Hydrol. Earth Syst. Sci. Discuss., 9, 485–539, 2012
www.hydrol-earth-syst-sci-discuss.net/9/485/2012/
doi:10.5194/hessd-9-485-2012

Hydrology and Earth System Sciences Discussions

Analysing inter-relationships among water, governance, human development variables in developing countries: WatSan4Dev database coherency analysis

C. Dondeynaz[1,2,3], C. Carmona Moreno[2], and J. J. Céspedes Lorente[3]

[1]Liverpool University, Liverpool, UK
[2]Institute for Environment and Sustainability, Joint Research Centre, Ispra, Italy
[3]Universidad de Almería, Almería, Spain

Received: 4 November 2011 – Accepted: 23 December 2011 – Published: 11 January 2012

Correspondence to: C. Dondeynaz (celinedondeynaz@yahoo.fr)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

The "Integrated Water Resources Management" principle was formally laid down at the International Conference on Water and Sustainable development in Dublin 1992. One of the main results of this conference is that improving Water and Sanitation Services (WSS), being a complex and interdisciplinary issue, passes through collaboration and coordination of different sectors (environment, health, economic activities, governance, and international cooperation). These sectors influence or are influenced by the access to WSS. The understanding of these interrelations appears as crucial for decision makers in the water sector. In this framework, the Joint Research Centre (JRC) of the European Commission (EC) has developed a new database (WatSan4Dev database) containing 45 indicators (called variables in this paper) from environmental, socio-economic, governance and financial aid flows data in developing countries.

This paper describes the development of the WatSan4Dev dataset, the statistical processes needed to improve the data quality; and, finally, the analysis to verify the database coherence is presented.

At the light of the first analysis, WatSan4Dev Dataset shows the coherency among the different variables that are confirmed by the direct field experience and/or the scientific literature in the domain. Preliminary analysis of the relationships indicates that the informal urbanisation development is an important factor influencing negatively the percentage of the population having access to WSS. Health, and in particular children health, benefits from the improvement of WSS. Efficient environmental governance is also an important factor for providing improved water supply services.

The database would be at the base of posterior analyses to better understand the interrelationships between the different indicators associated in the water sector in developing countries. A data model using the different indicators will be realised on the next phase of this research work.

HESSD

9, 485–539, 2012

1 Introduction

The experience of development cooperation in the water sector shows that only building water supply and sanitation infrastructure is inefficient in bringing sustainable water supply and sanitation services to the population. Water supply and sanitation is a complex issue that impact other sectors across the society, in first line, health and environment but also institutional capacities, and economic sectors such as agriculture and industries.

A multi-dimensional approach is needed and therefore, in 1992, principles for sustainable management in the water sector are formalized with the adoption of the integrated water resources management (IWRM) approach[1] by the international community. "IWRM" is defined as a "process aimed at ensuring that water is used more efficiently (economic dimension), promoting equitable access to water (social dimension) and guaranteeing sustainability (environmental dimension)" (EuropeAid, 2009). Concretely, this has led to a redefinition of strategies and ways of behaving in the water sector: improving efficiency and sustainable development of water and sanitation services should ensure the involvement of all stakeholders (institutional, civil society, suppliers, funders...) and sectors (education, heath, economic activities...) concerned to building appropriate facilities.

Millennium Development Goals initiative (MDG's) also calls for increasing efforts and finding solutions to extend the WSS coverage. Relevant for developing countries, the millennium goals were set by the international community in 2000 to foster efforts towards poverty reduction. Water and sanitation is concerned under the Target 7C: Halve, by 2015, the proportion of the population without sustainable access to safe drinking water and basic sanitation.

In term of research, this cross-cutting approach is translated by performing crossed analyses of the different dimensions of a question. For instance, Adler et al. (2009)

[1] Principles of DUBLIN international conference are available online (Dublin statement for water and sustainable developement, 1992).

have built a framework for the analysis human development index data, financial resources, and the MDG's targets. It allows the evaluation of country progress towards MDG's considering the development measure together with financial flows indicators. Botting et al. (2010) evaluated the impact of Official Development Aid (ODA) on water and sanitation coverage and infant and child mortality.

Applying the same approach, this research wants to identify the key indicators (variables) explaining the various levels of access to WSS at national scale. The aim of this research is to map the relationships between existing variables which influence or/and being influenced by the WSS. Having such a comprehensive view could support decision making at country scale and the international donors could have a general picture helping in orienting better their investments. This requires gathering various indicators on economic and social status, governance, and environment. Using the standard MDG indicators (the percentages of the population having access to improved water supply and sanitation), the objective is to build a methodological framework to analyze the variable behaviours and thus, to map them.

Within this framework, this paper presents the data sources and variable selection criteria that are included in the WaterSan4Dev database (Sect. 2), the methods and analysis performed to establish the coherence of the dataset (Sect. 3). It finally presents the results of the coherency verification process first for African countries and then extended to the rest of the developing countries for the year 2004 (Sect. 4). This year is taken as reference because of being the last release of the MDG's variables regarding water and sanitation when this work started. The results of the exploratory interpretation respectively the key variables impacted or being influenced by the level of WSS and the found countries profiles are also detailed in Sect. 4.

2 Database description

International institutions collect and provide data with detailed methodologies necessary to support and orient their actions. The indicators that monitor progresses towards

HESSD
9, 485–539, 2012

the MDG's are under the responsibility of the United Nations. These data are freely available and accessible through web online databases and can be considered as reference data.

The indicators used in this work are collected from the World Bank (WB), Organisation for Economic Co-operation and Development (OECD), Food and Agriculture Organization (FAO), World Health Organization (WHO), United Nations Department of Economic and Social Affairs (UN DESA), United Nations Development Programme (UNDP), United Nations Statistics Division (UNSD), United Nations Human Settlements Programme (UN-HABITAT) and the Joint Monitoring Programme (JMP). Some indicators come from research institutions such as Universities (Yale, Columbia, Harvard Universities), NGO (Transparency International) or Institutes (Wallingford, Centre for Ecology and Hydrology) that benefit from international recognition in the domain.

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 is chosen as most of the data are supplied at this level. Data sets for 2004 are used because the last release of the Joint Monitoring Programme (JMP) report on WSS access level at the beginning of this research. Moreover, the majority of other indicators present more completed datasets for 2002 to 2004.

The variables are chosen considering all the variables that can both result in and/or influence (double-way relationships) the WSS levels. These variables are thematically clustered under four main areas or pillars: environment (water resources variables), human pressure (demographic, human activity pressure variables), governance (variables measuring various aspects of governance including environmental aspects), human development (social and health variables).

Complementary to these four pillars, as developing countries are the main focus of this work, the Official development Aid delivery is included in the database. This last pillar represents the global disbursed official aid provided to the developing countries by donors.

HESSD
9, 485–539, 2012

The data collection covers countries worldwide. 132 variables were analysed based on the following main criteria: (i) relevance: the variable plays a potential role regarding the water supply or sanitation level; (ii) data availability: the dataset has enough observations (less than 50 percents of the dataset are missing); (iii) reliability: the variable is produced by trustful providers and fully described methods; and, (iv) inter-correlation: variables with high correlations with other variables (above 0.9) are removed.

After applying these different criteria to the whole set of variables, 45 variables are selected for this work (Table 1) and included in the WatSan4Dev database. The description of all variables is detailed in Appendix A.

The errors and incoherence of the dataset (the relationships between the variables and magnitudes of the values) are tracked through Principal Component Analysis (PCA). Because of the strong heterogeneity of the sources and the computation methods, the WatSan4Dev dataset should be considered by the community for qualitative analysis and not for quantitative interpretation purposes.

The number of countries is also reduced from 192 to 156: small states, islands and countries having more than 35 % of missing values are removed to limit analysis perturbations and, in particular, to avoid bias in the imputation process. Table 2 lists the countries being considered in the WatSan4Dev dataset.

In order to simplify the analysis and coherency of the WaterSan4Dev dataset, this is analysed in two phases: (1) in a first phase, the analysis is done for the African countries sub-set; and, (2) in a second phase, using the conclusions and analysis results from the African sub-set, the whole WatSan4dev database is analysed.

In the specific case of the African countries subset, the following indicators were also included because of their specific relevance regarding this continent: the Biological Oxygen Demand (BOD), the Children Diarrhoea proportion (Child diarrhoea), and the prevalence of Diarrhoea in slums (% diarrhoea slums). These indicators are excluded of the WatSan4Dev dataset because of too many missing values for the rest of the countries.

HESSD
9, 485–539, 2012

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

3 Methods

3.1 Normalization and data imputation

One of the major difficulties in implementing and analysing the WatSan4Dev dataset is missing data. The solution adopted is to process the missing data using multiple imputation methods (Horton and Lipsitz, 2001). Effectively, standard imputation methods such as mean, mode or nearest neighbour methods introduce important biases in the data distribution and therefore have an impact on the analysis and interpretations.

The imputation process can only be applied to data following a normal distribution. Therefore, previous to this imputation step, normalization of the distribution is performed on these 45 variables to allow better missing values imputation. Standard normalization processes are used according to the data distribution (square root, logarithmic and Ordinary Least Square (OLS) regression normalization). Complementary normalization tests were performed to verify the statistical stability of the variables meaning that the data distribution was not changed because of the normalisation process.

