODA CI
HDP	14.6%	1.8	100	3.1	8.7
ODA CI	14.3%	(=)	8.5	8.5	100
CEC	14.6%	1.8	3.1	100	8.7
AP	15.4%	100	2.4	2	(=)
WR	33.3%	(=)	(=)	(=)	(=)
WS	24.9%	40.3%	40.2%	40.2%	25 %
Variables S model	**Initial Probabilities**	**Simulation 1 AP**	**Simulation 2 HDP**	**Simulation 3 CEC**	**Simulation 4 ODA CI**
HDP	13%	4.1	100	5.6	10.3
ODA CI	14.3%	3.1	11.4	11.4	100
CEC	13%	4.1	5.6	100	10.3
AP	13.6%	100	4.2	4.2	2.9
WR	33.3%	(=)	(=)	(=)	(=)
S	12%	29.6%	30.9%	30.9%	25%

*values are intended as HIGH category for each variable. Negative variations are in red and positive variations are in green.

Table 6.12: Profile 5- WS and S simulations

Simulation 1 indicates that an increase in AP translates into an increase in WS of up to 40.3% and S of up to 29.6%. These countries are characterized by an abundance of water resources: 81% of the countries benefit from medium (48%) to high (33%) WR. All the countries (except Zimbabwe) withdraw a limited amount of water, all uses considered, relative to the full set.

63.5% of them have a low AP value, indicating that a significant part of their drawn WR is used for industrial and raw material exploitation. Their municipal uses (urban activities and population) absorb an important share of the remaining WR. As the total amount withdrawn is limited, the amount of water dedicated to domestic use appears relatively large. The agricultural share is therefore low—below 50%—and the share for irrigation practices is near 0. Equatorial Guinea, Central Africa, Lesotho, Papua New Guinea, Congo and Democratic Congo are examples of economies strongly oriented towards oil extraction and mining, with weak agriculture characterized by subsistence farming. Good WSS does not appear to be associated with industrial water demand, highlighting that there are no significant synergies between industrial and population supply (Chapter 6, sections 3.1 and 4.3). In this case, the natural resource exports have "counteracting effects on growth, by weakening the manufacturing and agricultural sectors" (Knack, 1999). Diversified economic activities and in particular strengthen the agricultural sector in these countries may evolve with improved WSS services (Chapter 6, section 4.3).

The countries with better AP values (16%) due to having a lower industrial share in water demand should focus on the direct reinforcement of WSS capacities. This will benefit the population and the agricultural sector (the main employer of the labour force[66]). Comoros, Cameroon, Burundi, Togo and Rwanda are representative of this category.

Simulation 2 shows that an increase in HDP induces an increase in WS and S of up to 40.2% and 0.9%, respectively. Poverty in these countries is a crucial issue, encouraging high demography, informal urban settle development and generally poor health, in particular in children, as well as poor average education. Advanced governance is globally weak (WGI RofL) and therefore the capacities of the states are small.

These states need to be supported by international donors, however their political instability (measured by WGI PS AV) acts as a barrier. In fact, only 15% of countries in profile 5 benefit from high ODA CI. The majority (57%) receive medium ODA CI and almost 30% receive low ODA CI. Simulation 4 shows that an increase in ODA CI produces an increase in HDP of up to 23.3% in both the WS and S models.

[66] Consult country profile from the CIA Factbook: https://www.cia.gov/library/publications/the-world-factbook/, last access 4th July 2013.

The improvement of CEC is associated with increased WSS, as shown in simulation 3: WS should reach 40.2%, and S 30.9%. Environmental management appears even more essential in economies focused on industrial/natural resource exploitation. CEC is generally low in these countries (48.5%), meaning that efforts need to be made to improve levels of civil society freedom and government accountability to citizens. Their basic governance needs to be strengthened.

6. DISCUSSION

In the majority of the countries assessed, access to an improved water supply rates higher than access to basic sanitation (Chapter 6, section 5.1). Reasons for this gap are the neglect of and cultural obstacles towards the development of sanitation. This dimension sits outside the parameters of our models, explaining the relatively similar results. These cultural psychological aspects are seen as crucial by the international community. Indeed, the "Sanitation for all" year (2008) highlights its "unpopularity" leading to a "sanitation crisis"[67]. Improved sanitation is often not considered necessary/vital and investment in it is not viewed as added-value spending; understanding the cultural/psychological factors behind these views is essential to its sustainable development (Haddad, 2005). Up to now, there has been no quantitative or qualitative method available to measure these effects. In numerous countries, awareness campaigns have proven to provide crucial leverage in convincing populations to implement and maintain sanitation facilities. The community-led total sanitation (CLTS) approach[68] developed and applied in Bangladesh since 1999 demonstrates the necessary cultural/educational work needed as part of sanitation programs. (Al Sa'ed & Mubarak, 2006) showed through a survey on sanitation facilities the difficulties to convince in that case, the rural population, to contribute to sanitation/wastewater infrastructure. In this particular case, the majority of respondents (55 %) were against new wastewater treatment and in favour of centralised wastewater infrastructure while only 18% were willing to contribute financially. This survey also showed the preference of the population to get the "standard sanitation" meaning a flush toilet connected to centralised

---

[67]International Year of Sanitation 2008: http://esa.un.org/iys/docs/IYS_flagship_web_small.pdf, last access 13th October 2013

[68] http://www.communityledtotalsanitation.org/, last access 13th October 2013

wastewater system. In South Africa, it has been reported that urban dwellers were reluctant to change their pit latrine for other low cost improved sanitation facilities, only accepting to exchange it for a standard flush toilet. Getting "standard toilet" has turn for instance in Cape town into a fight for equality between poor and rich, black and white people[69].Social acceptance of the sanitation technology is crucial (Mara et al., 2010). The implementation of sanitation alternatives should be supported by awareness and active involvement of the communities (Roma & Jeffrey, 2010). (Katukiza et al., 2010) measured slums residents' preference for urine diversion dry toilet (UDDT) and VIP latrine only after proper information and group discussion on the different technologies alternatives.

Although the models built show the limits and constraints of the accuracy of the WGI values, this research—together with the existing literature and records of field experience—demonstrates the crucial role of governance in sustainable WSS (Plummer & Cross, 2006). This is highlighted that specific studies related to governance in the water sector are not extensive and often not detailed at subsector level. For instance, corruption is measured across the water sector as a whole, often including WSS, agriculture and hydropower. A corruption analysis is not performed at a sub-sector level. The disaggregation of data by sub sectors, specific analyses and/or case studies—in particular as it relates to sanitation—is rare on a national scale. Descriptions of corruption mechanisms—such as those offered by Transparency International—are rejected for the whole water sector, with specific information on S often being neglected. At the very least, comparing case studies on corruption mechanisms in both sub-sectors could increase the accuracy of the interpretations made.

With the multivariate analyses (Chapter 4, section 4), this modelling work presents other margins for improvement: i) increasing the representativeness of the variables selected to define water resources and ii) include wealth inequality in HDP. The water quality indicators could not be included in the analyses and models, which perhaps explain the negligible weight of water resources for WSS. Such indicators could, at a national scale, allow the modelling of interactions

---

[69] Article in http://bigstory.ap.org/article/south-africa-toilet-talk-turns-political, last access 13th August 2014

between the environment and development through the analysis of the qualitative impacts on water resources. This would be particularly useful in cases where a lack of sanitation influence the water quality and thus health significantly. For instance, the highest rates of child mortality continue to be found in sub-Saharan Africa. Out of the four main diseases that cause children mortality, two (diarrhoea and malaria) are directly linked to WSS services access.

7. CONCLUSIONS

Identifying mechanisms that influence WSS is complex because of the cross-interaction of multiple factors and issues. The BNs methods were chosen as they measure probabilistic conditional dependence relationships between variables. The suggested models are efficient tools for identifying and measuring the influence of key elements on WSS levels (statistical performance between 80-95%).

As a first step, general models were built separately for WS and S so as to observe and identify potential differences. Thematic sub models allowed focusing on relationships between variables within each dimension. Based on the five thematic composite indicators (HDP, CEC, AP, ODA CI, WR), these general models synthesize the large mechanisms involved in WSS access that exist in the common field literature. The statistical classification error rates are less than 12% for Water Supply (WS) and 10% for Sanitation (S) while they range from 19% to 5% for the thematic models. Sensitivity analyses provided the key variables influencing the WSS variables (research question 2, Chapter 2, section 4.2) BNs has extended the results from linear regression analyses (Chapter 5, section 5) because considering the total variability of the data subset (Table 6.3).

The MDGs aim—alleviating poverty—was confirmed as the first priority by the modelling. Improving WSS access is as good mean to move toward this objective. In details, the HDP thematic sub-model estimates that key variables are household income (GDP per cap), urbanization processes where the reduction of informal settlements is crucial (Urban Pop, % slums, %Poverty) and finally health improvement (child mortality and malaria). The urbanization process facilitates WSS access except in cases of informal urban settlement development. It also emphasizes the importance of the control of disease (malaria) and the mortality of children. Governance provides an essential framework for sustainable WSS service delivery. The education of the population facilitates WSS development by creating specific demand and also through the provision of the necessary skills to maintain services. Gender issues are secondary.

The analysis highlighted that WS is significantly sensitive to environmental commitment and level of democracy/freedom with a positive change of 21 % while is induces only a +8% change on S. This behaviour can be an expression of the less consideration of sanitation as an issue to be handled collectively like water supply. In both cases, the CEC is first influenced by the environmental commitment of a country in front of the international community (Particip). High WSS is positively associated with this formal commitment. This environmental concern is supported by civil society where means for this support to be expressed are available. This framework is attractive to external donors (ODA CI).

In addition, the ODA CI appeared to influence significantly on both WS and S levels as confirmed by sensitivity analyses, even if this effect is relatively limited (between 10% to 7% probability changes and ranking 4th out of 5). Complying with multivariate analysis output, high ODA CI is estimated to be associated with poor WS (60% probability) and S (64% probability). In fact, ODA CI is provided preferentially—but not exclusively—to poor countries. The models showed that high amount ODA CI is highly probably provided (more than 60%) to country showing poor WSS accesses. Funds are still invested at high level, at 22% probability in relatively less poverty-stricken countries although they are still not considered developed.

In the attempt to go beyond the non-significant correlation between WSS and ODA-WSS (Chapter 4, sector 3.1), the use of BN methods and the aggregation of ODA WSS with the global ODA and the political stability of the country WGI PS AV has revealed the relationship between these two variables. The ODA-WSS is integrated in ODA CI and WSS are sensitive to this new ODA CI component however the association remains limited (Chapter 6, sections 3 and 4). Nevertheless, the modelling supports more the analysis of the most probable profile of a country receiving a high ODA inputs.

The AP thematic model highlighted synergies with the type economic development evaluated through water demand structure. The development of irrigated agriculture and the emergence of urban activities are favourable to WSS development and vice et versa. The positive interaction between WSS sector and agricultural sector rather occur in the context of rural organisation around irrigation. However, this is not systematic −33% of countries are estimated to have low WSS despite a high value in AP; nor exclusive – few countries show high AP value little developed irrigation area like Honduras, Nicaragua and Guatemala (Chapter 5, section 4.2, profile 3). The availability of WR has little influence on the pattern of activity development or the level of water demand. In fact, WR also counts for little when it comes to the population's access to WSS (ranking last in WSS sensitivity analyses). This element could have been better appreciated if indicators related to water quality had been included (Chapter 4, section 4).

In a second step, this general model was applied to the five country profiles built around the same five axes: HDP, CEC, AP, ODA CI and WR. The models and their behaviours highlighted divergences in key variables or processes and enabled us to test scenarios with the different profiles. In a general sense, sanitation access mainly falls behind water supply (-21 points of percentage on average). Countries from profiles 4 (essential external support) and 5 (primary resource consumption)—mainly sub-Saharan countries—are struggling with both water and sanitation access and have the lowest levels of both out of all the developing countries. Profiles 1 (on the way to well-being) and 2 (freedom/democracy black spot) are home to the more advanced countries, while profile 3 (agricultural economy) shows a spread of WSS distributions. The agricultural context has complex influence on WSS, which explains the diverse WSS levels among countries from profile 3.

In this chapter, models were adapted to suit groups of countries from profile 1, profile 4 and 5 and the results were then analysed. The comparative analysis can be scaled down within a profile, as was demonstrated using Brazil and its neighbour countries as an example.

The models for Profile 1 represent the reality of the data well, with error rates below 8%. Two simulations were run: 1) If all countries belonging to profile 1 had high S (above 58% access), their human development would increase by 5%, reducing the probability of their having poor HDP to 9%. This positive improvement was accompanied by a 10% growth of CEC. This highlights the wider benefits of ensuring access to sanitation services. 2) As governance has become important factor in country development, a simulation focused on this was run. Where there is advanced governance for all (100%), the HDP would be estimated to increase by 12 points of percentage (using the profile 1 group of countries), as would all the HDP sub-variables.

Countries belonging to profile 4 experience desertification effects in the context of limited water resource availability (67%). Therefore, the development of adapted or mitigation measures is crucial. As they benefit from minimum governance (political stability and civil society freedom) and concern for the environment, efforts should focus on HDP dimensions—the control of urbanization, the organization-reinforcement of health care and education, and the improvement of advanced governance as it relates to state effectiveness (corruption control, make and rule laws). The model indicates different behaviours when it factoring in the impact of CEC on sanitation. ODA CI appears to be more effective on S and CEC as observed in the general model. Further investigation into these specific countries may shed more light on this behaviour and why only S is affected.

Profile 5 countries show weaknesses in almost all areas, with low HDP (68-70%), CEC (48-50 %) and AP (65-67%) values indicating pressure from industrial activities and limited development of agriculture. Despite this, these countries only receive average levels of investment from donors (a medium level of ODA CI for 57% of countries)—a fact that can be often explained by the persistence of political instability in the given country. Therefore, ensuring basic governance conditions, stability and absence of violence and accountability to citizen is essential: 1) to foster economic diversification and reinforce rural development; 2) to sustain efforts made for poverty reduction; and 3) to appeal for additional external support if necessary. Reinforcing the agricultural sector is crucial; investing in the WSS sector may also support this process (building capacities, creating complementary market opportunities…)

The Bayesian modelling was applied and valid for the year 2004. The next step is to extend this modelling approach to several temporal points potentially available. The next chapter, Chapter 7, will look for complementary data assessing its completeness in order to add at least three temporal points to the WatSan4Dev database. The Bayesian networks will be described in their dynamic form providing their main advantages. The retrospective Bayesian approach will be applied on the dataset for 2000, 2004 and 2007 because it will allow to analysis these past 7-year period.

CHAPTER 7: TEMPORAL ANALYSIS AND MODELLING OF WATER SUPPLY AND SANITATION IN DEVELOPING COUNTRIES WITH BAYESIAN NETWORKS

1. INTRODUCTION

The previous chapter detailed the Bayesian Network modelling methodologies that were applied to the WatSan4Dev dataset for 2004. This chapter looks at extending the analysis and modelling to the time period 2000-2007 (Fig 7.1 and Appendix D, section 5). The retrospective modelling approach is adopted to integrate past records and simulate variables status to reach user-defined targets. It is defined by retrospective approach, the modelling of the existing records that aim at fitting the past reality to get lessons learnt and identify processes that occurred. This can be used for running simulation that will allow the analysis of the influence and sensitivity of the different variables involved in the water sector.

This chapter focuses on the development of the methodological approach because the international data providers recently updated the data used in this study. The data recorded at the two other temporal points, 2000 and 2007, have been retained. The two main indexes—WS and S—have been accumulating data since 1990 and are managed by the Joint Monitoring Programme, but it is only since the launch of the Millennium Development Goals initiative in 2000 that the data series have begun to reduce the data gaps (reducing missing values).

Therefore, sections 2 and 3 of this chapter summarise the review and the pre-processing of the data series for the different years (2000, 2004, 2007), while section 4 provides the analysis of the proposed temporal models. A retrospective approach is taken to simulate the state of explicative variables to reach the MDGs targets: at least 88% WS and 75% S access for the populations in all of the selected developing countries A complementary modelling approach, the predictive or forecasting approach, is not considered in this chapter but its added value, its limits and its implementation constrains will be discussed in the Chapter 8, section 2.2

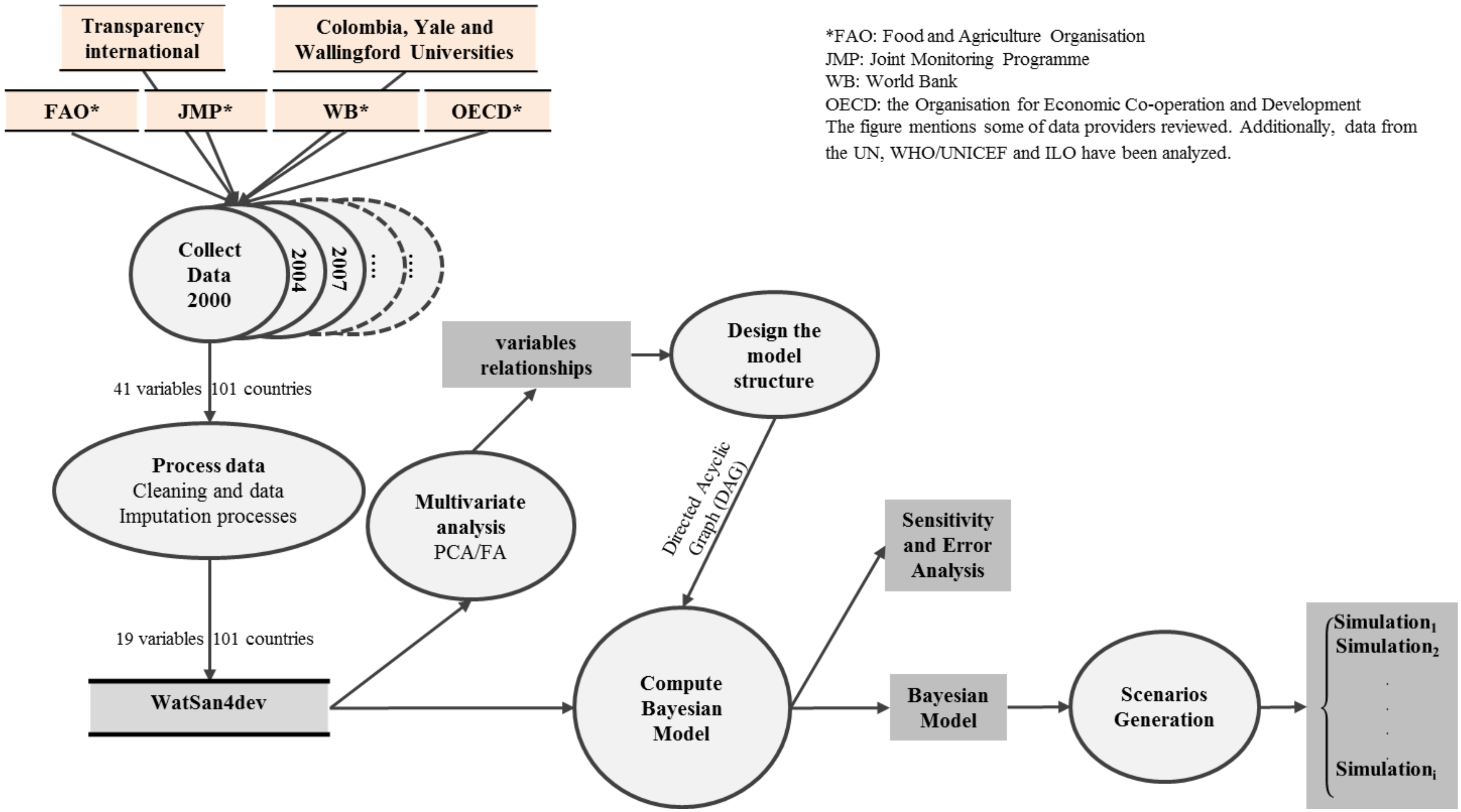


Figure 7.1: Functional Data flow Diagram in which a rectangle with two black borders represents a database, a circle indicates a process, and a grey rectangle a result. The external input data is in orange and the results of this research are in grey.

2. Data review

2.1 Variables review

The availability of the variables selected for the modelling in 2004 was limited (Table 7.1). The data used in this chapter were collected in 2009 and 2011. Since these dates, the FAO and JMP have enhanced the data available by adding information from additional years and updating the previous series. Therefore, the analysis presented in this chapter needs to be updated to reflect this improvement.

The period spans three temporal points: 2000, 2004, and 2007. Prior to 2000, data present high levels of missing data for developing countries, even for demographic and socio-economic variables. Environmental variables are scarce with only the datasets for 2004 showing acceptable missing data rates— in particular the water withdrawal variables for 2000 and 2007 presented missing data rates of above 70% when collected. Water resources are considered to have been constant over this 10 year period. The female economic activity rate is replaced by women economic activity rate because of absence of a dataset for 2000. The difference between the two is that the women economic activity rate reports on the total active population while the female economic activity rate reports on the female population.

The AP and CEC dimensions were the most affected by missing values (total withdrawals, water withdrawals for industries and municipal purposes) and by the absence of variables (water use intensity in agriculture and participation in international environmental agreements). The continuous updating of water demand data (withdrawal indicators) could address the high rate of missing data in the near future.

Dimension	Variables	Availability			Missing data (%)			Variable Units
		2000	2004	2007	2000	2004	2007	
	Water Supply (WS)	a	A	a	0	0	2	%
	Sanitation (S)	a	A	a	0	0	3	%
HDP	% poverty	na	A	na	na	34	na	%
	Urban Pop	a	A	a	0	0	0	%
	School enrol	a	A	a	9	22	6	%
	Women eco	a	A	a	2	1	1	%
	% slums	a	A	a	4	22	63	%
	Child Mortal-5	a	A	a	0	0	6	$^0/_{00}$
	School G/B	a	A	a	9	22	6	[0.63,1.02]
	GDP per cap	a	A	a	6	6	9	thousand $ per cap
	Malaria	a	A	a	32	22	14	$^0/_{00}$
	WGI.GE	a	A	a	0	0	1	[-2.5,+2.5]
CEC	Particip to IEA	na	Na	na	na	13	na	[0,1]
	WGI.V.A.	a	A	a	0	0	1	[-2.5,+2.5]
ODA CI	WGI.PS.AV	a	A	a	0	0	1	[-2.5,+2.5]
	ODA WSS	a	A	a	9	6	6	$ per cap
	ODA	a	A	a	0	0	4	$ per cap
AP	% irrigation	a	A	a	3	3	3	%
	Water Use int Agri	na	A	na	na	2	na	m^3 $ha^{-1}yr^{-1}$
	Total withdrawal	md	A	md	73	4	79	m^3 yr^{-1} per cap
	Withdrawal Municipal	md	A	md	75	1	83	%
	Withdrawal Industrial	md	A	md	76	1	83	%
WR	Precipit	c	C	c	0	c	c	mm yr^{-1}
	Desert risk	c	C	c	7	c	c	%
	TWRR	c	C	c	0	c	c	m^3yr^{-1} per cap

*na: not available, a: available, c :constant, md: high proportion of missing data

Table 7.1: Overview of data series availability and percentage of missing data for 2000, 2004 and 2007

In an attempt to find proxies, several additional indicators were studied with a view to substituting them for the missing variables for 2000 and 2007. This would aim to help in: 1) evaluating the human activity development quantitatively and by type of activity i.e. agriculture, industries and municipal economic activities; 2) qualifying the environmental commitment and management of the country. The two criteria against which the potential proxies were selected are low missing data and consistency with the variables selected for 2004.

Further to the first aim, the Livestock production index (Livestock.Prod), Agricultural Production index (Agri.Prod) (AQUASTAT database) and the estimations of added-valued amounts by sector (Agri$, Industry$ and Services$) in percentages of GDP (World Bank database), were analysed.

Within the 2004 dataset, these five new variables were included in the multivariate analysis (PCA using Pearson distance) so as allow the analysis of the relationships between the water withdrawals variables (municipal withdrawals, industrial withdrawals and total withdrawals).

- The livestock and agricultural production indexes have very little missing data. However, the PCA—performed on the WatSan4dev subset and these additional two variables—shows no significant correlations with other variables (Table 7.2). In fact, these two variables are isolated with high factor loadings on F6 (Table 7.3).
- The three indicators related to the added-values amounts provided by economic sectors offer a more accurate impression of a country’s human-economic development than proxies of water withdrawals variables. Agri$, Industry$ and Services$ are only slightly correlated with the total withdrawals and are not significantly correlated with the type of water withdrawals (-107, 0.112 and 0.016 correlation respectively with Withdrawal municipal and-0.098, 0.101 and -0.071 correlation respectively Withdrawal industrial). The multivariate analysis matrix shows high factor loads for these three variables on F1 and F3 components, while withdrawals variables load on F2 (Table 7.3).

Variables	Agri.prod.	livestock.prod	Industry$	Agri.$	Services$	Withdrawal municipal	Withdrawal industrial	Total withdrawals
Agri.prod	**1**	**0.292**	0.029	0.108	-0.128	-0.107	-0.098	0.065
livestock.prod	**0.292**	**1**	-0.121	0.102	-0.008	**-0.210**	-0.084	0.112
Industry$	0.029	-0.121	**1**	**-0.673**	**-0.299**	0.112	0.101	**0.304**
Agri$	0.108	0.102	**-0.673**	**1**	**-0.357**	**-0.300**	-0.095	**-0.326**
Services$	-0.128	-0.008	**-0.299**	**-0.357**	**1**	0.016	-0.071	**0.211**
Withdrawal municipal	-0.107	**-0.210**	0.112	**-0.300**	0.016	**1**	**0.514**	**-0.553**
Withdrawal industrial	-0.098	-0.084	0.101	-0.095	-0.071	**0.514**	**1**	**-0.251**
Total withdrawal	0.065	0.112	**0.304**	**-0.326**	**0.211**	**-0.553**	**-0.251**	1

*Values in bold are significantly different from 0, with a significance level alpha=0.05

Table 7.2: Pearson correlation matrix

Therefore, these indicators are left out of the database and no proxies are available for substitution.

Percentage of variance						
	F1	F2	F3	F4	F5	F6
Eigenvalue	10.660	3.151	2.515	2.400	1.671	1.259
Variability (%)	35.534	10.502	8.383	8.000	5.569	4.196
Cumulative %	35.534	46.036	54.419	62.420	67.989	72.184
PCA-factor loadings						
Variables	F1	F2	F3	F4	F5	F6
Agri.prod	-0.022	-0.211	0.013	0.110	-0.181	**0.661**
livestock.prod	0.041	-0.306	-0.147	0.184	0.039	**0.746**
Industry$	0.469	0.063	**0.590**	0.126	0.307	-0.090
Agri$	**-0.788**	-0.225	-0.315	0.191	-0.133	0.031
Services$	0.503	0.083	**-0.532**	-0.349	-0.257	-0.016
% Poverty	**-0.729**	-0.037	-0.125	0.037	-0.027	-0.069
Urban.Pop	**0.733**	0.248	0.272	-0.055	-0.079	0.048
School.enrol	**0.786**	0.164	-0.075	0.239	-0.046	-0.002
Femal.eco	**-0.600**	0.207	-0.195	0.111	-0.200	-0.035
% Slums	**-0.871**	-0.087	-0.033	0.035	0.056	-0.010
Withdrawal municipal	-0.003	**0.784**	0.358	-0.163	-0.216	0.142
Withdrawal industrial	0.058	**0.701**	0.112	0.267	-0.051	0.112
TWRR	-0.294	0.257	-0.382	**0.613**	0.090	-0.263
S	**0.834**	-0.037	0.051	0.081	-0.009	0.035
Child.Mortal.5	**-0.889**	0.003	0.014	-0.204	0.018	0.009
WS	**0.866**	0.064	-0.027	0.005	-0.150	0.057
school.G.B	**0.632**	0.175	-0.162	0.202	0.121	-0.027
Particip	0.354	0.160	-0.470	-0.051	**-0.581**	-0.029
GDP.per.cap	**0.907**	0.174	0.157	0.018	0.163	-0.052
% irrigation	**0.554**	**-0.552**	-0.253	0.274	0.031	-0.048
ODA.WSS	-0.041	0.128	-0.374	-0.366	**0.629**	0.182
ODA	-0.462	0.199	-0.365	-0.290	**0.488**	0.103
Malaria	**-0.839**	0.217	-0.106	0.024	0.083	-0.042
Water.Use.int.Agri	**0.646**	**-0.616**	-0.178	-0.024	0.102	-0.078
Precipit	-0.143	0.278	-0.338	**0.807**	0.086	-0.021
Desert	-0.214	-0.266	-0.068	**-0.616**	-0.328	-0.202
Total.withdrawals	**0.648**	**-0.532**	-0.147	0.192	0.106	-0.188
WGI.V.A.	0.420	0.420	**-0.589**	-0.094	-0.139	0.038
WGI.PS.AV	**0.439**	0.340	-0.355	-0.305	0.384	-0.046
WGI.GE	**0.768**	0.188	-0.268	-0.275	0.055	0.055

Table 7.3: Factor loadings on the first six components, representing 72.18% of the variability with overall KMO value of 0.842.

With regards to the second aim, no potential proxies were identified that could be substituted for the environmental governance index (Env gov) or for the level of country participation in environmental agreements (Particip). Both indicators Environmental Sustainability Index ESI and Environmental Performance Index (EPI) are indicators that are not comparable over time[70]. The sub indicators in ESI computation were not reused in EPI framework (Chapter 2, section 3.2.2).

2.2 Analysis of missing data patterns (Fig 7.2)

A study of the pattern of the missing values was performed to analyse the randomness of the missing data. This analysis revealed: i) a recurrent pattern of missing data over the three years and, ii) patterns for one specific year. The high correlations (above 0.5) between variables in 2000, 2004 and 2007 occurred for two reasons related to the indicator computation:

- Data collection and computation of the variables were performed at the same times by same provider. This relates to i) WS and S compiled by the JMP with a correlation between 0.81 to 0.91 depending on the year; ii) Worldwide Governance Indexes (only 2007 has a missing value) iii) ODA and ODA WSS collected by the OECD through the DAC system, with a correlation of 0.74 for 2007. The ODA WSS presents more missing data than general ODA.
- Variables are mathematically dependent. Water withdrawals by sector are computed as a share of the total withdrawal (Appendix B).

While significant correlations can be seen between missing data patterns, they are not recurrent over the 3 years, therefore the randomness of the missing data is ensured:

- For 2000: Missing data patterns for the proportion of slums (% slums) and women economic activity (Women eco) show a correlation of 0.700.
- For 2004: Missing data patterns for women economic activity (Women eco) and water withdrawals show a correlation of 0.700.
- For 2007: ODA, ODA WSS and GDP per capita show similar missing data patterns with correlations of 0.650 and 0.600, respectively. The ODA variables also present a significant correlation (0.490) with Malaria.

---

70 http://envirocenter.yale.edu/programs/environmental-performance-management/environmental-sustainability-index, last access 18th August 2014.

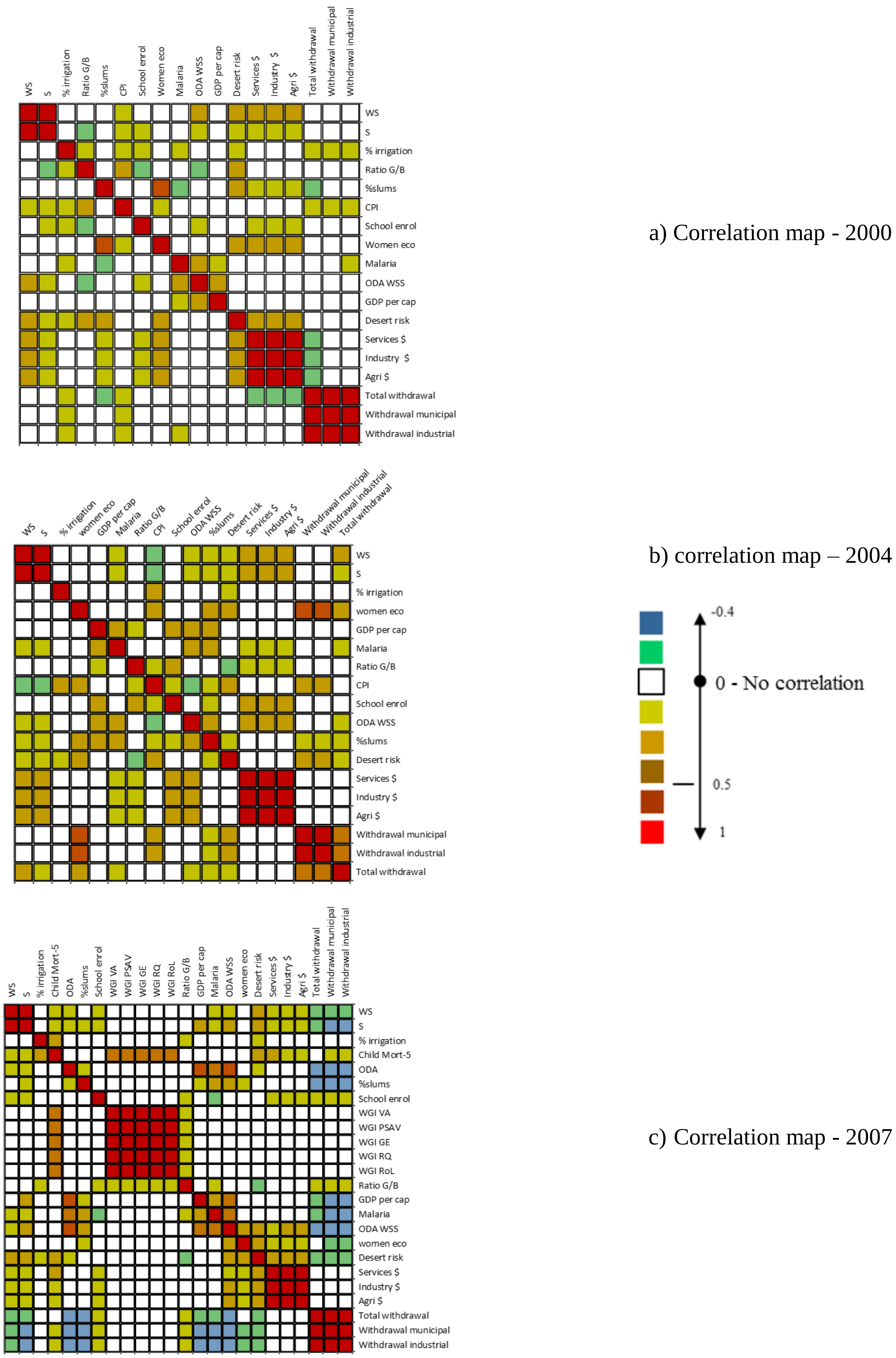


Figure 7.2: Missing data correlation maps for 2000, 2004 and 2007. Only variables with missing data are represented, to make the maps easily readable.

3. *DBNs Application*

The Bayesian Network general model is adapted to the variables availability, in keeping with the principles used in the 2004 model (Chapter 6).

3.1 Input data processing

In keeping with the methodology used for 2004 (Chapter 6, section 2), variables were grouped and computed using two composite indicators—HDP and ODA CI—and multivariate analyses for each year:

- HDP includes income (GDP per cap), the proportion of urban population (Urban pop), the girls' attendance at primary school (School G/B), the school enrolment rate for both sexes (School enrol), the women economic participation rate (women eco), the proportion of the population living in slums (% slums), the prevalence of malaria (Malaria), the child mortality rate under five years (Child Mortal-5) and the advanced governance index (WGI RofL). WSS are excluded from the composite indicator as for 2004.
- ODA CI is computed using the amount of Official Development Assistance (ODA), the specific amount of ODA dedicated to WSS (ODA WSS) and the political stability and absence of violence index (WGI PSAV).

Each variable is normalised with the appropriate standard method, logarithmic, Standard error transformation or Ordinary least Square regression. Urban Pop, School Enrol, Women eco, WGI-RofL and WGI PSAV are by virtue of their nature already normalised. These variables (from 2000 to 2007) are rescaled in a common scale.

The proportion of irrigation in agricultural areas (% irrigation) and the voice and accountability index (WGI VA) are treated as a single variable like the Water supply and Sanitation variables. Therefore, as the Bayesian Network method deals with non-normal distribution (Lee et al., 2005), these variables are kept in their original format. The other advantage resides in that worldwide MDG targets for WSS can be used as threshold for categorisation.

As for 2004, variables are discretised in three categories (High, Mid and Low). However, the clustering is manually defined to set common limits for the three years (Table 7.4) (Uusitalo, 2007). WSS variables are categorised according to the MDG targets: 88% of worldwide population having access to improved WS and 75% to basic sanitation (UN, 2012).

Variables	Categories				Variable Unit or intervals*
	LOW		MID	HIGH	
	Minimum	Upper limit	Upper limit	Maximum	
Water Supply (WS)	22.00	88	na	100.00	%
Sanitation (S)	9.00	75	na	100.00	%
% irrigation	0	2	10	100	%
WGI.V.A.	-2.14	-0.8	0.2	1.17	[-2.5,+2.5]
HDP	-4.56	-1.2	0.8	5.04	na
ODA CI	-2.35	-1.2	1.8	3.01	na
WR	-2.43	-0.2	1	2.87	na

Table 7.4: Category limits for each variable.