The multiple imputation methods compare country observations based on several indicators and impute missing data without modifying the statistical nature of the variables. Missing data computations are performed to obtain realistic values rather than accurate quantitative values because of the qualitative nature of the indicators collected. This will assure the coherency and significance of the analysis performed but quantitative interpretation is to be avoided.

In this work, the Expectation-Maximization Algorithm (EM) (Honaker and King, 2010) was chosen as a multiple imputation method. It 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 about the missing data.

HESSD
9, 485–539, 2012

The Amelia II[2] software used in this paper for data imputation implements the EMB algorithm (Expectation Maximization Bootstrap). This algorithm combines the classic EM algorithm with a bootstrap approach: for each draw, algorithm bootstraps the data to simulate estimation uncertainty and then run the EM algorithm to find the mode of the posterior for the bootstrapped data. Several completed sets are created and then combines under the analyst control. The assumptions are (1) the imputation model assumes that the complete data (that is, both observed and unobserved) are multivariate normal and (2) the data are missing at random (MAR). This algorithm is fully described in Honaker et al. (2011).

The dataset is incrementally completed by imputing missing data for variables with less missing data before processing more incomplete ones. This process is done on the worldwide dataset with 156 observations – countries. The imputation process provides satisfying results including manual verifications and specific case-by-case corrections.

3.2 Analysis methods

3.2.1 Principal Component Analysis (PCA)

The Principal Component Analysis (PCA) is a statistic description technique allowing an optimised graphic representation of multi-dimensional data. The representation allows 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 advantage of the method is the reduction of the initial N-dimensional space into a low dimensional map (the optimal view for a variability criterion) and build a set of P uncorrelated factors

[2]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).

HESSD
9, 485–539, 2012

(called Principal Components). The PCA technique is widely used in the multivariate analyses domain to derive dominant patterns of variability.

From a formal point of view, the PCA uses an orthogonal transformation converting a set of observations (M) of possibly correlated variables (N) into a set of values of uncorrelated variables called principal components. 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 (that is, accounts 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. This essential feature allows the representation on the Figs. 1 and 2 of only the three main factors of the PCA because gathering a significant level of variability. Therefore the interpretation is facilitated because being directly readable as the number of dimensions is reduced. The independency of the principal components is ensured as the data set is jointly normally distributed after the normalization phase. PCA computation details can be found in Pearson (1902) and Hotelling (1935).

PCA is used in this work to verify the coherency and robustness of the WatSanDev dataset tracking data errors and incoherencies among the variables. This analysis is suitable for understanding multi-dimensional objects, in this case the behaviour of 45 variables-indicators for 156 observations-countries. In this case the PCA provides a global view of the correlations among variables and/or countries, thus indicating areas for further analysis and orienting the modelling in the second phase of the work.

3.2.2 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 called factors. While the 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 produces error estimates by the analysis of the only shared variances (Bartholomew et al., 2008). The latter method allows the identification of the structure underlying such variables and to estimate scores to measure latent factors themselves. FA computation details can be found in Thurstone (1947).

In this paper, FA is used to confirm and refine the PCA results and analyses performed on the WatSan4dev database. The FA analysis is focused here on the four first components following the significant factors loadings revealed by the PCA. This analysis is also used to statically confirm the structure of the database which was done according to thematic criteria.

3.2.3 Agglomerative Hierarchical Clustering (AHC)

Agglomerative Hierarchical clustering is a bottom-up algorithm processing each observation of a dataset as a singleton cluster at the outset. Then, the AHC algorithm successively merges (or agglomerates) pairs of clusters until all clusters have been merged into a single cluster that contains all the observations. Further information about the AHC methods can be found in Ward (1963).

The AHC is used in this paper to identify homogenous country profiles according to the significant factors defined by the Factor Analysis.

3.2.4 Ordinary Least Square (OLS) linear regression analysis

Although, the modelling phase is not part of this paper as will be considered in a later stage of this work, the classical Ordinary Least Square (OLS) regression analysis is used here to provide an exploratory identification of the key explanatory variables influencing the water supply and sanitation access variables to be explained.

The OLS regression analysis is a method for predicting the value of dependent variables Y_i, based on the values of independent variables X_i. The equation system can be written as follow:

HESSD
9, 485–539, 2012

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

where Y_i are the dependent variables, $\beta_{0,i}$ are the intercepts of the model, $X_{j,i}$ corresponds to the jth 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.

The results from 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 for estimating the risk of assuming the null hypothesis. Finally, the relative influence of the explanatory variables are considered as significant if complying with the field experience, the classical cases studies or the scientific literature in the domain.

4 Results

The coherency validation process using PCA is performed in two main steps (1) the coherence of the database is first performed on African countries because of the homogeneity of the countries (clusters of countries – profiles – can be computed); (2) same analyses and methods are then applied and extended to the rest of the developing countries in WatSan4Dev. Being the most interesting output of this work, the coherence validation process over WatSan4Dev is completed and complemented by a factor analysis.

The WatSan4Dev variables in the dataset are considered as coherent if other than showing statistical coherency and robustness, they also comply with field experiences, cases studies and scientific literature.

An exploratory interpretation for both the water supply and the sanitation access levels is provided. Linear regression analyses help to identify the key variables impacting or being impacted by the WSS. A complementary clustering is performed over the WatSan4Dev database to define country profiles.

HESSD
9, 485–539, 2012

4.1 The African countries sub-set

Countries in the African subset show closer statistical behaviours, coherency and stability than the whole set of developing countries. The main aim of this process is to test methodologies using robust statistical analysis that will drive the analyses for the rest of developing countries in the WatSan4Dev dataset.

4.1.1 Coherency of the African countries sub-set

The Principal Component Analysis is applied to the variables selected (Table 1). Composite Indicators, namely the Worldwide Governance Indicators (WGI), Governance Index (GI Afr), Human development Index (HDI), Environmental Sustainability Index (ESI), Human Poverty Index (HPI), Water Poverty (WPI) are re-projected into the PCA space to facilitate the interpretation of the Principal Components and avoid bias produced by the partial overlap (autocorrelation) of their sub pillars with active variables.

The cumulated variability of the first three components is equal to 50.386 % (Table 3). Although this quantity of variability can be considered as low, if we take into account the heterogeneity, collection methods and the nature of the variables, the cumulated variability of these first three principal components can be considered as suitable for a first interpretation. However, caution is needed when interpreting the maps, as some information might be hidden in subsequent factors.

Figure 1 shows the relationships among the variables of the WatSan4Dev dataset – African sub-set. As displayed in Table 4, the first component (F1) shows that development indicators (Group 1) are negatively correlated with poverty indicators (Group 2). The second component (F2) represents the indicators related to water resources and water demand, (Groups 3 and 4).

Water supply and sanitation levels are negatively correlated with children mortality under 5 yr old (Child Mortal-5) and the fertility rates (Fertility) as demonstrated by the

HESSD
9, 485–539, 2012

experience from the field: without safe drinking water and/or proper sanitation, the first victims are children because of water and sanitation related diseases dissemination[3].

Governance variables, namely the Worldwide Governance Indicators (WGI), Corruption perception Index (CPI), Environmental governance and Governance Index (GI Afr) in Group 1, even collected through various institutions, methods and data sources, are coherent and highly correlated together (values from 0.645 to 0.852). They are positively correlated with the development of the country and the income (Gross Domestic Product – "GDP per cap", Health expenditure – "Health exp", Gross school enrolment – "School enrol" and the literacy rate of youth – "literacy Youth").

As well known, Water demand is in majority for agricultural purpose in developing countries (Hinrichsen et al., 1997). Almost 80 % of African countries considered (105 out of 136) dedicate water to agriculture at more than 50 %. Few countries showing very low total water withdrawals per capita, such as Congo, transfer water needs to domestic purposes. Therefore, total water withdrawal is negatively correlated to domestic and industrial withdrawals.

The Environmental Sustainability Index (ESI in Group 4) is a cross-cutting and a very complex index. It demonstrates the capacity of a country to manage its environment in a sustainable way. Therefore, as expected, this is highly correlated with many variables, but in particular positively correlated with the water resources availability (TWRR in Group 4), the level of urbanization (With Domestic in Group 4) and negatively correlated with agriculture pressure (Water Use int Agri). A correlation with the Domestic water demand (With Domestic) underlies an urban context because it considers withdrawals from water pipe infrastructure (Hinrichsen et al., 1997). The literature calls for caution with this index because of its multi-dimensional aspect and the sustainability concept itself. Lee et al. (2005) mentioned that an increase of income impacts positively on the control of the pollution but has a detrimental effect on most eco-efficiency measures of ESI.

[3]http://www.unicef.org/wash/index_healthandeducation.html.

HESSD
9, 485–539, 2012

The level of agricultural land equipped in irrigation system (% irrigation area) is more correlated with the level of development (human development index HDI, life expectancy at birth – "life expect birth") of the country than to the water resources availability (TWRR), the surface of water bodies (WB) or the amount of precipitations (Precipit). We advance the hypothesis here that 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 underlies investment, technological capacities and strong political will for agricultural development. Environmental conditions account for little in the choice of irrigation development in this case.

The Biological Oxygen Demand (BOD in Group 1), an indicator of the water surface quality, is correlated with development indicators (mainly Gross Domestic product – "GDP per cap", Tot Access to sanitation – "AIS", and the mortality rate of children under 5 – "Child Mortal-5"), the urbanization pressure (Growth Urban) as well as intensive agriculture (% irrigation areas) indicators. Indeed, high Biological Oxygen Demand (BOD) can be caused by wastewater discharge in urban and industrial areas or by intensive agriculture practices. Therefore, in this specific context, BOD measures the impact and degree of development of the human activities rather than being a specific environmental indicator.

The participation of women into economic activities is high correlated with the fertility rates, and the mortality of children under 5 yr old, and then to all other variables expressing the poverty. This correlation agrees with the well-known behaviour of female economic participation described in Boserup (1989), Mammen and Paxson (2000), and Beguy (2009): the percentage of women being economically active is high when the income per capita is low and drops with the increment of GDP of the country until the threshold of 2550 $ per capita (valid for the 1970–1985 period). After this threshold, the trend inverts and the female economic activity increases with the GDP (Appendix B). African countries are clearly coherent with this behaviour.

HESSD
9, 485–539, 2012

The variable on participation in international environmental agreements (Particip IEAg in Group 3) shows a negative correlation with the water resources available (TWRR in Group 4) and the amount of precipitation (Precipit in Group 4). This may indicate that countries having water scarcity issues are likely to engage themselves in international environmental strategy and initiative.

Central variables in PCA show no significant relationships with the rest of variables; however the following remarks can be made:

- Regarding the dam capacity, no significant correlations can be observed with any other variable except agricultural area. Here again, it is clearly shown that the capacity of countries to have reservoirs and dams neither impacts directly nor indirectly the level of access to WSS.

- The agricultural surface area is lowly correlated with National Biodiversity Index (NBI) (0.414) and dam capacity (Dam). No definitive conclusions can be made, as the factors loading of these variables are distributed on the three axes and close to the centre.

- The Children Diarrhoea Prevalence behaviour (Child Diarrhoea) is mainly caused by inverse negative correlation with the indicators of development (Group 1). Low correlations are observed with poverty (Group 2).

- Financial flow, either global aid (ODA) or financial flow specific to the WSS (ODA WSS), has few significant correlations in this African sub-set and is distributed among the three components with a low correlation with the other variables linked to WSS itself. Financial flows are close linked to governance indicators in African countries but this needs to be further analysed in a second stage of this work.

The variable correlations are coherent with the common scientific knowledge and the African data subset can be considered as consistent and coherence.

HESSD
9, 485–539, 2012

4.1.2 African countries sub-set: exploratory interpretation

This regression analysis should be considered as a confirmation of the relationships deduced from the PCA analysis and an exploratory interpretation phase rather than a modelling of the dataset. We remind that criteria for validating the "linear model" consider the significance of the variables, the confidence intervals, and the logic relying on at least on case studies and/or literature. Separated analyses are performed for water supply and sanitation level (Tables 5 and 6).

The key elements are children mortality, environmental governance, and water withdrawal due to industries that confirm coherency of the dataset by highlighting coherent relationships. The positive impact on mortality is well-known: the access to safe drinking water reduces diarrhoea prevalence and thus, mortality of the under five years-old children by 20 % (UNICEF, 2009). In the framework of the IWRM, good environmental governance and in particular water governance is encouraged as being a good leverage for implementing water supply facilities and their sustainability, ensuring water quality and equitable sharing between uses.

The link between industrial water needs and water supply level can be understood in two ways: first, an industrial water demand supposes that the existing water sector is well-organized and consolidated enough (meaning a high percentage of WSS access) to allow industrial development without hacking supplies for the population. On the contrary, an important need in terms of water for industrial activities may lead to creation of water infrastructure, water resource management capacities and thus, the population benefits from the industrial development. More investigation is needed to examine the positive but complex link between industrial development and water infrastructure at national scale, in particular in case of massive private investments in specific places.

Keys elements linked with sanitation are related to the advance of the society regarding health, governance and the level of urbanisation. In fact, basic sanitation is more likely to be available in a country that invests on health where environmental governance exists, and where corruption is low; allowing the setting up of water resources management.

Non appropriate sanitation and wastewater treatments lead to diseases development and dissemination however sanitation is neglected for decades. Countries that invest large funds in health in general are likely to include as well sanitation improvements. The increasing stress on water resources also fosters the re-use of wastewater sanitation products for agriculture, often called environmental sanitation. Many initiatives follow this approach that tackles the sanitation and wastewater treatment as well as food production issues (Scott et al., 2002).

The Non Governmental Organisation "Transparency International" has demonstrated the important impact of poor governance and corruption on water supply and sanitation services delivery in particular in slums (Transparency International, 2008). However, improved Sanitation facilities are expected to be created in urban areas where open defecation is a less possible and socially acceptable alternative, in addition to cost-effective advantages. This explains the gap between rural and urban population: sanitation figures are always far lower in rural areas than in urban areas. In southern Asia in 2005, this rate was 19 % in urban areas and only 6 % in rural areas, and in sub-saharian Africa respectively 31 % and 13 % (JMP, 2005).

The analysis performed in this section demonstrated the coherency of the African countries sub-set and brought first clues on key elements being influenced or influencing the water supply and sanitation services level.

4.2 Coherency validation of the WatSan4Dev dataset

This section presents the analyses of the WatSan4Dev dataset driven by the conclusions obtained from the African subset. The analyses performed include the rest of developing countries in the world (Table 2). The World Bank set the threshold at 12 276 current $ per capita but the Bahrain, Kuwait Quatar, United Emirates are included here because being considered as emerging and developing countries by the FMI. Countries from Eastern Europe and Central Asia are excluded because respectively be out of our target countries and a group with particular behaviour that should imply a specific analysis. Therefore, 101 developing countries are considered in these analyses.

HESSD
9, 485–539, 2012

The principal component analysis

The first four components show that 64.57 % of the variability of the variables is explained (Table 7), composite indicators are re-projected within the PCA (Fig. 2). Note that in this case, the variability explained (58.79 %) by the 3 first of the PCA is increased by 8 % compared to the African sub-set (50.86 %) due to the increment of observations-countries in the analysis. Several variables were excluded from this global analysis because of the important amount of missing values (greater than 30 %): Agricultural Production Index – "Agri Prod Index", Dam capacity – "DAM" and Children Diarrhoea prevalence – "Children Diarrhoea" and Prevalence Diarrhoea in slums – "% Diarrhoea slums".

The first component (F1) represents the variables linked to socio-economic and poverty conditions. The second component (F2) represents the environmental conditions and the third one (F3) the country industrialisation level (industrial and intensive agricultural pressure). Finally, the fourth factor (F4) represents the global disbursed official aid (Table 8).

As for the African sub-set, 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 – "AIS" and "AIWS", Human development Index – "HDI", life expectancy at birth – "Life expect Birth") as expected.

Poverty is also negatively correlated to governance variables (corruption perception index, environmental governance, Worldwide Governance) and at a lower level the education (School enrolment – "School enrol", ratio Girls to boys at school – "School G/B").

The evolution of rural/urban population (Growth Urb pop and Growth Rural Pop, Urban pop) and the interaction between these variables is rather complex implying demographic economic and urban transition processes explaining their position located

HESSD
9, 485–539, 2012
WatSan4Dev database coherency analysis
C. Dondeynaz et al.

in between poverty and development parameters[4]. Indeed, developing countries are in the "second wave" of this multi–dimensional transition. Being a long and complex process, countries can be at a diverse stage but should move towards an increment of urban population, with a lower growth of their total population due to the drop of the fertility rate and the increase of income (UNFPA, 2007).

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 to risk of desertification (Desert risk). As mentioned for Africa, the Environmental Sustainability Index (ESI) and Water poverty Index (WPI) include in their calculations development or governance aspects justifying their loads either in F2 and F1.

The third component F3 presents more complex but coherent correlations: water use intensity level in agriculture, the proportion of irrigated area and the total withdrawal[5] are highly correlated altogether and negatively with domestic and industrial withdrawal levels. We observe here the competition between the various uses of water, the agriculture being generally the first consumer in developing countries. The first group describing the intensive level of the agriculture is also linked with the level of development (positive side in F1).

Regarding financial flows, other than the stability and absent of violence (WGI PS AV) and other governance indicators already explained in the African sub-set, the global official development Aid (ODA) is more correlated with health (Child mortal-5, fertility), access to sanitation and water facilities (AIS and AIWS) and poverty variables (Poverty, % slums). The Official development aid specific on WSS is not significantly

[4]A complete analysis can be consulted in the World Urbanization Prospects, 2006 and in UNFPA, 2007.

[5]According to the FAO, except in Europe and North America, agriculture is by far the first user of water even if the trend over the last 60 yr is decreasing due to economic changes. It is visible in the data and mentioned in Sect. 3.1. Therefore agriculture water demand weights the most in the behaviour of the total withdrawal.

represented by the first three PCA components, but presents a high factor load (0.791) on the fourth's one. It presents correlation with governance index (stability and absent of violence – "WGI PS AV", corruption – "CPI" and "WGi CofC", and rule of law – "WGI RofL" as well with the growth of rural population – "Rural Growth Pop").

Finally, the proportion of the agricultural area is showing no significant factors loading in the principal components considered here, and thus will not be included for further analyses.