3.2 Structuring the models

Each time slice is modelled according to a structure adapted from 2004 modelling (Chapter 6, section 2.1) which is identical for the 3 temporal points. Indeed, the PCA (Fig 7.4 and Appendix D, section 5) confirms that variables organised themselves similarly over that period. The 2004 modelling mainly explained the contribution of WSS to other sectors in developing countries (Fig 7.3a). However, the literature also suggests that there can be multiple relationships between these sectors (measured by the composite indicators). Therefore, different directions of relationships are tested in this research using the general model, with a particular focus on the conditional link between CEC, AP and WSS. The WR is removed, having proven to have little influence on WSS and to be a constant over the different years. The model described in figure 7.3b performs better statistically and models these conditional relationships:

- In civil society, the existence of freedom of expression and/or authorities that are accountable to their citizens (WGI VA) are favourable conditions within which to increase and maintain efficient WSS services. Irrigation infrastructure is conditioned to skills, suppliers, financial capacities and availability. The opportunity and mobilisation of funds for irrigation development also benefits the WSS sector and households through its multi-purpose infrastructure (i.e. a reservoir provides water supply for crops and energy but also to surrounding villages).

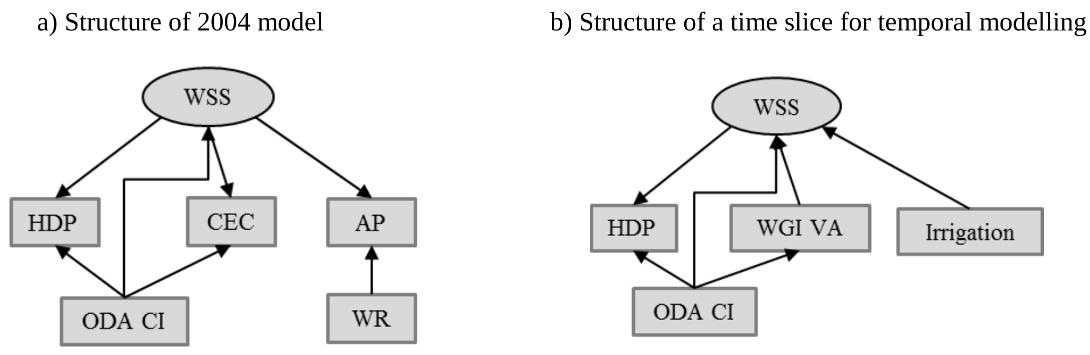


Figure 7.3: the two structures used for modelling WatSan4Dev subset.

The links between variable nodes are set according to the correlations observed. The PCA indicates a high correlation between variables across the three years (time slices). Linear regression analyses were performed to estimate the degree of linearity of the temporal evolution of variables using the goodness of the linear model (Table 7.5).

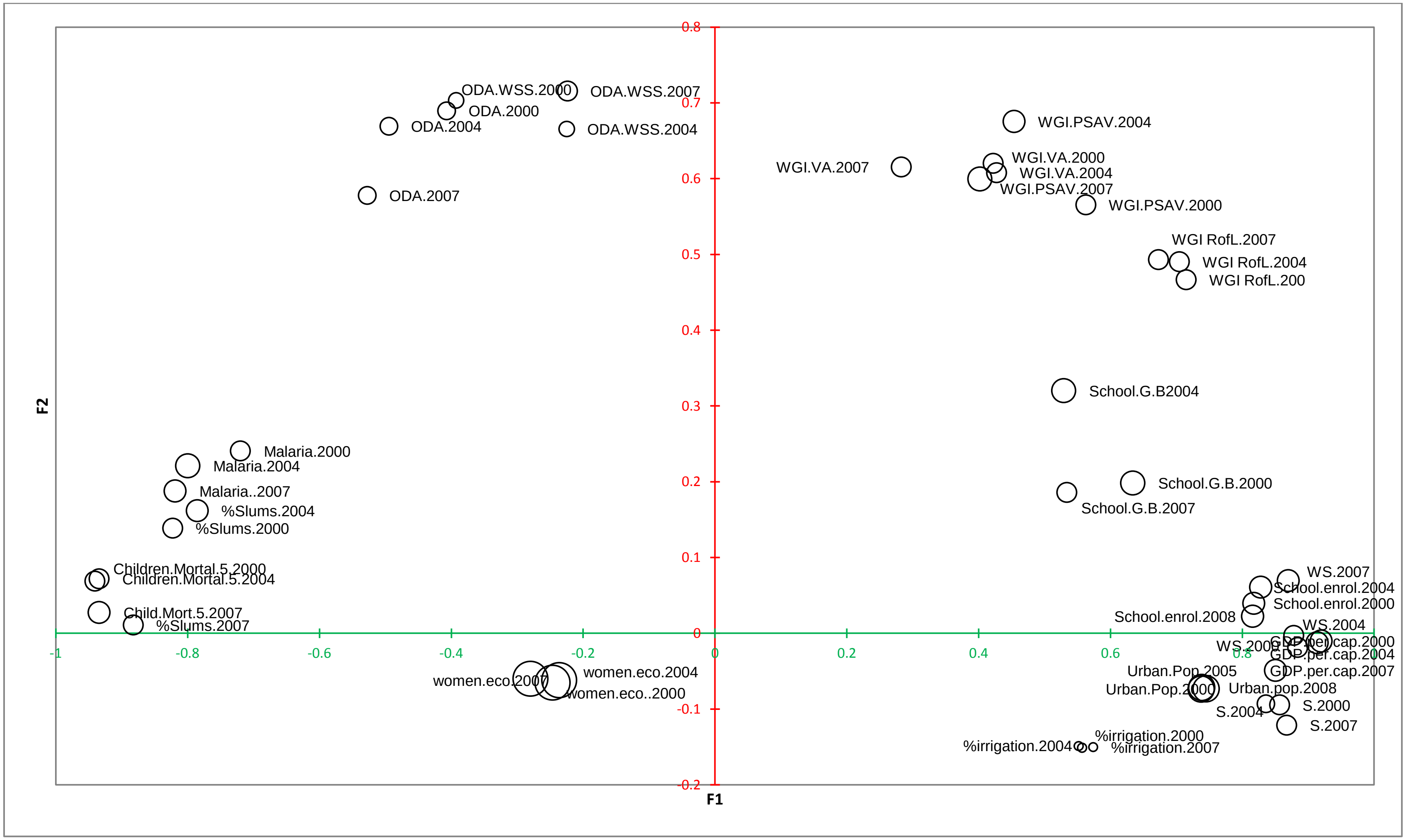
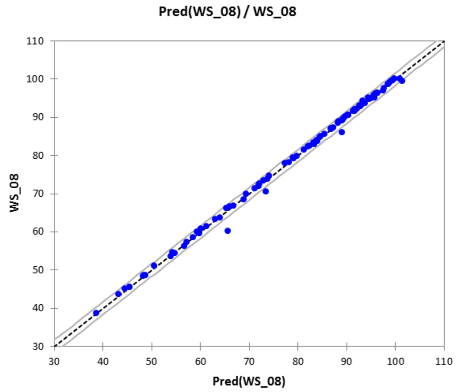


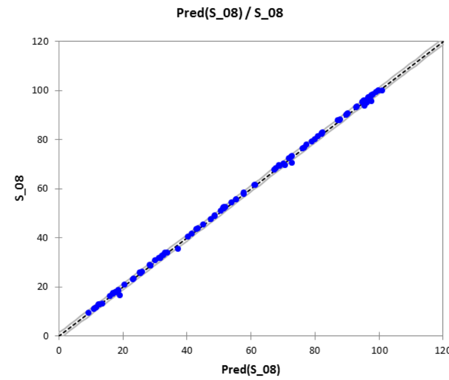
Figure 7.4: Principal component analysis including 2000, 2004 and 2007, where F1 is the green axis, F2 the red axis and F3 the size of the dots. The three first components explain 67.24% of the total variability, with KMO measures above 0.564.

$WS_{2007}=f(WS_{2004}, WS_{2000})$



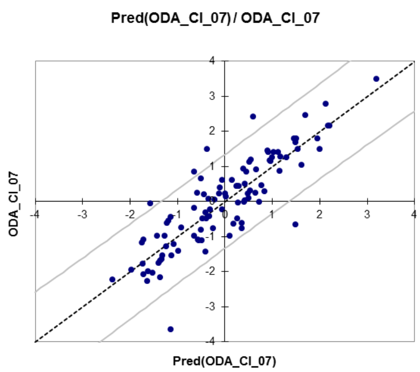
adjusted R^2=1

$S_{2007}=f(S_{2004}, S_{2000})$



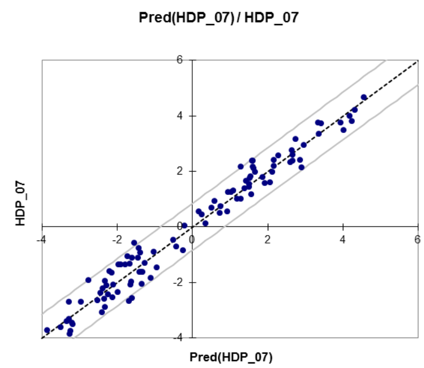
adjusted R^2=1

ODA $CI_{2007}=f(\text{ODA } CI_{2004}, \text{ODA } CI_{2000})$



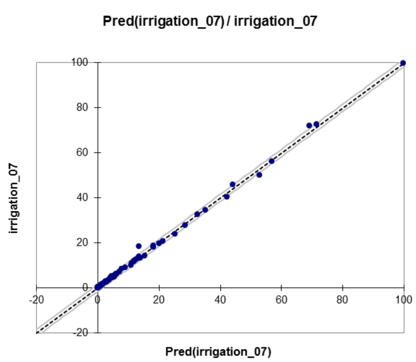
adjusted R^2=0.741

$HDP_{2007}=f(HDP_{2004}, HDP_{2000})$



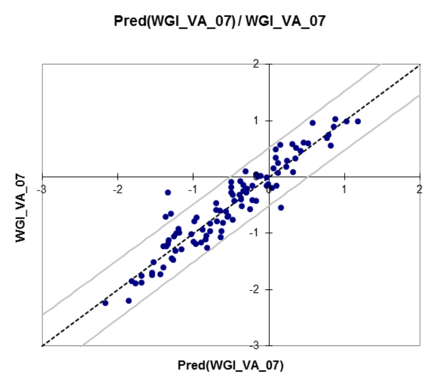
adjusted R^2=0.966

$Irrigation_{2007}=f(Irrigation_{2004}, Irrigation_{2000})$



adjusted R^2=0.998

WGI $WA_{2007}=f(\text{WGI } WA_{2004}, \text{WGI } VA_{2000})$



adjusted R^2=0.891

* Methods: OLS regression with best model selection according to adjusted R^2 and confidence interval of 95%.

Table 7.5: Goodness of the OLS model for each variable in 2007 which includes its previous states.

4. Analysis and modelling of 2000-2007 data using Bayesian networks.

A unique model was built to model the three temporal points (Fig 7.5), giving an immediate picture of the 2000-2007 period. This retrospective modelling approach allows analysing this past period in an easy and graphic way. It also offers the advantage to simulate what could have been done differently modifying the model parameters and to get lessons learnt based on the past real records. The S model performs better statistically than the WS model, showing a lower classification error rates (below 9.2 %) (Table 7.6). One explanation for this may be the fact that the threshold for delimiting categories has not be selected according to the distribution pattern, but rather has been set according to MDGs targets for 2015. The classification error rates are below 12%, which are considered good performance levels. The sensitivity analyses are close, with the first five variables displayed matching the key elements in the WS and S statuses in 2007. The next section details the structure and the performance of these two models.

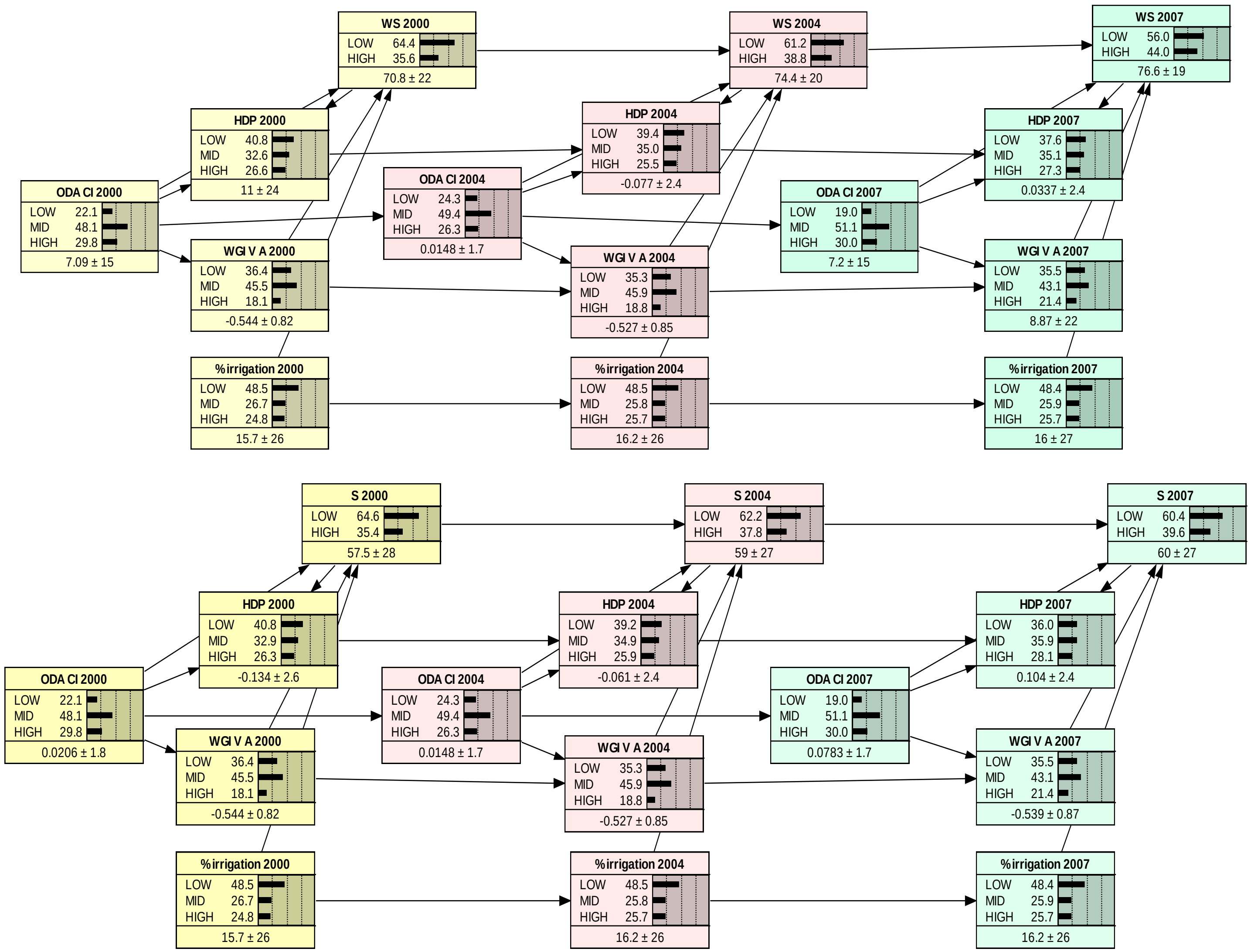
4.1 Description of the temporal models for WS and S

The performance for the WS model exceeds 87%, with a better estimation for the low categories than the high categories and the 2004 and 2007 years. The statistical performance is estimated by estimating the Error Rate, Logarithmic Loss (LL), Quadratic Loss (QL) and Spherical Payoff (SP) coefficients (Chapter 6, section 2.2). The main parameters that influence the 2007 WS variable are the previous state of WS (2004), HDP in 2007, Irrigation (2000 to 2007) and WS in 2000. The previous states of HDP and WGI VA are of medium importance. The external financial flow (ODA CI) has the least influence.

The performance level of the S model reaches at least 90%, with better estimations for 2007 and 2004 and the Low category. The sensitivity analysis is similar to the one from the WS model. The first six variables provided are, in order, S in 2004, HDP in 2007, Irrigation (2000 to 2007) and S in 2000. The ODA CI and WGI VA levels in 2000 come last. The main divergence with the WS model resides in the fact that ODA CI variable ranks slightly higher than WGI VA, in contrast to the WS sensitivity analysis.

WS Model	error rate	LL	QL	SP	S Model	error rate	LL	QL	SP
WS_00	12.12%	0.2943	0.1777	0.9026	S_00	9.18%	0.2557	0.1531	0.9148
WS_04	4.04%	0.1983	0.093	0.9536	S_04	2.04%	0.1835	0.08237	0.9604
WS_07	6.19%	0.2484	0.1316	0.9317	S_07	6.32%	0.2454	0.1302	0.9323
Sensitivity of WS_07	ER	BV	RMS	Rank	Sensitivity of S_07	ER	BV	RMS	rank
WS_04	26.49	0.027568	0.166	1	S_04	77.97	0.026346	0.1623	1
HDP_07	18.17	0.0189028	0.1375	2	HDP_07	63.37	0.022402	0.1497	2
irrigation_07	13.01	0.0135399	0.1164	3	irrigation_07	32.28	0.018614	0.1364	3
irrigation_04	12.1	0.012955	0.118	4	irrigation_04	30.97	0.017218	0.1312	4
irrigation_00	10.55	0.0109809	0.1048	5	irrigation_00	27.94	0.015078	0.1228	5
WS_00	3.5	0.0067641	0.08224	6	S_00	22.11	0.00729	0.08538	6
WGI_VA_07	4.46	0.0046405	0.06812	7	HDP_04	12.84	0.004401	0.06634	7
HDP_04	4.011	0.0041257	0.06461	8	HDP_00	8.452	0.002819	0.0531	8
WGI_VA_04	3.965	0.0041257	0.06423	9	ODA_CI_07	7.742	0.002776	0.05269	9
HDP_00	2.307	0.24007	0.049	10	ODA_CI_04	5.11	0.001919	0.04381	10
ODA_CI_07	1.857	0.0019329	0.04396	11	WGI_VA_07	2.891	0.001911	0.04371	11
WGI_VA_00	1.756	0.0018272	0.04275	12	WGI_VA_04	2.498	0.001857	0.0431	12
ODA_CI_04	1.195	0.0012441	0.03527	13	ODA_CI_00	2.145	0.00108	0.03286	13
ODA_CI_00	0.7436	0.0007741	0.02782	14	WGI_VA_00	1.466	0.000831	0.02883	14

Table 7.6: Performance parameters of WS and S models with results of the 2007 sensitivity analyses for WS and S

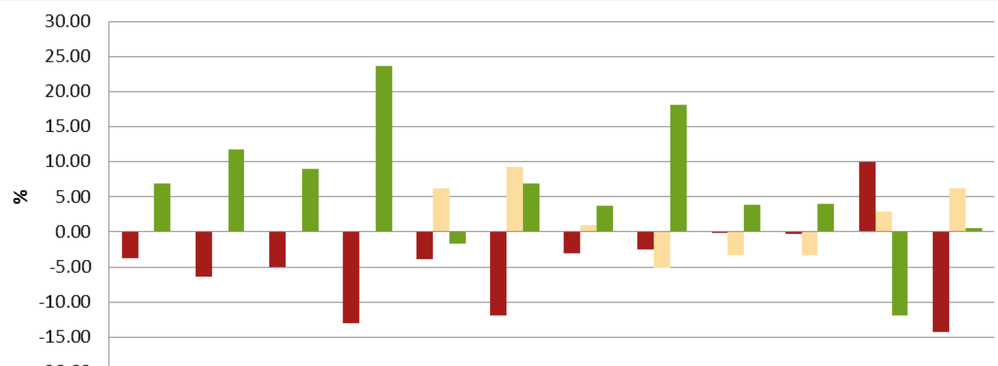


*nodes for 2000 are in yellow, nodes for 2004 are in red and nodes for 2007 are in green

Figure 7.5: DAG of W and S model and initial probabilities

4.2 Longitudinal analysis of WSS

Progress towards the MDG targets for WS improved between 2004 and 2007 (+14%) on performance in the preceding period (+9%) (Fig 7.6). However, the majority of countries (56%) were still experiencing fewer than 88% WS access in 2007. The UN estimated access to WS to be at an average of 61% for the population in sub-Saharan Africa in 2010, with South East Asia, Latin America and North Africa all reaching this threshold during this period (UN, 2012).



	S		WS		HDP		WGI VA		Irrigation		ODA CI	
	00-04	00-07	00-04	00-07	00-04	00-07	00-04	00-07	00-04	00-07	00-04	00-07
Low	-3.75	-6.44	-4.98	-13.04	-3.87	-11.85	-2.99	-2.53	-0.13	-0.25	9.89	-14.29
Mid	0.00	0.00	0.00	0.00	6.15	9.24	0.91	-5.21	-3.36	-3.27	2.83	6.24
High	6.84	11.74	9.02	23.63	-1.68	6.84	3.74	18.14	3.89	4.03	-11.90	0.53

Figure 7.6: Relative probability variations (reference year 2000) over the 2000-2004 and 2000-2007 periods within the three categories of High, Mid and Low.

The various levels of WSS progress are accompanied by the enhancement of HDP during this decade, going from low to medium/high. Over the first period (2000-2004), the improvement is concentrated in the medium category (+6%). This progression is less strong in the second period (+3%), but there is a significant reduction of Low HDP category probability (-8 %). As already demonstrated, the improvement of WSS directly contributes to incremental human development by encouraging better school attendance (Dreibelbis et al., 2013), better health (Turley et al., 2013) and indirectly increasing incomes and effective governance, at least at a local level (Chapter 6, section 3). The following information confirms that the relevant human development indicators largely reflect an upward trend over this period:

1. The median GDP in developing countries increased over the period, even in countries that experienced a slower GDP growth rate post 2005 (IMF, 2007, Fig 1.1 in Chapter 1)

2. Education improved, with primary school enrolment increasing from 59% to 77% in sub-Saharan Africa (UNESCO, 2011) on average.
3. The number of Malaria cases varies from year to year depending on seasonal conditions (high variability), but the number of deaths due to malaria is estimated to have reduced since 2000 in all the monitored regions: Africa (-13%), Latin America (-48%) and South East Asia (-17%). Prevention measures and access to appropriate care have been prioritised within the framework of the MDGs initiative (target 6.C[71]) (WHO, 2011, Chapter 7).
4. The under-five death rate has decreased globally since 2000-2010. However, "sub-Saharan Africa and South Asia are lagging behind other regions. The highest rates of under-five mortality continue to be in sub-Saharan Africa, where 1 child in 9 dies before age five" (Childinfo.org). This specific region reached 109 deaths per 1,000 live births in 2010 (versus 178 deaths per 1,000 births in 1990).The MDG-related targets may not be achieved by 2015, as only Latin America, East Asia and Northern Africa are on track. Trend-wise, fertility rates are decreasing while life expectancy at birth increases. (UN DESA, 2011) estimates that these demographic trends are going to continue over the next 30-40 years.
5. Girls' attendance at school progressed in all the developing countries over the decade (UN, 2012). Latin America and South-East Asia look like they will reach the 2015 target for primary school attendance (the parity index should be between 97 and 103). Sub-Saharan Africa, Western Asia and Northern Africa remain slightly under this target.

This favourable context (HDP) may be undermined by the development of informal settlements in response to a rapid urbanisation process. In slum areas improved WSS services are rarely available and the configuration of slums (their rough land morphology, dense population, extreme poverty leading to illegal plugging) makes WSS implementation and maintenance very complex. "The absolute size of urban dweller increments is unprecedented: Africa gained an average 13 million additional urban dwellers per year in 2005-2010" (UN DESA, 2011). The

[71] Refer to http://www.un.org/millenniumgoals/aids.shtml, last access on the 2nd September 2014.

percentage of the urban population living in slums continued to decrease but at a slower rate between 2005-2010 compared to the previous 5-year period in Africa and Latin America (UN HABITAT, 2010, Table 1.4). This phenomenon is explained by a growth in the proportion of poor people in urban areas. The development of informal urban areas appears as the main barrier to HDP progresses.

The model estimates that good governance—in terms of accountability and freedom of expression (WGI VA)—increased by 18% over the whole period, with a smooth increase (+4%) in 2000-2004 and faster growth (+14.5%) during the subsequent period. It is important to remember that WGI VA measures the practice of democratisation—how the voices of citizens can be expressed and included in country policies/strategies (Neumayer, 2002). Democratisation processes of this type have been observed in developing countries since 1989, following years of colonial and post-colonial centralisation (Olowu, 2003).

In keeping with the culture of intensive agriculture, the proportion of agricultural areas equipped for irrigation has remained stable over the decade (variations are estimated to be below 4%). In fact, irrigation practices little concern and are unequally distributed across our selected countries. Irrigated area as a share of "total cultivated area has reached only 6% in Africa, reaching much higher levels in Asia (37%) and Latin America (14%)" (You et al., 2011). More than two-thirds of Africa's existing irrigated area is concentrated in five countries—Egypt, Madagascar, Morocco, South Africa, and Sudan. (You et al., 2011) highlight that to "achieve the full potential of irrigation development will require significant efforts not only in direct investments but also in institution building, training, and market development". In the sensitivity analyses performed, the irrigation variable has a significant level of potential influence on WS and S coverage and represents good leverage for WSS improvement.

In terms of ODA CI in selected recipient countries, during the first period, the proportion of countries benefiting from medium ODA CI increased, as did the percentage related to Low ODA CI levels. Over the 2004-2007 period, there is a progression in the percentage of countries benefiting from medium to high ODA CI levels. The explanation appears to be linked to the ODA CI implantation geographical changes over the period.

The general ODA levels reflect continuous growth from 2000-2008. At a regional level, the ODA commitment repartition is less linear during the 2001-2008 period (Fig 7.7): commitments towards Africa (49% of selected countries) decreased between 2003-2005, as did those towards Latin America (25% of selected countries) between 2001-2003. Asia (26% selected countries), on the other hand, benefited from a temporary increase in assistance.

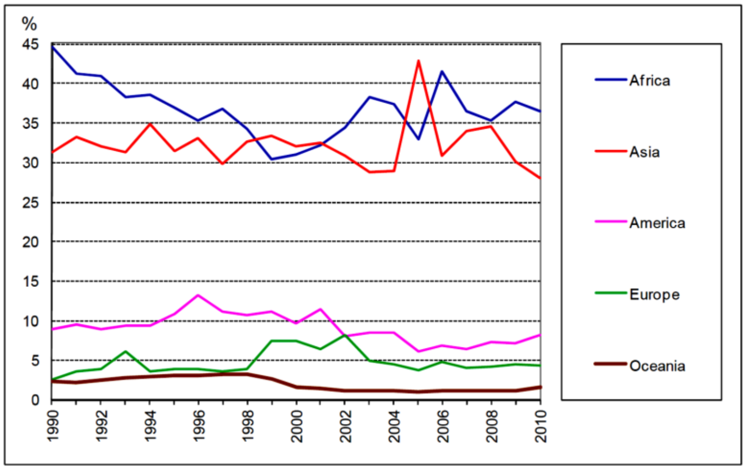


Figure 7.7: Regional shares of total net ODA as a percentage of total ODA (OECD, 2012)

The amount of ODA dedicated to WSS has increased smoothly since 2002 and remains below 7% of the Total ODA (Fig 7.8). (OECD, 2012) noted that there had been "no increased prioritisation of the water sector since 2000-2001". Africa and Latin America benefited from growing ODA invested in WSS during the period 2005-2008, while 2000-2002 saw the level of ODA WSS stabilise and even decrease in Latin America. The opposite trend occurred in Asia: 2002-2005 saw a growth of ODA (up to 43% of the total ODA), targeting WSS improvement; however since 2005 this commitment has been decreasing.

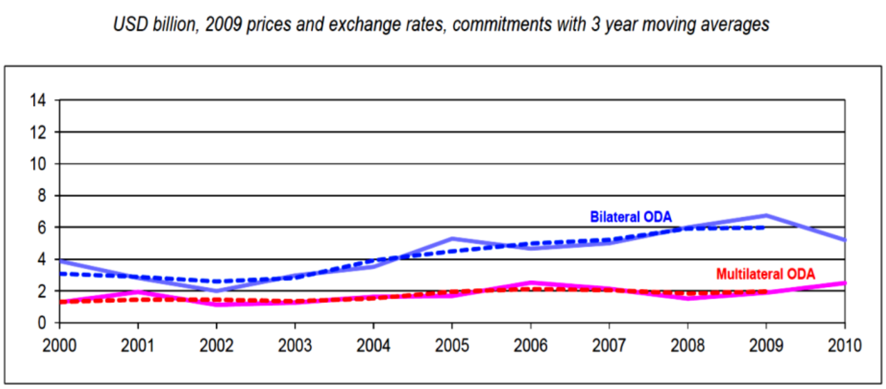


Figure 7.8: Total Assistance dedicated to Water for developing countries (OECD, 2012)

The last variable included in the ODA CI is the WGI VA (which measures political stability and the absence of violence). Over the recorded period, globally the selected countries increased their stability and fought against violence, recording increasingly frequent positive WGI values (above 0.5).

Since 2008, this Official Development Assistance has decreased in the face of the global economic crisis: "it fell by 4% in real terms in 2012, following a 2% fall in 2011. The continuing financial crisis and euro zone turmoil has led several governments to tighten their budgets" (OECD, 2013). Therefore, the increase of mid and high ODA CI categories probabilities may be slow down or reduced if remaining on the same trend in the near future.

4.3 The retrospective simulations

This type of model offers the advantage of being able to learn from the past by analysing past records. This tool allows a running retrospective simulation setting a user-defined target: in this research, 100% of the selected countries reached the targeted levels of access to improved WS and S (88% and 75%, respectively). Simulations 1 and 2 estimate the efforts that would have been needed and the positive influence that would have been had on HDP if this target had been fulfilled in 2007.

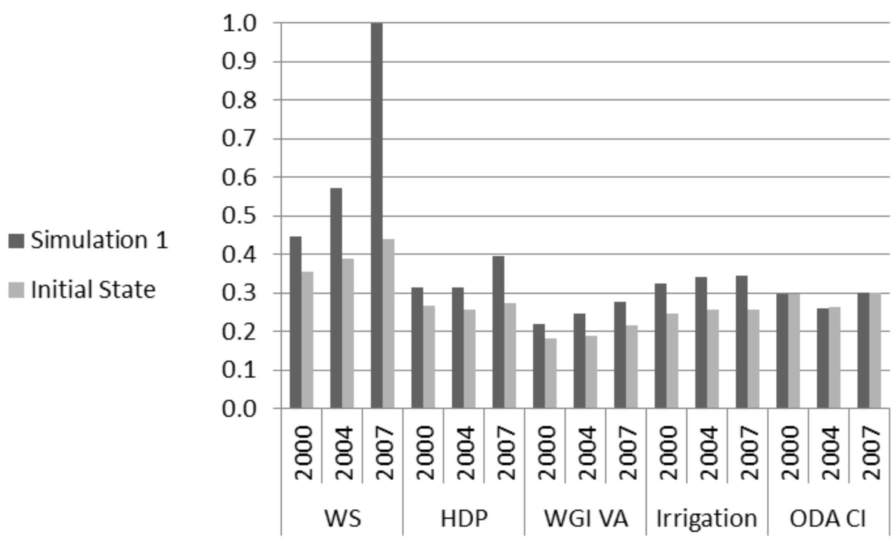
Simulation DBN-WS 1

In this simulation, the high category for WS 2007 is set to 100% and the simulated probabilities are compared with the initial state (IS) (Fig 7.9). The scaling down of the MDGs' global target to selected countries represents a huge improvement. The model estimated the proportion of developing countries with high WS levels to be at 44.5% in 2000 (versus 35.5% in IS), 57.2% in 2004 (versus 38.8% in IS) and suggested this could reach 100% by 2007-2008.

Such WS progress would prompt an increase in human development, ranging from +17% in 2000 to +44% in 2007 when compared to IS. This would signify improved household incomes, better access to education, better control and rehabilitation of informal settlements and better health strategies—in particular those targeting the poor. The development of irrigation also fosters WS access, as conformed to the sensitivity analysis (Table 7.6). The model estimates the high category probability at 32-35% (+32% on average compared to IS), which parallels a decrease in the low category to around 36-35% (-25% in average).

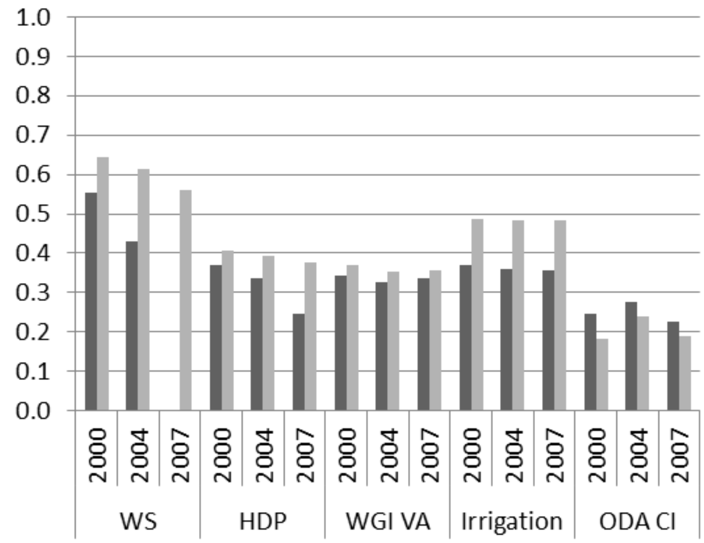
The democratisation and accountability of authorities high category (WGI VA) should be increased by +21% in 2000 and +29% in 2004 and 2007 when compared with IS. It is estimated that this trend will remain steady, reaching 22% in 2000 and 28% in 2007.

Finally, it is estimated that the external input (ODA CI) high category will remain stable over the period (ranging from 25-30%) and in this way will sustain needy countries. Improvements to WS and HDP will be accompanied by the reduction of ODA CI, with an 11-18% increment in the Low category when compared with IS. In fact, high investment of ODA CI first targets poor country with low WSS access. With the improvement of WSS access and generally, the country development, the ODA CI commitment is expected to diminish to reorient these resources to poorest countries.



a) Probabilities for High category

Continued…



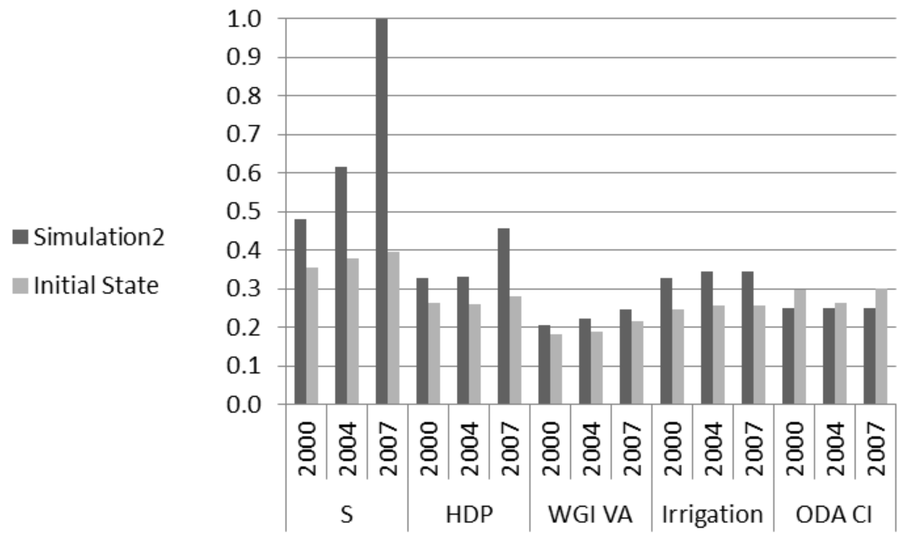
b) Probabilities for Low category

Figure 7.9: Probabilities for High and Low categories in the Initial State (IS) and in Simulation 1

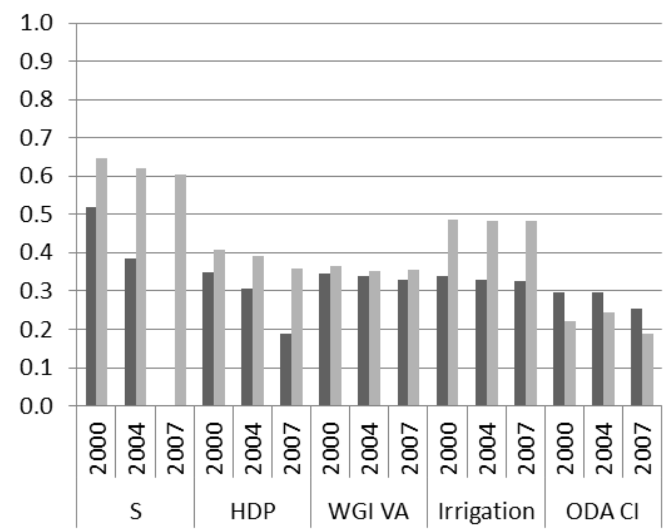
Simulation DBN-S 2

Achieving the S target (75% access) will require significant effort, following a similar process to WS (Fig 7.10). In simulation 2, the high S category is estimated to be at 48% in 2000 (versus 35.4% in IS), 61.6% in 2004 (versus 37.8% in IS) and to potentially reach 100% in 2007. The gap between the simulated level and IS was already estimated to be around 35% in 2000, which is larger than in WS model (+10 points).