Factor analysis

A factor analysis is performed to confirm the factors loading configuration outlined by the PCA. The four first factors explain 64.256 % of the variability and the PCA results are confirmed and refined (Table 9). The first component (F1) represents the socio-economic development of the country gathering all variables related to health, education, poverty and urbanisation. The second component (F2) represents governance either in these general aspects (various WGI variables) or specifically to the environmental sustainable management (Environmental Governance – "Env gov"). The third one (F3) represents water resources availability and environmental quality showing high loadings for Precipitations (Precipit), Total Water Renewable Resources (TWRR), Water Bodies surface (WB), the Desertification risk (desert risk), the Environmental Sustainability Index (ESI) and the National Biodiversity Index (NBI). Finally, the fourth component (F4) expresses the pressure put on water resources according to its various uses (industrial – "withdrawal industrial", domestic – "withdrawal domestic" and global amount of water need – "total withdrawal"). The intensive agriculture implying irrigation and drainage system (Irrigated surface – "% irrigated areas" and water use intensity – "Water Use Int Agri") competes with the others water consumers. This component representing a balance between water need, is more complex to interpret. It can be an indicator of the intensive agriculture, or by contrast, the level of industrialisation of a country.

The financial flows towards developing countries (ODA and ODA WSS) show different loads as mentioned in the PCA results: the general official development aid is linked to health and social factor (F1) while the water and sanitation aid is sensible to governance aspects (F2).

4.2.1 The WatSan4Dev database: exploratory interpretation

Identifying key variables

Linear regression is performed with either water supply (AIWS) or sanitation (AIS) as dependent variables: it provides key elements explaining the water supply and sanitation variables behaviours by selecting significant variables (Tables 10 and 11).

Key elements for interpretation are respectively urbanisation (% slums), health aspects (Malaria) and the environmental governance (Env gov). Developing countries, being in an important and often informal urbanisation, encounter difficulties in providing basics services in slums, explaining the negative correlation. The analysis also shows the impact of access to improved water supply on health, children mortality and propagation of malaria. Environmental Governance appears also in the model accordingly to the experience from the field.

The analysis presents similar relationships as for water supply that confirms the coherency of the WatSan4Dev database. In slums areas, the access to basic sanitation is rare and improving this status implies important efforts in terms of investments, control of corruption and affordability for the poorest. As already mentioned, the availability of sanitation utilities reduces significantly Children mortality as well as the dissemination of diseases.

Identifying country profiles

Based on the factor analysis (Sect. 4.2.2 Factor analysis), a two steps clustering is performed to explore the different country profiles. The aim is to identify potential

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

9, 485–539, 2012

WatSan4Dev database coherency analysis

C. Dondeynaz et al.

country profiles according to the significant factors, representing for the Factor 1) the socio-economic development; Factor 2) the governance efficiency; and Factor 3) water resources availability. The quality of the clustering is satisfying with a silhouette coefficient fair value of 0.4 (Kaufman and Rousseeuw, 1990, consider that an average silhouette around 0.5 indicates reasonable partitioning of data; less than 0.2, the data do not exhibit cluster structure). The fourth component is not taken into consideration in the interpretation because it requires caution and investigation before interpretation. Three main groups can be constituted with following features (Table 12):

- The first group presents the lower social and economic level and thus a low level of governance. These countries, mainly from the Saharian Africa and south East-Asia, benefit from a medium situation regarding water resource availability and environmental conditions (biodiversity).

- The second cluster concerns countries with the highest socio- economic development and therefore most efficient governance. However these countries, mainly from the Middle-East or northern Africa, encounter dry environment and scarce freshwater resources.

- The third cluster gathers countries with medium socio-economic and governance level. These countries, mainly from Latin America, benefit either from favourable environmental conditions and thus available water resources.

5 Conclusions and discussions

Water and Sanitation still represent a challenge and a good leverage for improving lives in developing countries. The objective of the work presented in this paper was to get a homogeneous and coherent database of variables linked to the WSS sector in developing countries to model the different variables using robust linear model in a second phase of this work.

HESSD
9, 485–539, 2012

In this way, this paper introduces a new consistent and coherent dataset of water, environmental and development indicators in developing countries (WatSan4Dev Dataset) gathering 45 variables coming from different national and international data providers (Table 1). Raw data have been cleaned and missing data processing applied. As a result of the statistical processing, the data of the WatSan4Dev dataset can be used for qualitative estimations and analysis. Quantitative estimations should be carefully managed as a consequence of the data imputation processes.

From a first statistical analysis, relationships between the different variables confirm the current knowledge and field experience in the water sector. These results show the consistency and coherence of the WatSan4Dev dataset. According to the factor analysis performed on developing countries, the variables organise themselves as follow:

The first component (F1) represents the socio-economic development of the country gathering all variables related to health, education, poverty and urbanisation. The second component (F2) represents governance either in these general aspects (various WGI variables) or specifically to the environmental sustainable management (Environmental Governance – "Env gov"). The third one (F3) represents water resources availability and environmental quality. Finally, the fourth component (F4) expresses the pressure put on water resources according to its various uses for intensive agriculture, industry or households. This latter component representing a balance between water need and uses is more complex to interpret.

Concerning financial flows invested in developing countries, the general Official Development Aid (ODA) is mainly correlated to health and poverty alleviation issues, and the specific Official Development Aid (ODA WSS) seems sensible to governance aspects.

In the exploratory phase, the WatSan4Dev database shows that urbanisation takes an important role in the access to water supply and sanitation facilities and more concretely the kind of urbanisation (urban versus slums). Informal urbanisation developments (slums) impact negatively the water supply and sanitation conditions because of the difficulty of local authorities in the cities to face and structure massive population

HESSD
9, 485–539, 2012

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◂◂ | ▸▸
◂ | ▸
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

flows from rural areas. The slums development is expected to rise up and by 2050, urban dwellers will likely account for 66 per cent of that in the less developed regions (UN, 2010). Therefore, extending the access to sanitation and water supply in these conditions can be hindered or slow down. However cities are an opportunity for investment into water supply and sanitation because the critical mass to create new infrastructures and management capacities is easily reachable.

The relationship between Children mortality under 5 yr (Child Mortality-5) and access to WSS (affecting the two components: AIWS and AIS) is clearly depicted with an important impact on children health. Indeed, more generally, WSS is clearly correlated with health which is an essential leverage for improving life-conditions.

Environmental conditions referring to water resources (total water renewable resources, the precipitations, water bodies or National biodiversity index) seem to be secondary factors in explaining the WSS level. Only the governance component, the way and the capacities engaged in the management of environment, has a positive relation with the water supply level (AIWS).

A new relationship between the variables comes also out: the participation of countries into international environmental consensus or initiatives is linked to the availability of water resources (TWRR and Precipitations) and the risk of desertification (desert risk). Therefore, the water scarcity seems to be the main motivation for developing countries to engage themselves into international political processes. This will be analysed and interpreted in deep in the second part of this work.

Finally, three main different behaviours/profiles among the developing countries can be observed regarding the first three factors (above mentioned as F1, F2, F3): (1) countries with highest socio economic level and governance but facing scarcity of water resources, typically Middle East countries; (2) countries with a medium development presenting a medium level of governance and with abundant water resources, typically, Latin American countries; (3) countries with the lowest level of socio economic and thus presenting weak governance, but with a medium level of water resources, typically, sub-saharian countries.

HESSD
9, 485–539, 2012

The next part of this research will introduce the different models of the data to produce a map of relationships between the different variables and countries reflecting the water sector knowledge. In fact, the various causal and consequential relationships are already identified throughout the multiple actors of the water sector. The model will measure and order the variables of the WatSan4Dev dataset according to their impact on the WSS and thus the results obtained from these models could be interpreted to improve the decision mechanisms into the policy making processes in the sector.

Appendix A

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, (b) permanent crops – land cultivated with long-term crops, (c) permanent meadows and pastures.

- **Agri Prod Index, Agricultural Production Index**

  This FAO indice shows the relative level of the aggregate volume of agricultural production for each year in comparison with the base period 1999–2001.

- **AIS, 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.

HESSD
9, 485–539, 2012
WatSan4Dev database coherency analysis
C. Dondeynaz et al.

– **AIWS, 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 1000 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 excluded. 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 and (5) Global Stewardship.

- **Femal eco, Female economic activity rates**

  This rate concerns women aged 15 and above and calculated on the basis of data on the 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.

- **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 US 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

HESSD
9, 485–539, 2012
WatSan4Dev database coherency analysis
C. Dondeynaz et al.

minus any subsidies not included in the value of the products. Data are in current international dollars.

- **GI afr – Governance Index for Africa**

  This 2009 Index of African Governance measures the degree to which the five categories of political goods, (i) safety and security (ii) rule of law transparency and corruption (iii) participation and human rights, (iv) sustainable economic opportunity (v) human development, are provided within Africa's fifty-three countries.

- **School G/B, Girls to Boys Ratio in Primary Education Enrolment**

  This is a mesure 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 aid 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

HESSD
9, 485–539, 2012

adult literacy rate (with two-thirds weight) and the combined primary, secondary and tertiary gross enrolment ratio (with one-third weight) (iii) adecent standard of living, as measured by GDP per capita (PPP US $).

- **HPI, Human Poverty Index**

  It takes into account three indicators qualifying deprivation.The first deprivation relates to survival: the likeliness of death at a relatively early age and is represented by the probability of not surviving to ages 40. The second dimension relates to knowledge: being excluded from the world of reading and communication and is measured by the percentage of adults who are illiterate. The third aspect relates to a decent standard of living, in particular, overall economic provisioning.

- **Life Expect Birth**, **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 cases

  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 5000 sq km are excluded. Overseas territories and dependencies are excluded from this column.

HESSD

9, 485–539, 2012

WatSan4Dev database coherency analysis

C. Dondeynaz et al.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◂ ▸
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

– **ODA, Official Development Aid**

Aid as a percent of government expenditure is the amount of official development assistance (ODA) received by a country as a percentage of its central government expenditure.

– **ODA-WSS Official Development Aid to the water sector**

Total of all donors disbursements of ODA towards all recipients related to Water supply and sanitation.

– **Particip IEA, 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.

– **% irrigated areas, 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.

HESSD
9, 485–539, 2012

– **Urban Pop, Urban Population – Rural Population**

Total population residing in urban areas – 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 is high, the lower is the water constraint is.

– **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.

HESSD
9, 485–539, 2012

- **WGI-V&A, 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, Worldwide 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.

HESSD
9, 485–539, 2012

– **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.

– **TOT-WITH, 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 + Domestic Water Use + Industrial Water Use.

– **With-Domestic, Water withdrawal for municipal purpose**

Annual quantity of water use for domestic 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.

– **With-Industrial, Water withdrawal for industrial purpose**

Annual quantity of water use by self-supplied industries not connected to any distribution network.

Appendix B

Understanding the female economic participation variable

Mammen and Paxson (2000) explained that the female participation in labour force shows a U-Shape pattern when crossed with the income level per capita for the

HESSD
9, 485–539, 2012

1970–1985 period (Fig. B1). Female economic participation shows high rates for low income countries and decreases up to around 2550 $ per capita for the period considered.

We can assume that the threshold (2550 $ per cap) raised up in absolute terms since 1985 but the U-shape is still valid. Therefore, the decreasing trend observed on our data corresponds to the first part of the U-shape (Fig. B2). This provides additional proofs on the coherency of the dataset.

The economic participation of women depends not only on the income but also on several other factors such as rural-urban context and fertility including social and cultural parameters that make more complex the explanation of this phenomenon (Ahn and Mira, 1998; Boserup, 1989; Beguy, 2009).

Acknowledgements. The present work benefited from the input from the various international institutions mentioned in this paper that provide the necessary datasets to handle this research. We would like to thank the team from University of Almeria for valuable comments and assistance regarding the statistical analyses. We also thank Alan Belward, head of the Global Environmental Monitoring Unit (JRC) and Daoyi Chen (Liverpool University) for their constant support to this research project. The comments of the reviewers are greatly appreciated.

References

Adler, N., Yahemsky, E., and Taverdyan, R.: A framework to measure the relative socio-economic performance of developing countries, Socio Econ. Plan. Sci., 44, 73–88, doi:10.1016/j.seps.2009.08.001, 2009.

AQUASTAT database: A global information system on water and agriculture, available at: http://www.fao.org/nr/water/aquastat/main/index.stm, 2011.

Bartholomew, D. J., Steele, F., Galbraith, J., and Moustaki, I.:Analysis of Multivariate Social Science Data (2 Edn.), New York, Chapman & Hall/Crc, 2008.

Beguy, D.: The impact of female employment on fertility in Dakar (Senegal) and Lome (Togo), Demogr. Res., 20, 97–127, doi:10.4054/DemRes.2009.20.7, 2009.

Boserup, E.: Population, the status of women and rural development, Popul. Dev. Rev., 15, Supplement: Rural development and population: institutions and policy, 45–60, 1989.
Botting, M. J., Porbeni, E. O., Joffres, M. R., Johnson, B. C., Black, R. E., and Mills, E. J.: Water and Sanitation infrastructures for health: the impact of foreign aid, Global. Health, 6, 12, doi:10.1186/1744-8603-6-12, 2010.
CPI database from Transparency International: available at: http://www.transparency.org/policy_research/surveys_indices/cpi/2004, 2004.
Dublin statement on water and sustainable development 1992: available at: http://www.wmo.int/pages/prog/hwrp/documents/english/icwedece.html, 2011.
Earth trends, the environmental information portal: available at: http://earthtrends.wri.org/searchable_db/index.php?action=select_theme&theme=, 2011.
Easterly, W.:The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics, The MIT Press, 2001.
Esty, D. C., Levy, M., Srebotnjak, T., and De Sherbinin, A.: Environmental Sustainability Index: Benchmarking National Environmental Stewardship, New Haven: Yale Centre for Environmental Law & Policy, available at: http://www.yale.edu/esi/, 2005.
EuropeAid: Water Sector Development and Governance: Complementarities and synergies between sector-wide Approach and Integrated Water Resource Management, European Commission publication Office, 44 pp., available at: http://ec.europa.eu/europeaid/infopoint/publications/europeaid/176a_en.htm, 2009.
FMI list of Emerging and developing countries for 2004: available at: http://www.imf.org/external/pubs/ft/weo/2004/02/data/groups.htm, 2011.
GEO Portal: available at: http://geodata.grid.unep.ch/#, 2011.
Hinrichsen, D., Robey, B., and Upadhyay, U. D.: Solutions for a Water-Short World, Population Reports, Series M, 14, Johns Hopkins School of Public Health, Baltimore, available at: http://www.k4health.org/pr/m14/m14chap2_2.shtml, 1997.
Honaker, J. and King, G.: What to Do about Missing Values in Time-Series Cross-Section Data, Am. J. Polit. Sci., 54, 561–581, available at: http://gking.harvard.edu/files/pr.pdf, 2010.
Honaker, J., King, G., and Blackwell, M.: Amelia II: A Program for Missing Data, Version 1.2-17,2011, available at: http://r.iq.harvard.edu/docs/amelia/amelia.pdf, 2011.
Horton, N. J. and Lipsitz, S. R.: Multiple Imputation in practice: Comparison of Software Packages for regression models with missing variables, Am. Stat., 55, 244–254, 2001.
Hotelling, H.: The most predictable criterion, J. Educ. Psychol., 26, 139–142, 1935.

Joint Monitoring Programme (JMP) Portal: available at: http://www.wssinfo.org/, 2011.
Kaufman, L. and Rousseeuw, P. J.: Finding groups in data: an introduction to Cluster Analysis, Ed. Wiley, New York, 1990.
Lee, H.-H., Kwon Chung, R., and Mo Koo, C.: On the Relationship between Economic Growth and Environmental Sustainability, The Eminent Environmental Economists Symposium, Seoul, Korea, 25 pp., 2005.
Mammen, K. and Paxson, C.: Women's Work and economic Development, J. Econ. Perspect., 14, 141–164, 2000.
OECD database: available at: http://stats.oecd.org/qwids/, 2011.
Pearson, E. S.: On lines and planes of closest fit to systems of points in space, Phil. Mag., 2, 559–572, 1902.
Reilly, M.: Data analysis using hot deck multiple imputation, Statistician, 42, 307–313, 1993.
Rotberg, R. I. and Gisselquist, R. M.: Strengthening African Governance: Index of African Governance, Results and Rankings, Cambridge, MA, available at: http://www.worldpeacefoundation.org/africangovernance.html, 2009.
Scott, C. A., Faruqui, N. F., and Raschid-Sally, L.: Wastewater Use in irrigated agriculture, CAB international, Hyderabad workshop, available at: http://web.idrc.ca/openebooks/112-4/#page_1, 2002.
Thurstone, L. L.: Multiple-Factor Analysis, University of Chicago Press, 1947.
Transparency international: Global corruption report: corruption in the water sector, Cambridge University Press, UK, 398 pp., ISSN 1749-3161, available at: http://www.transparency.org/publications/publications/global_corruption_report/gcr2008, 2008.
UN Department of Economic and Social Affairs, Population Division: World Urbanization Prospects: The 2005 Revision, Working Paper No. ESA/P/WP/200, 2006.
UNFPA: State of the world population 2007: Unleashing the Potential of Urban Growth, United Nations Population Fund publications, ISBN 978-0-89714-807-8, 100 pp., available at: http://www.unfpa.org/swp/swpmain.htm, 2007.
UNICEF, WHO: Diarhhoea: Why children are still dying and what can be done?, WHO publications, ISBN 978-92-4-159841-5, 57 pp., available at: http://whqlibdoc.who.int/publications/2009/9789241598415_eng.pdf, 2009.
United Nations, Department of Economic and Social Affairs, Population Division: World Urbanization Prospects, the 2009 Revision: Highlights, New York, available at: http://esa.un.org/unpd/wup/Documents/WUP2009_Highlights_Final.pdf, 2010.

HESSD
9, 485–539, 2012

United Nations Population Division: World Urbanization Prospects: The 2001 Revision, UNPD publications, 35–49, available at: http://www.un.org/esa/population/publications/wup2001/WUP2001_CH3.pdf, 2001.
Ward, J. H.: Hierarchical Grouping to Optimize an Objective Function, J. Am. Stat. Assoc., 58, 236–244, 1963.
World Bank database: available at: http://data.worldbank.org/, 2011.