In terms of poverty reduction, the percentage of countries with high HDP is estimated to grow up to 32.6% in 2000, 33.1% in 2004 and 45.6% in 2007. Irrigation levels are even higher, at 32% greater on average than the IS. Poor accountability and the weakness of civil society (Low WGI VA) drop by 5.5% on average, while good WGI VA is increased progressively (+14% in 2000, +18.5% in 2004, +15% in 2007). External Assistance (ODA CI) is estimated to decrease over the period, with an increase in the countries receiving low ODA CI (+33.4% in 2000, +22% in 2004, and +33% in 2007 in comparison to IS).



a) Probabilities for High category



b) Probabilities for Low category

Figure 7.10: Probabilities for High and Low categories in the Initial State (IS) and in Simulation 2

5. CONCLUSIONS AND DISCUSSION

These retrospective models are supporting tools that facilitate our understanding of multiple variables trends that occurred over the 2000-2007 period. Bayesian networks allow for the computing of the initial state (IS) so as to describe the probabilistic distribution of each variable for each time slice. Their statistical performance was good, with classification error rates of below 12%, and therefore the models fit the data well. WS and S are sensitive in a similar way, as the six first variables involved in the WSS changes are identical. The S model is slightly more sensitive to ODA CI, while WS is slightly more sensitive to WGI VA. The mechanisms behind WS and S seem to function similarly over time. This similarity may be due to the scale (101 countries) and synthetic nature of the indexes, which may mask potential divergences in detail (either within a country or within sub-indicators). The main issue is still that S is undervalued by authorities and donors.

The range of progress made over the period corresponds to the nature of the indicators: they measure, like HDP or WGI VA, complex mechanisms involving societal organisation and evolution. Therefore, the trends computed by the models show a relatively small range of changes. WS progressed more during 2004-2007 than in the previous period, while S showed a more regular trend. Access to S improved by 12% and by 23% for WS. HDP progressed in particular, with poor HDP reduced by 11%. Democracy and accountability (WGI VA) also increased by 18%. Irrigation remained stable over the period. The external assistance flows associated to political stability (ODA CI) progressed over the period: 2000-2004 saw their progress towards low-medium categories, while the 2007 probabilities are similar to those for 2000. Since 2007-2008, this level of support has fallen due to the global economic crisis. In addition, the political instability and conflicts have also increased in Africa and the Middle East where the Arab revolutions since 2011 are expected to reduce the WGI PSAV value.

Beyond the qualitative and probabilistic analysis of the dataset, the models allow running retrospective scenarios, modifying parameters of variables. When modifying one or several parameters, as in the 2004 models, the initial probabilities are recomputed to provide an estimation of the variables according to new conditions set. The MDGs targets for WSS (88% WS and 75% S) are set as targets at a global level for 2007 for our countries. As a demonstration, the example of WSS targets being met by all the countries considered in 2007 was tested. The aim was to estimate the effort required and potential consequences of such achievement. For instance, reaching 88% WS access for all countries represents a huge amount of effort over seven/eight years. This improvement is estimated to increase HDP by 17-44% when compared to IS. Irrigation practices (high probability) should rise from 32-35% (+32% on average compared to IS), combined with a parallel decrease in the low category of around 36-35% (-25% in average). The WGI VA should progress steadily and 28% of countries should reach high WGI VA levels.

The number of temporal points included in such a model is crucial to how well it can identify mid-long term trends. The limited time period implies high probability changes to achieve the targets set. These estimated probabilities are therefore difficult to implement in reality. The extension of the observation period would reinforce the progressivity of the changes needed and provide achievable estimations. The temporal points are considered every 3-4 years, as annual reporting is not always possible. In addition, annual evaluations would not necessarily add value to the modelling process for two reasons: 1) yearly changes are limited, as most of the indicators measure societal or environmental aspects that follow progressive patterns of change i.e. education or water demand (only ODA CI is subject to a more volatile evolution); 2) yearly modelling would rapidly increase the complexity of the model by multiplying the number of nodes. The analysis cannot therefore be simplified.

The retrospective approach is a robust approach allowing better understanding the complex mechanisms involved between the different variables. The development of a predictive/forecasting approach could complement this in an interesting way. This approach aims at generating the variables probabilistic distributions at time *t*. A review of existing Bayesian predictive/forecasting modelling, constraints and potential paths is offered in the concluding remarks (Chapter 8, section 2.2).

CHAPTER 8: CONCLUSIONS AND FURTHER RESEARCH

This chapter first summarizes the outcomes of the work and highlights their interest regarding the gaps identified and the research questions. Its section 2 details the main limits of this work and explores the way forward to extend the work done.

1. CONCLUSIONS

Despite the progress made towards the MDGs WSS targets, WS and in particular S are far from being available to all as measured by the Joint Monitoring Programme. The 2012 MDG report highlighted that the water supply target was achieved in 2010, with 88% of the worldwide population having access to an improved water source (UN, 2012). South Asia, Eastern Asia and Latin America slightly passed their regional targets. With an S regional target set to 75%, Sub Saharan Africa is not on track, despite significant efforts made over the last two decades (+12 points of percentages of the access to pipe and improved water supply). The sanitation situation in 2010 is even worse (UN, 2012). Only Latin America, Northern Africa and Western and Central Asia reached at regional level relatively high sanitation access rates (above 80%).

Starting from this statement, this research aimed at designing a methodology for the analysis of key variables in WSS delivery and country development at worldwide level. This research attempted to respond to several gaps identified in the literature review (Chapter 2, section 4.1).

- Gap 1: the cross-cutting analysis of water supply and sanitation services delivery, and more generally, of water management, is mainly performed at river-basin and/or local scale but not for all river basins to allow the aggregation of national and continental figures (Chapter 2, section 2.1).

Building on (Sullivan et al., 2003) and (Giné Garriga, et al., 2009), the WSS analysis was scaled up to national country level respecting the multi-dimensional characteristic of the WSS sector. Official Development Assistance was added in the analytical framework following the principles stated in (Sullivan et al., 2003). Therefore, this work built a multi-dimensional database to cover socio-economic development, governance, official development assistance and environmental indicators extending in scope and scale these available frameworks (research question 1, Chapter 2, section 4.2). (Giné Garriga, et al., 2009) developed a BNs model to support the integrated water resource management of a Kenyan river basin. This type of modelling is a useful tool to support strategy/decision making because it provides understandable information on complex issues. It also manages missing data and complex (linear and non-linear) behaviours.

A tailored analytical framework was built that combined multivariate analyses and clustering methods with Bayesian Networks modelling to analyse the gathered data (research question 2, Chapter 2, section 4.2). This framework extended the existing methodological approaches, from composite indicator (Sullivan, 2002), (Foster et al., 2005), (Alrike & Santos, 2010), (Mihci et al., 2012) to correlation and regression methods (Rodgers, 1979) (Nyong & Kanaroglou, 1999), (Wolf, 2007), (McGarvey et al., 2008), (Botting et al., 2010), (Kiendrebeogo, 2012), (Dreiblebis et al., 2012), (El Fadel et al.,2013), (Tigabu et al., 2013). The relationships between the variables involved in WSS delivery to the population have been mapped considering the whole variability of the data (research question 3, Chapter 2, section 4.2). In addition, this new Bayesian modelling enabled to simulate the probabilistic distributions of variables according to user-defined parameters. This scenario feature supports decision making through the estimations of WSS probabilistic changes according to potential actions foreseen. This framework was also extended to the temporal analysis of 2000-2007 period. The retrospective modelling developed provided a longitudinal analysis of the dataset. This model allowed to simulate the efforts required to reach the MDGs WSS targets: 88% of the population having access to improved water supply and 75% with access to sanitation facilities in 100% of select countries.

- Gap 2: the Official Development Aid has been studied with specific variables but not with the full scope of variables involved in WSS delivery (research question 4, Chapter 2, section 4.2).

(Botting et al, 2010) found a correlation between the ODA-WSS and Water supply access, but not with Sanitation. Additionally, a positive association was observed between sanitation and the reduction of children mortality. (Wolf, 2007) observed that the ODA volatility is significantly and positively associated to WSS delivery (regression coefficient between 0.002 to 0.005 with 1 percent significance level), while the correlation with governance is limited (the control of corruption index is not significantly correlated with WSS while the Press freedom index showed a significant correlation, 0.002 coefficient with 1 percent significance level). According to our results, in 2004, the multivariate analysis and the BNs modelling indicated that ODA overall flow is positively correlated with poverty aspects and poor WSS coverage (Chapter 5, section 3.4 and Chapter 6, section 3). This means that ODA mainly targeted weak countries (low HDP at 45% probability) with poor access to WSS services (between 60-65% probability) in 2004. Completing (Wolf, 2007) and (Witford et al, 2010), ODA WSS is found significantly correlated with political stability and absence and violence (WGI PS AV), however no significant correlation with WSS services is observed (as Botting, et al., 2010). This means that the ODA WSS was not really addressed to support WSS and efforts in this way are to be done.

- Gap 3: The ODA is extensively studied from the donor point of view while the analysis of recipient countries, that can provide insights on where ODA is committed and needed, remains limited.

(Tjonneland, 1998), (Tuman & Ayoub, 2004), (Larru & Tezanos-Vasquez, 2012) among others studied the amount, targets, type etc. of ODA provided. (Knack, 1999) and (Botting et al, 2010) examined the impact of the aid concentring on a specific aspect, respectively governance and children mortality. Only (Adler, et al., 2010a) looked at ODA from the country recipient point of view and classified the recipient countries according to their efficiency in using their financial resources (including ODA) to achieve MDGs targets. Following this view, this work has looked at analysing the recipient countries situation on a wider cross-cutting and common base (research question 5, Chapter 2, section 4.2). In this way, five country profiles have been defined according to the level of development, the commitment to environment, the ODA received, the water resources availability and the water demand. Countries from profiles 4 and 5 are identified as weak countries in terms of WSS delivery and widely in terms of development. These results are further developed in section 1.4 in this chapter.

The following sections summarize the main outputs of this research.

1.1 The development of an analysis system for the water sector

Within the framework of this research (Fig 8.1), the WatSan4Dev dataset is a new development database (collecting socio-economic, environmental and governance indicators) that brings together the available indexes across different international institutions. The data collection process involved was extensive, including up to 130 indicators (Appendix A). From a preliminary analysis (of missing values, relevance and reliability of the variables), only 41 variables were included in the WatSan4dev database. The database was created on a national scale because of data availability, the fact that ODA is computed at country national scale, and above all, to allow cross-country analyses—the developing countries being a heterogeneous group. They qualified based on their socio-economic country status, the water demand of their population, water resource estimates, the official development assistance flows received and governance indexes. Because of the different origins and gathering methods, raw data was processed and robust missing data processing methods applied to fill the gaps (Honaker & King, 2010). As a result of this statistical processing, the coherence and robustness of the different indicators increased.

However, the WatSan4Dev dataset should only be used for qualitative estimations and analysis. Of the variables in WatSan4Dev, 25 out of 41 have proven to be independent, coherent with the literature and relevant knowledge from the field (Dondeynaz et al., 2012).

This research identified and analysed both the key variables involved in WSS service development and the different profiles among the selected countries. A Principal Component analysis (Jolliffe, 2002) and a Factor Analysis (Thurstone, 1947) allowed for the measuring of relationships between variables (Chapter 5, section 3). Combined with clustering methods (k-mean clustering), five country profiles have been characterised. This analysis identified the less advantaged countries and any weaknesses that could be used to focus the efforts of authorities and donors. The visual graphs and maps provided understandable overview of a country or a profile according to the five dimensions considered.

Building on these multivariate analyses, Bayesian Network (BNs) models were developed separately for WS and S to provide an easy tool with which to run scenarios. BNs offer several advantages with regards to the WatSan4Dev dataset—they use conditional probability and they can cope with diverse variable types, non-normal distribution and missing data (Lee et al., 2005), (Aguilera et al., 2011). In essence, the BNs models provide estimations of potential improvements to WSS and, more widely, of human development according to the input parameters set. Several simulations were run and are detailed in the following paragraphs.

The robustness, simplicity and flexibility of the combined methods were the main drivers of this research, ensuring the replication of the methodological process according to the variables and series available. In fact, in this study the analyses and modelling were mainly applied to the 2004 dataset, because it was the most extensive dataset available at the beginning of this work. However, this work has also explored ways of extending the analysis and modelling to apply to a longer period—in this work 2000 to 2007. A retrospective modelling approach was tested using DBNs and validated. The extension of the observation period is possible with the provision of additional data by providers.

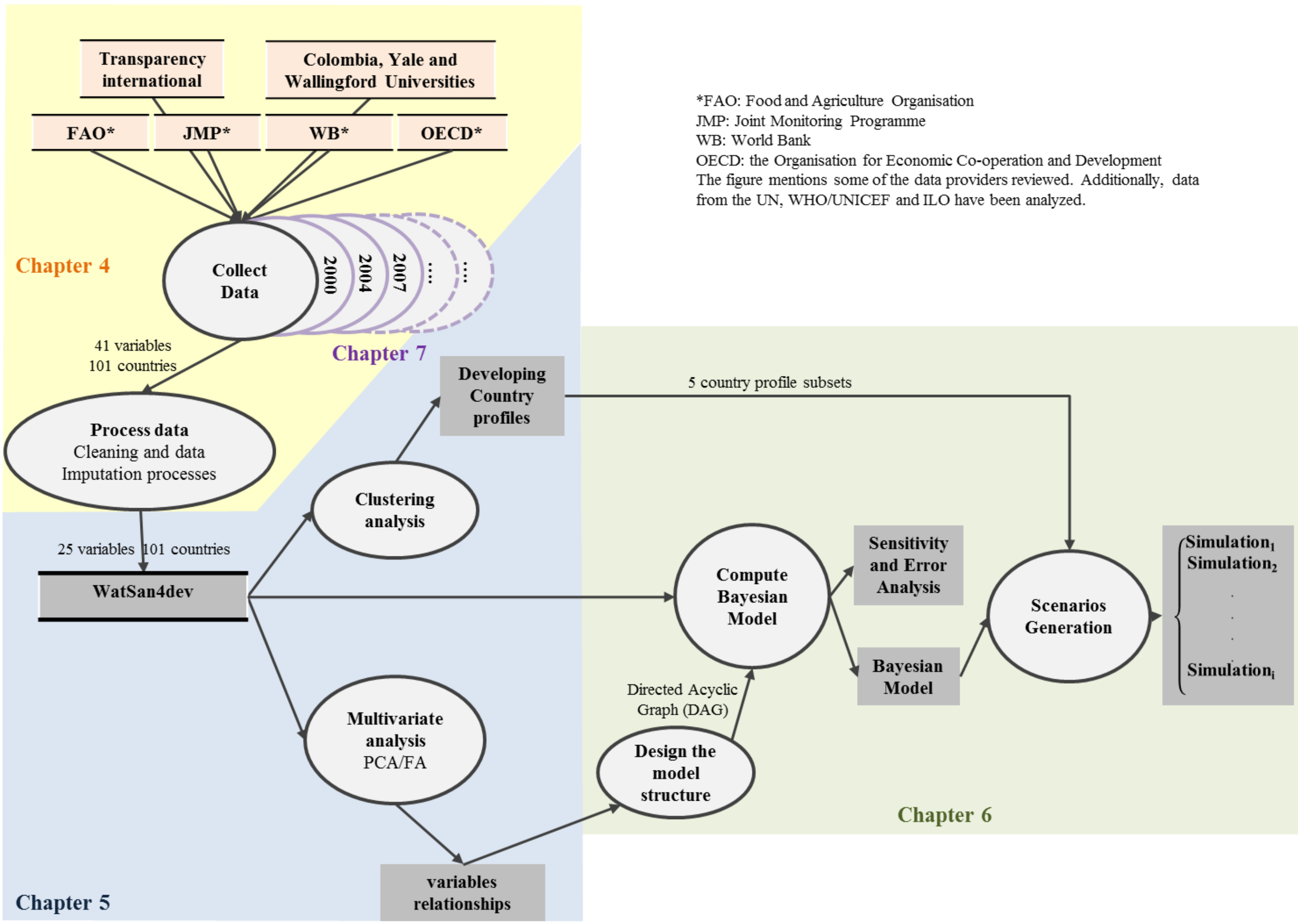


Figure 8.1: Recapitulative Functional Data flow Diagram where a rectangle with two black borders represents a database, a circle indicates a process and a grey rectangle a result. The external input data is in orange and the results of this research are in grey.

1.2 The analysis of 25 variables and 101 developing countries

The multivariate analysis was performed to respond to a twofold aim: i) to establish the relationships between the variables of the WatSan4Dev; ii) to synthetize the important volume of information into few and coherent factors. This facilitates the geographical analysis and simplifies the interpretation of the results. The 25 variables subset is organised around five main factors.

Factor 1 (HDP) represents the human development in a country not only as a function of income, life expectancy and education (as is the case in the HDI index), but also taking population and living conditions (Urban Pop, % Slums, WS, S), and advance governance aspects (WGI RofL) into account. It extends the concept developed by the Human Development Index (HDI). **Factor 2 (AP)** expresses how economic activities—namely agricultural, industrial and municipal activities—put pressure on water resources. **Factor 3 (WR)** gathers variables quantifying the amount of water resources available. **Factor 4 (ODA CI)** represents the external financial flows ODA and ODA WSS conditioned by the level of political stability and absence of violence in the country. **Factor 5 (CEC**) links the involvement of the country in International Environmental Agreements together with governance related to Voice and Accountability indicator (WGI VA).

Constructed around these five factors, five different country profiles are proposed in this work, with Profiles 4 and 5 being the most disadvantaged and Profile 1 is the most favourable. This new classification—based on multiple indicators—is a different way to prioritize countries in difficulty indicating weaknesses hindering development and well-being. The EU institutions and other donors could use these profiles to establish better cooperation strategies and to simulate the impact of their policies. The EU institutions are interested in cooperation programs with countries from profiles 3, 4 and 5 (Chapter 8, section 1.4) (European Commission, 2011).

In fact, the countries in Profiles 4 and 5 are less advanced with regards to WSS access (and socio-economic development is a priority), but the efforts need to be put in different areas according to models built: a) countries in Profile 4 are facing water stress due to scarcity in a context of low development. Efficient resource management, mitigation measures and adapted solutions are crucial to improving WSS; b) in the case of profile 5, efforts should be concentrated on ensuring political stability and a reduction in violence to combat low development (natural resource-centred economy, disorganised, or weak authorities) implied by such a context. This is an a-priori condition to ensure the multiplication of individual/collective initiatives towards sustainable development. It was also highlighted that despite their situation, profile 5 countries received in 2004 few external support (ODA CI).This is probably due to the risk of implementing investments in countries with low governability.

Countries in Profile 3 need to make sanitation a priority. Within an irrigated agricultural context water supply services are often ensured, while basic sanitation development remains neglected. Rural areas continue with feasible/acceptable non-improved practices such as open defecation, with investments in basics sanitation facilities not seen as a priority for poor households. It is worth noting that a majority of countries in Profile 2 benefited from a significant level of ODA (-0.036 centroid value), while their WSS level is almost 100%. A detailed analysis of the external assistance provided to these countries should be considered to better understand on what kind of assistance was being provided.

From the 2004 dataset, urbanisation plays an essential role in access to improved water supply and sanitation facilities, factoring in the kind of urbanisation (formal versus slums). This generalises to selected developing countries the results found in local-regional studies, for instance in Africa (Wambui Kimani-Murage & Ngindu, 2007), (Foppen & Kansiime, 2009), or in Asia (Shakar & Rahman, 2007). Slums developments are expected to increase and, by 2050, urban dwellers will likely account for 66% of the population in less developed regions (UN, 2010). Extending access to sanitation and water supply in these conditions may be hindered or slow down in a near future. The positive correlation between access to water supply and sanitation (WSS) and the reduction of Child mortality under 5 years (Child Mortality-5) was made clear in this work (Chapter 5, section 5), in compliance with (UNICEF & WHO, 2009), (Botting et al., 2010), (Butala et al, 2010). The positive correlation of WS services coverage with the level of income was also confirmed as shown by (Wolf, 2007). Irrigation was found to encourage water supply improvements, most probably thanks to multi-use of the irrigation infrastructure (i.e. irrigation scheme associating reservoirs and networks used for both the crops and the rural community supply). These findings emphasised the potential role of irrigation in term of the agricultural sector organisation and it positive effects on WSS access complementing (Kiendrebeogo, 2012). The linear models showed that environmental conditions referring to water resources (total water renewable resources, precipitations, desert risk) are secondary when it comes to factors explaining the WSS level.

This regression analysis was considered useful even explaining 68-70 % of the total variability of the dataset. This can be considered enough as the results are cross-validated by other research works in the domain. Part of the relationship between WS and S is non-linear (Fig 6.8) which is confirmed by the 2012 MDGs report (UN, 2012): S access is lower than WS access (-21 points of percentage in average). Therefore, Bayesian networks modelling were developed to explain this non-linearity behaviour of the variables.

1.3 Bayesian networks modelling for 2004

The proposed models are efficient tools for identifying and measuring probabilistic changes to the thematic variables and WSS levels (statistical performances range from 80 to 95%) (Dondeynaz et al., 2013). The general models synthesise and estimate the general behaviours involved with WSS access using composite indicators (CIs), while thematic sub-models measure the role of each sub-variable into the CI to scale down the analysis/modelling to each thematic dimension.

The general models were built separately for both Water Supply and Sanitation to observe and identify potential differences between both variables. The statistical classification error rates are less than 12% for Water Supply (WS) and 10% for Sanitation (S). The thematic models show error rates ranging from 5 % to 20%.

First, BNs models confirmed that improving WSS access is good leverage for the alleviation of poverty. The WSS access is strongly associated to the country development (around +35 % probability change of WSS) and is first sensitive, as expected, to the income level (Wiltford et al., 2010). The benefit of WSS on country development is coherent with the reduction to 10% of probability to have low human development in case of 100% WS/or S coverage. The urbanisation level is the second strong factor associated to development with the limit of slums development (Butala et al, 2010) (Turley et al, 2013). Health care and advanced governance (WGI RofL) complete these key factors. The latter stresses the role of “good of governance” in the sustainable development of WSS services that included in this work government effectiveness, control of corruption, regulatory quality (Witford et al., 2010), and rule of law capacities. These results follow (TI, 2008) and (Davis, 2004) empirical observations regarding the negative consequences of corruption on WSS services management. These findings diverge from (Wolf, 2007) that found no significant interaction between WSS and control of corruption when analysing the African dataset.

The models highlighted that WS is more sensitive to environmental commitment and level of democracy/freedom with a positive change of 21 % than S with +8% change. This behaviour can be an expression of the less consideration of sanitation as an important environmental issue. Sanitation also has suffered of cultural/psychological barriers that make sanitation a private issue and limit the collective initiative/management to fulfil the sanitation gap. More evidences and research should be found to support this hypothesis.

The ODA CI has significant effect on both Water supply and Sanitation access (between 7 and 10 points of probability changes) whereas (Wolf, 2007) and (Botting, et al., 2010) found only significant correlation between ODA-WSS and WS. This divergence is due to the fact that this research has aggregated the overall ODA flow, with the specific ODA-WSS and the WGI PS AV. High-level ODA CI first benefits to poor and middle development countries at 77% probability. Funds are also invested in relatively less poverty-affected countries (at 21% probability) probably due to the geo-strategy of the donors. The high ODA CI also has positive effects on CEC fostering environmental commitment and civil freedom up to medium level (at 44% probability). This estimation is higher for countries in profile 4. More investigation is required on these specific countries to identify potential explanation of this.

The development of irrigated agriculture (AP) is also associated with WSS development (around 20 points of probability change). The positive interaction between WSS sector and agricultural sector rather occur in the context of rural organisation around irrigation. The irrigation scheme favours the multiple use of the water supply infrastructure. This is confirmed by (Hinrichsen et al., 1997) and (Rosegrant & Ringler, 1998). This positive interaction between agriculture and WSS may be increased through the reuse of human excreta or/and wastewater sludge treatment once the cultural/psychological barrier is overcome. (Haddad, 2005), (Al-Sa'ed & Mubarak, 2006) demonstrated the existence of a cultural /psychological barrier with regard to the reuse of treated water in agriculture. In these studies, the majority of rural populations declared to not accept such practices even if this option is not fully rejected. The conditions of the treatment appeared key in the wastewater reuse acceptance. The expansion of such practices is little documented and measured at worldwide scale, hence the difficulty to confirm this complementarity (Raschid-Sally & Jayakody, 2008).

Lastly, this analysis highlighted that water availability (WR) counts for little in the context of population's access to WSS and in the water demand distribution by economic activities. Table 8.1 recapitulates the probability change of WS, S and composites indicators according to the 25 selected variables. This table offers a synthetic view of the probabilistic effects of each sub-variable (i.e. Particip, WGI VA, and ODA CI for CEC model) on the tested variable (i.e. CEC composite index). The table 8.1 provides the probability change of the observed variable for a specific category (i.e. high category) according to the parent nodes settings (i.e. HDP high category is set to 100% probability).

Variables	Global models		
	Suitable Level	High WS probability change	High S probability change
HDP	HIGH	35.9	34.5
AP	HIGH	18.1	21.3
CEC	HIGH	21.7	8.1
ODA CI	LOW	10.8	7.3
	Suitable level	High HDP probability change	
GDP per cap	HIGH	53	
Urban pop	HIGH	43.1	
Child Mortal -5	LOW	33.7	
% poverty	LOW	33.2	
% slums	LOW	31.9	
Malaria	LOW	27.8	
WGI RofL	HIGH	24.3	
School enrol	HIGH	24.4	
School G/B	HIGH	18	
Femal eco	LOW	16.6	
ODA CI	LOW	0.3	
	Suitable level	High CEC probability change	
Particip	HIGH	39.4	
WGI VA	HIGH	23.8	
ODA CI	LOW	5.6	
	Suitable level	High AP probability change	
% irrigation	HIGH	59.3	
Water Use in Agriculture	HIGH	35.7	
Total Withdrawal	HIGH	14.7	
WR	LOW[a]	-1.7	

[a]The probability change of water demand (AP) is estimated in case of reduction of WR that could potentially occur in areas subject to precipitation pattern changes.

Table 8.1: Recapitulative table on probability changes (differences between recalculated and initial probability) when setting each variable at 100 % probability, i.e. high WS probability is incremented by 35.9 percentage points when HDP is set at 100% high probability.

1.4 Country profiles

The general model was extended to country profiles, with their specific behaviours analysed and divergences highlighted. These models provide the advantage to deeper the analysis of each group of country and run specific on-demand simulations.

Countries belonging to Profile 4 experience desertification effects in a context of limited water resource availability (67% of countries have low water resources). Therefore, the development of water stress/scarcity adapted or mitigation measures can be crucial. These countries benefits from minimum governance (political stability and civil society freedom) and concern about the environment. Models suggested that ODA CI has a better efficiency either on S and/or CEC than in a general behaviour certainly due to this favourable context. However, significant efforts are still needed regarding the HDP dimension—the control of urbanisation, the organisation/reinforcement of health care and education and, the improvement of “advanced governance” (that relates to effectiveness, corruption control, making and ruling laws).

Profile 5 countries have weaknesses in almost all dimensions: low HDP (68-70%), low CEC (48-50 %), low values in AP (65-67%). Despite their less advanced situation, these countries benefit from average levels of donor investment (mainly medium levels of ODA CI for 57% of countries), which can often be explained by the persistence of political instability. Therefore, ensuring basic governance conditions, stability and the absence of violence and accountability within the citizen environment is essential to: 1) fostering economic development and reinforcing rural development which support poverty alleviation and, 2) appealing for additional external ODA if necessary.

The countries belonging to profiles 4 and 5 should have the priority of the international support as they are the less favourable with regards with the variables selected. The EC countries priority in 2006[72] included in majority the countries from profile 3, 4, 5 from African, Caribbean and Pacific (ACP) region where access to WSS services are low and HDP limited. However, Mediterranean (Morocco, Egypt, Tunisia, Jordan, Algeria, Lebanon)

[72] 2006 EC priority list for Aid effectiveness: http://ec.europa.eu/europeaid/how/ensure-aid-effectiveness/documents/2006-list-of-46-ec-priority-countries_en.pdf, last access on the 11th August 2014.

countries from profile 2 are being supported by the EC, probably for geo-strategic and proximity reasons.

1.5 Extending the modelling for the 2000-2007 period

The two models built for WS and S allowed for measuring the evolution of the variables over the period 2000-2007 in an easy and readable way. The number of variables was reduced to 19 because of missing data or unavailability of data: i.e. water withdrawals by use represent 75% of the missing values. For 2004 the modelling was adapted and modified as follows: i) MDGs targets for WSS (88% WS and 75% S) access were used as the limits of the High and Low categories; ii) the environmental commitment (CEC) and the distribution of economic activities (AP) indicators were conditionally linked as a favourable framework for WSS development.

The range of types of progress made during the whole period corresponds to the nature of the indicators: they summarise complex mechanisms, touching on societal evolution, as HDP or WGI VA. Therefore, the trends computed by the models show a relatively small range of changes. Between 2000 and 2007, the coverage of WSS improved by 11% for Sanitation and 23% for Water supply. WS progressed more during 2004-2007 than the previous period, while S shows a more regular trend. Access to S improved by 12% and WS by 23% over the whole period. HDP was reduced by 11%. Democracy and accountability (WGI VA) also increased by 18%. Irrigation practice remains stable over this period. The external assistance flow (ODA CI) shows a more volatile evolution: 2000-2004 sees the progress of low and medium categories while 2007's probabilities are similar to 2000.

Beyond this quantitative description of the dataset, this retrospective approach allowed the estimation of explicative variables considering the hypothesis of the achievement of WSS targets in 2007. The main added value of this modelling approach resides in that past records are the basis of simulation, therefore, contributes to its robustness. WS and S variables were sensitive to the similar six variables: i) 2004 WS or S state ii) the HDP 2007 state iii) irrigation development state for 2007, 2004 and 2000 and, iv) 2007 WS or S state (Table 7.6). The S model appeared slightly more sensitive to External financial Assistance (ODA CI), while WS was slightly more sensitive to democratisation and accountability variable (WGI VA). The latter two variables came at a secondary level but their roles should not be underestimated as they implied probability changes between 15 to 30 % depending on the simulations performed.

Better levels of commitment to sanitation were and are still needed. International Sanitation year was in 2008, when the international community raised awareness around this gap. The effects are not visible in the 2010 statistics, thus efforts should be continued.

2. Main limits and Further Research

2.1 The limitations

The data availability and harmonisation were the first limit of this work as for other research studies (Adler et al, 2010a), (Botting et al., 2010), (Gine Garriga, 2009) etc. The missing data was more linked to the non-measurement of relevant aspects (as for water quality) than on the number of observations selected (Chapter 4, section 4).

The first dimension that is not included in the WatSan4Dev relates to the water quality measurement. In fact, the link between water resources and WSS was not observed in the multivariate analysis (Chapter 5, section 3.3). It is supposed that including water quality indicators would have revealed this relationship. This work suggests to dedicate efforts in developing and consolidating such indicators, and to integrate water quality in the MDGs' monitoring process. The analytical framework can be scaled down to one of the five pilot countries that benefited from the "Rapid Assessment of Drinking Water Quality" (JMP, 2010) pilot project. This case study at national scale may complete the relationships identified in this work by insights on the water quality of the water supply source.

Additionally, the inequality of wealth through the GINI index was not included because of missing data in this variable and the fact that this indicator was only available for 48 countries for the year 2004. As discussed in Chapter 4, section 4, this indicator (when further developed) should complete the results presented here by enriching the development dimension regarding the inequalities inside the countries. The positive effect of HDP on WSS may be limited according to the level of the inequality. Considering the GINI data availability, a specific study might be performed for Latin America, where data is available for most of the countries in this continent.

Finally, cultural practices and psychological barriers associated to sanitation and hygiene, are missing and measured mainly at local scale through specific households surveys (Haddad, 2005), (Al Sa'ed & Mubarak, 2006). These factors are from importance for Sanitation services extension and its sustainability (Chapter 6, section 6). As shown in cases studies, the cultural/psychological context is important in selecting the sanitation technology while it can be essential in the willingness to pay for the construction and the maintenance of the infrastructure/service. The inclusion of such indicator could imply divergences between WS and S models.

The availability of data was also an issue over time, only two other temporal points were collected (2000 and 2007). More temporal points might be included in further research. The temporal retrospective modelling offer the advantages to 1) provide an overview of the decade in a graphic and readable manner; and, 2) to compute simulations based on the past records. The extension of the observation period adding temporal points would enhance the quality of simulations. A complementary predictive modelling would be interesting to develop. It will allow simulating WSS level at *t* time according to set prior parameters. The quality of the prediction is expected to reach a lower level than the retrospective modelling. The implementation of this approach is discussed in next section.

The number of variables (indicators) versus the number of observations (countries) included in the WateSan4Dev database that were used for the multivariate analyses (PCA and FA) could be a second limitation (Chapter 4 and 5) but this is balanced here by other statistical indicators (commonalities, KMO indicator, factors loading as detailed in Chapter 5, section 2). According to the literature, it is recommended: i) a minimum sample size from 100 observations to more than 500 (Chapter 3, section 3.2); and, ii) a STV ratio from 3:1 to 20:1 or more (Rouquette & Falissard, 2011). Since the minimal number of observations (101) was used in the analyses, the number of variables explaining the high heterogeneity of the phenomena is too high (41 variables, in Chapter 4). The risk is to obtain components or factors that (Chapter 3, section 3.2): 1) on one side, they could be not representative enough because the number of variables per component is too low (less than 3 variables per component); or, 2) on another side, they could be too complex because an overload of variables per component (more than 6-7 variables per component).

In Chapter 5, the number of variables is reduced to 25 increasing the STV ratio to 4:1. These statistics could be slightly improved by increasing the number of observations (developing countries) but the main issue will be the increment of missing data for those countries. Since the STV ratio continues to be low, the objective of these linear analyses is to drive the

Bayesian Network modelling phase that will take into account the whole variability of the dataset (linear and non-linear) through probabilistic modelling (Chapter 6). In addition, the results of both the multivariate (linear) and BN models are coherent with the scientific literature and the experience reported from the field and, therefore, cross-validate the results obtained.

2.2 Way forward

The WSS situation is most critical in rural areas where the pipe network is not adapted because of the dissemination of the habitat. Non-improved drinking water sources (unprotected well, surface water source) and open defecation often remain the only option and are thus common practice (UN, 2012). In 2010, the MDGs progress report highlighted that, worldwide, the number of people in rural areas without an improved water source is still five times greater than in urban areas, even if the gap is decreasing. "This rural sanitation crisis persists even in regions with high coverage of improved drinking water: 17 per cent of rural dwellers in Latin America and the Caribbean and 9 per cent in Northern Africa still resort to open defecation" (UN, 2012). In sub-Saharan Africa, of the rural population, 61% of the 20% poorest practice open-defecation, while only 49% of the 20% richest in rural areas benefit from improved sanitation.

Modelling the difference between rural and urban dynamics related to WSS and country development will suppose a breakdown of the relevant variables. For instance, governance or education variables are not collected according to urban-rural context by international providers. However, applying these methods on a local/regional scale could create the possibility of collecting this appropriate data through country statistics, alternative sources or even directly. This scaling down of the data collection would also facilitate data comparability, as the data collected would concern only one country. Two cases studies, one in a slum area and one in a rural area, could highlight the potential differences.

Complementing the urban/rural gap, research is needed to further identify the cultural/ psychological and barriers that limit the commitment to sanitation (Chapter 6, section 6). In fact, there is a need to study the reasons and develop specific indicators to explain why sanitation access is progressing slowly worldwide and, generally, steps behind the water supply service coverage in particular in rural and slums areas.

Another way to explore could be scaling down the methods and analysis to one country or few case studies (i.e. one per country profile). The modelling could be transposed on a national scale, where the regional/provincial or district observations are substituted for the country observations. In this case, the country profile analysis and BNs could support the development of strategies at a national scale, identifying weak areas and running models for these specific areas. On this scale, several opportunities could be opened:

- At this level, the variable selection could be enlarged according to the data collection made within the country. For instance, qualitative indicators of the performance of infrastructure, water quality and types of management that are not available at global level could be added. This kind of variable could orient the modelling to highlight what is needed—what is conditioning the improvement of WSS access.
- A multi-level model could be created based on district/local information and regional aggregates. The scenarios could be run either at the national level to provide a global picture or on a regional/local scale (modifying the input data accordingly).
- The clustering analysis could also be performed for regional/local observations to help prioritise the most needed areas according to selected variables, but also the challenges they face. In fact, grouping of similar observations could extract and synthesise the main the characteristics within an extensive amount of information. This analysis could be included in a Geographic Information System (GIS) to create a layer for geographic analyses to support national strategy and planning. It will also enable the running of specific probabilistic scenarios for each group and testing for ways to improve.

The BNs methods offer the flexibility to adapt to variables' availability within the country. It would be crucial to ensure data comparability at national level and on sub-scales (harmonised household survey of all administrative scales, centralised information system at national level). The number of variables could also be adjusted and organised in the DAG according to the existing knowledge within the country. The categorisation of variables is adaptable to the country context through the number and type (qualitative or numerical scale) of categories and the definition of their limits. Qualitative variables and constant variables could also be included in addition to quantitative variables.

A case study may also target countries from profile 4 with the aim of analysing the benefits and the efficiency in implementing ODA. Indeed, the model built suggested that the ODA CI flows support environmental commitment and civil freedom (CEC) at a higher level than the general behaviour (Chapter 6, section 5.3.1). Such study can support or complement these results.

In this study, a retrospective approach to modelling 2000-2007 is detailed. The main advantage of this is that the modelling provides realistic analyses and scenarios based on a previous period to be considered in the future. Another complementary approach could be explored—that of predictive/forecasting modelling. The literature on this use of Bayesian methods is scarce, with the papers available related to the computer science, health or commercial domains (Chapter 2, section 2.3.3). (Oracle, 2006) has compared the Bayesian (his own Bayesian analytical forecast) to the Best fit approach (linear and logarithmic regressions) in forecasting selling data. The Bayesian approach appeared in this specific case to generate less error.

Dynamic Bayesian Network methods offer the opportunity to generate probabilistic estimations in *t* time based on the defined behaviours of variables. This requires defined equations for each variable according to the DAG of the model (Fig 8.2). For instance:

$$S_{2007} = f(S_{2004}, \text{WGI VA}_{2007}, \text{Irrigation}_{2007}, \text{ODA CI}_{2007}). \quad (6)$$

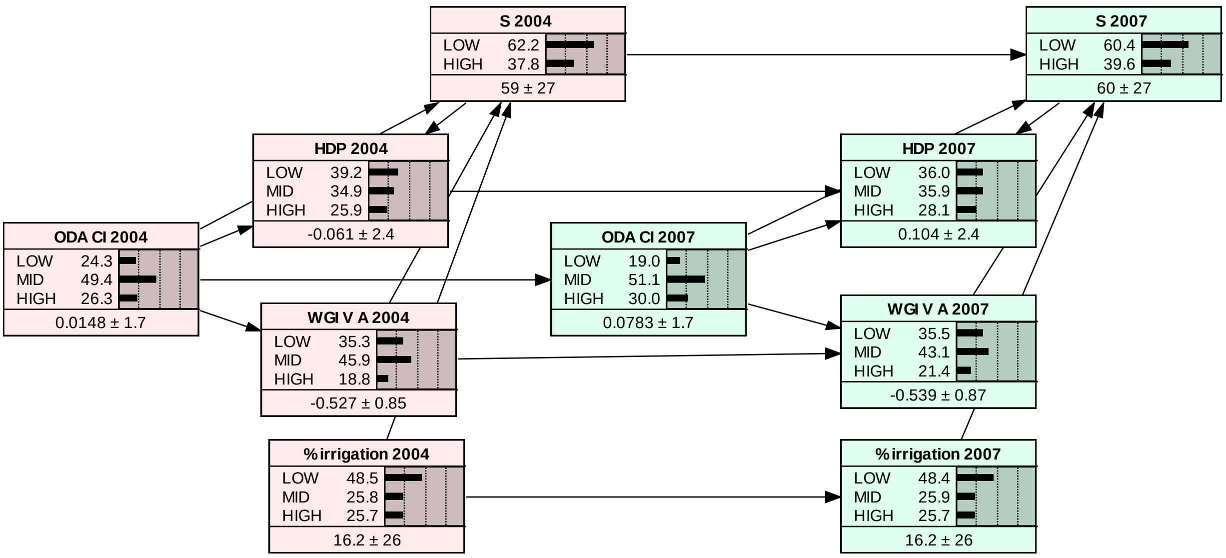


Figure 8.2: DAG of two time slices: 2004 and 2007

According to the function and the definition of the time unit, BNs compute and generate the probabilistic distribution of each variable at time *t*, *t+1*, *t+2*,….*t+n*

The main issue with implementing this option is the definition of the temporal equations. The fact that the relationships between the variables are not fully linear needs to be explored and estimated, even if the temporal behaviour is mostly linear (Chapter 7, section 2.2).

Initially linear equations can be used, as they are the simplest and most robust. Therefore, a linear equation could be defined to express the variable at time *t* in the function of its previous states. For instance, using OLS regression:

$$WS_{2007} = \beta_{intercept+}\beta_1 WS_{2004} + \beta_2 WS_{2000} + \varepsilon_{ws} \quad (7)$$

where the data consists of **n** observations of the variables $WS_{2007}, WS_{2004}, WS_{2000}$, β is a p×1 vector of unknown parameters, p the predictors and ε_{ws} are unobserved scalar random errors.

However, this approach does not consider non-linear behaviour between variables inside the time slice. An evaluation of the representativeness of the model would be essential.

A second approach could be the generation of a variable probabilistic distribution at *t* time according to the variable status at that time *t* and its progress over a reference period (in this study, the 2000-2007 period). In this way, the model generates estimations of probabilistic distributions at *t+1*, *t+2*,….*t+n*, learning from the trends observed during the reference period.

Both options allow for the computation of the probabilistic distribution of each variable, taking into consideration the reference period. The challenge remains in modelling the relationship between variables within a time slice or generating a projection of the whole time slice.

Finally, the period of reference should be extended to improve the quality of the model. Typically, Official Development Assistance—both generally and specific to WSS—is subject to annual fluctuations according to donors' political commitments. The extension of the reference period would refine and reinforce the projection. The international data providers regularly update the indicators that will allow including additional temporal points (i.e. 2010-2012) in the near future.

APPENDIX A: INDICATORS ANALYSED FOR INCLUSION IN THE WATSAN4DEV DATABASE

The data collection has been carried out at the beginning of 2009. Since this date, the datasets could have been updated and/or completed.

ID	name	Included in WatSan4dev for preliminary analysis	Reason of exclusion or complementary information	Data availability for 2004	Missing data rate in % (based on 237 UN countries)	Source
1	Environmental Sustainability index (ESI)	YES	Composite index used as a reference	146 observations	38	Yale and Colombia Universities, available online http://www.yale.edu/esi/
2	Urban Population (% of the total population living in urban areas)	YES		223 observations	6	UN HABITAT
3	Number of urban agglomeration per regions	NO	These are regional aggregates of the size and number of existing agglomerations	Na	na	UN HABITAT
4	GDP-PPP- Income per capita	YES		192 observations	18	WORLD BANK
5	National Poverty rates (% of the population below the national poverty line)	YES		99 observations (combination of years)	58	WORLD BANK
6	Expenditure on health (as % of the GDP)	YES	not used for the modelling - high correlation with GDP per cap	175 observations	26	Human Development Report 2008 table 6
7	Physicians (per 100 000 people)	NO	Limited relevance	174 observations	26	Human Development Report 2008 table 6
8	Prevalence of HIV [15-49 years]	NO	Limited relevance	162 observations	31	Human Development Report 2008 table 9
9	Life expectancy at birth for both sexes (in years)	YES	Variable not used for the modelling because of high correlation with fertility and child mort-5 variables	194 observations	18	Human Development Report 2008 table 10

ID	name	Included in WatSan4dev	Reason of exclusion or complementary information	data availability for 2004	% missing data	Source
10	Total population affected by natural origin disaster	NO	limited relevance and high rate of 0 value (158 in 2004)	232 observations	2	Centre for Research on the Epidemiology of Disasters (CRED) - Emergency Events Database (EM-DAT)
11	Loss of lives dues to disaster	NO	limited relevance and high rate of 0 value (164 in 2004)	232 observations	2	Centre for Research on the Epidemiology of Disasters (CRED)
12	Population affected by conflicts	NO	limited relevance and regional aggregates	na	na	UN HABITAT -TABLE A5
13	Number of refugees and asylum seekers (by country of origin - country of asylum)	NO	limited relevance			UN HABITAT- report on Human settlements in 2007 table B7
14	Total Population growth	NO	dependent indicator to indicator n. 34	225 observations	5	UN Population Division
15	Total Population density	NO	dependent indicator to indicator n. 34	221 observations	7	UN Population Division
16	Urban Population growth in %	YES	not used for the modelling because calling for complex interpretation on urbanization dynamics.	220 observations	7	UN Population Division
17	Rural Population growth in %	YES	not used for the modelling because calling for complex interpretation on urbanization dynamics.	221 observations	7	UN Population Division
18	Proportion of the Urban population living in slums	YES	Slums mainly concern our target group of countries (22% missing values) therefore kept.	79 observations in 2005	66	UNHABITAT
19	% of the population having access to improved Water supply	YES	The datasets computed for rural and urban areas are excluded, focusing on the total access to Water Supply and Sanitation	184 observations	22	Joint Monitoring Program
20	% of the population having access to improved Sanitation	YES	The datasets computed for rural and urban areas are excluded, focusing on the total access to Water Supply and Sanitation in this research.	173 observations	27	Joint Monitoring Program

ID	Name	Included in WatSan4dev	Reason of exclusion or complementary information	data availability for 2004	% missing data	Source
21	Proportion of Household connection for drinking water	NO	Household connection is too restrictive when considering WSS services in the developing world.	183 observations	21	Joint Monitoring Program
22	Internally displaced person	NO	The data concern a too limited number of countries. They are very broad estimates as mentioned in the metadata. Its relevance is limited to qualify governance and human development.	49 observations	80	Internal Displacement Monitoring Centre, Norwegian Refugee Council
23	Net Migrants rate (state and estimate)	NO	5-years estimates, the relevance to WSS is limited	203 observations	14	UN Population Division
24	People killed by wind storms	NO	High rate of 0 values			Centre for Research on the Epidemiology of Disasters (CRED) taken from http://geodata.grid.unep.ch
25	People killed by Drought	NO	Very few values different from 0			Centre for Research on the Epidemiology of Disasters (CRED) taken from http://geodata.grid.unep.ch
26	People killed by Floods	NO	High rate of 0 values	47 observations different from 0	80	Centre for Research on the Epidemiology of Disasters (CRED) http://geodata.grid.unep.ch
27	GINI index	NO	High missing data rate, and few values are related to developing countries. Data update available with a total of 48 observations	23 observations	90 in 2009, 80 in 2014	WORLD BANK
28	Proportion of population living on less1.25$/d	NO	high missing data rate	28 observations	90	WORLD BANK
29	Proportion of population living on less 2$/d	NO	high missing data rate	28 observations	94	WORLD BANK
30	Total land and total surface areas	Na	only used to weight datasets	na	na	FAO
31	Arable land	NO	Sub-indictor to indicator 30	216 observations	8	FAO
32	Permanent crops	No	Sub-indictor to indicator 30	204 observations	14	FAO
33	% Agricultural area of land area	YES	Excluded because no added value in PCA results	222 observations	6	FAO
34	Total Population	NO	used only to weight datasets			UN Population Division
35	Rural Population	NO	dependent indicator (urban population)	220 observations	7	UN Population Division

ID	Name	Included in WatSan4dev	Reason of exclusion or complementary information	data availability for 2004	% missing data	Source
36	Urban population	YES		220 observations	7	UN Population Division
37	Women Economically active population	YES		172 observations	27	ILO (women economic participation rate) or Human Development Report 2008 table 31
38	Proportion of labour force in Agriculture	NO	Excluded because the agricultural water demand index is found more relevant	217 observations	8	FAO
39	Human Development Index	YES	used as reference indicator then removed	177 observations	25	http://hdr.undp.org/en/statistics/
40	Services, value added to GDP(%)	NO	Excluded as the municipal water demand index is found more relevant	185 observations	22	WORLD BANK
41	Industries, value added to GDP(%)	NO	Excluded as the industrial water demand index is found more relevant	185 observations	22	WORLD BANK
42	Agriculture, value added to GDP (%)	NO	Excluded as the agricultural water demand is found more relevant	185 observations	22	WORLD BANK
43	Average precipitation in depth	YES	This unit is chosen	216 observations	8	FAO
44	Average precipitation in volume	NO	Same index as previous indicator	216 observations	8	FAO
45	Water resources: Total renewable (actual) per capita)	YES	The global estimate of water resources is kept. Sub-indicators by water sources are excluded	177 observations	24	FAO
46	Water resources: Total internal renewable or per capita	NO	Sub-indicator to index 45	178 observations	24	FAO
47	Water resources: Total external renewable	NO	Sub-indicator to index 45	182 observations	23	FAO
48	Water resources: total exploitable	NO	Sub-indicator to index 45			FAO
49	Surface water: total renewable (actual)	NO	Sub-indicator to index 45			FAO
50	Surface water: total external renewable (actual)	NO	Sub-indicator to index 49	162 observations	32	FAO
51	Surface water: produced internally	NO	Sub-indicator to index 49			FAO
52	Surface water: inflow not submitted to treaties (actual)	NO	Sub-indicator to index 49			FAO
53	Surface water: inflow secured through treaties	NO	Sub-indicator to index 49			FAO

ID	Name	Included in WatSan4dev	Reason of exclusion or complementary information	data availability for 2004	% missing data	Source
54	Surface water: accounted inflow (actual)	NO	Sub-indicator to index 49			FAO
55	Surface water: total flow of border rivers (actual)	NO	Sub-indicator to index 49			FAO
56	Surface water: accounted flow of border rivers (actual)	NO	Sub-indicator to index 49			FAO
57	Surface water: accounted part of border lakes (actual)	NO	Sub-indicator to index 49			FAO
58	Surface water: total entering and bordering the country (actual)	NO	Sub-indicator to index 49			FAO
59	Surface water: outflow secured through treaties (actual)	NO	Sub-indicator to index 49			FAO
60	Groundwater: total renewable (actual)	NO	Sub-indicator to index 45	159 observations	33	FAO
61	Groundwater: produced internally	NO	Sub-indicator to index 60			FAO
62	Groundwater: entering the country (natural)	NO	Sub-indicator to index 60			FAO
63	Groundwater: leaving the country (natural)	NO	Sub-indicator to index 60			FAO
64	Overlap between surface water and groundwater	NO	Sub-indicator to index 60			FAO
65	Dependency ratio	NO	Sub-indicator to index 60			FAO
66	Total dam capacity	YES	high missing data, included in preliminary analysis but excluded because limited added value in PCA/FA	57 observations	76	Dam
67	Agricultural water withdrawal (absolute value, in % of the total withdrawal or per capita)	YES		162 observations	32	FAO
68	Municipal water withdrawal (absolute value, in % of the total withdrawal or per capita)	YES		162 observations	32	FAO
69	Industrial water withdrawal (absolute value, in % of the total withdrawal or per capita)	YES		163 observations	32	FAO
70	Total water withdrawal (absolute value, or per capita in percent of the total internal resources)	YES		162 observations	32	FAO
71	Total Surface water withdrawal	NO	Sub-indicator to index 70			FAO
72	Total Groundwater withdrawal	NO	Sub-indicator to index 70			FAO
73	Desalinated water produced	NO	marginal phenomenon in developing countries			FAO
74	Wastewater: produced volume	NO	high missing data rate	16 observations	94	FAO
75	Wastewater: treated volume	NO	high missing data rate	17 observations	94	FAO

ID	name	Included in WatSan4dev	Reason of exclusion or complementary information	data availability for 2004	% missing data	Source
76	Total Area equipped for irrigation: absolute values or % of the Agricultural area	YES		182 observations	2	FAO
77	Area equipped for irrigation: full control surface irrigation	NO	Sub-indicator to index 76			FAO
78	Area equipped for irrigation: full control sprinkler irrigation	NO	Sub-indicator to index 76			FAO
79	Area equipped for irrigation: full control localized irrigation	NO	Sub-indicator to index 76			FAO
80	Area equipped for irrigation: full control - total	NO	Sub-indicator to index 76			FAO
81	Area equipped for irrigation: equipped lowland areas	NO	Sub-indicator to index 76			FAO
82	Area equipped for irrigation: spate irrigation	NO	Sub-indicator to index 76			FAO
83	Flood recession cropping area non-equipped	NO	Sub-indicator to index 76			FAO
84	Cultivated wetlands and inland valley bottoms non-equipped	NO	Sub-indicator to index 76			FAO
85	Agricultural water managed area: total	NO	high missing rate and proxy of index 76	102 observations	57	FAO
86	Area equipped for irrigation as % of agricultural water managed area	NO	similar indicator as Total Area equipped for irrigation			FAO
87	Area equipped for irrigation: actually irrigated	NO	breakdown between ground and surface water is not considered			FAO
88	Area equipped for full control irrigation by surface water	NO	breakdown between ground and surface water is not considered			FAO
89	Area equipped for full control irrigation by groundwater	NO	breakdown between ground and surface water is not considered			FAO
90	Area equipped for full control irrigation by other sources	NO	breakdown between ground and surface water is not considered			FAO
91	Area equipped for irrigation: power irrigated area	NO	Sub-indicator to index 76			FAO
92	Power irrigated area as % of area equipped for irrigation	NO	Sub-indicator to index 76			FAO
93	Irrigated grain production as % of total grain production	NO	limited relevance to developing countries			FAO

ID	Name	Included in WatSan4dev	Reason of exclusion or complementary information	data availability for 2004	% missing data	Source
94	Area equipped for irrigation as % of irrigation potential	NO	Sub-indicator to index 76			FAO
95	Area equipped for irrigation as % of cultivated area	NO	similar indicator as Total Area equipped for irrigation			FAO
96	Drained cultivated area: total	NO	high missing data rate	18 observations	92	FAO
97	Area equipped for irrigation: drained	NO	Sub-indicator to index 96			FAO
98	Drained area as % of area equipped for full control surface irrigation	NO	Sub-indicator to index 96			FAO
99	Non-irrigated cultivated area: drained	NO	Sub-indicator to index 96			FAO
100	Drained cultivated area as % of total cultivated area	NO	Sub-indicator to index 96			FAO
101	Conservation agriculture area: >30% ground cover	NO	Sub-indicator to index 96			FAO
102	Area salinized by irrigation	NO	high missing data rate	few value available mainly in central Asian countries, India and China.		FAO
103	Salinized area as % of area equipped for full control irrigation	NO	Similar index as previous one			FAO
104	Area waterlogged by irrigation or not irrigated	NO	Sub-indicator			FAO
105	Population affected by water related disease	NO	high missing data rate	5 observations	97	WHO
106	Water bodies	YES	Included in preliminary multivariate analysis and removed because low added value in PCA	220 observations	7	FAO
107	Human poverty index - HP-1	YES	Included in preliminary multivariate analysis and removed because similar behaviour as HDI	114 observations	51	INDEX-http://geodata.grid.unep.ch
108	Total fertility rate	YES	excluded because of collinearity with children mortality rate and birth expectancy	194 observations	18	UN Population Division
109	Malaria reported cases	YES		94 observations	60	WHO
110	% population in dryland	NO	Similar to index 111	163 observations		FAO
111	Proportion of area under desert risk	YES		163 observations		FAO
112	Children Mortality under 5 years	YES		192 observations	19	WHO/UNICEF

ID	name	Included in WatSan4dev	Reason of exclusion or complementary information	data availability for 2004	% missing data	Source
113	Water Poverty index	YES	used as reference indicator then removed	147 observations	38	Sullivan 2000
114	Agricultural production index	YES	Included in preliminary multivariate analysis then removed because low significance	212 observations	10	FAO
115	Water use intensity for agriculture	YES		159 observations	33	FAO
116	Girls to Boys Ratio in Primary Education Enrolment (primary school)	YES		161 observations	32	HDR
117	Combined gross enrolment ratio in education (both sexes) (%)	YES		184 observations	22	HDR
118	Literacy rate of youth 15-24	NO	high missing data rate	83 observations	65	HDR
119	Proportion of Children <5 with diarrhoea	NO	high missing data rate	29 observations	87	UNICEF/WHO
120	Demographic and disease variables specific to population living in slums	NO	urban-rural breakdown not considered			UN HABITAT
121	Worldwide Governance indexes	YES		between 203 and 208 observations	2 to 4	WORLD BANK
122	Index of African Governance	YES	scope on Africa - relevant when focusing on this continent	52 observations	4	Harvard university - http://belfercenter.ksg.harvard.edu/publication/18541/strengthening_african_governance.html
123	Corruption Perception Index	YES		146 observations	38	Transparency international - http://www.icgg.org/corruption.cpi_2004.html
124	BOD demands in lakes-groundwater-rivers	YES	Included in preliminary multivariate analysis , then removed because of limited added value in PCA/FA	74 observations in 2000	69	UNEP http://geodata.grid.unep.ch
125	ODA general	YES		171 observations	28	OECD DAC database http://stats.oecd.org/qwids/
126	ODA WSS total	YES		141 observations	40	OECD DAC database http://stats.oecd.org/qwids/
127	ODA WSS breakdown by sub sector	NO	Sub-indicators to index 126			OECD DAC database http://stats.oecd.org/qwids/
128	Remittances received per country	NO	Not considered in this work	188 observations	20	WORLD BANK-IMF

Table A.1: List of indicators reviewed for inclusion in the WatSan4Dev Database

APPENDIX B: WATSAN4DEV INDICATORS: DESCRIPTION (IN ALPHABETIC ORDER: SHORT NAME, FULL NAME)

% agri area, Agricultural area

Agricultural area is the sum of areas under a) arable land - land under temporary agricultural crops (multiple-cropped areas are counted only once), temporary meadows for mowing or pasture, land under market and kitchen gardens and land temporarily fallow (less than five years).

Agri Prod Index, Agricultural Production Index

This FAO index shows the relative level of the aggregate volume of agricultural production for each year in comparison with the base period 1999-2001.

S, Sanitation Access

Percentage of the population having access to improved sanitation source. For MDG monitoring, an improved sanitation facility is defined as one that hygienically separates human excreta from human contact.

WS, Water supply Access

Percentage of the population having access to improved water supply source. An improved drinking-water source is defined as one that, by nature of its construction or through active intervention, is protected from outside contamination, in particular from contamination with faecal matter

BOD, Biological Oxygen Demand release

Emissions of organic water pollutants are measured in terms of biochemical oxygen demand, which refers to the amount of oxygen that bacteria in water will consume in breaking down waste. It covers rivers, lakes and groundwaters.

Child Diarrhoea, Children with Diarrhoea

Proportion of children under 5 having or who have had diarrhoea within the two weeks before the survey.

Child Mortal-5, Under-five Mortality Rate

Probability of dying between birth and exact age 5. It is expressed as deaths per 1,000 births.

CPI – Corruption Perception Index

CPI measures the perceived level of public-sector corruption in 180 countries and territories around the world. The CPI is a "survey of surveys", based on 13 different expert and business surveys.

Desert risk, Dyland Area

The concept of drylands continues to be debated. In this data set, drylands are taken as areas with a potential hazard of desertification. The hyperarid zone is not subject to desertification and is therefore are excluded (assimilated at 0). Hence drylands are defined as the arid, semi-arid and dry subhumid zones, or areas with lengths of growing periods of 1-179 days.

ESI, Environmental Sustainability Index

ESI benchmarks the ability of nations to protect the environment over the next several decades. It does so by integrating 76 data sets – tracking natural resource endowments, past and present pollution levels, environmental management efforts, and the capacity of a society to improve its environmental performance –into 21 indicators of environmental sustainability. These indicators permit comparison across a range of issues that fall into the following five broad categories:1) Environmental Systems 2) Reducing Environmental Stresses 3)Reducing Human Vulnerability to Environmental Stresses 4) Societal and Institutional Capacity to Respond to Environmental Challenges 5) Global Stewardship .

Femal eco/Women Eco, Female economic activity rate

This rate concerns women aged 15 and above and calculated on the basis of data on the female economically active population (person looking for or having an occupation) and male population from ILO (International Labour Organization). Generally, Students, retired people and persons not looking for an occupation are excluded.

The women economic activity (Women eco) rate is used in the temporal modelling replacing the female economic activity rate. This is the ratio between the female economically active population and the total population (ILO).

Fertility, Fertility Rates

Total fertility rate is an estimate of the number of children an average woman would have if current age-specific fertility rates remained constant during her reproductive years.

GDP per cap, Gross Domestic Product – Purchasing Power Parity

Gross Domestic Product (Purchasing Power Parity) is gross domestic product converted to GDP per capita based on purchasing power parity (PPP). PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as the U.S. dollar has in the United States. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. Data are in current international dollars.

School G/B, Girls to Boys Ratio in Primary Education Enrolment

This is a measure of the attendance of girls at primary school. The core at this level consists of education provided for children, the customary or legal age of entrance being not younger than five years or older than seven years. This level covers in principle six years of full-time schooling.

School enrol, Gross enrolment ratio at school

It is calculated by expressing the number of students enrolled in primary, secondary and tertiary levels of education, regardless of age, as a percentage of the population of official school age for the three levels.

Health exp, Health Expenditure (PPP- Capita)

This is the sum of public and private health expenditure (in PPP, International $) divided by population. Health expenditure includes the provision of health services, family planning activities, nutrition activities and emergency assistance designated for heath, but excludes the provision of water and sanitation.

HDI, Human Development Index

Summering human development, it measures the average achievements in a country in three basic dimensions of human development: i) A long and healthy life, as measured by life expectancy at birth ii) Knowledge, as measured by the adult literacy rate (with two-thirds weight) and the combined primary, secondary and tertiary gross enrolment ratio (with one-third weight) iii) A decent standard of living, as measured by GDP per capita (PPP US$).

Life Expect, Life expectancy at birth

Life expectancy at birth, both sexes is the average number of years that a newborn baby is expected to live if the age-specific mortality rates effective at the year of birth apply throughout his or her lifetime.

Literacy Youth, Literacy rate of Youth

It is the percentage of people ages 15 -24 who can, with understanding, both read and write a short, simple statement related to their everyday life.

Malaria, Malaria prevalence,

Number of reported cases per 1000 persons in country

NBI, National Biodiversity Index

This index is based on estimates of country richness and endemism in four terrestrial vertebrate classes and vascular plants; vertebrates and plants are ranked equally; index values range between 1.000 (maximum: Indonesia) and 0.000 (minimum: Greenland). Countries with land area less than 5,000 sq km are excluded. Overseas territories and dependencies are excluded from this column.

ODA, Official Development Assistance

Net official development assistance (ODA) consists of disbursements of loans made on concessional terms (net of repayments of principal) and grants by official agencies of the members of the Development Assistance. The ODA disbursements of all donors are expressed by recipient country and in USD by capita.

ODA-WSS Official Development Assistance to the water sector

Total of all donors' disbursements of ODA towards all recipients related to Water supply and sanitation, expressed in USD per capita.

Particip, Participation to international environmental agreements

It is calculated taking into account the participations to Framework Convention on Climate Change (UNFCCC), Vienna Convention on the Protection of the Ozone Layer, Convention on the Trade in Endangered Species (CITES), Basel Convention on the Transboundary Movement of Hazardous Waste and United nations.

%poverty, Poverty Rate

National poverty rate is the percentage of a country's population living below the country's established national poverty line.

% irrigation, Total surface in irrigation

Area equipped to provide water (via irrigation) to the crops. It includes areas equipped for full and partial control irrigation, equipped lowland areas, pastures, and areas equipped for spate irrigation.

TWRR, Total Water Renewable Resources

This is an estimate of the surface water resources available for use in a country corresponding to the sum of the internal renewable surface water resources and the total external actual renewable surface water resources.

Urban Pop, Urban Population -Rural Population

Total population residing in urban areas. Because of national differences in the characteristics that distinguish urban from rural areas, the distinction between urban and rural population is not amenable to a single definition that would be applicable to all countries. National definitions are most commonly based on size of locality. Population which is not urban is considered rural (World Urbanization Prospects: The 2005 Revision).

% slums, Urban Slum population

Proportion of the urban population living in slums (A slum is a contiguous settlement where the inhabitants are characterized as having inadequate housing and basic services.)

WB, Water Bodies Surface

It's the ratio of Water bodies regarding the total country surface. Water bodies are oceans, seas, lakes, reservoirs, and rivers. They can be either fresh or salt water bodies.

WPI, Water Poverty Index

WPI expresses an interdisciplinary measure which links household welfare with water availability and indicates the degree to which the water scarcity impacts on population .WPI has of five component indices: Resources, Access, Capacity, Use, and Environment. The more this index, the lower is the water constraint (Sullivan, 2003).

Water Use Int Agri, Water use intensity for agriculture

This is the amount of water used in the agricultural sector per hectare of temporary and permanent cropland in the year specified. This indicator shows a country's dependence on irrigation for agricultural production

WGI VA, Worldwide Governance Index Voice and accountability

This index captures perceptions of the extent to which country'citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association and free media.

WGI PS AV, Worldwide Governance Index Political Stability and Absence of Violence

This index captures perceptions of the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means including politically-motivated violence and terrorism.

WGI GE, Worldwide Governance Index Government effectiveness

This index captures perceptions of the quality of the public services, the quality of the civil services, and the degree of its independence from political pressure, the quality of policy formulation and implementation and the credibility of the government's commitments to such policies.

WGI RQ, Worldwide Governance Index Regulatory Quality

This index captures perceptions of the ability of the government to formulate and implement sound policies and regulations to permit and promote private sector development

WGI RofL, Worlwide Governance Index Rule of Law

This index captures perceptions of the extent to which agents have confidence in and abide by the rule of the society and in particular the quality of the contract enforcement, property rights, the police and the courts as well as the likelihood of crime and violence.

WGI CofC, Worldwide Governance Index Control of Corruption

This index captures perceptions of the quality of public power is exercised for private gain, including both petty and grand forms of corruption, as well as capture of the states by elites and privates interests.

Total withdrawals, Water Withdrawal Total

Annual gross quantity of water produced and used for agricultural, industrial and domestic purposes. It does not include other in situ-uses: energy, mining, recreation, navigation, fisheries and the environment, which are typically non consumptive uses of water. The typology of water use is independent from the source of water.

Total Water Use = Agricultural Water Use+ Municipal Water Use+ Industrial Water Use.

Withdrawal municipal, Water withdrawals for municipal purpose

Annual quantity of water use for municipal purposes. It is usually computed as the total amount of water supplied by public distribution networks, and usually includes the withdrawal by those industries connected to public networks.

Withdrawals industrial, Water withdrawals for industrial purpose

Annual quantity of water use by self-supplied industries not connected to any distribution network.

APPENDIX C: LIST OF COUNTRIES EXCLUDED

Country	Criteria of exclusion
French Guiana	non independent territory
French Polynesia	non independent territory
Gibraltar	non independent territory
Greenland	non independent territory
Guadeloupe	non independent territory
Martinique	non independent territory
Mayotte	non independent territory
New Caledonia	non independent territory
Reunion	non independent territory
Saint Pierre and Miquelon	non independent territory
Wallis and Futuna	non independent territory
Western Sahara	non independent territory
Puerto Rico	non independent territory
Australia	developed country
Austria	developed country
Belgium	developed country
Canada	developed country
Denmark	developed country
Finland	developed country
France	developed country
Germany	developed country
Greece	developed country
Holy See	developed country
Iceland	developed country
Ireland	developed country
Israel	developed country
Italy	developed country
Japan	developed country
Liechtenstein	developed country
Luxembourg	developed country
Monaco	developed country
Netherlands	developed country
New Zealand	developed country
Norway	developed country
Portugal	developed country
Russian Federation	developed country
Spain	developed country
Sweden	developed country
Switzerland	developed country
United Kingdom of Great Britain and Northern Ireland	developed country
United States of America	developed country

Continued…

Country	Criteria of exclusion
Antarctic	Pop <150,000
American Samoa	Pop <150,000
Andorra	Pop <150,000
Anguilla	Pop <150,000
Antigua and Barbuda	Pop <150,000
Aruba	Pop <150,000
Bermuda	Pop <150,000
British Virgin Islands	Pop <150,000
Cayman Islands	Pop <150,000
Christmas Island	Pop <150,000
Cocos (Keeling) Islands	Pop <150,000
Cook Islands	Pop <150,000
Falkland Islands (Malvinas)	Pop <150,000
Faroe Islands	Pop <150,000
Grenada	Pop <150,000
Guernsey	Pop <150,000
Isle of Man	Pop <150,000
Jersey	Pop <150,000
Johnston Atoll	Pop <150,000
Kiribati	Pop <150,000
Malta	Pop <150,000
Marshall Islands	Pop <150,000
Micronesia (Federated States of)	Pop <150,000
Midway Islands	Pop <150,000
Montserrat	Pop <150,000
Nauru	Pop <150,000
Niue	Pop <150,000
Norfolk Island	Pop <150,000
Northern Mariana Islands	Pop <150,000
Palau	Pop <150,000
Pitcairn Island	Pop <150,000
Saint Helena	Pop <150,000
Saint Kitts and Nevis	Pop <150,000
Saint Lucia	Pop <150,000
Saint Vincent and the Grenadines	Pop <150,000
San Marino	Pop <150,000
Seychelles	Pop <150,000
Singapore	Pop <150,000
Solomon Islands	Pop <150,000
Svalbard and Jan Mayen Islands	Pop <150,000
Tokelau	Pop <150,000
Tonga	Pop <150,000
Turks and Caicos Islands	Pop <150,000
Tuvalu	Pop <150,000
United States Virgin Islands	Pop <150,000
Wake Island	Pop <150,000

Continued….

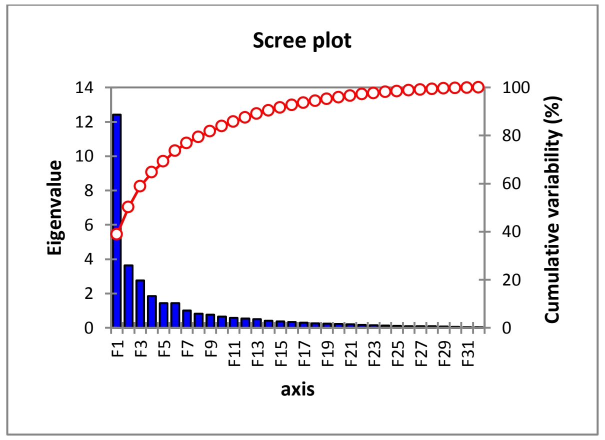
Country	Criteria of exclusion	Country	Criteria of exclusion
Albania	Non-cooperation priority	Afghanistan	>35% missing data
Armenia	Non-cooperation priority	Brunei Darussalam	>35% missing data
Azerbaijan	Non-cooperation priority	Timor-Leste	>35% missing data
Belarus	Non-cooperation priority	Vanuatu	>35% missing data
Bosnia and Herzegovina	Non-cooperation priority	Iraq	>35% missing data
Bulgaria	Non-cooperation priority	Democratic People's Republic of Korea	>35% missing data
China	Non-cooperation priority	Guam	>35% missing data
Croatia	Non-cooperation priority	Occupied Palestinian Territory	>35% missing data
Cyprus	Non-cooperation priority	Somalia	>35% missing data
Czech Republic	Non-cooperation priority	Bahamas	>35% missing data
Estonia	Non-cooperation priority	Barbados	>35% missing data
Fiji	Non-cooperation priority	Maldives	>35% missing data
Georgia	Non-cooperation priority	Montenegro	>35% missing data
Hungary	Non-cooperation priority	Samoa	>35% missing data
Kazakhstan	Non-cooperation priority	Dominica	>35% missing data and Pop <150,000
Kyrgyzstan	Non-cooperation priority	Netherlands Antilles	>35% missing data
Latvia	Non-cooperation priority		
Lithuania	Non-cooperation priority		
Moldova	Non-cooperation priority		
Mongolia	Non-cooperation priority		
Poland	Non-cooperation priority		
Republic of Korea	Non-cooperation priority		
Romania	Non-cooperation priority		
Serbia	Non-cooperation priority		
Slovakia	Non-cooperation priority		
Slovenia	Non-cooperation priority		
Tajikistan	Non-cooperation priority		
The former Yugoslav Republic of Macedonia	Non-cooperation priority		
Trinidad and Tobago	Non-cooperation priority		
Turkey	Non-cooperation priority		
Turkmenistan	Non-cooperation priority		
Ukraine	Non-cooperation priority		
Uzbekistan	Non-cooperation priority		

Table C.1. List and criteria of exclusion of countries

APPENDIX D: CORRELATION MATRICES AND RELATED STATISTICAL TESTS OF PRINCIPAL COMPONENTS ANALYSIS PERFORMED

1. PRINCIPAL COMPONENT ANALYSIS OF THE WHOLE WATSAN4DEV FOR 2004

Variables	KMO measure	Variables	KMO measure	Variables	KMO measure
WS	0.976	CPI	0.874	ODA WSS	0.424
S	0.893	TWRR	0.586	ODA	0.765
%Poverty	0.927	Health expend	0.908	Malaria	0.917
% Agri Area	0.575	Life Expect Birth	0.937	Water Use int Agri	0.869
Urban Pop	0.863	Child Mortal-5	0.925	WB	0.430
Env gov	0.839	School G/B	0.942	Precipit	0.613
School enrol	0.911	Rural Growth	0.764	Desert risk	0.753
Femal eco	0.885	Urban Growth	0.835	Total withdrawals	0.834
% Slums	0.916	NBI	0.743	**KMO**	**0.865**
Fertility Rate	0.930	Particip	0.711		
Withdrawal municipal	0.661	GDP per cap	0.910		
Withdrawal industrial	0.750	%irrigation	0.857		



	F1	F2	F3	F4	F5	F6
Eigenvalue	12.419	3.647	2.762	1.852	1.451	1.436
Variability (%)	38.808	11.396	8.632	5.788	4.534	4.488
Cumulative %	38.808	50.205	58.836	64.624	69.158	73.647

Table D.1: Kaiser-Meyer-Olkin (KMO) measures of sampling adequacy for active variables database, scree plot and eigenvalues of the first factors- PCA of the WatSan4Dev whole dataset.