HESSD
9, 485–539, 2012

WatSan4Dev database coherency analysis

C. Dondeynaz et al.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀I ▶I
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

Table 1. Indicators-Variables implemented in the WatSan4Dev dataset.

Variables to be explained			Data provider
Access to Water Supply and Sanitation services		Access to improved Water source (AIWS) Access to improved Sanitation (AIS)	Joint Monitoring Programme Joint Monitoring Programme
Pillars	**Sub-pillars**	**Indicators**	
Human development	Income	Human Development Index (HDI) Gross Domestic Product – Purchasing Power Parity (GDP per cap) Poverty Rate (Poverty) Human Poverty index (HPI) Female economic activity rate (femal eco)	UNDP World Bank World Bank UNDP ILO
	Health	Malaria cases (Malaria) Fertility Rate (fertility) Children with Diarrhoea (Child Diarrhoea) Prevalence Diarrhoea in slums (% Diarrhoea slums) Mortality Rate for children under 5 (Child Mortal-5) Life expectancy at birth (life Expect Birth) Health expenditure(Health exp)	WHO/UNICEF WHO/UNICEF WHO/UNICEF UN HABITAT UNICEF UN DESA Population WHO
	Education	Ratio Girls to boys in preliminary school (School G/B) Literacy rate of Youth (Literacy Youth) Gross enrolment at primary, secondary and tertiary school (School enrol)	World Bank World Bank UNESCO 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 quality	Biological Oxygen Demand (BOD)	FAO
	Water resources availability	Total renewable water resources (TWRR)	FAO

Table 1. Continued.

Pillars	Sub-pillars	Indicators	
Human pressure	Water demand	Total Withdrawals (Total with) Withdrawal by sectors of activities: (With domestic, With industrial)	FAO FAO
	Agricultural pressure	Agriculture area (% Agri area) Agriculture production index (Agri Prod Index) Total surface in irrigation (% irrigated areas) Water use intensity in Agriculture (Water Use Int Agri)	FAO FAO FAO FAO
	Demographic pressure	Urban population (Urban Pop) Urban Slum population (% slums)	UN- HABITAT UN- HABITAT
Governance	Governance efficiency	Voice and accountability (WGI AV) Political Stability and Absence of violence (WGI PS AV) Government effectiveness (WGI GE) Regulatory Quality (WGI RQ) Rule of Law (WGI RofL) Control of corruption (WGI CofC) Corruption Perception Index (CPI) Governance Index for Africa (GI afr)	World Bank World Bank World Bank World Bank World Bank World Bank Transparency international Harvard university
	Environmental concern	Environmental governance (Env Gov) Participation to International Environmental agreements (Particip IEAg)	World Economic Forum International conventions and agreements.
Financial flow	Aid flow	Official Development Aid (ODA) Disbursement for the water sector, breakdown by subsectors (ODA-WSS)	OECD

HESSD
9, 485–539, 2012
WatSan4Dev database coherency analysis
C. Dondeynaz et al.

Table 2. Countries included in the WatSan4Dev database.

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
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

HESSD
9, 485–539, 2012

Table 2. Continued.

WatSan4Dev countries	GDP per cap 2004	WatSan4Dev countries	GDP per cap 2004
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	Malaysia	4875
South Asia	628	Gabon	5340
Pakistan	629	Botswana	5425
Nigeria	644	Lebanon	5453
Papua New Guinea	660	Libya	5906
India	668	Chile	5929
Yemen, Rep.	693	Mexico	7224
Sao Tome & Principe	710	Equatorial Guinea	8886
Moldova	721	Oman	10 374
Senegal	760	Saudi Arabia	10 784
Nicaragua	834	Bahrain*	16 726
Djibouti	839	Kuwait*	27 148
Cote d'Ivoire	873	Qatar*	44 292
Cameroon	919	United Arab Emirates*	28 371
Bolivia	977		
Sri Lanka	1054		
Guyana	1057		

* Countries considered by the FMI as emerging and developing countries (FMI, 2004).

HESSD

9, 485–539, 2012

WatSan4Dev database coherency analysis

C. Dondeynaz et al.

Table 3. PCA parameters on African Dataset.

	F1	F2	F3
Variability (%)	30.269	12.748	7.369
Cumulative %	30.269	43.017	50.386

HESSD
9, 485–539, 2012

Table 4. Factors loading for the first three components (50.38 %).

PCA – Factor loadings			
Variables*	F1	F2	F3
Fertility	**0.937**	0.075	−0.018
Child Mortal-5	**0.821**	0.185	−0.088
Life Expect Birth	**−0.615**	−0.354	−0.029
% Agri. Area	0.135	−0.098	**0.540**
UrbanPop	**−0.617**	0.143	−0.358
WB	0.178	0.451	**0.532**
Precipit	0.237	**0.769**	0.037
Growth Rural Pop	**0.689**	−0.233	0.096
Growth Urban Pop	**0.636**	−0.046	0.187
AIWS	**−0.789**	0.119	0.192
AIS	**−0.745**	−0.079	0.072
House connect	**−0.844**	−0.093	−0.001
ODA	0.277	0.182	0.366
GDP per cap	**−0.894**	−0.019	−0.097
School G/B	**−0.641**	0.160	0.438
With Domestic	−0.233	**0.741**	−0.150
With Industrial	−0.244	**0.700**	−0.098
Agri Prod Index	0.148	**−0.283**	−0.150
% Irrigation	**−0.497**	−0.142	0.102
Water Use Int Agri	−0.255	**−0.719**	0.238
Desert risk	0.175	**−0.403**	0.337
TWRR	0.292	**0.726**	−0.127
Tot Withdrawals	−0.165	**−0.464**	−0.030
CPI	**−0.571**	−0.121	0.370
ODA WSS	−0.146	−0.212	**0.349**
DAM	−0.160	−0.096	**0.275**
BOD	**−0.567**	0.245	−0.094
Poverty	**0.522**	−0.095	0.396
Literacy	**−0.601**	0.356	0.049
Malaria	0.173	**0.606**	0.375
% Children diarrhea	**0.378**	−0.137	−0.168
% diarrhea slums	**0.769**	−0.363	−0.045
NBI	0.406	0.375	**0.455**
Env gov	**−0.561**	0.284	0.353
Particip to IEA	−0.065	−0.324	**0.524**
School enroll	**−0.761**	0.105	0.157
Health exp	**−0.874**	0.055	0.125
Femal eco	**0.722**	0.166	0.307
WGI.V.A.	−0.252	0.081	**0.451**
WGI.PS.AV.	**−0.411**	0.040	0.347
WGI.GE	**−0.619**	−0.181	0.415
WGI.RQ.	**−0.429**	−0.093	0.464
WGI.RofL.	**−0.598**	−0.185	0.468
GI Afr	**−0.631**	−0.034	0.423
HDI	**−0.925**	−0.004	−0.004
WPI	**−0.633**	0.378	−0.074
ESI	−0.277	**0.543**	0.055
HPI	**0.804**	−0.160	−0.018

* In bold the main factor load for the variable.

HESSD
9, 485–539, 2012

WatSan4Dev database coherency analysis

C. Dondeynaz et al.

Table 5. Model parameters with water supply as dependent variable (adjusted $R^2 = 0.629$).

Water supply							
	Unstandardized Coefficients		Standard Coefficient	t	Sig.	95 % Confidence Interval for β	
	β	Std. Dev.	β			Lower bound	Upper bound
Child Mortal-5	−0.593	0.093	−0.572	−6.391	0.000	−0.779	−0.406
Env. gov.	0.406	0.115	0.326	3.526	0.001	0.175	0.638
With industrial	0.195	0.078	0.221	2.505	0.016	0.039	0.352

HESSD
9, 485–539, 2012

WatSan4Dev database coherency analysis

C. Dondeynaz et al.

Table 6. Model parameters with sanitation as dependent variable (adjusted $R^2 = 0.555$).

Sanitation level							
	Unstandardized Coefficients		Standard Coefficient	t	Sig.	95 % Confidence Interval for β	
	β	Std. Dev.	β			Lower bound	Upper bound
Health exp	0.354	0.132	0.376	2.685	0.010	0.089	0.620
Water Use int. agri.	0.441	0.114	0.376	3.881	0.000	0.212	0.669
Urban pop.	0.311	0.109	0.296	2.848	0.007	0.091	0.530
Env. gov.	0.478	0.155	0.369	3.085	0.003	0.166	0.790
CPI	−0.427	0.179	−0.309	−2.382	0.021	−0.788	−0.066

HESSD
9, 485–539, 2012
WatSan4Dev database coherency analysis
C. Dondeynaz et al.

Table 7. PCA parameters for WatSan4Dev dataset.

	F1	F2	F3	F4
Eigenvalue	12.407	3.647	2.758	1.852
Variability (%)	38.771	11.396	8.620	5.786
Cumulative %	38.771	50.168	58.788	64.574

HESSD
9, 485–539, 2012

Table 8. Factors loading for the first four components (64.57 %).

PCA – Factor loadings				
Variables*	F1	F2	F3	F4
Poverty Rate	**−0.741**	0.149	0.051	−0.047
% Agri Area	−0.249	−0.161	0.052	**−0.505**
Urban Pop	**0.776**	−0.105	0.194	−0.107
Env. gov.	**0.662**	−0.324	0.140	0.180
School enroll	**0.806**	0.239	−0.062	−0.110
Femal eco	**−0.567**	0.307	0.215	−0.067
% Urban Slums	**−0.870**	0.092	0.044	0.095
Fertility	**−0.856**	−0.162	0.277	0.128
With domestic	0.167	−0.010	**0.890**	−0.073
With industrial	0.220	0.380	**0.618**	0.066
CPI	**0.745**	−0.195	0.015	0.262
TWRR	−0.150	**0.792**	0.052	0.035
Health exp	**0.912**	0.008	0.044	−0.029
AIS	**0.821**	−0.029	−0.185	−0.008
Life Expect Birth	**0.857**	0.047	−0.220	−0.035
Child Mortal-5	**−0.873**	−0.103	0.231	0.040
AIWS	**0.860**	−0.061	−0.073	−0.107
School G/B	**0.636**	0.258	−0.068	0.031
Growth Rural Pop	**−0.454**	−0.209	−0.115	0.426
Growth Urb Pop	**−0.509**	−0.169	0.065	0.177
NBI	−0.046	**0.785**	0.095	−0.293
Particip to IEA	0.335	0.152	0.021	**−0.392**
GDP per cap	**0.927**	−0.075	0.002	0.069
% irrigated areas	0.442	0.135	**−0.735**	−0.046
ODA WSS	−0.031	−0.088	0.001	**0.791**
ODA	−0.418	0.045	0.163	**0.567**
Malaria	**−0.786**	0.213	0.308	0.111
Water Use int Agri	0.480	−0.175	**−0.751**	−0.026
WB	0.038	**0.519**	−0.023	0.014
Precipit	−0.070	**0.927**	−0.009	0.028
Desert risk	−0.288	**−0.542**	−0.053	−0.303
Total with	0.536	0.006	**−0.708**	−0.062
WPI	**0.683**	0.492	−0.123	0.012
ESI	0.275	**0.420**	0.132	0.104
WGI.V.A.	**0.428**	0.203	0.137	0.018
WGI.PS.AV	**0.454**	−0.083	0.100	0.385
WGI.GE	**0.753**	−0.166	−0.003	0.167
WGI.RQ	**0.640**	−0.116	0.087	0.169
WGI.RofL.	**0.693**	−0.282	−0.057	0.242
HDI	**0.945**	0.073	−0.121	−0.024
WGI-CofC-2004	**0.687**	−0.278	0.004	0.250

* In bold the main factor load for the variable.

HESSD
9, 485–539, 2012

WatSan4Dev database coherency analysis

C. Dondeynaz et al.

Table 9. Component matrix: the four components gathering 64.256 % of the variability.

Component Matrix[a]				
Components	F1	F2	F3	F4
Poverty	**−0.671**	−0.316	0.107	−0.033
Urban Pop	**0.747**	0.224	−0.054	−0.243
Env gov	0.415	**0.691**	−0.225	−0.089
School enrol	**0.764**	0.238	0.291	0.006
Femal eco	**−0.597**	−0.104	0.291	−0.167
% Urban Slums	**−0.787**	−0.382	0.000	−0.010
Fertility	**−0.837**	−0.228	−0.207	−0.219
Withdrawal domestic	0.108	0.071	−0.052	**−0.902**
Withdrawal industrial	0.143	0.067	0.331	**−0.622**
CPI	0.452	**0.800**	−0.063	0.044
TWRR	−0.190	−0.076	**0.863**	−0.038
Health exp	**0.794**	0.435	0.099	−0.076
AIWS	**0.781**	0.261	0.016	0.144
Life Expect	**0.817**	0.296	0.065	0.176
Child Mortal-5	**−0.831**	−0.297	−0.109	−0.180
AIWS	**0.798**	0.345	−0.015	0.032
School G/B	**0.532**	0.305	0.292	0.057
Growth Rural Pop	**−0.511**	0.025	−0.256	0.200
Growth Urb Pop	**−0.515**	−0.076	−0.271	0.003
NBI	−0.003	−0.180	**0.750**	−0.121
Particip to IEA	0.248	**0.297**	0.256	−0.029
GDP per cap	**0.822**	0.433	−0.010	−0.031
% irrigated areas	0.495	0.031	0.124	**0.688**
ODA WSS	−0.344	**0.516**	−0.032	.149
ODA	**−0.648**	0.265	0.072	−0.041
Malaria	**−0.813**	−0.179	0.215	−0.247
Water Use int Agri	0.497	0.183	−0.135	**0.728**
WB	0.000	0.009	**0.455**	0.003
Precipit	−0.061	−0.176	**0.870**	−0.016
Desert risk	−0.280	0.054	**−0.426**	0.093
Total withdrawal	0.572	0.115	0.090	**0.676**
WPI	**0.648**	0.159	0.582	0.082
ESI	0.089	0.357	**0.611**	−0.094
WGI.V.A.	0.150	**0.696**	0.362	−0.100
WGI.PS.AV	0.118	**0.789**	0.043	−0.011
WGI.GE	0.464	**0.822**	−0.036	0.057
WGI.RQ	0.346	**0.806**	0.001	−0.035
WGI.RofL.	0.394	**0.845**	−0.167	0.123
HDI	**0.890**	0.333	0.112	0.069
WGI-CofC-2004	0.389	**0.836**	−0.162	0.063
Extraction Method: principal Component Analysis.				
Rotation Method: Varimax with Kaiser Normalization.				

[a] Rotation converged in 6 iterations.

HESSD

9, 485–539, 2012

WatSan4Dev database coherency analysis

C. Dondeynaz et al.

Table 10. Model parameters with water supply as dependent variable – method stepwise (adjusted $R^2 = 0.723$).

Water supply level							
	Unstandardized Coefficients		Standard Coefficient	T	Sig.	95 % Confidence Interval for β	
	β	Std. Dev.	β			Lower bound	Upper bound
Child Mortal-5	−0.095	0.032	−0.297	−2.985	0.004	−0.158	−0.032
% Slums	−0.216	0.051	−0.347	−4.208	0.000	−0.318	−0.114
Env. gov.	0.310	0.135	0.148	2.298	0.024	0.042	0.578
Malaria	−0.105	0.046	−0.206	−2.271	0.025	−0.197	−0.013

HESSD

9, 485–539, 2012

WatSan4Dev database coherency analysis

C. Dondeynaz et al.

Table 11. Model parameters with Sanitation as dependent variable, stepwise method (adjusted $R^2 = 0.680$).

	Sanitation level						
	Unstandardized Coefficients		Standard Coefficient	T	Sig.	95 % Confidence Interval for β	
	β	Std. Dev.	β			Lower bound	Upper bound
Child Mortal-5	−0.164	0.058	−0.302	−2.826	0.006	−0.278	−0.049
% Slums	−0.377	0.086	−0.356	−4.363	0.000	−0.548	−0.205
Malaria	−0.229	0.084	−0.264	−2.721	0.008	−0.396	−0.062

HESSD
9, 485–539, 2012

Title Page
Abstract Introduction
Conclusions References
Tables Figures
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

Table 12. Clustering of countries based on factor analysis: characteristics.

		Factor score 1	Factor score 2	Factor score 3	Factor score 4
	Outlier Cluster	Mean	Mean	Mean	Mean
TwoStep Cluster Number	1	−0.79766	−0.32632	−0.17893	0.03059
	2	1.04923	0.33891	−1.36709	0.06500
	3	0.60710	0.29224	0.91452	−0.07440

Countries of Cluster 1	Countries of Cluster 2	Countries of Cluster 3
Sudan, Benin, Burkina Faso, Chad, Eritrea, Ethiopia, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Malawi, Mali, Mauritania, Mozambique, Niger, Rwanda, Senegal, Sierra Leone, Tanzania, Uganda, Zambia, Burundi, Comoros, Cape Verde, Zimbabwe, Angola, Cameroon, Central African Rep, Congo, Congo Dem Rep, Côte d'Ivoire, Equatoria Guinea, Liberia, Nigeria, Togo, Bangladesh, Cambodia, India, Nepal, Swaziland, Bhutan, Madagascar, Lao People's Dem Rep, Myanmar, Pakistan	South Africa, Djibouti, Syrian Arab Rep, Bahrain, Egypt, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Qatar, Saudi Arabia, Tunisia, United Arab Emirates, Iran, Islamic Rep, Algeria	Namibia, Botswana, Papua New Guinea, Guyana, Sri Lanka, Nicaragua, Sao Tome & Principe, Chile, El Salvador, Mauritius, Cuba, Ecuador, Guatemala, Indonesia, Peru, Philippines, Suriname, Thailand, Uruguay, Vietnam, Belize, Bolivia, Costa Rica, Malaysia, Argentina, Brazil, Colombia, Gabon, Honduras, Jamaica, Panama, Paraguay, Venezuela, Dominican Rep, Mexico

HESSD
9, 485–539, 2012
WatSan4Dev database coherency analysis
C. Dondeynaz et al.