Variables	%Poverty	% Agri Area	Urban Pop	Env gov	School enrol	Femal eco	% Slums	Fertility	Withdrawal municipal	Withdrawal industrial	CPI	TWRR	Health exp	S	Life Expect Birth	Child Mortal-5	WS	school G/B	Rural Growth	Urban Growth	NBI	Particip	GDP per cap	%irrigation	ODA WSS	ODA	Malaria	Water Use int Agri	WB	Precipit	Deser risk	Total withdrawal
%Poverty	**1**	**0.275**	**-0.493**	**-0.536**	**-0.510**	**0.460**	**0.655**	**0.672**	-0.047	-0.099	**-0.562**	**0.239**	**-0.630**	**-0.548**	**-0.632**	**0.611**	**-0.600**	**-0.326**	**0.272**	**0.295**	0.165	**-0.216**	**-0.708**	**-0.371**	0.009	**0.408**	**0.617**	**-0.351**	0.044	0.158	0.176	**-0.408**
% Agri Area	**0.275**	**1**	-0.187	-0.133	-0.155	0.116	0.130	0.180	0.015	-0.139	**-0.208**	-0.194	-0.149	-0.182	**-0.222**	**0.215**	-0.072	-0.122	0.017	0.086	0.076	0.130	**-0.262**	-0.168	-0.154	-0.014	0.082	-0.080	0.086	-0.082	**0.243**	-0.193
Urban Pop	**-0.493**	-0.187	**1**	**0.434**	**0.607**	**-0.465**	**-0.650**	**-0.545**	**0.335**	0.141	**0.528**	-0.107	**0.683**	**0.624**	**0.680**	**-0.630**	**0.643**	**0.392**	**-0.469**	**-0.317**	-0.116	**0.208**	**0.727**	0.133	-0.146	**-0.305**	**-0.616**	**0.311**	0.004	-0.160	-0.172	**0.294**
Env gov	**-0.536**	-0.133	**0.434**	**1**	**0.338**	**-0.323**	**-0.572**	**-0.467**	**0.211**	0.051	**0.753**	**-0.352**	**0.629**	**0.535**	**0.448**	**-0.456**	**0.570**	**0.387**	-0.188	-0.164	-0.192	**0.203**	**0.637**	**0.211**	0.130	-0.145	**-0.430**	**0.295**	-0.069	**-0.308**	0.007	**0.251**
School enrol	**-0.510**	-0.155	**0.607**	**0.338**	**1**	**-0.411**	**-0.667**	**-0.719**	0.059	**0.227**	**0.473**	0.019	**0.798**	**0.636**	**0.697**	**-0.736**	**0.671**	**0.636**	**-0.473**	**-0.511**	0.140	**0.348**	**0.688**	**0.401**	-0.067	**-0.348**	**-0.591**	**0.406**	**0.197**	0.099	**-0.284**	**0.488**
Femal eco	**0.460**	0.116	**-0.465**	**-0.323**	**-0.411**	**1**	**0.511**	**0.452**	0.087	0.018	**-0.374**	**0.297**	**-0.492**	**-0.426**	**-0.483**	**0.477**	**-0.498**	**-0.205**	**0.234**	**0.371**	**0.349**	0.033	**-0.566**	**-0.420**	-0.027	**0.200**	**0.589**	**-0.446**	**0.216**	**0.254**	0.071	**-0.435**
% Slums	**0.655**	0.130	**-0.650**	**-0.572**	**-0.667**	**0.511**	**1**	**0.746**	-0.083	-0.121	**-0.653**	0.181	**-0.786**	**-0.740**	**-0.707**	**0.714**	**-0.775**	**-0.560**	**0.437**	**0.469**	0.085	**-0.352**	**-0.801**	**-0.415**	0.042	**0.405**	**0.638**	**-0.490**	0.029	0.144	0.121	**-0.488**
Fertility	**0.672**	0.180	**-0.545**	**-0.467**	**-0.719**	**0.452**	**0.746**	**1**	0.091	-0.059	**-0.561**	-0.003	**-0.765**	**-0.738**	**-0.779**	**0.827**	**-0.735**	**-0.610**	**0.457**	**0.570**	-0.099	**-0.311**	**-0.780**	**-0.614**	0.128	**0.468**	**0.758**	**-0.546**	-0.041	-0.096	**0.262**	**-0.650**
Withdrawal municpal	-0.047	0.015	**0.335**	**0.211**	0.059	0.087	-0.083	0.091	**1**	**0.514**	0.047	-0.079	0.152	0.005	-0.013	0.052	0.093	0.039	-0.168	-0.003	0.111	0.101	0.131	**-0.500**	-0.054	0.031	0.094	**-0.535**	0.052	-0.009	-0.112	**-0.553**
Withdrawal industrial	-0.099	-0.139	0.141	0.051	**0.227**	0.018	-0.121	-0.059	**0.514**	**1**	0.092	**0.275**	0.174	0.059	0.077	-0.113	0.139	0.117	-0.103	-0.173	**0.317**	0.065	0.144	**-0.223**	0.004	-0.061	0.101	**-0.443**	0.128	**0.282**	**-0.276**	**-0.251**
CPI	**-0.562**	**-0.208**	**0.528**	**0.753**	**0.473**	**-0.374**	**-0.653**	**-0.561**	0.047	0.092	**1**	-0.160	**0.673**	**0.548**	**0.593**	**-0.640**	**0.623**	**0.395**	**-0.197**	**-0.315**	-0.179	**0.239**	**0.716**	**0.251**	**0.199**	-0.112	**-0.517**	**0.403**	-0.056	**-0.226**	-0.054	**0.410**
TWRR	**0.239**	-0.194	-0.107	**-0.352**	0.019	**0.297**	0.181	-0.003	-0.079	**0.275**	-0.160	**1**	-0.080	**-0.199**	-0.145	0.122	**-0.232**	0.033	-0.179	-0.188	**0.589**	0.095	-0.141	-0.087	-0.037	0.180	**0.358**	**-0.298**	0.141	**0.728**	**-0.241**	0.003
Health exp	**-0.630**	-0.149	**0.683**	**0.629**	**0.798**	**-0.492**	**-0.786**	**-0.765**	0.152	0.174	**0.673**	-0.080	**1**	**0.708**	**0.719**	**-0.750**	**0.763**	**0.589**	**-0.462**	**-0.487**	-0.014	**0.313**	**0.899**	**0.344**	-0.016	**-0.354**	**-0.675**	**0.394**	-0.005	-0.056	**-0.221**	**0.481**
S	**-0.548**	-0.182	**0.624**	**0.535**	**0.636**	**-0.426**	**-0.740**	**-0.738**	0.005	0.059	**0.548**	**-0.199**	**0.708**	**1**	**0.754**	**-0.770**	**0.738**	**0.515**	**-0.238**	**-0.429**	-0.058	**0.224**	**0.711**	**0.471**	0.004	**-0.450**	**-0.735**	**0.515**	0.063	-0.066	**-0.208**	**0.574**
Life ExpectBirth	**-0.632**	**-0.222**	**0.680**	**0.448**	**0.697**	**-0.483**	**-0.707**	**-0.779**	-0.013	0.077	**0.593**	-0.145	**0.719**	**0.754**	**1**	**-0.889**	**0.752**	**0.491**	**-0.253**	**-0.339**	0.005	**0.353**	**0.765**	**0.531**	-0.065	**-0.458**	**-0.832**	**0.554**	0.120	-0.013	**-0.290**	**0.583**
Child Mortal-5	**0.611**	**0.215**	**-0.630**	**-0.456**	**-0.736**	**0.477**	**0.714**	**0.827**	0.052	-0.113	**-0.640**	0.122	**-0.750**	**-0.770**	**-0.889**	**1**	**-0.778**	**-0.552**	**0.306**	**0.388**	-0.033	**-0.286**	**-0.786**	**-0.556**	0.126	**0.460**	**0.808**	**-0.564**	-0.121	-0.046	**0.335**	**-0.570**
WS	**-0.600**	-0.072	**0.643**	**0.570**	**0.671**	**-0.498**	**-0.775**	**-0.735**	0.093	0.139	**0.623**	**-0.232**	**0.763**	**0.738**	**0.752**	**-0.778**	**1**	**0.519**	**-0.329**	**-0.364**	-0.033	**0.328**	**0.770**	**0.464**	-0.098	**-0.412**	**-0.731**	**0.498**	0.098	-0.108	-0.182	**0.480**
Ratio school G/B	**-0.326**	-0.122	**0.392**	**0.387**	**0.636**	**-0.205**	**-0.560**	**-0.610**	0.039	0.117	**0.395**	0.033	**0.589**	**0.515**	**0.491**	**-0.552**	**0.519**	**1**	**-0.359**	**-0.325**	0.160	**0.265**	**0.577**	**0.311**	0.023	-0.165	**-0.386**	**0.364**	0.135	0.152	**-0.275**	**0.392**
Rural Growth	**0.272**	0.017	**-0.469**	-0.188	**-0.473**	**0.234**	**0.437**	**0.457**	-0.168	-0.103	**-0.197**	-0.179	**-0.462**	**-0.238**	**-0.253**	**0.306**	**-0.329**	**-0.359**	**1**	**0.488**	-0.151	**-0.335**	**-0.367**	-0.189	**0.358**	0.131	**0.243**	-0.165	-0.058	-0.130	0.088	**-0.204**
Urban Growth	**0.295**	0.086	**-0.317**	-0.164	**-0.511**	**0.371**	**0.469**	**0.570**	-0.003	-0.173	**-0.315**	-0.188	**-0.487**	**-0.429**	**-0.339**	**0.388**	**-0.364**	**-0.325**	**0.488**	**1**	-0.041	**-0.211**	**-0.434**	**-0.273**	0.086	0.171	**0.338**	-0.192	0.053	-0.077	0.037	**-0.436**

Values in bold are significantly different from 0 with a significance level alpha=0.05

Table D.2: Correlation matrix (Pearson (n)) of the PCA performed on the WatSan4Dev variables –part 1

Continued…

Variables	%Poverty	% Agri Area	Urban Pop	Env gov	School enrol	Femal eco	% Slums	Fertility	Withdrawal municipal	Withdrawal industrial	CPI	TWRR	Health exp	S	Life Expect Birth	Child Mortal-5	WS	school G/B	Rural Growth	Urban Growth	NBI	Particip	GDP per cap	%irrigation	ODA WSS	ODA	Malaria	Water Use int Agri	WB	Precip	Desert risk	Total withdrawal
NBI	0.165	0.076	-0.116	-0.192	0.140	**0.349**	0.085	-0.099	0.111	**0.317**	-0.179	**0.589**	-0.014	-0.058	0.005	-0.033	-0.033	0.160	-0.151	-0.041	**1**	**0.259**	-0.115	0.038	**-0.227**	-0.127	**0.230**	**-0.203**	**0.348**	**0.732**	-0.191	-0.041
Particip	**-0.216**	0.130	**0.208**	**0.203**	**0.348**	0.033	**-0.352**	**-0.311**	0.101	0.065	**0.239**	0.095	**0.313**	**0.224**	**0.353**	**-0.286**	**0.328**	**0.265**	**-0.335**	**-0.211**	**0.259**	**1**	0.180	**0.216**	-0.068	-0.186	**-0.259**	0.151	0.177	0.031	0.111	0.195
GDP per cap	**-0.708**	**-0.262**	**0.727**	**0.637**	**0.688**	**-0.566**	**-0.801**	**-0.780**	0.131	0.144	**0.716**	-0.141	**0.899**	**0.711**	**0.765**	**-0.786**	**0.770**	**0.577**	**-0.367**	**-0.434**	-0.115	0.180	**1**	**0.347**	-0.030	**-0.388**	**-0.723**	**0.446**	-0.097	-0.110	**-0.285**	**0.522**
%irrigation	**-0.371**	-0.168	0.133	**0.211**	**0.401**	**-0.420**	**-0.415**	**-0.614**	**-0.500**	**-0.223**	**0.251**	-0.087	**0.344**	**0.471**	**0.531**	**-0.556**	**0.464**	**0.311**	-0.189	**-0.273**	0.038	**0.216**	**0.347**	**1**	-0.065	**-0.290**	**-0.563**	**0.733**	**0.198**	0.106	-0.125	**0.687**
ODA WSS	0.009	-0.154	-0.146	0.130	-0.067	-0.027	0.042	0.128	-0.054	0.004	**0.199**	-0.037	-0.016	0.004	-0.065	0.126	-0.098	0.023	**0.358**	0.086	**-0.227**	-0.068	-0.030	-0.065	**1**	**0.534**	0.079	-0.016	0.000	-0.076	-0.014	-0.050
ODA	**0.408**	-0.014	**-0.305**	-0.145	**-0.348**	**0.200**	**0.405**	**0.468**	0.031	-0.061	-0.112	0.180	**-0.354**	**-0.450**	**-0.458**	**0.460**	**-0.412**	-0.165	0.131	0.171	-0.127	-0.186	**-0.388**	**-0.290**	**0.534**	**1**	**0.486**	**-0.278**	0.096	0.048	0.066	**-0.389**
Malaria	**0.617**	0.082	**-0.616**	**-0.430**	**-0.591**	**0.589**	**0.638**	**0.758**	0.094	0.101	**-0.517**	**0.358**	**-0.675**	**-0.735**	**-0.832**	**0.808**	**-0.731**	**-0.386**	**0.243**	**0.338**	**0.230**	**-0.259**	**-0.723**	**-0.563**	0.079	**0.486**	**1**	**-0.587**	0.033	**0.228**	0.167	**-0.568**
Water Use int Agri	**-0.351**	-0.080	**0.311**	**0.295**	**0.406**	**-0.446**	**-0.490**	**-0.546**	**-0.535**	**-0.443**	**0.403**	**-0.298**	**0.394**	**0.515**	**0.554**	**-0.564**	**0.498**	**0.364**	-0.165	-0.192	**-0.203**	0.151	**0.446**	**0.733**	-0.016	**-0.278**	**-0.587**	**1**	0.030	-0.193	0.001	**0.785**
WB	0.044	0.086	0.004	-0.069	**0.197**	**0.216**	0.029	-0.041	0.052	0.128	-0.056	0.141	-0.005	0.063	0.120	-0.121	0.098	0.135	-0.058	0.053	**0.348**	0.177	-0.097	**0.198**	0.000	0.096	0.033	0.030	**1**	**0.439**	**-0.308**	-0.112
Precip	0.158	-0.082	-0.160	**-0.308**	0.099	**0.254**	0.144	-0.096	-0.009	**0.282**	**-0.226**	**0.728**	-0.056	-0.066	-0.013	-0.046	-0.108	0.152	-0.130	-0.077	**0.732**	0.031	-0.110	0.106	-0.076	0.048	**0.228**	-0.193	**0.439**	**1**	**-0.531**	-0.079
Desert risk	0.176	**0.243**	-0.172	0.007	**-0.284**	0.071	0.121	**0.262**	-0.112	**-0.276**	-0.054	**-0.241**	**-0.221**	**-0.208**	**-0.290**	**0.335**	-0.182	**-0.275**	0.088	0.037	-0.191	0.111	**-0.285**	-0.125	-0.014	0.066	0.167	0.001	**-0.308**	**-0.531**	**1**	-0.026
Total withdrawal	**-0.408**	-0.193	**0.294**	**0.251**	**0.488**	**-0.435**	**-0.488**	**-0.650**	**-0.553**	**-0.251**	**0.410**	0.003	**0.481**	**0.574**	**0.583**	**-0.570**	**0.480**	**0.392**	**-0.204**	**-0.436**	-0.041	0.195	**0.522**	**0.687**	-0.050	**-0.389**	**-0.568**	**0.785**	-0.112	-0.079	-0.026	**1**
WPI	**-0.432**	**-0.366**	**0.493**	**0.252**	**0.696**	**-0.355**	**-0.598**	**-0.669**	-0.019	**0.251**	**0.398**	**0.451**	**0.643**	**0.568**	**0.579**	**-0.600**	**0.561**	**0.540**	**-0.413**	**-0.549**	**0.332**	**0.229**	**0.659**	**0.381**	-0.083	**-0.314**	**-0.363**	**0.329**	0.121	**0.403**	**-0.401**	**0.573**
ESI	-0.169	**-0.248**	**0.259**	0.180	**0.307**	0.143	**-0.268**	**-0.273**	0.007	0.186	**0.321**	**0.635**	**0.327**	**0.215**	0.166	-0.168	0.156	**0.230**	-0.193	**-0.267**	**0.283**	**0.250**	**0.225**	-0.047	0.061	0.004	0.029	-0.104	0.084	**0.341**	-0.142	0.133
WGI.V.A.	**-0.239**	0.022	**0.246**	**0.371**	**0.411**	-0.065	**-0.377**	**-0.322**	0.110	**0.201**	**0.540**	0.182	**0.458**	**0.242**	**0.330**	**-0.339**	**0.391**	**0.323**	**-0.222**	**-0.305**	0.193	**0.528**	**0.358**	0.097	0.144	-0.002	-0.177	0.111	**0.208**	0.174	-0.026	0.101
WGI.PS.AV	**-0.357**	-0.104	**0.315**	**0.521**	**0.288**	-0.165	**-0.366**	**-0.304**	0.060	0.102	**0.658**	0.000	**0.445**	**0.290**	**0.367**	**-0.367**	**0.309**	**0.325**	-0.076	-0.151	**-0.202**	0.120	**0.492**	0.041	**0.363**	0.168	**-0.228**	0.172	0.022	-0.068	-0.028	0.173
WGI.GE	**-0.604**	-0.137	**0.455**	**0.807**	**0.518**	**-0.342**	**-0.656**	**-0.593**	0.035	0.099	**0.881**	-0.193	**0.703**	**0.559**	**0.601**	**-0.624**	**0.647**	**0.482**	**-0.229**	**-0.267**	-0.122	**0.407**	**0.724**	**0.300**	0.177	-0.172	**-0.536**	**0.441**	-0.032	**-0.199**	-0.051	**0.395**
WGI.RQ	**-0.394**	-0.095	**0.404**	**0.717**	**0.474**	**-0.251**	**-0.539**	**-0.424**	0.118	0.107	**0.788**	-0.174	**0.640**	**0.439**	**0.529**	**-0.527**	**0.570**	**0.401**	-0.146	**-0.198**	-0.053	**0.410**	**0.615**	0.190	0.185	-0.103	**-0.429**	**0.329**	0.034	-0.154	-0.098	**0.244**
WGI.RofL.	**-0.550**	-0.121	**0.411**	**0.792**	**0.443**	**-0.390**	**-0.634**	**-0.505**	0.011	0.003	**0.874**	**-0.301**	**0.606**	**0.513**	**0.585**	**-0.605**	**0.621**	**0.427**	-0.121	-0.194	**-0.268**	**0.326**	**0.674**	**0.334**	**0.227**	-0.088	**-0.530**	**0.459**	-0.027	**-0.284**	0.012	**0.373**
HDI	**-0.690**	**-0.255**	**0.754**	**0.541**	**0.810**	**-0.506**	**-0.788**	**-0.844**	0.067	0.128	**0.650**	-0.114	**0.855**	**0.791**	**0.918**	**-0.902**	**0.806**	**0.625**	**-0.372**	**-0.433**	0.013	**0.279**	**0.903**	**0.478**	-0.095	**-0.475**	**-0.787**	**0.531**	0.112	-0.011	**-0.346**	**0.593**
WGI-CofC	**-0.513**	-0.096	**0.484**	**0.760**	**0.408**	**-0.363**	**-0.623**	**-0.502**	0.050	0.014	**0.913**	**-0.260**	**0.611**	**0.506**	**0.567**	**-0.602**	**0.606**	**0.400**	-0.120	-0.170	**-0.229**	**0.271**	**0.681**	**0.236**	**0.223**	-0.075	**-0.504**	**0.429**	-0.102	**-0.276**	-0.034	**0.339**

Values in bold are significantly different from 0 with a significance level alpha=0.05, values in blue correspond to supplementary variables

Table D.2: Correlation matrix (Pearson (n)) of the PCA performed on the WatSan4Dev variables –part 2

Continued…

Variables	WPI	ESI	WGI.V.A.	WGI.PS.AV	WGI.GE	WGI.RQ	WGI.RofL.	HDI	WGI-CofC
%Poverty	**-0.432**	-0.169	**-0.239**	**-0.357**	**-0.604**	**-0.394**	**-0.550**	**-0.690**	**-0.513**
% Agri Area	**-0.366**	**-0.248**	0.022	-0.104	-0.137	-0.095	-0.121	**-0.255**	-0.096
Urban Pop	**0.493**	**0.259**	**0.246**	**0.315**	**0.455**	**0.404**	**0.411**	**0.754**	**0.484**
Env gov	**0.252**	0.180	**0.371**	**0.521**	**0.807**	**0.717**	**0.792**	**0.541**	**0.760**
School enrol	**0.696**	**0.307**	**0.411**	**0.288**	**0.518**	**0.474**	**0.443**	**0.810**	**0.408**
Femal eco	**-0.355**	0.143	-0.065	-0.165	**-0.342**	**-0.251**	**-0.390**	**-0.506**	**-0.363**
% Slums	**-0.598**	**-0.268**	**-0.377**	**-0.366**	**-0.656**	**-0.539**	**-0.634**	**-0.788**	**-0.623**
Fertility	**-0.669**	**-0.273**	**-0.322**	**-0.304**	**-0.593**	**-0.424**	**-0.505**	**-0.844**	**-0.502**
Withdrawal municipal	-0.019	0.007	0.110	0.060	0.035	0.118	0.011	0.067	0.050
Withdrawal industrial	**0.251**	0.186	**0.201**	0.102	0.099	0.107	0.003	0.128	0.014
CPI	**0.398**	**0.321**	**0.540**	**0.658**	**0.881**	**0.788**	**0.874**	**0.650**	**0.913**
TWRR	**0.451**	**0.635**	0.182	0.000	-0.193	-0.174	**-0.301**	-0.114	**-0.260**
Health exp	**0.643**	**0.327**	**0.458**	**0.445**	**0.703**	**0.640**	**0.606**	**0.855**	**0.611**
S	**0.568**	**0.215**	**0.242**	**0.290**	**0.559**	**0.439**	**0.513**	**0.791**	**0.506**
Life ExpectBirth	**0.579**	0.166	**0.330**	**0.367**	**0.601**	**0.529**	**0.585**	**0.918**	**0.567**
Child Mortal-5	**-0.600**	-0.168	**-0.339**	**-0.367**	**-0.624**	**-0.527**	**-0.605**	**-0.902**	**-0.602**
WS	**0.561**	0.156	**0.391**	**0.309**	**0.647**	**0.570**	**0.621**	**0.806**	**0.606**
Ratio school G/B	**0.540**	**0.230**	**0.323**	**0.325**	**0.482**	**0.401**	**0.427**	**0.625**	**0.400**
Rural Growth	**-0.413**	-0.193	**-0.222**	-0.076	**-0.229**	-0.146	-0.121	**-0.372**	-0.120
Urban Growth	**-0.549**	**-0.267**	**-0.305**	-0.151	**-0.267**	**-0.198**	-0.194	**-0.433**	-0.170
NBI	**0.332**	**0.283**	0.193	**-0.202**	-0.122	-0.053	**-0.268**	0.013	**-0.229**
Particip	**0.229**	**0.250**	**0.528**	0.120	**0.407**	**0.410**	**0.326**	**0.279**	**0.271**
GDP per cap	**0.659**	**0.225**	**0.358**	**0.492**	**0.724**	**0.615**	**0.674**	**0.903**	**0.681**
%irrigation	**0.381**	-0.047	0.097	0.041	**0.300**	0.190	**0.334**	**0.478**	**0.236**
ODA WSS	-0.083	0.061	0.144	**0.363**	0.177	0.185	**0.227**	-0.095	**0.223**
ODA	**-0.314**	0.004	-0.002	0.168	-0.172	-0.103	-0.088	**-0.475**	-0.075
Malaria	**-0.363**	0.029	-0.177	**-0.228**	**-0.536**	**-0.429**	**-0.530**	**-0.787**	**-0.504**
Water Use int Agri	**0.329**	-0.104	0.111	0.172	**0.441**	**0.329**	**0.459**	**0.531**	**0.429**
WB	0.121	0.084	**0.208**	0.022	-0.032	0.034	-0.027	0.112	-0.102
Precip	**0.403**	**0.341**	0.174	-0.068	**-0.199**	-0.154	**-0.284**	-0.011	**-0.276**
Desert risk	**-0.401**	-0.142	-0.026	-0.028	-0.051	-0.098	0.012	**-0.346**	-0.034
Total withdrawal	**0.573**	0.133	0.101	0.173	**0.395**	**0.244**	**0.373**	**0.593**	**0.339**
WPI	**1**	**0.507**	**0.343**	**0.226**	**0.404**	**0.323**	**0.292**	**0.699**	**0.283**
ESI	**0.507**	**1**	**0.421**	**0.315**	**0.262**	**0.245**	0.156	**0.250**	**0.223**
WGI.V.A.	**0.343**	**0.421**	**1**	**0.524**	**0.663**	**0.698**	**0.586**	**0.364**	**0.553**
WGI.PS.AV	**0.226**	**0.315**	**0.524**	**1**	**0.617**	**0.533**	**0.711**	**0.414**	**0.671**
WGI.GE	**0.404**	**0.262**	**0.663**	**0.617**	**1**	**0.876**	**0.906**	**0.664**	**0.895**
WGI.RQ	**0.323**	**0.245**	**0.698**	**0.533**	**0.876**	**1**	**0.826**	**0.568**	**0.805**
WGI.RofL.	**0.292**	0.156	**0.586**	**0.711**	**0.906**	**0.826**	**1**	**0.616**	**0.916**
HDI	**0.699**	**0.250**	**0.364**	**0.414**	**0.664**	**0.568**	**0.616**	**1**	**0.613**
WGI-CofC	**0.283**	**0.223**	**0.553**	**0.671**	**0.895**	**0.805**	**0.916**	**0.613**	**1**

Values in bold are significantly different from 0 with a significance level alpha=0.05, values in blue correspond to supplementary variables

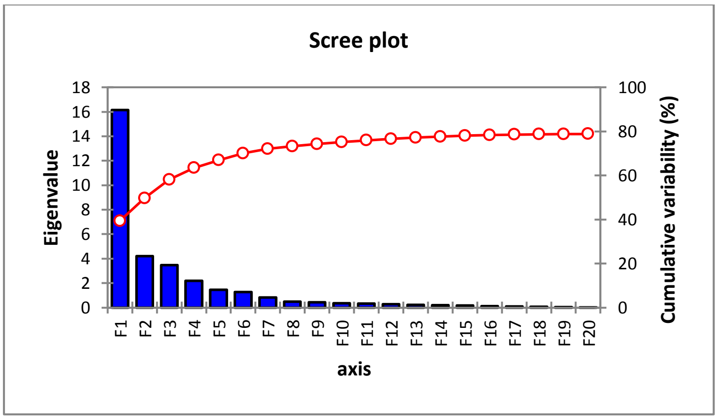
Table D.2: Correlation matrix (Pearson (n)) of the PCA performed on the WatSan4Dev variables –part 3

2. Factor Analysis of the WatSan4Dev dataset for 2004: statistical tests

The correlation matrix is the same as for the PCA (Pearson's), the Cronbach's alpha value is 0.73 and, the extraction method of factors is the principal components.

Variables	KMO	Variables	KMO
Poverty	0.91843287	WGI.GE	0.87557974
% Agri Area	0.56673442	WGI.RQ	0.91130702
Urban Pop	0.85455146	WGI.RofL.	0.85503032
Env gov	0.82302474	S	0.92068874
School enrol	0.86015812	Life Expect	0.90190477
Femal eco	0.90105474	Child Mortal-5	0.93456033
% Slums	0.9146056	WS	0.96191123
Fertility Rate	0.9453419	School G/B	0.94172534
Withdrawal municipal	0.59211156	Rural Growth	0.78425918
Withdrawal industrial	0.54677203	Urban Growth	0.80356871
CPI	0.88406625	NBI	0.58576772
TWRR	0.62454713	Particip	0.68768418
Health exp	0.86881389	GDP per cap	0.88335311
Malaria	0.91121241	%irrigation	0.88220889
Water Use int Agri	0.85430995	ODA WSS	0.54126553
WB	0.43028111	ODA	0.75994136
Precipit	0.68049014	HDI	0.89386673
Desert risk	0.60450985	WGI-CofC-2004	0.92684256
Total withdrawals	0.83324795	**KMO**	**0.85836502**
WPI	0.87523086		
ESI	0.65743787		
WGI.V.A.	0.81846525		
WGI.PS.AV	0.87218145		

Table D.3: Kaiser-Meyer-Olkin (KMO) measures of sampling adequacy –FA of the WatSan4Dev variables



	F1	F2	F3	F4	F5	F6
Eigenvalue	16.345	4.455	3.716	2.430	1.887	1.643
Variability (%)	39.867	10.865	9.063	5.927	4.602	4.008
Cumulative %	39.867	50.732	59.795	65.722	70.324	74.331

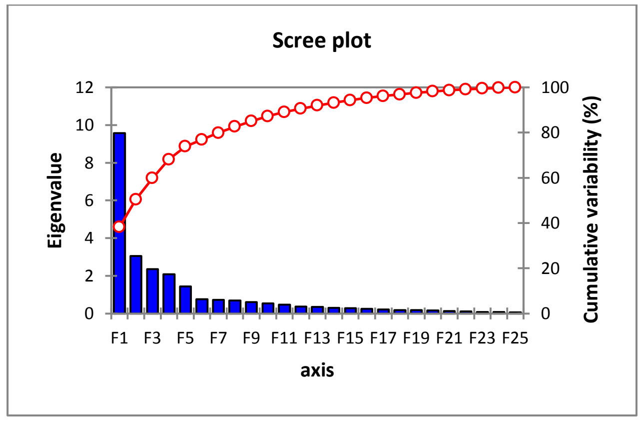
.

Variable	Final communality	Variable	Final communality
%Poverty	0.608	Particip	0.633
% Agri Area	0.567	GDP per cap	0.900
Urban Pop	0.697	%irrigation	0.778
Env gov	0.722	ODA WSS	0.661
School enrol	0.764	ODA	0.836
Femal eco	0.648	Malaria	0.817
%slums	0.780	Water Use int Agri	0.851
Fertility	0.846	WB	0.699
Withdrawal municipal	0.851	Precipit	0.856
Withdrawal industrial	0.576	Desert risk	0.699
CPI	0.866	Total withdrawals	0.863
TWRR	0.918	WPI	0.860
Health exp	0.846	ESI	0.728
S	0.711	WGI.V.A.	0.762
Life Expect Birth	0.838	WGI PS AV	0.671
Child Mortal-5	0.868	WGI GE	0.920
WS	0.782	WGI RQ	0.816
School G/B	0.512	WGI RofL	0.924
Rural Growth	0.686	HDI	0.943
Urban Growth	0.679	WGI CofC	0.889
NBI	0.766		

Table D.4: Eigenvalues for the first factors and communalities of the WatSan4Dev variables

3. PRINCIPAL COMPONENT ANALYSIS OF THE WATSAN4DEV SUBSET (25 VARIABLES) FOR 2004

Variables	KMO measure	Variables	KMO measure
% Poverty	0.942	GDP per cap	0.897
Urban Pop	0.840	%irrigation	0.865
School enrol	0.938	ODA WSS	0.515
Femal eco	0.884	ODA	0.761
% Slums	0.916	Malaria	0.901
Withdrawal municipal	0.624	Water Use int Agri	0.819
Withdrawal industrial	0.694	Precipit	0.515
TWRR	0.519	Desert risk	0.711
S	0.903	Total withdrawal	0.802
Child Mortal-5	0.927	WGI.V.A.	0.764
WS	0.970	WGI.PS.AV	0.821
School G/B	0.910	WGI.RofL	0.869
Particip	0.645	**KMO**	**0.847**



	F1	F2	F3	F4	F5
Eigenvalue	9.567	3.042	2.352	2.081	1.432
Variability (%)	38.269	12.169	9.409	8.324	5.727
Cumulative %	38.269	50.438	59.846	68.170	73.898

Table D.5: Kaiser-Meyer-Olkin (KMO) measures of sampling adequacy, scree plot and eigenvalues of the first factors - PCA for the WatSan4Dev subset

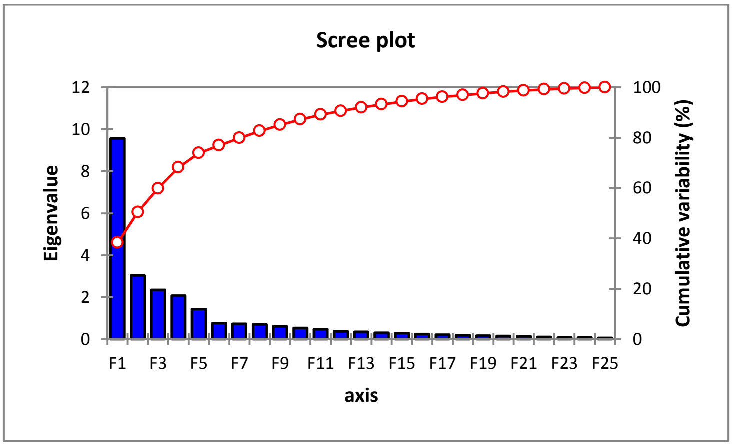
Variables	%Poverty	Urban Pop	School enrol	Femal eco	%slums	Withdrawal municipal	Withdrawal industrial	TWRR	S	Child Mortal-5	WS	School G/B	Particip	GDP per cap	%irrigation	ODA WSS	ODA	Malaria	Water Use int Agri	Precipit	Desert risk	Total withdrawals	WGI.V.A.	WGI.PS.AV	WGI.RofL.
%Poverty	**1**	**-0.493**	**-0.510**	**0.460**	**0.655**	-0.047	-0.099	**0.257**	**-0.548**	**0.611**	**-0.600**	**-0.326**	**-0.216**	**-0.708**	**-0.371**	0.009	**0.408**	**0.617**	**-0.351**	0.158	0.176	**-0.408**	**-0.239**	**-0.357**	**-0.550**
Urban Pop	**-0.493**	**1**	**0.607**	**-0.465**	**-0.650**	**0.335**	0.141	-0.134	**0.624**	**-0.630**	**0.643**	**0.392**	**0.208**	**0.727**	0.133	-0.146	**-0.305**	**-0.616**	**0.311**	-0.160	-0.172	**0.294**	**0.246**	**0.315**	**0.411**
School enrol	**-0.510**	**0.607**	**1**	**-0.411**	**-0.667**	0.059	**0.227**	0.000	**0.636**	**-0.736**	**0.671**	**0.636**	**0.348**	**0.688**	**0.401**	-0.067	**-0.348**	**-0.591**	**0.406**	0.099	**-0.284**	**0.488**	**0.411**	**0.288**	**0.443**
Femal eco	**0.460**	**-0.465**	**-0.411**	**1**	**0.511**	0.087	0.018	**0.314**	**-0.426**	**0.477**	**-0.498**	**-0.205**	0.033	**-0.566**	**-0.420**	-0.027	**0.200**	**0.589**	**-0.446**	**0.254**	0.071	**-0.435**	-0.065	-0.165	**-0.390**
%slums	**0.655**	**-0.650**	**-0.667**	**0.511**	**1**	-0.083	-0.121	**0.205**	**-0.740**	**0.714**	**-0.775**	**-0.560**	**-0.352**	**-0.801**	**-0.415**	0.042	**0.405**	**0.638**	**-0.490**	0.144	0.121	**-0.488**	**-0.377**	**-0.366**	**-0.634**
Withdrawal municipal	-0.047	**0.335**	0.059	0.087	-0.083	**1**	**0.514**	-0.077	0.005	0.052	0.093	0.039	0.101	0.131	**-0.500**	-0.054	0.031	0.094	**-0.535**	-0.009	-0.112	**-0.553**	0.110	0.060	0.011
Withdrawal industrial	-0.099	0.141	**0.227**	0.018	-0.121	**0.514**	**1**	**0.251**	0.059	-0.113	0.139	0.117	0.065	0.144	**-0.223**	0.004	-0.061	0.101	**-0.443**	**0.282**	**-0.276**	**-0.251**	**0.201**	0.102	0.003
TWRR	**0.257**	-0.134	0.000	**0.314**	**0.205**	-0.077	**0.251**	**1**	**-0.219**	0.150	**-0.249**	0.025	0.070	-0.163	-0.105	-0.041	**0.200**	**0.380**	**-0.312**	**0.742**	**-0.250**	-0.017	0.146	-0.026	**-0.344**
S	**-0.548**	**0.624**	**0.636**	**-0.426**	**-0.740**	0.005	0.059	**-0.219**	**1**	**-0.770**	**0.738**	**0.515**	**0.224**	**0.711**	**0.471**	0.004	**-0.450**	**-0.735**	**0.515**	-0.066	**-0.208**	**0.574**	**0.242**	**0.290**	**0.513**
Child Mortal-5	**0.611**	**-0.630**	**-0.736**	**0.477**	**0.714**	0.052	-0.113	0.150	**-0.770**	**1**	**-0.778**	**-0.552**	**-0.286**	**-0.786**	**-0.556**	0.126	**0.460**	**0.808**	**-0.564**	-0.046	**0.335**	**-0.570**	**-0.339**	**-0.367**	**-0.605**
WS	**-0.600**	**0.643**	**0.671**	**-0.498**	**-0.775**	0.093	0.139	**-0.249**	**0.738**	**-0.778**	**1**	**0.519**	**0.328**	**0.770**	**0.464**	-0.098	**-0.412**	**-0.731**	**0.498**	-0.108	-0.182	**0.480**	**0.391**	**0.309**	**0.621**
School G/B	**-0.326**	**0.392**	**0.636**	**-0.205**	**-0.560**	0.039	0.117	0.025	**0.515**	**-0.552**	**0.519**	**1**	**0.265**	**0.577**	**0.311**	0.023	-0.165	**-0.386**	**0.364**	0.152	**-0.275**	**0.392**	**0.323**	**0.325**	**0.427**
Particip	**-0.216**	**0.208**	**0.348**	0.033	**-0.352**	0.101	0.065	0.070	**0.224**	**-0.286**	**0.328**	**0.265**	**1**	0.180	**0.216**	-0.068	-0.186	**-0.259**	0.151	0.031	0.111	0.195	**0.528**	0.120	**0.326**
GDP per cap	**-0.708**	**0.727**	**0.688**	**-0.566**	**-0.801**	0.131	0.144	-0.163	**0.711**	**-0.786**	**0.770**	**0.577**	0.180	**1**	**0.347**	-0.030	**-0.388**	**-0.723**	**0.446**	-0.110	**-0.285**	**0.522**	**0.358**	**0.492**	**0.674**
%irrigation	**-0.371**	0.133	**0.401**	**-0.420**	**-0.415**	**-0.500**	**-0.223**	-0.105	**0.471**	**-0.556**	**0.464**	**0.311**	**0.216**	**0.347**	**1**	-0.065	**-0.290**	**-0.563**	**0.733**	0.106	-0.125	**0.687**	0.097	0.041	**0.334**
ODA WSS	0.009	-0.146	-0.067	-0.027	0.042	-0.054	0.004	-0.041	0.004	0.126	-0.098	0.023	-0.068	-0.030	-0.065	**1**	**0.534**	0.079	-0.016	-0.076	-0.014	-0.050	0.144	**0.363**	**0.227**
ODA	**0.408**	**-0.305**	**-0.348**	**0.200**	**0.405**	0.031	-0.061	**0.200**	**-0.450**	**0.460**	**-0.412**	-0.165	-0.186	**-0.388**	**-0.290**	**0.534**	**1**	**0.486**	**-0.278**	0.048	0.066	**-0.389**	-0.002	0.168	-0.088
Malaria	**0.617**	**-0.616**	**-0.591**	**0.589**	**0.638**	0.094	0.101	**0.380**	**-0.735**	**0.808**	**-0.731**	**-0.386**	**-0.259**	**-0.723**	**-0.563**	0.079	**0.486**	**1**	**-0.587**	**0.228**	0.167	**-0.568**	-0.177	**-0.228**	**-0.530**
Water Use int Agri	**-0.351**	**0.311**	**0.406**	**-0.446**	**-0.490**	**-0.535**	**-0.443**	**-0.312**	**0.515**	**-0.564**	**0.498**	**0.364**	0.151	**0.446**	**0.733**	-0.016	**-0.278**	**-0.587**	**1**	-0.193	0.001	**0.785**	0.111	0.172	**0.459**
Precipit	0.158	-0.160	0.099	**0.254**	0.144	-0.009	**0.282**	**0.742**	-0.066	-0.046	-0.108	0.152	0.031	-0.110	0.106	-0.076	0.048	**0.228**	-0.193	**1**	**-0.531**	-0.079	0.174	-0.068	**-0.284**
Desert risk	0.176	-0.172	**-0.284**	0.071	0.121	-0.112	**-0.276**	**-0.250**	**-0.208**	**0.335**	-0.182	**-0.275**	0.111	**-0.285**	-0.125	-0.014	0.066	0.167	0.001	**-0.531**	**1**	-0.026	-0.026	-0.028	0.012
Total withdrawals	**-0.408**	**0.294**	**0.488**	**-0.435**	**-0.488**	**-0.553**	**-0.251**	-0.017	**0.574**	**-0.570**	**0.480**	**0.392**	0.195	**0.522**	**0.687**	-0.050	**-0.389**	**-0.568**	**0.785**	-0.079	-0.026	**1**	0.101	0.173	**0.373**
WGI.V.A.	**-0.239**	**0.246**	**0.411**	-0.065	**-0.377**	0.110	**0.201**	0.146	**0.242**	**-0.339**	**0.391**	**0.323**	**0.528**	**0.358**	0.097	0.144	-0.002	-0.177	0.111	0.174	-0.026	0.101	**1**	**0.524**	**0.586**
WGI.PS.AV	**-0.357**	**0.315**	**0.288**	-0.165	**-0.366**	0.060	0.102	-0.026	**0.290**	**-0.367**	**0.309**	**0.325**	0.120	**0.492**	0.041	**0.363**	0.168	**-0.228**	0.172	-0.068	-0.028	0.173	**0.524**	**1**	**0.711**
WGI.RofL.	**-0.550**	**0.411**	**0.443**	**-0.390**	**-0.634**	0.011	0.003	**-0.344**	**0.513**	**-0.605**	**0.621**	**0.427**	**0.326**	**0.674**	**0.334**	**0.227**	-0.088	**-0.530**	**0.459**	**-0.284**	0.012	**0.373**	**0.586**	**0.711**	**1**

Values in bold are significantly different from 0 with a significance level alpha=0.05

Table D.6: Correlation matrix (Pearson (n)) of the PCA performed on the WatSan4Dev subset

4. FACTOR ANALYSIS OF THE WATSAN4DEV SUBSET (25 VARIABLES) FOR 2004

The correlation matrix and the KMO measures are the same as for the PCA (Pearson's), the Cronbach's alpha value is 0.37, and the extraction method of factors is the principal components.