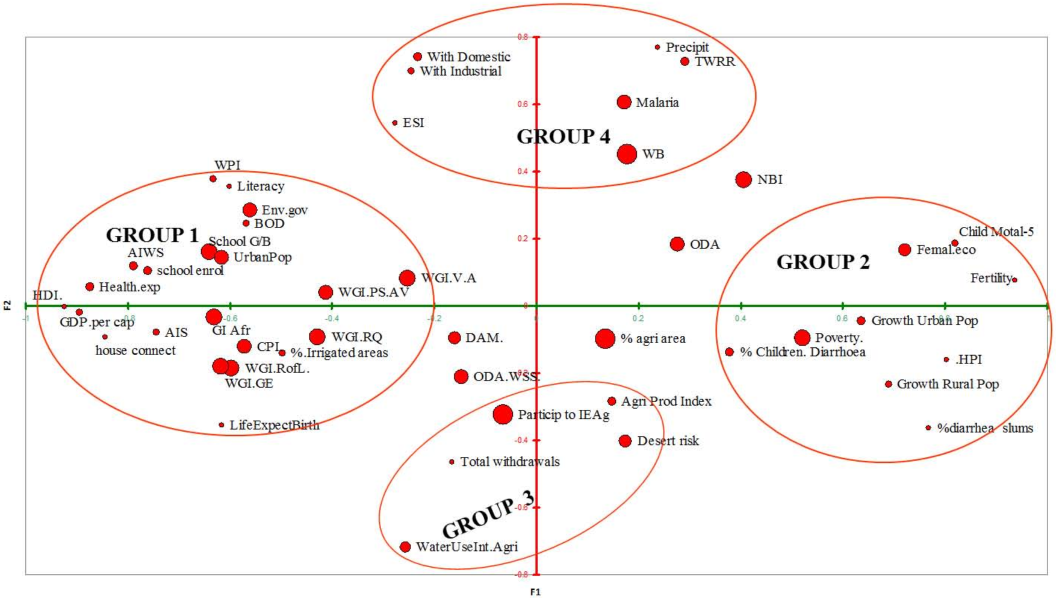

Fig. 1. Three first components of PCA on the African dataset (50.38 %).

HESSD
9, 485–539, 2012

WatSan4Dev database coherency analysis

C. Dondeynaz et al.

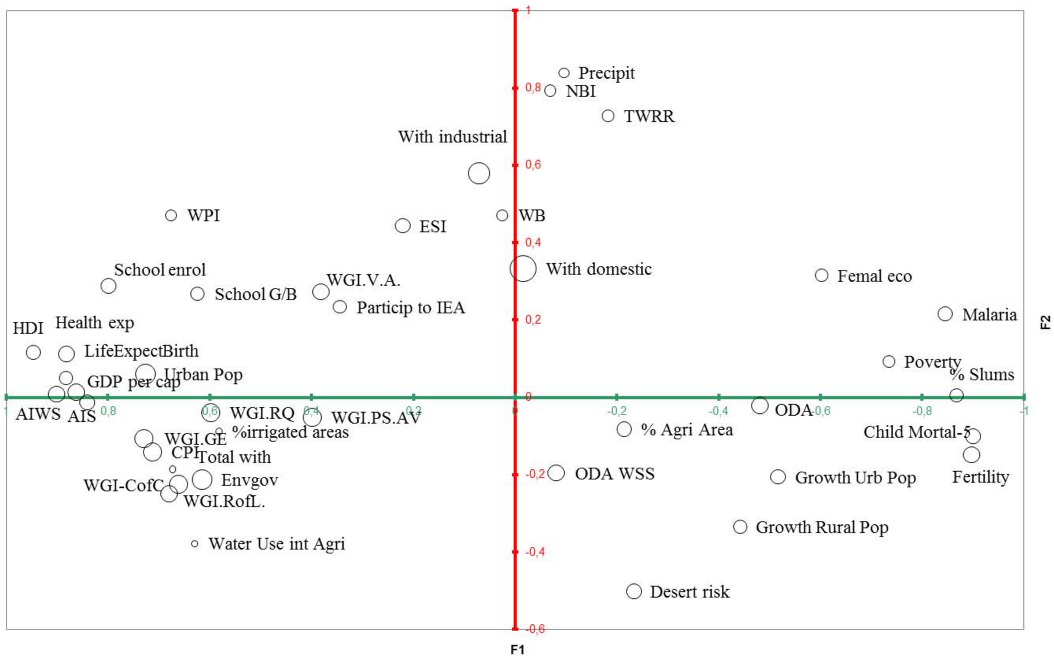

Fig. 2. Three first components of the PCA performed on the WatSan4Dev dataset. In green the first component F1, in red the second component F2 and the size of the point represents the factor loading in the third component F3.

HESSD
9, 485–539, 2012

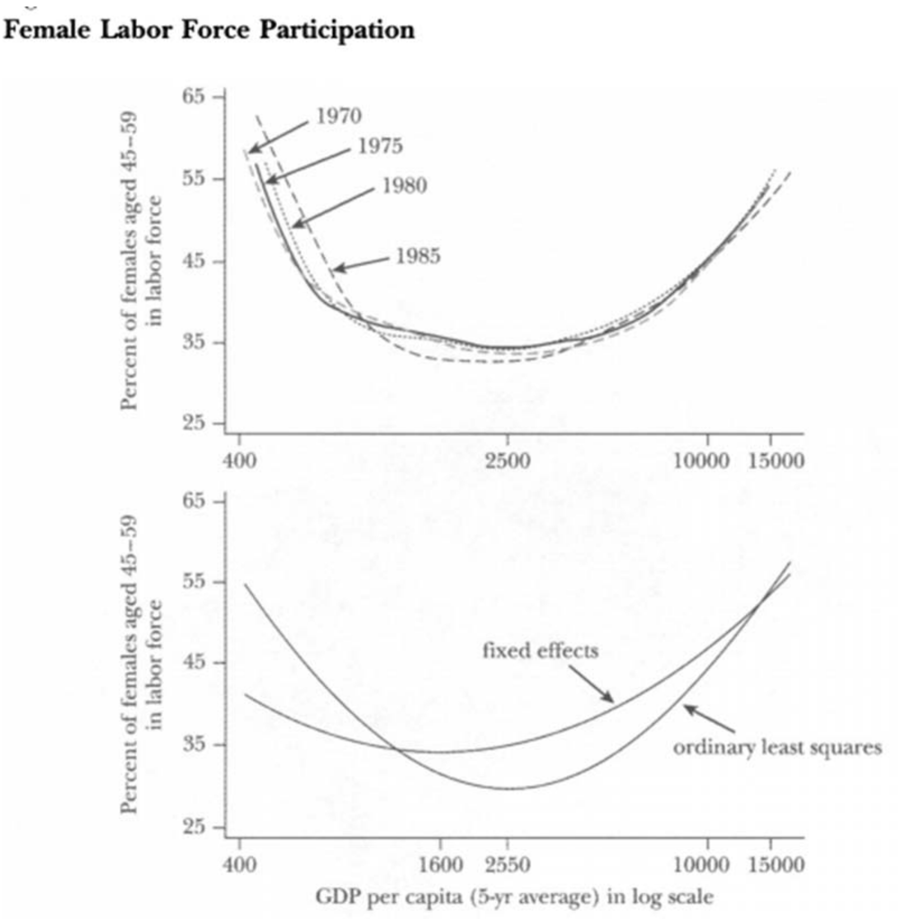

Fig. B1. Distribution shape of female participation in labour force (for females aged 45–59 – but still true for other ages).

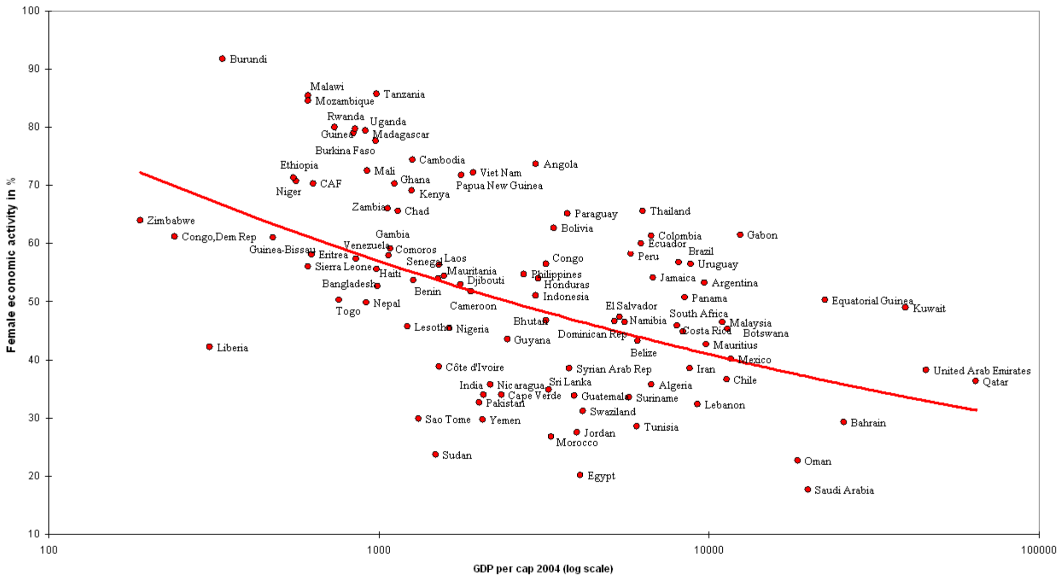

Fig. B2. Female economic participation versus income per capita, distribution and trendline for 2004.

HESSD
9, 485–539, 2012

WatSan4Dev database coherency analysis

C. Dondeynaz et al.