	F1	F2	F3	F4	F5
Eigenvalue	9.567	3.042	2.352	2.081	1.432
Variability (%)	38.269	12.169	9.409	8.324	5.727
Cumulative %	38.269	50.438	59.846	68.170	73.898

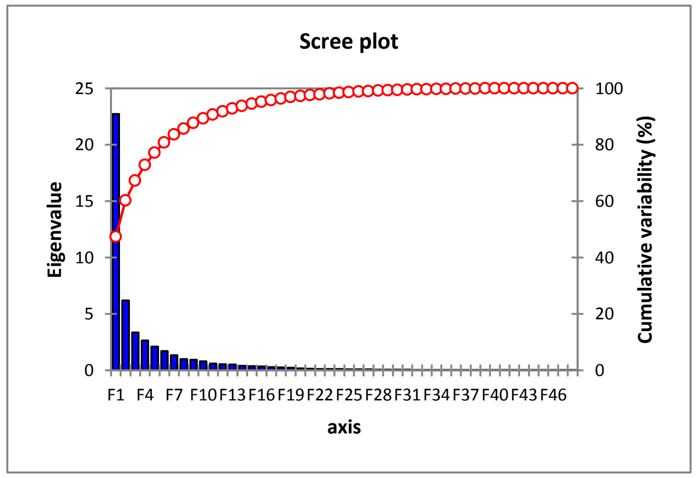
Variable	Final communality
%Poverty	0.573
Urban Pop	0.673
School enrol	0.713
Femal eco	0.566
%slums	0.775
Withdrawal municipal	0.834
Withdrawal industrial	0.604
TWRR	0.753
S	0.717
Child Mortal-5	0.842
WS	0.783
School G/B	0.510
Particip	0.747
GDP per cap	0.864
%irrigation	0.747
ODA WSS	0.690
ODA	0.724
Malaria	0.793
Water Use int Agri	0.847
Precipit	0.871
Desert risk	0.708
Total withdrawals	0.799
WGI.V.A.	0.776
WGI.PS.AV	0.726
WGI.RofL.	0.841

Table D.7: Eigenvalues of the first factors and communalities - FA of the WatSan4Dev subset

5. PRINCIPAL COMPONENT ANALYSIS OF THE WATSAN4DEV SUBSET FOR 2000-2007 PERIOD

The multivariate analyses aimed at looking for errors in the dataset following the methodology used for year 2004.

	F1	F2	F3	F4	F5	F6
Eigenvalue	22.725	6.193	3.360	2.647	2.118	1.725
Variability (%)	47.344	12.902	7.000	5.515	4.412	3.593
Cumulative %	47.344	60.247	67.247	72.762	77.174	80.768



Variable	KMO measure	Variable	KMO measure	Variable	KMO measure	Variable	KMO measure
S.2000	0.900	women.eco.2004	0.570	%irrigation.2007	0.787	WGI.PSAV.2000	0.896
S.2004	0.907	women.eco.2007	0.564	ODA.WSS.2000	0.843	WGI.PSAV.2004	0.805
S.2007	0.951	%Slums.2000	0.888	ODA.WSS.2004	0.716	WGI.PSAV.2007	0.764
WS.2000	0.902	%Slums.2004	0.893	ODA.WSS.2007	0.761	Children.Mortal.5.2000	0.927
WS.2004	0.898	%Slums.2007	0.940	ODA.2000	0.840	Children.Mortal.5.2004	0.914
WS.2007	0.932	School.G.B.2000	0.862	ODA.2004	0.848	Child.Mort.5.2007	0.933
Urban.Pop.2000	0.872	School.G.B2004	0.857	ODA.2007	0.814	WGI RofL.2007	0.832
Urban.Pop.2005	0.890	School.G.B.2007	0.895	Malaria.2000	0.943	WGI RofL.2004	0.864
Urban.pop.2008	0.900	GDP.per.cap.2000	0.927	Malaria.2004	0.933	WGI RofL.200	0.926
School.enrol.2000	0.928	GDP.per.cap.2004	0.920	Malaria..2007	0.897	**KMO**	**0.882**
School.enrol.2004	0.894	GDP.per.cap.2007	0.920	WGI.VA.2000	0.858		
School.enrol.2008	0.896	%irrigation.2000	0.869	WGI.VA.2004	0.820		
women.eco..2000	0.681	%irrigation.2004	0.812	WGI.VA.2007	0.734		

Table D.8: Kaiser-Meyer-Olkin (KMO) measures of sampling adequacy, scree plot and eigenvalues of the first factors - PCA of the WatSan4Dev variables from 2000 to 2007

Variables	S.2000	S.2004	S.2007	WS.2000	WS.2004	WS.2007	Urban.Pop.2000	Urban.Pop.2005	Urban.pop.2008	School.enrol.2000	School.enrol.2004	School.enrol.2008	women.eco..2000	women.eco.2004	women.eco.2007
S.2000	**1**	**0.977**	**0.912**	**0.765**	**0.742**	**0.731**	**0.667**	**0.654**	**0.641**	**0.669**	**0.674**	**0.667**	**-0.238**	**-0.196**	-0.181
S.2004	**0.977**	**1**	**0.893**	**0.740**	**0.730**	**0.708**	**0.623**	**0.612**	**0.597**	**0.651**	**0.675**	**0.662**	**-0.249**	**-0.206**	-0.190
S.2007	**0.912**	**0.893**	**1**	**0.775**	**0.763**	**0.768**	**0.636**	**0.625**	**0.622**	**0.703**	**0.695**	**0.707**	**-0.203**	-0.167	-0.158
WS.2000	**0.765**	**0.740**	**0.775**	**1**	**0.984**	**0.921**	**0.677**	**0.659**	**0.668**	**0.684**	**0.694**	**0.661**	**-0.248**	**-0.230**	**-0.226**
WS.2004	**0.742**	**0.730**	**0.763**	**0.984**	**1**	**0.924**	**0.645**	**0.631**	**0.641**	**0.679**	**0.689**	**0.658**	**-0.226**	**-0.213**	**-0.211**
WS.2007	**0.731**	**0.708**	**0.768**	**0.921**	**0.924**	**1**	**0.649**	**0.643**	**0.640**	**0.682**	**0.686**	**0.656**	-0.142	-0.122	-0.118
Urban.Pop.2000	**0.667**	**0.623**	**0.636**	**0.677**	**0.645**	**0.649**	**1**	**0.993**	**0.989**	**0.609**	**0.619**	**0.652**	-0.174	-0.141	-0.132
Urban.Pop.2005	**0.654**	**0.612**	**0.625**	**0.659**	**0.631**	**0.643**	**0.993**	**1**	**0.986**	**0.603**	**0.603**	**0.649**	-0.159	-0.126	-0.116
Urban.pop.2008	**0.641**	**0.597**	**0.622**	**0.668**	**0.641**	**0.640**	**0.989**	**0.986**	**1**	**0.606**	**0.617**	**0.649**	-0.153	-0.121	-0.111
School.enrol.2000	**0.669**	**0.651**	**0.703**	**0.684**	**0.679**	**0.682**	**0.609**	**0.603**	**0.606**	**1**	**0.918**	**0.908**	-0.167	-0.142	-0.138
School.enrol.2004	**0.674**	**0.675**	**0.695**	**0.694**	**0.689**	**0.686**	**0.619**	**0.603**	**0.617**	**0.918**	**1**	**0.933**	-0.178	-0.147	-0.145
School.enrol.2008	**0.667**	**0.662**	**0.707**	**0.661**	**0.658**	**0.656**	**0.652**	**0.649**	**0.649**	**0.908**	**0.933**	**1**	-0.150	-0.115	-0.111
women.eco..2000	**-0.238**	**-0.249**	**-0.203**	**-0.248**	**-0.226**	-0.142	-0.174	-0.159	-0.153	-0.167	-0.178	-0.150	**1**	**0.991**	**0.981**
women.eco.2004	**-0.196**	**-0.206**	-0.167	**-0.230**	**-0.213**	-0.122	-0.141	-0.126	-0.121	-0.142	-0.147	-0.115	**0.991**	**1**	**0.997**
women.eco.2007	-0.181	-0.190	-0.158	**-0.226**	**-0.211**	-0.118	-0.132	-0.116	-0.111	-0.138	-0.145	-0.111	**0.981**	**0.997**	**1**

Values in bold are significantly different from 0 with a significance level alpha=0.05

Table D.9: Correlation matrix (Pearson (n)) of the PCA performed on the WatSan4Dev subset for the 2000-2007 period –part 1

Continued…

Variables	S.2000	S.2004	S.2007	WS.2000	WS.2004	WS.2007	Urban.Pop.2000	Urban.Pop.2005	Urban.pop.2008	School.enrol.2000	School.enrol.2004	School.enrol.2008	women.eco..2000	women.eco.2004	women.eco.2007
%Slums.2000	**-0.772**	**-0.760**	**-0.760**	**-0.739**	**-0.724**	**-0.725**	**-0.604**	**-0.595**	**-0.589**	**-0.626**	**-0.636**	**-0.632**	**0.213**	0.164	0.151
%Slums.2004	**-0.725**	**-0.738**	**-0.732**	**-0.691**	**-0.688**	**-0.683**	**-0.495**	**-0.489**	**-0.480**	**-0.621**	**-0.610**	**-0.638**	**0.227**	0.183	0.167
%Slums.2007	**-0.800**	**-0.778**	**-0.759**	**-0.755**	**-0.752**	**-0.756**	**-0.669**	**-0.673**	**-0.665**	**-0.660**	**-0.658**	**-0.648**	0.192	0.162	0.149
School.G.B.2000	**0.540**	**0.514**	**0.518**	**0.535**	**0.559**	**0.497**	**0.399**	**0.402**	**0.395**	**0.651**	**0.647**	**0.636**	-0.076	-0.051	-0.042
School.G.B2004	**0.437**	**0.433**	**0.432**	**0.510**	**0.535**	**0.506**	**0.302**	**0.296**	**0.305**	**0.555**	**0.581**	**0.506**	0.144	0.158	0.163
School.G.B.2007	**0.461**	**0.453**	**0.497**	**0.513**	**0.543**	**0.478**	**0.242**	**0.231**	**0.238**	**0.450**	**0.496**	**0.450**	-0.143	-0.134	-0.137
GDP.per.cap.2000	**0.754**	**0.713**	**0.790**	**0.787**	**0.776**	**0.787**	**0.746**	**0.739**	**0.741**	**0.765**	**0.734**	**0.754**	**-0.203**	-0.170	-0.161
GDP.per.cap.2004	**0.750**	**0.709**	**0.791**	**0.777**	**0.767**	**0.775**	**0.742**	**0.734**	**0.735**	**0.763**	**0.734**	**0.758**	**-0.203**	-0.173	-0.166
GDP.per.cap.2007	**0.742**	**0.708**	**0.789**	**0.711**	**0.703**	**0.701**	**0.736**	**0.733**	**0.742**	**0.752**	**0.733**	**0.749**	**-0.232**	**-0.198**	-0.189
%irrigation.2000	**0.508**	**0.530**	**0.529**	**0.462**	**0.462**	**0.451**	0.161	0.154	0.149	**0.418**	**0.428**	**0.389**	**-0.251**	**-0.245**	**-0.238**
%irrigation.2004	**0.504**	**0.528**	**0.522**	**0.476**	**0.479**	**0.467**	0.163	0.156	0.153	**0.413**	**0.431**	**0.391**	**-0.253**	**-0.247**	**-0.240**
%irrigation.2007	**0.516**	**0.538**	**0.538**	**0.503**	**0.505**	**0.493**	0.180	0.172	0.170	**0.428**	**0.451**	**0.412**	**-0.255**	**-0.251**	**-0.246**
ODA.WSS.2000	**-0.364**	**-0.324**	**-0.341**	**-0.336**	**-0.312**	**-0.256**	**-0.417**	**-0.416**	**-0.419**	**-0.291**	**-0.256**	**-0.301**	0.035	0.017	0.013
ODA.WSS.2004	-0.158	-0.112	-0.174	**-0.228**	**-0.212**	-0.176	**-0.273**	**-0.284**	**-0.284**	-0.168	-0.136	-0.139	-0.035	-0.041	-0.035
ODA.WSS.2007	**-0.214**	**-0.237**	**-0.264**	-0.165	-0.168	-0.119	-0.179	-0.186	-0.179	-0.124	-0.121	-0.157	0.044	0.030	0.023
ODA.2000	**-0.337**	**-0.324**	**-0.339**	**-0.372**	**-0.346**	**-0.267**	**-0.333**	**-0.327**	**-0.339**	**-0.232**	**-0.265**	**-0.287**	0.041	0.022	0.021
ODA.2004	**-0.419**	**-0.415**	**-0.434**	**-0.448**	**-0.431**	**-0.365**	**-0.324**	**-0.314**	**-0.331**	**-0.330**	**-0.360**	**-0.324**	0.032	0.021	0.024
ODA.2007	**-0.390**	**-0.372**	**-0.446**	**-0.462**	**-0.459**	**-0.380**	**-0.328**	**-0.321**	**-0.334**	**-0.335**	**-0.341**	**-0.371**	0.058	0.038	0.043
Malaria.2000	**-0.570**	**-0.547**	**-0.574**	**-0.580**	**-0.573**	**-0.551**	**-0.504**	**-0.513**	**-0.506**	**-0.542**	**-0.540**	**-0.542**	**0.266**	**0.239**	**0.229**
Malaria.2004	**-0.651**	**-0.650**	**-0.689**	**-0.684**	**-0.686**	**-0.668**	**-0.561**	**-0.558**	**-0.558**	**-0.584**	**-0.576**	**-0.559**	**0.305**	**0.286**	**0.281**
Malaria..2007	**-0.667**	**-0.657**	**-0.707**	**-0.654**	**-0.654**	**-0.644**	**-0.584**	**-0.583**	**-0.582**	**-0.617**	**-0.634**	**-0.625**	**0.256**	**0.226**	**0.214**
WGI.VA.2000	**0.204**	**0.215**	**0.198**	**0.387**	**0.388**	**0.393**	**0.230**	**0.214**	**0.228**	**0.373**	**0.419**	**0.361**	-0.121	-0.107	-0.099
WGI.VA.2004	**0.244**	**0.255**	**0.216**	**0.405**	**0.400**	**0.408**	**0.257**	**0.246**	**0.255**	**0.368**	**0.429**	**0.375**	-0.181	-0.162	-0.154
WGI.VA.2007	0.116	0.129	0.064	**0.287**	**0.285**	**0.280**	0.162	0.151	0.161	**0.307**	**0.333**	**0.299**	-0.150	-0.133	-0.123
WGI.PSAV.2000	**0.361**	**0.344**	**0.345**	**0.452**	**0.450**	**0.522**	**0.390**	**0.388**	**0.405**	**0.392**	**0.421**	**0.367**	-0.170	-0.160	-0.154
WGI.PSAV.2004	**0.300**	**0.283**	**0.297**	**0.290**	**0.297**	**0.402**	**0.304**	**0.315**	**0.313**	**0.313**	**0.325**	**0.322**	-0.102	-0.086	-0.078
WGI.PSAV.2007	**0.279**	**0.248**	**0.280**	**0.232**	**0.234**	**0.319**	**0.319**	**0.335**	**0.336**	**0.314**	**0.310**	**0.348**	-0.095	-0.076	-0.066
Child.Mortal.5.2000	**-0.775**	**-0.758**	**-0.810**	**-0.809**	**-0.810**	**-0.817**	**-0.653**	**-0.646**	**-0.646**	**-0.770**	**-0.785**	**-0.770**	**0.219**	0.186	0.175
Child.Mortal.5.2004	**-0.767**	**-0.750**	**-0.803**	**-0.793**	**-0.796**	**-0.805**	**-0.627**	**-0.620**	**-0.623**	**-0.771**	**-0.779**	**-0.767**	**0.212**	0.180	0.169
Child.Mortal.5.2007	**-0.773**	**-0.760**	**-0.808**	**-0.798**	**-0.803**	**-0.801**	**-0.631**	**-0.616**	**-0.623**	**-0.764**	**-0.785**	**-0.763**	**0.248**	**0.216**	**0.206**
WGI RofL.2007	**0.508**	**0.477**	**0.503**	**0.551**	**0.556**	**0.606**	**0.391**	**0.397**	**0.386**	**0.457**	**0.451**	**0.445**	**-0.211**	**-0.200**	-0.194
WGI RofL.2004	**0.533**	**0.502**	**0.530**	**0.586**	**0.594**	**0.640**	**0.409**	**0.411**	**0.408**	**0.467**	**0.485**	**0.458**	**-0.242**	**-0.230**	**-0.225**
WGI RofL.200	**0.508**	**0.483**	**0.501**	**0.617**	**0.626**	**0.661**	**0.422**	**0.416**	**0.425**	**0.475**	**0.506**	**0.458**	-0.172	-0.166	-0.165

Values in bold are significantly different from 0 with a significance level alpha=0.05

Table D.9: Correlation matrix (Pearson (n)) of the PCA performed on the WatSan4Dev subset for the 2000-2007 period –part 2

Continued…

Variables	%Slums.2000	%Slums.2004	%Slums.2007	School.G.B.2000	School.G.B2004	School.G.B.2007	GDP.per.cap.2000	GDP.per.cap.2004	GDP.per.cap.2007	%irrigation.2000	%irrigation.2004	%irrigation.2007	ODA.WSS.2000	ODA.WSS.2004	ODA.WSS.2007
S.2000	**-0.772**	**-0.725**	**-0.800**	**0.540**	**0.437**	**0.461**	**0.754**	**0.750**	**0.742**	**0.508**	**0.504**	**0.516**	**-0.364**	-0.158	**-0.214**
S.2004	**-0.760**	**-0.738**	**-0.778**	**0.514**	**0.433**	**0.453**	**0.713**	**0.709**	**0.708**	**0.530**	**0.528**	**0.538**	**-0.324**	-0.112	**-0.237**
S.2007	**-0.760**	**-0.732**	**-0.759**	**0.518**	**0.432**	**0.497**	**0.790**	**0.791**	**0.789**	**0.529**	**0.522**	**0.538**	**-0.341**	-0.174	**-0.264**
WS.2000	**-0.739**	**-0.691**	**-0.755**	**0.535**	**0.510**	**0.513**	**0.787**	**0.777**	**0.711**	**0.462**	**0.476**	**0.503**	**-0.336**	**-0.228**	-0.165
WS.2004	**-0.724**	**-0.688**	**-0.752**	**0.559**	**0.535**	**0.543**	**0.776**	**0.767**	**0.703**	**0.462**	**0.479**	**0.505**	**-0.312**	**-0.212**	-0.168
WS.2007	**-0.725**	**-0.683**	**-0.756**	**0.497**	**0.506**	**0.478**	**0.787**	**0.775**	**0.701**	**0.451**	**0.467**	**0.493**	**-0.256**	-0.176	-0.119
Urban.Pop.2000	**-0.604**	**-0.495**	**-0.669**	**0.399**	**0.302**	**0.242**	**0.746**	**0.742**	**0.736**	0.161	0.163	0.180	**-0.417**	**-0.273**	-0.179
Urban.Pop.2005	**-0.595**	**-0.489**	**-0.673**	**0.402**	**0.296**	**0.231**	**0.739**	**0.734**	**0.733**	0.154	0.156	0.172	**-0.416**	**-0.284**	-0.186
Urban.pop.2008	**-0.589**	**-0.480**	**-0.665**	**0.395**	**0.305**	**0.238**	**0.741**	**0.735**	**0.742**	0.149	0.153	0.170	**-0.419**	**-0.284**	-0.179
School.enrol.2000	**-0.626**	**-0.621**	**-0.660**	**0.651**	**0.555**	**0.450**	**0.765**	**0.763**	**0.752**	**0.418**	**0.413**	**0.428**	**-0.291**	-0.168	-0.124
School.enrol.2004	**-0.636**	**-0.610**	**-0.658**	**0.647**	**0.581**	**0.496**	**0.734**	**0.734**	**0.733**	**0.428**	**0.431**	**0.451**	**-0.256**	-0.136	-0.121
School.enrol.2008	**-0.632**	**-0.638**	**-0.648**	**0.636**	**0.506**	**0.450**	**0.754**	**0.758**	**0.749**	**0.389**	**0.391**	**0.412**	**-0.301**	-0.139	-0.157
women.eco..2000	**0.213**	**0.227**	0.192	-0.076	0.144	-0.143	**-0.203**	**-0.203**	**-0.232**	**-0.251**	**-0.253**	**-0.255**	0.035	-0.035	0.044
women.eco.2004	0.164	0.183	0.162	-0.051	0.158	-0.134	-0.170	-0.173	**-0.198**	**-0.245**	**-0.247**	**-0.251**	0.017	-0.041	0.030
women.eco.2007	0.151	0.167	0.149	-0.042	0.163	-0.137	-0.161	-0.166	-0.189	**-0.238**	**-0.240**	**-0.246**	0.013	-0.035	0.023
%Slums.2000	**1**	**0.931**	**0.840**	**-0.568**	**-0.355**	**-0.367**	**-0.784**	**-0.777**	**-0.692**	**-0.427**	**-0.437**	**-0.438**	**0.366**	0.184	**0.201**
%Slums.2004	**0.931**	**1**	**0.817**	**-0.495**	**-0.298**	**-0.331**	**-0.741**	**-0.738**	**-0.633**	**-0.489**	**-0.498**	**-0.500**	**0.346**	0.145	**0.268**
%Slums.2007	**0.840**	**0.817**	**1**	**-0.536**	**-0.438**	**-0.409**	**-0.847**	**-0.832**	**-0.745**	**-0.469**	**-0.474**	**-0.478**	**0.338**	0.136	**0.198**
School.G.B.2000	**-0.568**	**-0.495**	**-0.536**	**1**	**0.671**	**0.646**	**0.579**	**0.571**	**0.537**	0.139	0.141	0.155	-0.152	-0.012	0.097
School.G.B2004	**-0.355**	**-0.298**	**-0.438**	**0.671**	**1**	**0.603**	**0.454**	**0.448**	**0.360**	**0.305**	**0.309**	**0.319**	0.032	-0.019	**0.222**
School.G.B.2007	**-0.367**	**-0.331**	**-0.409**	**0.646**	**0.603**	**1**	**0.415**	**0.408**	**0.334**	**0.246**	**0.251**	**0.270**	-0.036	0.024	0.056
GDP.per.cap.2000	**-0.784**	**-0.741**	**-0.847**	**0.579**	**0.454**	**0.415**	**1**	**0.992**	**0.897**	**0.387**	**0.373**	**0.388**	**-0.371**	**-0.221**	**-0.241**
GDP.per.cap.2004	**-0.777**	**-0.738**	**-0.832**	**0.571**	**0.448**	**0.408**	**0.992**	**1**	**0.899**	**0.394**	**0.378**	**0.394**	**-0.362**	**-0.225**	**-0.234**
GDP.per.cap.2007	**-0.692**	**-0.633**	**-0.745**	**0.537**	**0.360**	**0.334**	**0.897**	**0.899**	**1**	**0.345**	**0.319**	**0.339**	**-0.358**	**-0.213**	**-0.270**
%irrigation.2000	**-0.427**	**-0.489**	**-0.469**	0.139	**0.305**	**0.246**	**0.387**	**0.394**	**0.345**	**1**	**0.992**	**0.990**	-0.129	-0.076	-0.195
%irrigation.2004	**-0.437**	**-0.498**	**-0.474**	0.141	**0.309**	**0.251**	**0.373**	**0.378**	**0.319**	**0.992**	**1**	**0.997**	-0.147	-0.090	-0.181
%irrigation.2007	**-0.438**	**-0.500**	**-0.478**	0.155	**0.319**	**0.270**	**0.388**	**0.394**	**0.339**	**0.990**	**0.997**	**1**	-0.150	-0.090	-0.185
ODA.WSS.2000	**0.366**	**0.346**	**0.338**	-0.152	0.032	-0.036	**-0.371**	**-0.362**	**-0.358**	-0.129	-0.147	-0.150	**1**	**0.732**	**0.630**
ODA.WSS.2004	0.184	0.145	0.136	-0.012	-0.019	0.024	**-0.221**	**-0.225**	**-0.213**	-0.076	-0.090	-0.090	**0.732**	**1**	**0.584**
ODA.WSS.2007	**0.201**	**0.268**	**0.198**	0.097	**0.222**	0.056	**-0.241**	**-0.234**	**-0.270**	-0.195	-0.181	-0.185	**0.630**	**0.584**	**1**
ODA.2000	**0.373**	**0.378**	**0.354**	-0.123	0.083	-0.076	**-0.364**	**-0.365**	**-0.387**	-0.184	**-0.200**	**-0.212**	**0.844**	**0.615**	**0.634**
ODA.2004	**0.428**	**0.430**	**0.436**	-0.117	-0.030	-0.138	**-0.440**	**-0.432**	**-0.430**	**-0.309**	**-0.324**	**-0.335**	**0.767**	**0.657**	**0.689**
ODA.2007	**0.446**	**0.413**	**0.445**	-0.189	-0.098	**-0.249**	**-0.490**	**-0.483**	**-0.456**	**-0.321**	**-0.331**	**-0.341**	**0.677**	**0.617**	**0.610**
Malaria.2000	**0.585**	**0.567**	**0.665**	**-0.503**	**-0.277**	**-0.417**	**-0.655**	**-0.645**	**-0.542**	**-0.345**	**-0.354**	**-0.366**	**0.455**	**0.295**	**0.335**
Malaria.2004	**0.630**	**0.618**	**0.663**	**-0.386**	**-0.227**	**-0.370**	**-0.729**	**-0.728**	**-0.628**	**-0.544**	**-0.550**	**-0.566**	**0.379**	**0.234**	**0.315**
Malaria..2007	**0.629**	**0.625**	**0.676**	**-0.444**	**-0.280**	**-0.349**	**-0.742**	**-0.730**	**-0.681**	**-0.542**	**-0.552**	**-0.568**	**0.407**	**0.237**	**0.298**
WGI.VA.2000	**-0.202**	-0.168	**-0.275**	**0.325**	**0.367**	**0.313**	**0.352**	**0.333**	**0.284**	0.131	0.151	0.154	**0.198**	**0.201**	**0.263**
WGI.VA.2004	**-0.231**	-0.187	**-0.291**	**0.316**	**0.307**	**0.296**	**0.333**	**0.322**	**0.319**	0.090	0.111	0.115	0.159	0.173	**0.228**
WGI.VA.2007	-0.117	-0.104	-0.170	**0.254**	**0.245**	**0.198**	0.191	0.180	0.162	0.021	0.045	0.043	**0.196**	**0.207**	**0.275**
WGI.PSAV.2000	**-0.282**	**-0.262**	**-0.496**	**0.325**	**0.414**	**0.286**	**0.545**	**0.527**	**0.458**	**0.241**	**0.227**	**0.240**	0.111	0.148	0.149
WGI.PSAV.2004	**-0.249**	**-0.241**	**-0.426**	**0.352**	**0.417**	**0.239**	**0.464**	**0.468**	**0.380**	0.096	0.085	0.093	**0.215**	**0.248**	**0.262**
WGI.PSAV.2007	**-0.249**	**-0.244**	**-0.428**	**0.341**	**0.323**	0.188	**0.416**	**0.417**	**0.362**	-0.021	-0.032	-0.025	0.135	**0.263**	**0.212**
Children.Mortal.5.2000	**0.801**	**0.765**	**0.805**	**-0.613**	**-0.480**	**-0.500**	**-0.850**	**-0.842**	**-0.754**	**-0.580**	**-0.590**	**-0.597**	**0.376**	**0.264**	**0.211**
Children.Mortal.5.2004	**0.789**	**0.762**	**0.794**	**-0.609**	**-0.483**	**-0.508**	**-0.840**	**-0.831**	**-0.743**	**-0.596**	**-0.607**	**-0.616**	**0.377**	**0.260**	**0.211**
Child.Mort.5.2007	**0.776**	**0.744**	**0.791**	**-0.604**	**-0.502**	**-0.536**	**-0.832**	**-0.827**	**-0.750**	**-0.597**	**-0.605**	**-0.615**	**0.341**	**0.208**	0.174
WGI RofL.2007	**-0.437**	**-0.410**	**-0.627**	**0.433**	**0.418**	**0.423**	**0.605**	**0.609**	**0.521**	**0.298**	**0.304**	**0.316**	-0.041	0.112	0.121
WGI RofL.2004	**-0.461**	**-0.413**	**-0.648**	**0.445**	**0.455**	**0.462**	**0.624**	**0.622**	**0.558**	**0.330**	**0.337**	**0.350**	-0.024	0.102	0.132
WGI RofL.200	**-0.459**	**-0.417**	**-0.648**	**0.433**	**0.475**	**0.448**	**0.652**	**0.646**	**0.564**	**0.340**	**0.349**	**0.363**	-0.052	0.070	0.099

Values in bold are significantly different from 0 with a significance level alpha=0.05

Table D 9: Correlation matrix (Pearson (n)) of the PCA performed on the WatSan4Dev subset for the 2000-2007 period –part 3and 4

Continued…

Variables	ODA.2000	ODA.2004	ODA.2007	Malaria.2000	Malaria.2004	Malaria..2007	WGI.VA.2000	WGI.VA.2004	WGI.VA.2007	WGI.PSAV.2000	WGI.PSAV.2004	WGI.PSAV.2007	Child.Mort5.2000	Child.Mort.5.2004	Child.Mort.5.2007	WGI RofL.2007	WGI RofL.2004	WGI RofL.200
S.2000	**-0.337**	**-0.419**	**-0.390**	**-0.570**	**-0.651**	**-0.667**	**0.204**	**0.244**	0.116	**0.361**	**0.300**	**0.279**	**-0.775**	**-0.767**	**-0.773**	**0.508**	**0.533**	**0.508**
S.2004	**-0.324**	**-0.415**	**-0.372**	**-0.547**	**-0.650**	**-0.657**	**0.215**	**0.255**	0.129	**0.344**	**0.283**	**0.248**	**-0.758**	**-0.750**	**-0.760**	**0.477**	**0.502**	**0.483**
S.2007	**-0.339**	**-0.434**	**-0.446**	**-0.574**	**-0.689**	**-0.707**	**0.198**	**0.216**	0.064	**0.345**	**0.297**	**0.280**	**-0.810**	**-0.803**	**-0.808**	**0.503**	**0.530**	**0.501**
WS.2000	**-0.372**	**-0.448**	**-0.462**	**-0.580**	**-0.684**	**-0.654**	**0.387**	**0.405**	**0.287**	**0.452**	**0.290**	**0.232**	**-0.809**	**-0.793**	**-0.798**	**0.551**	**0.586**	**0.617**
WS.2004	**-0.346**	**-0.431**	**-0.459**	**-0.573**	**-0.686**	**-0.654**	**0.388**	**0.400**	**0.285**	**0.450**	**0.297**	**0.234**	**-0.810**	**-0.796**	**-0.803**	**0.556**	**0.594**	**0.626**
WS.2007	**-0.267**	**-0.365**	**-0.380**	**-0.551**	**-0.668**	**-0.644**	**0.393**	**0.408**	**0.280**	**0.522**	**0.402**	**0.319**	**-0.817**	**-0.805**	**-0.801**	**0.606**	**0.640**	**0.661**
Urban.Pop.2000	**-0.333**	**-0.324**	**-0.328**	**-0.504**	**-0.561**	**-0.584**	**0.230**	**0.257**	0.162	**0.390**	**0.304**	**0.319**	**-0.653**	**-0.627**	**-0.631**	**0.391**	**0.409**	**0.422**
Urban.Pop.2005	**-0.327**	**-0.314**	**-0.321**	**-0.513**	**-0.558**	**-0.583**	**0.214**	**0.246**	0.151	**0.388**	**0.315**	**0.335**	**-0.646**	**-0.620**	**-0.616**	**0.397**	**0.411**	**0.416**
Urban.pop.2008	**-0.339**	**-0.331**	**-0.334**	**-0.506**	**-0.558**	**-0.582**	**0.228**	**0.255**	0.161	**0.405**	**0.313**	**0.336**	**-0.646**	**-0.623**	**-0.623**	**0.386**	**0.408**	**0.425**
School.enrol.2000	**-0.232**	**-0.330**	**-0.335**	**-0.542**	**-0.584**	**-0.617**	**0.373**	**0.368**	**0.307**	**0.392**	**0.313**	**0.314**	**-0.770**	**-0.771**	**-0.764**	**0.457**	**0.467**	**0.475**
School.enrol.2004	**-0.265**	**-0.360**	**-0.341**	**-0.540**	**-0.576**	**-0.634**	**0.419**	**0.429**	**0.333**	**0.421**	**0.325**	**0.310**	**-0.785**	**-0.779**	**-0.785**	**0.451**	**0.485**	**0.506**
School.enrol.2008	**-0.287**	**-0.324**	**-0.371**	**-0.542**	**-0.559**	**-0.625**	**0.361**	**0.375**	**0.299**	**0.367**	**0.322**	**0.348**	**-0.770**	**-0.767**	**-0.763**	**0.445**	**0.458**	**0.458**
women.eco..2000	0.041	0.032	0.058	**0.266**	**0.305**	**0.256**	-0.121	-0.181	-0.150	-0.170	-0.102	-0.095	**0.219**	**0.212**	**0.248**	**-0.211**	**-0.242**	-0.172
women.eco.2004	0.022	0.021	0.038	**0.239**	**0.286**	**0.226**	-0.107	-0.162	-0.133	-0.160	-0.086	-0.076	0.186	0.180	**0.216**	**-0.200**	**-0.230**	-0.166
women.eco.2007	0.021	0.024	0.043	**0.229**	**0.281**	**0.214**	-0.099	-0.154	-0.123	-0.154	-0.078	-0.066	0.175	0.169	**0.206**	-0.194	**-0.225**	-0.165
%Slums.2000	**0.373**	**0.428**	**0.446**	**0.585**	**0.630**	**0.629**	**-0.202**	**-0.231**	-0.117	**-0.282**	**-0.249**	**-0.249**	**0.801**	**0.789**	**0.776**	**-0.437**	**-0.461**	**-0.459**
%Slums.2004	**0.378**	**0.430**	**0.413**	**0.567**	**0.618**	**0.625**	-0.168	-0.187	-0.104	**-0.262**	**-0.241**	**-0.244**	**0.765**	**0.762**	**0.744**	**-0.410**	**-0.413**	**-0.417**
%Slums.2007	**0.354**	**0.436**	**0.445**	**0.665**	**0.663**	**0.676**	**-0.275**	**-0.291**	-0.170	**-0.496**	**-0.426**	**-0.428**	**0.805**	**0.794**	**0.791**	**-0.627**	**-0.648**	**-0.648**
School.G.B.2000	-0.123	-0.117	-0.189	**-0.503**	**-0.386**	**-0.444**	**0.325**	**0.316**	**0.254**	**0.325**	**0.352**	**0.341**	**-0.613**	**-0.609**	**-0.604**	**0.433**	**0.445**	**0.433**
School.G.B2004	0.083	-0.030	-0.098	**-0.277**	**-0.227**	**-0.280**	**0.367**	**0.307**	**0.245**	**0.414**	**0.417**	**0.323**	**-0.480**	**-0.483**	**-0.502**	**0.418**	**0.455**	**0.475**
School.G.B.2007	-0.076	-0.138	**-0.249**	**-0.417**	**-0.370**	**-0.349**	**0.313**	**0.296**	**0.198**	**0.286**	**0.239**	0.188	**-0.500**	**-0.508**	**-0.536**	**0.423**	**0.462**	**0.448**
GDP.per.cap.2000	**-0.364**	**-0.440**	**-0.490**	**-0.655**	**-0.729**	**-0.742**	**0.352**	**0.333**	0.191	**0.545**	**0.464**	**0.416**	**-0.850**	**-0.840**	**-0.832**	**0.605**	**0.624**	**0.652**
GDP.per.cap.2004	**-0.365**	**-0.432**	**-0.483**	**-0.645**	**-0.728**	**-0.730**	**0.333**	**0.322**	0.180	**0.527**	**0.468**	**0.417**	**-0.842**	**-0.831**	**-0.827**	**0.609**	**0.622**	**0.646**
GDP.per.cap.2007	**-0.387**	**-0.430**	**-0.456**	**-0.542**	**-0.628**	**-0.681**	**0.284**	**0.319**	0.162	**0.458**	**0.380**	**0.362**	**-0.754**	**-0.743**	**-0.750**	**0.521**	**0.558**	**0.564**
%irrigation.2000	-0.184	**-0.309**	**-0.321**	**-0.345**	**-0.544**	**-0.542**	0.131	0.090	0.021	**0.241**	0.096	-0.021	**-0.580**	**-0.596**	**-0.597**	**0.298**	**0.330**	**0.340**
%irrigation.2004	**-0.200**	**-0.324**	**-0.331**	**-0.354**	**-0.550**	**-0.552**	0.151	0.111	0.045	**0.227**	0.085	-0.032	**-0.590**	**-0.607**	**-0.605**	**0.304**	**0.337**	**0.349**
%irrigation.2007	**-0.212**	**-0.335**	**-0.341**	**-0.366**	**-0.566**	**-0.568**	0.154	0.115	0.043	**0.240**	0.093	-0.025	**-0.597**	**-0.616**	**-0.615**	**0.316**	**0.350**	**0.363**
ODA.WSS.2000	**0.844**	**0.767**	**0.677**	**0.455**	**0.379**	**0.407**	**0.198**	0.159	**0.196**	0.111	**0.215**	0.135	**0.376**	**0.377**	**0.341**	-0.041	-0.024	-0.052
ODA.WSS.2004	**0.615**	**0.657**	**0.617**	**0.295**	**0.234**	**0.237**	**0.201**	0.173	**0.207**	0.148	**0.248**	**0.263**	**0.264**	**0.260**	**0.208**	0.112	0.102	0.070
ODA.WSS.2007	**0.634**	**0.689**	**0.610**	**0.335**	**0.315**	**0.298**	**0.263**	**0.228**	**0.275**	0.149	**0.262**	**0.212**	**0.211**	**0.211**	0.174	0.121	0.132	0.099
ODA.2000	**1**	**0.888**	**0.795**	**0.494**	**0.447**	**0.453**	0.088	0.035	0.104	0.096	**0.216**	0.177	**0.374**	**0.374**	**0.343**	-0.053	-0.049	-0.086
ODA.2004	**0.888**	**1**	**0.864**	**0.535**	**0.518**	**0.515**	0.003	-0.007	0.086	-0.020	0.165	0.158	**0.479**	**0.487**	**0.449**	-0.085	-0.093	-0.161
ODA.2007	**0.795**	**0.864**	**1**	**0.574**	**0.549**	**0.530**	-0.019	-0.018	0.092	-0.083	0.052	0.055	**0.522**	**0.533**	**0.496**	-0.148	**-0.203**	**-0.239**
Malaria.2000	**0.494**	**0.535**	**0.574**	**1**	**0.787**	**0.754**	-0.130	-0.101	-0.003	**-0.321**	**-0.226**	**-0.254**	**0.690**	**0.691**	**0.665**	**-0.405**	**-0.425**	**-0.417**
Malaria.2004	**0.447**	**0.518**	**0.549**	**0.787**	**1**	**0.875**	**-0.200**	-0.181	-0.039	**-0.361**	**-0.228**	-0.156	**0.799**	**0.802**	**0.789**	**-0.458**	**-0.485**	**-0.497**
Malaria..2007	**0.453**	**0.515**	**0.530**	**0.754**	**0.875**	**1**	**-0.283**	**-0.251**	-0.103	**-0.373**	**-0.261**	**-0.203**	**0.832**	**0.833**	**0.833**	**-0.449**	**-0.495**	**-0.503**
WGI.VA.2000	0.088	0.003	-0.019	-0.130	**-0.200**	**-0.283**	**1**	**0.897**	**0.860**	**0.611**	**0.520**	**0.369**	**-0.371**	**-0.366**	**-0.365**	**0.537**	**0.562**	**0.630**
WGI.VA.2004	0.035	-0.007	-0.018	-0.101	-0.181	**-0.251**	**0.897**	**1**	**0.944**	**0.524**	**0.524**	**0.411**	**-0.345**	**-0.338**	**-0.353**	**0.565**	**0.586**	**0.589**
WGI.VA.2007	0.104	0.086	0.092	-0.003	-0.039	-0.103	**0.860**	**0.944**	**1**	**0.415**	**0.425**	**0.371**	**-0.209**	**-0.203**	**-0.205**	**0.445**	**0.433**	**0.444**
WGI.PSAV.2000	0.096	-0.020	-0.083	**-0.321**	**-0.361**	**-0.373**	**0.611**	**0.524**	**0.415**	**1**	**0.836**	**0.714**	**-0.471**	**-0.467**	**-0.487**	**0.710**	**0.729**	**0.789**
WGI.PSAV.2004	**0.216**	0.165	0.052	**-0.226**	**-0.228**	**-0.261**	**0.520**	**0.524**	**0.425**	**0.836**	**1**	**0.864**	**-0.349**	**-0.349**	**-0.382**	**0.699**	**0.711**	**0.692**
WGI.PSAV.2007	0.177	0.158	0.055	**-0.254**	-0.156	**-0.203**	**0.369**	**0.411**	**0.371**	**0.714**	**0.864**	**1**	**-0.287**	**-0.289**	**-0.298**	**0.653**	**0.617**	**0.575**
Child.Mort.5.2000	**0.374**	**0.479**	**0.522**	**0.690**	**0.799**	**0.832**	**-0.371**	**-0.345**	**-0.209**	**-0.471**	**-0.349**	**-0.287**	**1**	**0.991**	**0.967**	**-0.577**	**-0.613**	**-0.630**
ChildMort.5.2004	**0.374**	**0.487**	**0.533**	**0.691**	**0.802**	**0.833**	**-0.366**	**-0.338**	**-0.203**	**-0.467**	**-0.349**	**-0.289**	**0.991**	**1**	**0.970**	**-0.575**	**-0.612**	**-0.628**
Child.Mort.5.2007	**0.343**	**0.449**	**0.496**	**0.665**	**0.789**	**0.833**	**-0.365**	**-0.353**	**-0.205**	**-0.487**	**-0.382**	**-0.298**	**0.967**	**0.970**	**1**	**-0.610**	**-0.645**	**-0.650**
WGI RofL.2007	-0.053	-0.085	-0.148	**-0.405**	**-0.458**	**-0.449**	**0.537**	**0.565**	**0.445**	**0.710**	**0.699**	**0.653**	**-0.577**	**-0.575**	**-0.610**	**1**	**0.964**	**0.917**
WGI RofL.2004	-0.049	-0.093	**-0.203**	**-0.425**	**-0.485**	**-0.495**	**0.562**	**0.586**	**0.433**	**0.729**	**0.711**	**0.617**	**-0.613**	**-0.612**	**-0.645**	**0.964**	**1**	**0.948**
WGI RofL.200	-0.086	-0.161	**-0.239**	**-0.417**	**-0.497**	**-0.503**	**0.630**	**0.589**	**0.444**	**0.789**	**0.692**	**0.575**	**-0.630**	**-0.628**	**-0.650**	**0.917**	**0.948**	**1**

APPENDIX E: FULL NAMES AND ABBREVIATIONS OF THE WATSAN4DEV VARIABLES

The variables list and their respective abbreviations can be kept open while reading to facilitate the understanding of the text.

Variable Full name	Abbreviation
Access to Improved Water source	WS
Access to Improved Sanitation	S
Human Development Index (HDI)	HDI
Gross Domestic Product – Purchasing Power Parity	GDP per cap
Proportion of population under the national poverty line	%Poverty
Female economic activity rate	Female Eco
Malaria cases	Malaria
Fertility Rate	Fertility
Mortality Rate for children under 5	Child Mortal-5
Life expectancy at birth	Life Expect Birth
Health expenditure per capita	Health exp
Ratio girls to boys in primary school	School G/B
Gross enrolment at primary, secondary and tertiary school	School enrol
Environmental Sustainability Index	ESI
Water Poverty Index	WPI
Percentage of land under risk of desertification (FAO arid areas)	Desert Risk
Surface of Water bodies	WB
Estimation of the total water withdrawals	Total withdrawals
Agricultural water withdrawals	Withdrawal agri
Municipal water withdrawals	Withdrawal municipal
Industrial water withdrawals	Withdrawal industrial
Agriculture area in percent of the total area	%Agri area
Total surface of irrigation in percentage	% irrigation
Urban population	Urban Pop
Urban population living in Slums	% slums
Urban population growth	Urban growth
Rural population growth	Rural growth
Governance index: Voice and accountability	WGI VA
Governance index: Political Stability and Absence of violence	WGI PS AV
Governance index: Government effectiveness	WGI GE
Governance index: Regulatory Quality (WGI RQ)	WGI RQ
Governance index: Rule of Law	WGI RofL
Governance index: Control of corruption	WGI CofC
Corruption Perception Index	CPI
Environmental governance index	Env Gov
Participation in International Environmental agreements	Particip
Total Official Development Assistance	ODA
ODA Disbursements for the water sector	ODA WSS
Human Development and Poverty composite index	HDP
Country Environmental Commitment composite index	CEC
Official Development Assistance composite index	ODA CI
Human Activity Pressure on water resources composite index	AP
Water resources availability composite index	WR

APPENDIX F: EU PRIORITY COUNTRIES IN 2006

2006 Priority Countries for Aid Effectiveness
80% of EuropeAid's expenditures (average of 2003, 2004 and 2005) were in the countries below

No	Europe, Southern Med, Middle East (Dir A)	Latin America (Dir B)	Sub-Saharan Africa, Caribbean and Pacific (Dir C)	Asia and Central Asia (Dir D)
1	Morocco			
2				Afghanistan
3	Palestinian Admin Areas			
4			South Africa	
5			~~Tanzania~~	
6	Egypt			
7			Ethiopia	
8			Dem Rep of Congo	
9			Sudan	
10			Mozambique	
11				India
12			Mali	
13			Madagascar	
14			Zambia	
15	Tunisia			
16			Burkina Faso	
17			Niger	
(50%)	**4**	**0**	**11**	**2**
18			Kenya	
19			Nigeria	
(60%)	**4**	**0**	**13**	**2**
20				Bangladesh
21			Angola	
22		Nicaragua		
23	Jordan			
24			Cameroon	
25				China
26			Chad	
27		Bolivia		
(70%)	**5**	**2**	**16**	**4**
28			Burundi	
29	Algeria			
30	Lebanon			
31				Indonesia
32			Jamaica	
33			Somalia	
34		Honduras		
35			Mauritania	
36				Vietnam
37		Peru		
(80%)	**7**	**4**	**20**	**6**
38			Haiti	
39			Guinea	
40				Pakistan
41			Zimbabwe	
42	Moldova			
43			Liberia	
44			Ivory Coast	
45			PNG	
46	Ukraine			
	9	**4**	**26**	**7**

Table F 1: European Commission priority countries for 2003-2005 period.

REFERENCES

Abida, C. L., & Girod, D. M. (2011). Do migrants improve their Hometowns? Remittances and Access to public Services in Mexico, 1995-2000. *Comparative Political Studies, 44*(3), 3-27.

Addo, H. O., Addo, K. K., & Langbong, B. (2013). Sanitation and its impact on the bacteriological quality of water : a study in three communities in Ghana. *African Journal of Food, Agriculture , Nutrition and Development, 13*(5), 8258-8272.

Ademiluyi, I. A., & Odugbesan, J. A. (2008). Sustainability and impact of the community water supply and sanitation programmes in Nigeria: an overview. *African journal of Agricultural Research, 3*(12), 811-817.

Adler, N., Yahemsky, E., & Taverdyan, R. (2010a). A framework to measure the relative socio – economic performance of developing countries. *Socio-Economic Planning Sciences, 44*(2), 73-78. doi: 10.1016/j.seps.2009.08.001

Adler, N., & Yazhemsky, E. (2010b). Improving discrimination on data envelopment analysis: PCA-DEA or variable reduction. *European Journal of Operational Research, 202,* 273-284.

African Union. (2008). *The eThekwini Declaration and AfricaSan Action Plan,* Durban: WSP Africa, Retrieved from http://www.wsp.org/sites/wsp.org/files/publications/eThekwiniAfricaSan.pdf.

Aguilera, P. A., Fernandez, A., Fernandez, R., Rumi, R., & Salmeron, A. (2011). Bayesian networks in environmental modelling. *Environmental Modelling and Software, 26*(12), 1376-1388. doi: 10.1016/j.envsoft.2011.06.004

Ahn, N., & Mira, P. (1999). *A Note on the Changing Relationship Between Fertility and Female Employment Rates in Developed Countries,*. FEDEA Working paper, (99-09).

Al-Sa'ed, R., & Mubarak, S. (2006). Sustainability assessment of onsite sanitation facilities in Ramallah-Albireh district with emphasis on technical, socio-cultural and financial aspects. *Management of Environmental Quality, 17*(2), 140-154.

Albuquerque Sardinha, R. M. (2008). Dryland Management and Combating Desertification Through Development. *Silva Lusitana, 16*(1), 21-44.

Alcamo, J., Doll, P., Kaspar, F., & Siebert.S. (1997). Global change and global scenarios of water use and availability: an application of waterGAP 1.0 (pp. 1-44): University of Kassel.

Alegre, H., Baptista, J. M., Cabrera, J. E., Cubillo, F., Duarte, P., Hirner, W., Parena, R. (2006). *Performance indicators for water supply services*. London: IWA Publishing.

Alesina, A., & Dollar, D. (2000). Who gives foreign aid to whom and why? *Journal of Economic Growth, 5*, 33-63.

Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., & Pelletier, F. (2012). Estimating trends in the total fertility rate with uncertainty using imperfect data: Examples from West Africa.*Demographic Research, 26*, 331-362. doi: 10.4054/DemRes.2012.26.15

Alkire, S. (2007). The missing dimensions of Poverty data: Introduction to the special issue. *Oxford Development Studies, 35*(4). doi: 10.1080/13600810701701863

Alkire, S., & Santos, M. E. (2010). Acute multidimensional poverty: A new index for developing countries. *OPHI Working paper* (pp. 139).

Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *The Annals of*

Mathematics, 34, 122-148.

Apaza, C. R. Measuring Governance and Corruption through the Worldwide Governance Indicators: Critiques, Responses, and Ongoing Scholarly Discussion. *Political Science & Politics, 42*, 4. doi: doi:10.1017/S1049096509090106.

Arnell, N. W., & King, R. (1998). Implications of Climate Change for global water resources. Report to departement of the Environment,Transport and the Regions. UK: Departement of Geography,University of Southampton.

Arrindell, W. A., & Van der Ende, J. (1985). An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement, 9*, 165 - 178.

Bandyopadbyay, S., & Wall, H. (2006). The determinants of aid in the post-Cold War era *Working Paper Series* (pp. 533-548): Federal Reserve Bank of St Louis.

Banes, C. (2000). How Swaziland Is Upgrading Its Slums. Retrieved from: Retrieved from http://www.citiesalliance.org/sites/citiesalliance.org/files/Civis2English%5B2%5D.pdf

Baran, E., & Jantunen, T. (2004). *Stakeholder consultation for Bayesian decision support systems in environmental management*. Paper presented at the Regional conference on Ecological and Environmental Modelling, Penang, Malaysia.

Barrett, S., & Graddy, K. (2000). Freedom, Growth, and the Environment. *Environment and Development Economics, 5*, 433-456.

Bartholomew, D. J., Steele, F., Galbraith, J., & Moustaki, I. (2008). *Analysis of Multivariate Social Science Data* (2nd Edition ed.). New York: Chapman&Hall/Crc.

Bates, B. C., Kundzewicz, Z. W., Wu, S., & Palutikof, J. P. (2008). *Climate Change and Water*. Geneva: Intergovernmental Panel on Climate Change Secretariat.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London, 53*, 370-418.

Bayyurt, N., & Yilmaz, S. (2012). *The Impacts of Governance and Education on Agricultural Efficiency: An International Analysis* Paper presented at the 8th International Strategic Management Conference.

Beall, J., Guha-Khasnobis, B., & Kanbur, R. (2010). *Urbanization and Development: Multidisciplinary Perspectives*: Oxford Scholarship.

Beguy, D. (2009). The impact of female employment on fertility in Dakar (Senegal) and Lome (Togo). *Demographic Research, 20*, 97-127. doi: 10.4054/DemRes.2009.20.7

Biswas, A. K. (2006). Water Management for Major Urban Centres. *International Journal of Water Resources Development, 22*(2), 183-197. doi: 10.1080/07900620600690789

Black, M. (1998). 1978-1998, Learning What Works: A 20 Year Retrospective View on International Water and Sanitation Cooperation (pp. 77). Washington, USA.

Bloom, D. E., Canning, D., Fink, G., Khanna, T., & Salyer, P. (2010). Urban Settlement: Data, Measures, and Trends *Urbanization and Development: Multidisciplinary Perspectives*. New York: Oxford University Press.

Borsuk, M., Stow, C., & Reckhow, K. (2004). A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling, 173*(2-3), 219-239. doi: 10.1016/j.ecolmodel.2003.08.020

Bos, R. (1999). Water resources development and health: the policy perspective. In B. H. Kay (Ed.), *Water Resources: Health, Environment and Development* (pp. 31-46). Florence, USA:CRC Press.

Boserup, E. (1989). Population, the status of women and rural development. *Population and* *development review, 15*(Supplement : rural development and population : institution and policy), 45-60.

Bossio, D., Erkossa, T., Dile, Y., McCartney, M., Killiches, F., & Hoff, H. (2012). Water implications of foreign direct investment in Ethiopia's agricultural sector. *Water alternatives, 5*(2), 223-242.

Botting, M. J., Porbeni, E. O., Joffres, M. R., Johnston.Bradley.C., Black, R. E., & Mills, E. J. (2010). Water and sanitation infrastructure for health: The impact of foreign aid. *Globalization and Health, 6*(12). doi: 10.1186/1744-8603-6-12

Braat, L. C., & van Lierop, W. J. F. (1987). *Economic-ecological modeling*. Amsterdam, The Netherland: Elsevier Science Ltd.

Brabazza, E., & Tello, J. E. (2014). A review of health governance: definitions, dimensions and tools to govern. *Health Policy, 116*(1), 11. doi: 10.1016/j.healthpol.2014.01.007

Brikmann, J. (2006). Indicators and criteria for measuring vulnerability, theoretical bases and requirements *Measuring the vulnerability to Natural Hazards* (pp. 55-77). New York: United Nations University Press

Bromley, J., Jackson, N., Clymer, O., Giacomello, A., & Jensen, F. (2005). The use of Hugin to develop Bayesian Network as an aid to integrated water resource planning. *Environmental Modelling and Software, 20*, 231-242. doi: 10.1016/j.envsoft.2003.12.021

Brouwer, R., & Hofkes, M. (2008). Integrated hydro-economic modelling: Approaches, key issues and future research directions. *Ecological Economics, 66*(1), 16-38.

Brown, C. M., Arbour, J. H., & Jackson, D. A. (2012). Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology, 61*(6), 941-954. doi: 10.1093/sysbio/sys047

Bryant, F. B., & Yarnold, P. R. (1995). Principal components analysis and exploratory and confirmatory factor analysis. *Reading and understanding multivariate statistics*. Washington, DC.: American Psychological Association.

Budds, J., & McGranahan, G. (2003). Are the debates on the water privatization missing the point: Experiences from Africa, Asia and Latin America. *15*, 87-114. doi: 10.1177/095624780301500222

Bues, A. (2011). *Agricultural Foreign Direct Investment and Water Rights: An Institutional Analysis from Ethiopia.* Paper presented at the International Conference on Global Land Grabbing UK. Retrieved from http://www.iss.nl/fileadmin/ASSETS/iss/Documents/Conference_papers/LDPI/83_Andrea_Bues.pdf

Burtaine, M. T., & Parks, B. C. (2013). When Do Environmentally-Focused Aid Projects Achieve their Objectives? Evidence from World Bank Post-Project Evaluations. *Global Environmental Politics,13*(2), 65-88.

Butala, N. M., VanRooyen, M. J., & Bhailal Patel, R. (2010). Improved health outcomes in urban slums through infrastructure upgrading. *Social Science and Medicine, 71*, 935-940.

Cai, X., McKinney, D. C., & Lasdon, L. S. (2003). Integrated hydrologic-agronomic-economic model for river basin management. *Journal of Water Resources Planning and Management, 129*(1), 16-22.

Cairncross, S. (1992). Sanitation and Water Supply: Practical Lessons from The Decade. *Water and Sanitation Discussion Paper Series* (pp. 63): The International Bank for Reconstruction and Development/The World Bank.

Canache, D., & Allison, M. E. (2005). Perceptions of Political Corruption in Latin American

Democracies. *Latin American Politics and Society, 47*(3), 91-111.
Canning, D. (2011). The Causes and Consequences of the Demographic Transition. *Working Paper Series* (pp. 1-22). USA: Program on the Global Demography of Aging.
Carbonnier, G. (2010). Official development assistance once more under fire from critics. *Revue internationale de politique de developpement, 1*, 137-142.
Carmona, G., Varela-Ortega, C., & Bromley, J. (2011). The use of participatory object-oriented Bayesian networks and agro-economic models for groundwater management in Spain. *Water Resources Management, 25*, 1509-1524. doi: 10.1007/s11269-010-9757-y
Carson, C. S. (2001). Toward a framework for Assessing Data Quality. *Working Paper*: IMF.
Castelletti, A., & Soncini-Sessa, R. (2007a). Coupling real-time control and socio-economic issues in participatory river basin planning. *Environmental Modelling and Software, 22*, 1114-1128. doi: 10.1016/j.envsoft.2006.05.018
Castelletti, A., & Soncini-Sessa, R. (2007b). Bayesian Networks and participatory modelling in water resource management. *Environmental modelling and Software, 22*, 1075-1088. doi: 10.1016/j.envsoft.2006.06.003
Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276.
Cattell, R. B. (1978). *The Scientific Use of Factor Analysis*. New York: Plenum.
Ceriani, L., & Verme, P. (2012). The origins of the Gini index: extracts from variabilità e Mutabilità(1912) by Corrado Gini. *Journal of the economic inequality, 10*(3), 421-423.
Chambers, R. (2009). Going to Scale with the Community-led Total Sanitation: Reflections on experiences, Issues and Ways Forward. *Practice Paper* (pp. 1-52). Brighton: Institute of Development Studies.
Chandra Prasad, R. P., Bhole, V., & Dutt, C. B. S. (2009). Is rapid urbanization leading to loss of water bodies? *Journal of Spatial Science, 2*(2), 43-52.
Chen, S. H., & Pollino, C. A. (2012). Good practice in Bayesian network modelling. *Environmental Modelling & Software, 37*, 134-145.
ChildInfo.org (Producer). (2012). Statistics: Children survival and health. Retrieved from http://www.childinfo.org/mortality_underfive.php
Clark, W. M., Varma, M. G. R., & Yates, R. A. (1986). Irrigation Practices: Peasant-Farming Settlement Schemes and Traditional Cultures [and Discussion]. *Philosophical transactions of the Royal Society, 316*, 229-244. doi: 0.1098/rsta.1986.0006
Clarke, P., & Hardy, R. (2007). Methods for handling missing data. *Epidemiological Methods in Life Course Research* (pp. 25): Oxford University Press.
Cohen, B. (2006). Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability. *Technology in Society, 28*(1-2), 63-80. doi: 10.1016/j.techsoc.2005.10.005
Comrey, A. L., & Lee, H. B. (1992). *A first Course in Factor Analysis*. Hillsdale, NJ: Erlbaum.
Cooper, G. F., Horvitz, E. J., & Heckerman, D. E. (1989). A method for temporal probabilistic reasoning. *Technical Report KSL 88-30* (pp. 1-24). Stanford, USA: Knowledge Systems laboratory.
Cortina, J. (1993). What is coefficient alpha: an examination of theory and applications. *Journal of applied psychology, 78*, 98-104.
Costello, A. B., & Osborne, J. W., (2005). Best Practices in exploratory factor analysis: four recommendations for getting the most. *Practical Assessment Research & Evaluation, 10*(7).

The Paris Declaration on Aid Effectiveness and Accra Agenda for Action (2005 and 2008).

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., & Spiegelhalter, D. J. (1999). *Probabilistic networks and expert systems*. Harrisonburg: Springer.

Cox, R. T. (1946). Probability, frequency and reasonable expectation. *American Journal of Physics, 14*, 1-13.

Crawford, S. L., Tennstedt, S. L., & McKindlay, J.B. (1995). A comparison of analytic methods for non-random missingness of outcomes data. *Journal of Clinical Epidemiology, 48*, 209-219.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychomerika, 16*(297-334).

Cureton, E. E., & d'Agostino, R. B. (1987). *Factor analysis: An applied approach*. Hillsdale: Lawrence Erlbaum Associates.

Davis, J. (2004). Corruption in Public Service Delivery: Experience from South Asia's Water and Sanitation Sector. *World Development, 32*(1), 53-71.

Dawsey**,** W. J., Minsker, B. S., & Amir, E. (2007). *Real time assessment of drinking water systems using Dynamic Bayesian Network*. Paper presented at the Restoring our Habitat-World Environmental and Water Resources Congress, Tampa, USA. http://ascelibrary.org/doi/book/10.1061/9780784409275

De Roo, A., Bouraoui, F., Burek, P., Bisselink, B., Vandecasteele, I., Mubareka, S., Lavalle, C. (2012). Current water resources in Europe and Africa, Matching water supply and water demand (pp. 41). Ispra: Joint Research Centre.

De Roo, A. P. J., Wesseling, C. G., & van Deursen, W. P. A. (1998). *Physically-based River Basin Modelling within a GIS: the LISFLOOD Model*. Paper presented at the GeoComputation Conference, University of Bristol, Bristol, United Kingdom. http://www.geocomputation.org/1998/06/gc_06.htm

Dean, T., & Kanazawa, K. (1988). *Probabilistic temporal reasoning*. Paper presented at the Seventh National Conference on Artificial Intelligence, Saint Paul, USA. Retrieved from http://www.aaai.org/Papers/AAAI/1988/AAAI88-093.pdf

Dean, T., & Kanazawa, K. (1989). A model for reasoning about persistence and causation *Techical Report CS-89-04* (pp. 1-23). Providence, USA: Brown University.

Dempster, A. P., Laird, N. M., & Rubin, D. B. Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, 39*, 1-38.

Dixon, J. R. (1964). *A programmed introduction to probability*. New York: John Wiley & Sons.

Dondeynaz, C., Carmona Moreno, C., & Cespedes Lorente, J. J. (2012). Analysing inter-relationships among water, governance, human development variables in developing countries. *Hydrology and Earth System Sciences, 16*, 3791-3816. doi: 10.5194/hess-16-3791-2012

Dondeynaz, C., López Puga, J., & Carmona Moreno, C. (2013). Bayesian networks modelling in support to cross-cutting analysis of water supply and sanitation in developing countries. *Hydrology and Earth System Sciences, 16*, 3791-3816. doi: 10.5194/hess-17-3397-2013

Dreibelbis, R., Greene, L. E., Freeman, M. C., Saboori, S., Chase, R. P., & Rheigans, R. (2013). Water, Sanitation and primary school attendance: multi-level assessment of determinants of household-reported absence in Kenya. *International Journal of Educational Development, 33*, 457-465. doi: 10.1016/j.ijedudev.2012.07.002

Driffield, N., & Jones, C. (2013). Impact of FDI, ODA and Migrant Remittances on Economic Growth in Developing Countries: A Systems Approach. *European Journal of Development Research*, *25*, 173-196. doi: 10.1057/ejdr.2013.1
Du, N. (2010). *Integrating surface water management in urban and regional planning: Case study of Wuhan river in China.* (PhD), University of Utrech, Enschede. Retrieved from http://www.itc.nl/library/papers_2010/phd/du.pdf (pp 164).
Du Toit, D., & Pollard, S. (2008). Updating public participation in IWRM: A proposal for a focused and structures engagement with catchment Management strategies. *Water Sa*, *34*(6), 6.
Duffy, F. H., Jones, K. J., McAnulty G.B., & Albert, M. S. (1995). Spectral coherence in normal adults: unrestricted principal component analysis; relation of factors to age, gender, and neuropsychologic data. *Clin Electroencephalogr, 26*(1).
Dyson, T. (2011). The role of the demographic transition in the process of Urbanization. *Population and Development Review*, *37* (Supplement), 34-54.
Easterly, W. (2002). *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics,*The MIT Press.
Edwards, W. (1998). Hailfinder, Tools for and experiences with Bayesian normative modelling. *American Psychologist*, *53*(4), 416-428. doi: 10.1037/0003-066X.53.4.416
El Mikawy, N., & Oia, I. Understanding and programming for linkage: democratic governance and development. *Oslo Governance Centre Discussion Paper* (pp. 14): Oslo Governance Centre.
El-Fadel, M., Maroun, R., Quba'a, R., Mawla, D., Sayess, R., Massoud, M. A., & Jamali, I. (2014). Determinants of diarrhoea prevalence in urban slums: a comparative assessment towards enhanced environmental management. *Environmental Monitoring Assessment*, *186*, 665-677. doi: 10.1007/s10661-013-3406-x
Estache, A., & Goicoechea, A. (2005). A Research Database on Infrastructure Economic Performance. *World Bank Policy Working Paper* (pp. 1-152).
Estache, A., Perelman, S., & Trujillo, L. (2005). Infrastructure Performance and Reform in Developing and Transition Economies: Evidence from a Survey of Productivity Measure. *World Bank Policy Research Working Paper* (pp. 1-27).
Esty, D. C., Levy, M., Srebotnjak, T., & De Sherbinin, A. (2005). Environmental Sustainability Index: Benchmarking National Environmental Stewardship (pp. 403). New Haven: Yale Centre for Environmental Law and Policy.
Esty, D. C., Levy, M. A., Srebotnjak, T., de Shebinin, A., Kim, C. H., & Anderson, B. (2006). Pilot 2006 Environmental Performance Index (pp. 1-362). New Haven: Yale Centre for Environmental Law & Policy.
EuropeAid. (2009). Water Sector Development and Governance: Complementarities and synergies between sector-wide Approach and Integrated Water Resource Management (pp. 44): European Commission.
European Commission, Increasing the impact of EU Development Policy: an Agenda for Change (2011).
Evans, P. (1992). Paying the Piper, An overview of community financing of water and sanitation. *Occasional Paper Series* (pp. 1-46). The Netherlands: IRC International Water and Sanitation Centre.
Everitt, S. (1975). Multivariate analysis: The need for data, and other problems. *British Journal of Psychiatry*, *126*, 227-240.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods, 4*(3), 272-299. doi: 1082-989X/99/S3.00

FAO. AQUASTAT: A global information system on water and agriculture. http://www.fao.org/nr/water/aquastat/main/index.stm

FAO. (1987). *Consultation on irrigation in Africa*. Lomé, Togo: Food and Agriculture Organisation.

FAO. (2000). Land resource potential and constraints at regional and country levels. *World Soil Resources Report* (pp. 1-122). Rome: FAO.

FAO. (2004). IWRM for sustainable use of water.50 years of international experience with the concept of the integrated water management. *Background document for the FAO/The Netherlands conference on Water for food and ecosystems* (pp. 16). Wageningen, The Netherlands: Ministry of the Agriculture, Nature and Food Quality.

FAO. (2011). Municipal and industrial water withdrawal modelling for the years 2000 and 2005 using statistical methods (pp. 15). Roma, Italy: Aquastat.

Fisman, R., & Gatti, R. (2002). Decentralization and Corruption: Evidence Across Countries. *Journal of Public economics, 83*, 325-345.

Foppen, J. W., & Kansiime, F. SCUSA: integrated approaches and strategies to address the sanitation crisis in unsewered slum areas in African mega-cities. *Reviews in Environmental Sciences and Biotechnology, 8*, 305-311. doi: 10.1007/s11157-009-9174-y.

Foster, J. E., Lopez-Calva, l. F., & Szekely, M. (2005). Measuring the Distribution of human development: methodology and an application to Mexico. *Journal of Human Development*, *6*(1), 5-25.

Froot, K. A. (1994). *Foreign Direct Investment* (K. A. Froot Ed. National Bureau of Economic Research Project Report ed.). USA: The University of Chicago Press.

Garson, D. G. (2008). Factor Analysis: Statnotes. (pp. 1-36): North Carolina State University Public Administration Program.

Gleditsch, N. P., & Sverdrup, B. O. (1996). *Democracy and the Environment*. Paper presented at the 36 the Annual Convention of the International Studies Association, Chicago.

Gerdung, A. (2008). Global Environmental Governance and the Role of Civil Society Groups. *New Zealand Journal of Environmental Law, 8*, 55-98.

Gillespie, B. (2005). Financing Urban water supply and sanitation. *International Review for Environmental Strategies, 5*(2), 449-471.

Giné Garriga, R., Perez Foguet, A., Molina, J. L., Bromley, J., & Sullivan, C. (2009). *Application of Bayesian Networks to assess water poverty*. Paper presented at the II International Conference on Sustainability Measurement and Modelling, Terrassa. http://upcommons.upc.edu/e-prints/bitstream/2117/7804/1/bayesiannetworks.pdf

Galtung, F. (2005). Measuring immeasurable: Boundaries and foundations of Macro Corruption Indices. In C. Sampford, A. Shacklock, & F. Galtung (Eds.), *Measuring Corruption* (pp. 101-130). London: Ashgate.

Gleick, P. H. (1996). Basic water requirements for human activities: Meeting basic needs. *Water International, 21*(2), 83-92. doi: 10.1080/02508069608686494

Glewwe, P., & van der Graag, J. (1988). Confronting Poverty in Developing Countries: definitions, Information and policies. *Living Standards Measurement Study. Working Paper* (pp. 9). Washington DC: The World Bank.

Gorsuch, R. L. (1974). *Factor Analysis*. Philadelphia: W.B.Sanders Co.

Gorsuch, R. L. (1983). *Factor Analysis* (2nd Edition ed.). Hillsdale,NJ: Erlbaum.

Gower, J. C. (1966). Some properties of root and vector methods used in multivariate analysis. *Biometrika, 53*, 325-338.

Greenland, S., & Finkle, W. (1995). A critical look at methods for handling covariates in epidemiologic regression analysis. *American Journal of Epidemiology, 142*, 1255-1268.

Guilford, J. P. (1954). *Psychometric methods* (2nd Edition ed.). New York: McGraw-Hill.

Haddad, M. (2005). Public Attitudes towards Socio-Cultural Aspects of Water Supply and Sanitation Services: Palestine as a Case Study. *Canadian Journal of Environmental Education, 10*, 195-211.

Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate Data Analysis* (7th Edition ed.): Prentice Hal.

Hall, A. W. (2007). Global Experience on Governance. In A. R. Turton, H. J. Hattingh, G. A. Maree, D. J. Roux, M. Claassen, & W. F. Strydom (Eds.), *Governance as a Trialogue: Government-Society-Science in Transition* (pp. 29-38).

Harman, H. H. (1976). *Modern factor analysis* (3rd Edition ed.). Chicago: University of Chicago Press.

Heckerman, D. (1995). Tutorial on learning with Bayesian Networks. *Technical Report* (pp. 56). Redmon: Microsoft.

Heidecke, C. (2006). Development and evaluation of a regional water poverty index for Benin. *Environment and production technology discussion paper* (pp. 1-55): International food Policy Research Institute.

Helmreich, B., & Horn, H. (2009). Opportunities in rain water harvesting. *Desalinisation, 248*(1-3), 118-124. doi: 10.1016/j.desal.2008.05.046

Hettige, H., Mani, M., & Wheeler, D. (1997). Industrial Pollution in Economic Development: Kuznets Revisited. *Policy Research Working Paper Series* (pp. 1-36): Development Research Group World Bank.

Hinrichsen, D., Robey, B., & Upadhyay, U. D. (1997). Solutions for a Water-Short World. *Population Reports, Serie M* (pp. 31): Johns Hopkins School of Public Health.

Hoeting, J. A., Madigan, D., & Raftery, A. E. (1999). Bayesian model averaging: a tutorial *Statistical Science, 14*, 382-401.

Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement, 65*, 202-226.

Honaker, J., & King, G. (2010). What to Do about Missing Values in Time-Series Cross-Section Data. *American Journal of Political Science, 54*, 561-581.

Honaker, J., King, G., & Blackwell, M. (2012). Amelia II: A Program for Missing Data (pp. 56).

Horton, N. J., & Lipsitz, S. R. (2001). Multiple Imputation in practice: Comparison of Software Packages for regression models with missing variables. *The American Statistician, 55*, 244-254.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*(417-441), 498-520.

Houthakker, H. S. (1959). Education and Income. *The Review of Economics and Statistic, 41*(1), 24-28.

Hudson, R. A., & Hanratty, D. M. (1989). *Bolivia: a Country Study*. Washington: GPO for the

Library of Congress.
Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. Thousand Oaks, CA: Sage Publications.
Hutton, G., Haller, L., & Bartram, J. (2007). Global cost-benefit analysis of water supply and sanitation interventions. *Journal of Water and Health, 5*(4), 481-502.
Ibrahim, J. G., Chen, M. H., Lipsitz, S. R., & Herring, A. H. (2005). Missing-data methods for generalized linear models: a comparative review. *Journal of the American statistical Association, 100*, 332-346.
IFAD. (2009). Fighting water scarcity in the Arab countries (pp. 1-8). Rome: International Fund for Agricultural Development.
IMF (Producer). (2004). List of Emerging and developing countries for 2004. Retrieved from http://www.imf.org/external/pubs/ft/weo/2004/02/data/groups.htm
IPCC. (2008). *Climate Change 2007: the Physical Science Basis, Contribution of Working Group I to the Fourth Assessment report of the Intergovernmental Panel on Climate Change* (S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, & H. L. Miller Eds.). Cambridge, UK and New York, USA: Cambridge University Press.
Jansen, R., Yu, H., Greenbaum, D., Kluger, N., Krogan, N. J., Chung, S., Gerstein, M. A Bayesian network approach for predicting protein interactions from genomic data. *Science, 302*(5644), 449-453.
Jeffers, J. N. R. (1967). Two case studies in the application principal component analysis. *Applied Statistics, Wiley for the Royal Statistical Society, 16*, 225-236. doi: 10.2307/2985919
Jemmali, H., & Matoussi, M. S. (2013). A multidimensional analysis of water poverty at local scale; application of improved water poverty index for Tunisia. *Water Policy, 15*(1), 98-116. doi: 10.2166/wp.2012.043
Jemmali, H., & Sullivan, C. (2014). Multidimensional Analysis of Water Poverty in MENA regiona: an empirical comparison with Physical Indicators. *Social Indicators Research, 115*(1), 257-282.
JMP. Joint Monitoring Programme Portal. from WHO/UNICEF http://www.wssinfo.org/
JMP. (2010). Meeting report: JMP Technical Task Force Meeting on Monitoring Drinking-water Quality (pp. 44). Villié-Morgon, France.
JMP. (2011). Drinking Water Equity, safety and sustainability. *JMP Thematic Report on Drinking Water* (pp. 1-64): UNICEF and World Health Organization.
Johnson, S. C. Hierarchical Clustering Schemes. *Psychometrika, 2*, 241-254.
Jolliffe, I. T. (2002). *Principal Component Analysis*. New York: Springer.
Joseph, F. D. J. (2005). An inventory and comparison of globally consistent geospatial databases and libraries, Geo-Spatial Data and Information (pp. 1-200). Rome: FAO.
Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement, 20*(1), 141-151. doi: 10.1177/001316446002000116
Kaiser, H. F. (1970).A second generation little-jiffy. *Psychometrika, 35*, 401-415.
Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark IV. *Education and Psychological Measurement, 34*(1), 111-117.
Karar, E. (2009). Integrated water resource management (IWRM): lessons from implementation in developing countries. *Water SA, 38*(6), 4.
Kass, M. A., & Li, Y. (2012). *Quantitative Analysis and Interpretation of Transient*

Electromagnetic Data via Principal Component Analysis. Paper presented at the Geoscience and Remote Sensing.

Katukiza, A. Y., Ronteltap, M., Oleja, A., Niwagaba, C. B., Kansiime, F., Lens, P. N. L., & 52–62, S. o. t. T. E. (2010). Selection of sustainable sanitation technologies for urban slums, a case of Bwaise III in Kampala. *Science of the Total Environnment, 409*, 52-62.

Kaufman, D., Kraay, A., & Mastruzzi, M. (2009a). Governance Matters VIII: Aggregate and Individual Governance Indicators, 1996-2008. *WB Policy Research Working Paper* (pp. 105): the World Bank.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2009b). *Response to "What Do the Worldwide Governance Indicators Measure?".* Retrieved from http://info.worldbank.org/governance/wgi/KKMResponseEJDR2Final.pdf

Kiendrebeogo, Y. (2012). Access to Improved Water Sources and Rural Productivity: Analytical framework and Cross-Country Evidence. *African Development Review, 24*(2), 153-166.

Kirmanoglu, H. (2003). *Political Freedom and Economic Well-Being: A Causality Analysis.* Paper presented at the International Conference on Policy Modelling, Istanbul, Turkey.

Kline, P. (1979). Psychometrics and psychology. London.

Knack, S. (1999). Aid Dependence and the Quality of Governance: A Cross-Country Empirical Analysis (pp. 38).

Konradsen, F., Van der Hoek, W., Amerasinghe, F. P., Mutero, C., & Boelee, E. (2004). Engineering and Malaria control: learning from the past 100 years. *Acta Tropica, 89*(2), 99-108. doi: 10.1016/j.actatropica.2003.09.013

Kontkanen, P., Myllymaki, P., & Silander, T. (1997). *Comparing predictive inference methods for discrete domains*. Paper presented at the Proceedings of the sixth International Workshop on Artificial Intelligence and Statistics, Ft. Lauderdale.

Kunce, J. T., Cook, W. D., & Miller, D. E. (1975). Random variables and correlational overkill. *Educational and Psychological Measurement, 35*, 529-534.

Kweku Baah Inkoom, D., & Zomanaa Nanguo, C. (2011). Utilisation of irrigation facilities towards poverty reduction. *Journal of Sustainable Development in Africa, 13*, 335-351.

Kyureghian, G., Capps, O. J., & Nayga, R. M. J. (2011). A missing variable imputation methodology with an empirical application. In D. M. Drukker (Ed.), *Advances in Econometrics, Volume 27 : Missing Data Methods : Cross-Sectional Methods and Applications.* Bradford, GBR: Emerald Insight.

Lambsdorff, J. G. (2006). The Methodology of the Corruption Perception Index (pp. 12): Transparency International.

Langbein, L., & Knack, S. (2010). The Worldwide Governance Indicators: Six, One, or None?. *Journal of Development Studies, 46*(2), 350-370.

Larru, J. M., & Tezanos Vasquez, S. (2012). Ayuda oficial española al desarrollo: Los retos de la especialización geográfica y sectorial. *Estudios de Economia applicada, 30*(3), 889-914.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London: Butterworth and Co.

Lawrence, P., Meigh, J., & Sullivan, C. (2002). The Water Poverty Index: an international comparison. *Keele Economic Research working paper* (pp. 24). Keele: Department of Economics Keele University.

Lee, S. M., Abbott, P., & Johantgen, M. (2005). Logistic regression and Bayesian networks to study outcomes using large data sets. *Nursing Research, 2*, 133-138.

Lescot, J. M., Leccia, O., & Vernier, F. *Challenges for integrated assessment and Cost-*

Effectiveness analysis of mitigation measures for controlling water pollution, Paper presented at the TWAM 2013 International Conference, Aveiro, Portugal. http://hal.archives-ouvertes.fr/docs/00/78/85/97/PDF/bx2013-pub00037762.pdf

Lewis, L. (2011). Cholera, Dengue Fever, and Malaria: The Unquestionable Link to Water, Retrieved from http://thewaterproject.org/cholera-dengue-fever-malaria-water.php

Li, H., Xu, L. C., & Zou, H. (2000). Corruption, Income Distribution and Growth. *Economics and Politics,12*(2), 155-185.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd Edition ed.). Hoboken: Wiley.

Lorez, M., & Bordoni, A. (2011). *Enhancing cancer registration datasets: Comparison of algorithms for Multiple Imputation of missing values,* Paper presented at the International Association of Cancer Registries Conference, Mauritius. Retrieved from http://www.nicer.org/assets/files/IACR%202011_ML%20poster.pdf

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research, 36*, 611-637.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84-99.

MacQueen, J. B. (1967). *Some Methods for classification and Analysis of Multivariate Observations,* Paper presented at the 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley. Retrieved from http://projecteuclid.org/euclid.bsmsp/1200512992

Madhulatha, T. S. (2012). An overview on clustering methods. *Journal of Engineering, 2*(4), 719-725.

Makoni, F. S., Manase, G., & Ndamba, J. (2004). Patterns of domestic water use in rural areas of zimbabwe, gender roles and realities. *Physics and Chemistry of the Earth, Parts A/B/C, 29*(15-18), 1294-1298. doi: 10.1016/j.pce.2004.09.013

Mammen, K., & Paxson, C. (2000). Women's Work and economic Development. *Journal of Economic Perspectives, 14*, 141-164.

Mara, D., Lane, J., Scott, B., & Trouba, D. (2010). Sanitation and Health. *PLoS Med,7*(11). doi: e1000363. doi:10.1371/journal.pmed.1000363

Marascuilo, A., & Levin, J. R. (1983). *Multivariate statistics in the social sciences*. Monterey, CA: Brooks/Cole.

Martın de Santa Olalla, F., Dominguez, A., Ortega, F., Artigao, A., & Fabeiro, C. (2007). Bayesian networks in planning a large aquifer in Eastern Mancha, Spain. *Environmental Modelling and Software, 22*(8), 1089-1100. doi: 10.1016/j.envsoft.2006.05.020

Martín, I. (2006). Female Employment in Mediterranean Arab countries: Much More than an Economic Issue. *IEMed Yearbook 2006* (pp. 145-149). Girona: European Institute of the Mediterranean.

Matsuno, Y., Konradsen, F., Tasumi, M., Van Der Hoek, W., Amerasinghe, F. P., & Amerasinghe, P. H. (1999). Control of Malaria Mosquito Breeding through Irrigation Water Management. *International Journal of Water Resources Development, 15*(1-2), 93-105. doi: 10.1080/07900629948952

Matteucci, M. A tutorial on clustering Algorithms: Politecnico di Milano. Retrieved from http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/index.html.

Mavrotas, G., & McGillivray, M. (2009). *Development Aid, A fresh look* (G. Mavrotas & M. McGillivray Eds.): Palgrave Macmillan.

McGavrey, S., Buszin, J., Reed, H., Smith, D. C., Rahman, Z., Andrezejewski, C., White, M. J. (2008). Community and Household determinants of water quality. *Journal of Water and Health, 6*(3), 339-349.

Meigh, J. R., McKenzie, A. A., & Sene, K. J. (1999). A grid-based approach to water scarcity estimates for eastern and southern Africa. *Water Resources Management, 13*, 85-115.

Meinzen-Dick, R., & Appasamy, P. P. (2002). Urbanization and Intersectoral Competition for Water (pp. 24): Wilson Centre.

Meulders, D., Rycx, F., & Plasman, R. A. (2004). Earnings inequality: gender, race and sexual orientation. *International Journal of Manpower, 25*(3-4), 244-250.

Mihajlovic, V., & Petkovic, M. (2001). Dynamic Bayesian Networks: A state of the Art. (pp 37): The Netherlands. Retrieved from http://www.ub.utwente.nl/webdocs/ctit/1/0000006a.pdf.

Mihci, H., Tolga Taner, M., & Sezen, B. (2012). Employment-adjusted human Development index. *South East European journal of Economics and Business, 7*(2), 115-137. doi: 10.2478/v10033-012-0020-8,

Molina, J.-L., Pulido-Velazquez, D., Garcia-Arostegui, J.-L., & Pulido-Velazquez, M. (2013). Hydro-economic river basin modelling: the application of a holistic surface-groundwater model to assess opportunity costs of water use in Spain. *Journal of Hydrology, 479*, 113-129. doi: 10.1016/j.ecolecon.2007.12.016

Molina, J. L., Bromley, J., Garcia-Arostegui, J. L., Sullivan, C. A., & Benavente, J. (2010). Integrated water resources management of overexploited hydrogeological systems using Object-Oriented Bayesian Networks. *Environmental Modelling and Software, 25*(8), 383-397. doi: 10.1016/j.envsoft.2006.05.020

Morita, S., & Zaelke, D. (2005). *Rule of Law, Good Governance, andl Sustainable Development,* Paper presented at the Seventh International Conference on Environmental Compliance and Enforcement, Marrackech, Morocco. Retrieved from http://www.inece.org/conference/7/vol1/

Murphy, K. P. (2002). *Dynamic Bayesian networks: representation, inference and learning.* (Doctor of Philosophy), University of California, Berkeley, USA. Retrieved from http://www.ee.uwa.edu.au/~roberto/research/projectsbiblio/10.1.1.93.778.pdf

Murtagh, F. (1983). A Survey of Recent Advances in Hierarchical Clustering Algorithms. *The Computer Journal, 26*(4), 354-359. doi: doi:10.1093/comjnl/26.4.354

Mutie, S. M., Mati, B., Home, P., Gadain, H., & Gathenya, J. (2006). *Evaluating land use change effects on river flow using USGS geospatial stream flow model in Mara river Basin,Kenya*. Bonn.

Nadkarni, S., & Shenoy, P. P. (2001). A Bayesian Network approach to making inferences in causal maps. *European Journal of Operational Research, 128*, 479-498.

Nadkarni, S., & Shenoy, P. P. (2004). A causal mapping approach to constructing Bayesian Networks. *Decision Support Systems, 38*, 259-281.

Neanidis, K. C., & Varvarigos, D. (2009). The allocation of volatile aid and economic growth: theory and evidence. *European journal of Political Economy, 25*(4), 447-462. doi: 10.1016/j.ejpoleco.2009.05.001

Neapolitan, R. E. (1990). *Probabilistic reasoning in expert systems: Theory and algorithms*, New York: John Wiley & Sons.

Neopalitan, R. E., & Morris, S. B. (2004). Probabilistic modeling with Bayesian Networks. In D. Kaplan (Ed.), *The sage Handbook of Quantitative Methodology for the Social Sciences* (pp. 371-390). Thousand Oaks, CA: Sage.

Neumayer, E. (2002). Do Democracies Exhibit Stronger International Commitment? A cross - Country Analysis. *Journal of Peace Research, 39*(2), 139-164.
Neumayer, E. (2003a). The Determinants of Aid allocation by regional multilateral development banks and United National agencies. *International Studies Quarterly*, *47*.
Neumayer, E. (2003b). *The pattern of Aid giving, the Impact of good governance on development*. London: Routledge.
Nicholson, A. E., & Korb, K. B. (2004). *Bayesian AI Tutorial*. Clayton, Australia.
Nicholson, S. E. (2011). Introduction to dryland environments *Dryland Climatology* (pp. 3-23). Cambridge: Cambridge University Press.
Norusis, M. J. (2005).SPSS 13.0 Statistical Procedures Companion. Chicago.
Notesein, F. (1945). Population: the long view *Food for the World* (pp. 35-57). Chicago: Chicago University of Chicago Press.
Nunnally, J. C. (1978). *Psychometric theory* (2nd Edition ed.). New York: McGraw-Hill.
Nyong, A. O., & Kanaroglou, P. S. (1999). Domestic water demand in rural semiarid northeastern Nigeria: identification of determinants and implications for policy. *Environnment & planning, 31*(12), 2127-2144.
O'Reilly, K. (2010). Combining Sanitation and Women's Participation in Water Supply: an Example from Rajasthan. *Development in Practice, 20*(1), 45-60. doi: 10.1080/09614520903436976
OECD. QWIDS: Query Wizard for International Development Statistics. http://stats.oecd.org/qwids/
OECD. (2009). *Managing Water for All: An OECD Perspective on Pricing and Financing*: OECD publishing.
OECD. (2007). Measuring Aid to the Water sector (pp. 4).Retrieved from http://www.oecd.org/dac/stats/40162562.pdf
OECD. (2011a). *Benefits of Investing in Water and Sanitation, an OECD Perspective*. (pp 148): OECD publishing.
OECD. (2011b). Quality Framework and guidelines for OECD Statistical Activities, version 2011/1 (pp. 69): OECD.
OECD. (2012). Development Aid at a glance, statistics by regions (pp. 1-17).
OECD. (2013). Aid to poor countries slips further as governments tighten budgets. Retrieved from http://www.oecd.org/development/stats/aidtopoorcountriesslipsfurtherasgovernmentstightenbudgets.htm
Olowu, D. (2003). Challenge of multi-level governance in developing countries and possible GIS applications. *Habitat International, 27*(4), 501-522.
Oluyemo, C. A. (2012). Women's empowerment as determinant for maintenance and sustainability of portable water and sanitation facilities in Ekiti State, Nigeria. *International Journal of Interdisciplinary Social Sciences, 6*(4), 49-69.
Omran, M. G. H., Engelbrech, A. P., & Salman, A. (2007). An overview of clustering methods. *Intelligent Data Analysis, 11*(6), 583-605.
Onda, K., LoBuglio, J., & Bartram, J. (2012). Global Access to Safe Water: Accounting for Water Quality and the Resulting Impact on MDG Progress. *International Journal of Environmental Research and Public Health, 9*(3), 880-894. doi: 10.3390/ijerph9030880
Onestini, M. (2011). Water Quality and Health in Poor Urban Areas of Latin America. *Water Resources Development, Vol. 27*(No. 1), 219–226.

Ongley, E. D. (1996). *Control of water pollution from agriculture. FAO irrigation and drainage paper,55,* Retrieved from http://www.fao.org/docrep/w2598e/w2598e00.htm#Contents

Ongley, E. D. (2001). Water Quality Programs in Developing Countries. *Water International, 26*(1), 14-23. doi: 10.1080/02508060108686883

Oracle. (2006). The Bayesian Approach to Forecasting. *White paper* (pp. 1-7): Oracle corporation.

Ordóñez Galán, C., Matías, J. M., Rivas, T., & Bastante, F. G. (2009). Reforestation planning using Bayesian Networks. *Environmental Modelling & Software, 24*, 1285-1292.

Orshansky, M. (1969). How Poverty is measured?. *Monthly Labor review, 92*(2), 37-41.

Pahl-Wostl, C., Arthington, A., Janos Bogardi, J., Bunn, S. E., Hoff, H., Lebel, L., Tsegai, D. (2013). Environmental flows and water governance: managing sustainable water uses. *Current Opinion in Environmental Sustainability, 5*, 341-351.

Pandey, V., Manandhar, S., & Kazama, F. (2012). Water Poverty Situation of Medium- sized river Basins in Nepal. *Water Resources Management, 26*(9), 2475-2490.

Pearl, J. (1978). An economic basis for certain methods of evaluating probabilistic forecasts. *International Journal of Man-Machine, 10*(2), 175-183. doi: 10.1016/S0020-7373(78)80010-8

Pearl, J. (1991). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo: Morgan kaufmann.

Pearl, J. (2001). Bayesian networks, causal inference and knowledge discovery. *Technical report* (pp. 9).

Pearson, E. S. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 6*(2), 559-572.

Pearson, R. (1999). Environmental indicators of healthy water resources. In B. H. Kay (Ed.), *Water Resources : Health, Environment and Development.* (pp. 13-30). Florence, USA: CRS Press.

Pellegrini, L., & Gerlagh, R. Causes of Corruption: A survey of cross-country Analyses and Extended Results. *Economics of Governance, 9*, 245-263.

Plummer, J., & Cross, P. (2006). Tackling corruption in the water and sanitation sector in Africa: starting the dialogue. *Working Paper* (pp. 41): Water and Sanitation Program (WSP).

Pradhan, M., Henrion, M., Provan, G., del Favero, B., & Huang, K. (1996). The sensitivity of belief network to imprecise probabilities: an experimental investigation. *Artificial intelligence, 85*(1-2), 363-397. doi: 10.1016/0004-3702(96)00002-1

Preacher, K. J., & MacCallum, R. C. (2002). Exploratory Factor Analysis in Behavior Genetics Research: Factor Recovery with Small Sample Sizes. *Behavior Genetics, 32*, 153-161.

Pulido-Velázquez, M., Andreu, J., Sahuquillo, A., & Pulido-Velázquez, D. (2008). Hydro-economic river basin modelling: the application of a holistic surface-groundwater model to assess opportunity costs of water use in Spain. *Ecological Economics, 66* (Special Issue Integrated Hydro-Economic Modelling), 51-65.

Rajan, R. G., & Subramanian, A. (2008). Aid and growth: what does the cross-country evidence really show? *Review of Economics and Statistics, 90*(4).

Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhya,A., 26*.

Raschid-Sally, L., & Jayakody, P. (2008). Drivers and characteristics of wastewater agriculture in developing countries: Results from a global assessment. *IWMI Research Report* (pp. 35). Colombo, Sri Lanka: International Water Management Institute.

Raskin, P. H., E.Zhu, Z.Stavisky, D. (1992). Simulation of water supply and demand in the Aral Sea region. *Water International, 17*(2), 55-67.

Ray, S. C. (2004). *Data envelopment analysis: theory and techniques for economics and operations research*. Cambridge, UK: Cambridge University Press.

Razeen, S. (2010). Regional Economic Integration in Asia: the track record and prospects. *Occasional Paper* (pp. 1-26): European Centre for International Political Economy.

Reed, B., & Coates, S. (2003). *Engineering and gender issues: evidence from low-income countries.* Paper presented at the The Institution of civil engineers.

Reich, P. (Cartographer). (1998). Global Desertification Vulnerability Map. Retrieved from http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/use/?cid=nrcs142p2_054003

Rencher, A. C., & Christensen, W. F. (2012). *Probability and Statistics: Methods of Multivariate Analysis* (3nd Edition ed.): John Wile & Sons.

Rodgers, G. B. (1979). Income and inequality as determinants of mortality: an international cross-section analysis. *Population Studies, 33*(2), 343-351.

Rogers, P., & Hall, A. W. (2003). Effective Water Governance. *TEC Background Papers Series* (Vol. 7, pp. 44). Sweden: Global Water Partnership Technical Committee.

Roma, E., & Jeffrey, P. (2010). Evaluation of community participation in sanitation scheme implementation. *Water Science and Technology, 62*(5).

Rosegrant, M., & Ringler, C. (1998). Impact on food security and rural development of transferring water out of a agriculture. *Water Policy, 1*, 567-586.

Rotberg, R. I., & Gisselquist, R. M. (2008). Strengthening African Governance, Ibrahim Index of African Governance: Results and Rankings (pp. 264). Cambridge, Massachusetts: Harvard College.

Rouquette, A., & Falissard, B. (2011). Sample size requirements for the internal validation of psychiatric scales. *International Journal of Methods in Psychiatric Research, 20*(4), 235-249. doi: 10.1002/mpr.352

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Saha, S., Gounder, R., & Su, J.-J. (2012). Is there a 'consensus' towards transparency international's Corruption Perceptions Index?. *International Journal of Business Studies, 20*(1), 1-10.

Salkind, N. J. (2014). *Statistics for People Who (Think They) Hate Statistics* (Fifth Editon ed.): Sage publications, Inc.

Santiso, C. (2001). International co-operation for democracy and good governance: moving toward a second generation? *European Journal of Development Research, 13*, 154-180.

Sarkar, S. K. A., & Rahman, M. (2005). Water Supply and Sanitation Condition of Slum Areas in Dhaka City. *Asian Journal of Water, Environment and Pollution, 5*(1), 13-15.

Sbrana, G. (2009). Technical note on the determinants of water and sanitation in Yemen. *Technical Note* (pp. 1-7). New York: Department of Economic and Social Affairs (DESA).

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*, 350-353.

Schouten, T., & Moriarty, P. (2003). Community water, community management: from system to service in rural areas (pp. 1-192): International Research Centre for Water and Sanitation.

Schreiner, B., & Hassan, R. (2011). *Transforming water management in South Africa: designing and implementing a new policy framework* (Vol. 2): Springer-Verlag.

Shihab, K., & Chalabi, N. (2007). Dynamic modeling of ground-water quality using Bayesian

techniques. *Journal of the American Water Resources Association, 43*(3), 664-674.
Shiklomanov, I. A. (1997). Section 1: Assessment of water resources and water availability of the world. *Comprehensive Assessment of the freshwater resources of the world*. Geneva: World Meteorological Organisation.
Sohaib Zubair, S., & Khan, M. A. (2014). Good Governance: Pakistan's Economic Growth and Worldwide Governance Indicators. *Pakistan Journal of Commerce and Social Sciences, 8*(1), 258- 271.
Sorlini, S., Palazzini, D., Mbawala, M., Ngassoum, M. B., & Collivagnarelli, M. C. (2013). Is drinking water from "Improved sources" really safe? A case study in the Logone valley (Chad-Cameroon). *Journal of Water and Health, 11*(4), 748-761. doi: 10.2166/wh.2013.017
Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences*. Mahwah: Lawrence Erlbaum.
Stone, M., & Brooks, R. J. (1990). Continuum regression: cross validated sequentially constructed prediction embracing ordinary least squares and principal components regression (with discussion). *Journal of the Royal Statistical Society, 52*, 237-269.
Sullivan, C. (2002). Calculating a Water Poverty Index. *World Development, 30*(7), 1195-1210.
Sullivan, C. A., Meigh, J. R., Giacomello, A. M., Fediw, T., Lawrence, P., Samad, M., Steyl, I. (2003). The water poverty index: Development and application at the community scale. *Natural Resources Forum, 27*(3), 189-199.
Susnik, J., Vamvakeridou-Lyroudia, L. S., Savic, D. A., & Kapelan, Z. (2012). Integrated System Dynamics Modelling for water scarcity assessment: Case study of the Kairouan region. *Sci. Total. Environ, 440*, 290-306. doi: 10.1016/j.scitotenv.2012.05.085
Tacke, T., & Waldmann, R. J. (2013). Infant mortality, relative income and public policy. *Applied Economics, 45*(22), 3240-3254.
Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53-55.
Temkin, L. S. (1993). *Inequality*. New York: Oxford University Press.
Thomas, M. (2009). What do the worldwide governance indicators measure? *European Journal of Development Research, 22*, 31-54. doi: doi:10.1057/ejdr.2009.32
Thurstone, L. L. (1947). *Multiple-Factor Analysis*: University of Chicago Press.
TI. Corruption Perception Index database. http://www.transparency.org/research/cpi/overview
TI. (2008). Global Corruption report 2008 (pp. 365): Transparency International.
TI. (2012). Corruption Perceptions Index 2012: Technical Methodology Note, (pp. 1-3).
Tigabu, A. D., Nicholoson, C. F., Collick, A. S., & Steenhuis, T. S. (2013). Determinants of household participation in the management of rural water supply systems: A case from Ethiopia. *Water Policy, 15*(6), 985-1000. doi: 10.2166/wp.2013.160.
Tilman, D. C., Kenneth G. Matson, Pamela A. Naylor, Rosamond. Polasky, Stephen. (2002). Agricultural sustainability and intensive production practices. *Nature, 418*, 6. doi: 10.1038/nature01014
Tjonneland, E. (1998). Aid, development and politics in Southern Africa: a critical look at new conditionalities in official development assistance. *Development Southern Africa, 15*(2), 185-200. doi: 10.1080/03768359808440005
Tuman, J. P., & Ayoub, A. S. (2004). The determinants of the Japanese official development assistance in Africa: A pooled time series analysis. *International Interactions, 30*, 47-52. doi: 10.1080/03050620490279346

Turley, R., Saith, R., Bhan, N., Rehfuess, E., & Carte, B. (2013). The effect of slum upgrading on slum dwellers' health, quality of life and social wellbeing. *Cochrane Database of Systematic Reviews, 1*. doi: 10.1002/14651858.CD010067.pub2#sthash.yaxtseoY.dpuf

UN. (2008). *The Millennium Development Goals Report 2008*, New York, USA: United Nations Department of Economic and Social Affairs.

UN. (2012). *The Millennium Development Goals Report 2012*, New York, USA.

UN-DESA. (2006). World Urbanization Prospects: The 2005 Revision *Working paper*. New York: UN Department of Economic and Social Affairs.

UN-DESA. (2009). World Urbanization Prospects, the 2009 Revision: Highlights, (pp. 45): UN Department of Economic and Social Affairs.

UN DESA. (2011). Population Distribution, Urbanization, Internal Migration and Development: An International Perspective (pp. 378): The United Nations Department of Economic and Social Affairs.

UN ESCWA. (2004). Enhancing the application of the Integrated Water resources Management in ESCWA region (pp. 42): United Nations Economic and Social Commission for the Western Asia.

UN HABITAT. (2003). *The Challenge of Slums: Global Report on Human Settlements*. London.

UN HABITAT (Producer). (2006). Slum trends in Asia. Retrieved from http://mirror.unhabitat.org/documents/media_centre/APMC/Slum%20trends%20in%20Asia.pdf

UN HABITAT. (revised version - 2010). Development Context and the Millennium Agenda. *The Challenge of Slums: Global Report on Human Settlements 2003* (pp. 1-23).

UNDP. (2013). Human Development Report: technical notes (pp. 8).

UNDP, & UNICEF. (2006). Human Development Report 2006: Children and Water, Sanitation and Hygiene: The Evidence. *Occasional paper* (pp. 9).

UNEP. Geodata portal. http://geodata.grid.unep.ch/

UNESCO. (2011). *Global education Digest 2011:Comparing Education Statistics Across the World*. Montreal, Canada: UNESCO Institute for statistics.

UNESCO-IHP., & GTZ. (2006). *Capacity building for ecological sanitation, Concepts for ecologically sustainable sanitation in formal and continuing education*.

UNFSA. (2007). State of the world population 2007: Unleashing the Potential of Urban Growth (pp. 100): The United Nations Population Fund.

UNICEF, & WHO. (2008). Progress on Drinking water and Sanitation: special focus on Sanitation, (pp. 58). UNICEF, New York and WHO, Geneva.

UNICEF, & WHO. (2009). *Diarrhoea: Why children are still dying and what can be done?* Geneva: WHO publications.

UNICEF, & WHO. (2012). Progress on Drinking water and Sanitation ,Update 2012 (pp. pp 66). New York, Geneva: UNICEF and WHO.

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological modelling, 203*, 312-318.

Van der Hoek, W., Konradsen, F., & Jehangir, W. A. (1999). Domestic use of irrigation water: Health hazard or opportunity? *International Journal of Water Resources Development, 15*, 107-119.

Varis, O., & Kuikka, S. (1997). Bayesian approach to expert judgment elicitation with case studies on climatic change impact assessment on surface waters. *Climate Change, 37*(3), 539-563. doi: 10.1023/A:1005358216361

Velicer, W. F., & Fava, J. L. (1988). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods, 3*, 231-251.

Wambui Kimani-Murage, E., & Ngindu, A. M. (2007). Quality of Water the Slum Dwellers Use: The Case of a Kenyan Slum. *Journal of Urban Health: Bulletin of the New York Academy of Medicine, 84*(6), 829-838. doi: 10.1007/s11524-007-9199-x

Warne, R. T., & Larsen, R. (2014). Evaluating a proposed modification of the Guttman rule for determining the number of factors in an exploratory factor analysis. *Psychological Test and Assessment Modelling, 56*, 104-123.

Watts, S. (2004). *Women, Water Management, and health*. Paper presented at the International Conference on Women and Infectious Diseases (ICWID).

WB. The World Bank open data portal. http://data.worldbank.org/

WB. (2005). *Remittances: Development Impact and Future Prospects*, (S. Munzele Maimbo & D. Ratha Eds.). Washington DC: The International Bank for Reconstruction and Development/the World Bank.

Whitford, A. B., Smith, H., & Mandawat, A. (2010). Disparities in access to clean water and sanitation: institutional causes. *Water policy, 12*(Supplement 1), 165-176.

WHO. (1981). Global Strategy for health for all by the Year 2000 (pp. 1-90). Geneva: World Health Organisation.

WHO. (2011). Impact of malaria control: World malaria report 2011. (pp. 51-79). Geneva: World Health Organisation publications.

WHO., & UNICEF. (2006). Core questions on drinking-water and sanitation for household surveys, (pp. 24). Geneva: World Health Organization and UNICEF.

Wilcox, R. R. (2009). *Basic Statistics: Understanding Conventional Methods and Modern Insights*. Cary, NC, USA: Oxford University Press.

Wolf, S. (2007). Does Aid Improve Public Service Delivery? *Review of the World Economics, 143*(4), 650-672. doi: 10.1007/s10290-007-0126-8

WSSCC. (2006). For her it's a big issue: Putting women at the centre of water supply, sanitation and hygiene, (pp. pp 36). Geneva: Water Supply and Sanitation Collaborative Council.

WSSCC-B. (2009). Grassroots in Bangladesh: targeting space for the marginalized in Sanitation movement in Bangladesh. *Technical Report* (pp. 1-6). Dhaka: Water Supply and Sanitation Collaborative Council- Bangladesh-WSSCC-B.

Xiang, Y. (2002). *Probabilistic reasoning in multi-agent systems, A graphical models approach*. Cambridge: Cambridge University Press.

Xie, H., You, L., Wielgosz, B., & Ringler, C. (2014). Estimating the potential for expanding smallholder irrigation in Sub-Saharan Africa. *Agricultural Water Management*, *131*, 183-193.

You, L., Ringler, C., Wood-Sichra, U., Robertson, R., Stanley., W., Zhu, T., Sun, Y. (2011). What is the irrigation potential for Africa? A combined biophysical and socioeconomic approach. *Food Policy, 36*(6), 770-782.

Zhang, R., Duan, Z., Tan, M., & Chen, X. (2012). The assessment of water stress with the Water Poverty Index in the Shiyang river Basin in China. *Environmental Earth Sciences, 67*(7), 5. doi: 0.1007/s12665-012-1655-6

Zhao, N. (Producer). (2009). The minimum Sample Size in Factor Analysis. Retrieved from https://www.encorewiki.org/display/~nzhao/The+Minimum+Sample+Size+in+Factor+Analysis.

Zorrilla, P., Carmona, G., De la Hera, A., Varela-Ortega, C., Martínez-Santos, P., Bromley, J., &

Henriksen, H. J. (2009). Evaluation of Bayesian Networks in Participatory Water Resources Management. *Ecology and Sociology, 15*, pp 17.

Zweig, G. G. (1998). *Speech recognition with dynamic Bayesian networks.* (Doctor of Philosophy), University of California, Berkeley, USA. Retrieved from http://www.icsi.berkeley.edu/ftp/pub/speech/papers/zweig_thesis.pdf.