**Hydrology and Earth System Sciences**

# Analysing inter-relationships among water, governance, human development variables in developing countries

**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

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

**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 42 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. Based on 25 relevant variables, the relationships between variables are described and organised into five factors (HDP – Human Development against Poverty, AP – Human Activity Pressure on water resources, WR – Water Resources, ODA – Official Development Aid, CEC – Country Environmental Concern). Linear regression methods are used to identify key variables having influence on water supply and sanitation. First analysis indicates that the informal urbanisation development is an important factor negatively influencing the percentage of the population having access to WSS. Health, and in particular children's health, benefits from the improvement of WSS. Irrigation is also enhancing

Water Supply service thanks to multi-purpose infrastructure. Five country profiles are also created to deeper understand and synthetize the amount of information gathered. This new classification of countries is useful in identifying countries with a less advanced position and weaknesses to be tackled. The relevance of indicators gathered to represent environmental and water resources state is questioned in the discussion section. The paper concludes with the necessity to increase the reliability of current indicators and calls for further research on specific indicators, in particular on water quality at national scale, in order to better include environmental state in analysis to WSS.

## 1 Introduction

The experience of development cooperation in the water sector shows that only the building of water supply and sanitation (WSS) infrastructures is inefficient in bringing sustainable water supply and sanitation services to the population. Water Supply (WS) and Sanitation (*S*) are complex issues that impact other sectors across the society, first and foremost, health and environment but also institutional capacities, and economic sectors such as agriculture and industry.

A multi-dimensional approach is needed and therefore, in 1992, principles for sustainable management in the water sector are formalised with the adoption of the integrated water resources management (IWRM) approach[1] by

---

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

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)" (Europe Aid, 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 (e.g. institutional, civil society, suppliers, funders) and sectors (e.g. education, heath, economic activities) concerned to building appropriate facilities.

Millennium Development Goals' initiative (MDGs) 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 are concerned under the Target 7C: Halve, by 2015, the proportion of the population without sustainable access to safe drinking water and basic sanitation.

In terms of research, this cross-cutting approach is translated by performing crossed analyses of the different dimensions of a question. For instance, Adler et al. (2010) have built a framework for the analysis of human development index data, financial resources, and the MDG's targets. It allows the evaluation of a country's progress towards MDGs, considering the development measures 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. Having such a comprehensive view could support decision making at country scale and the international donors have a general picture helping in better orienting their investments. For instance, in the case of Africa, talking about prioritisation for allocating funds "*it is unclear if the EU donors are focusing their commitments and disbursements on those countries that most lack water supply and sanitation*" (Fonsceca et al., 2008). Beyond this mismatching between the level of WSS and funds affected, it is useful to identify the main reasons explaining a low level of access. In fact a similar access may result from different factors that include roughly: those related to the infrastructure itself ("hardware") and those related to its sustainable management ("software"). This software component is related to the context in which the infrastructure should function and can include governance conditions, constraints on water resources, political commitment towards its population and the environment and the availability of necessary skills. Identifying country weaknesses (within hard or soft factors) should follow the same framework for all developing countries.

The actual fact is that the data on the variables influencing WSS are currently disseminated among international and/or national providers, making difficult their direct use to support decision makers. For this reason, one of the objectives of this
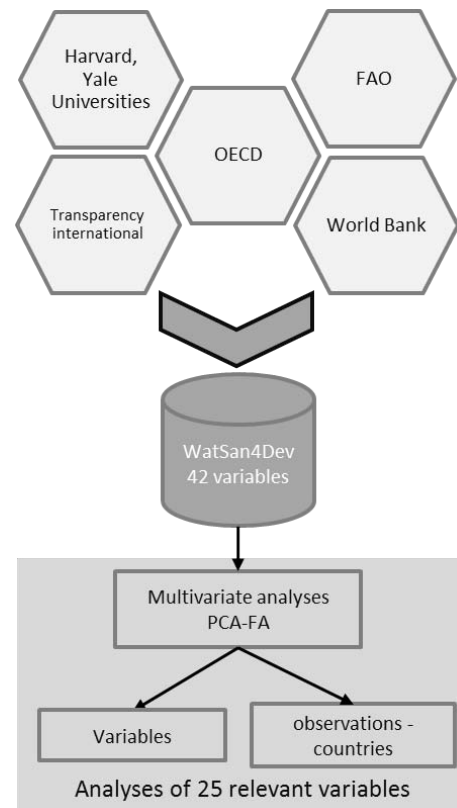


**Fig. 1.** Approach followed that provides two main results: (1) the identification of key variables impacting the level of WSS services and (2) geographical analysis identifying several country profiles.

work is to look into building a common framework around a coherent database (WatSan4Dev) based on the available variables collected by official international providers.

Once gathering various indicators on economic and social status, governance, and environmentally related WSS services at national scale, the reliability and coherence of the indicators, as well as their relevance toward WSS issues are to be verified before making the database available. In this way, the work proposed here presents several statistical methodologies to pre-process the indicators and analyse the different variables and observations. MDG indicators and the percentage of the population having access to improved water supply and sanitation are the references representing WSS.

The analysis of the different variables are performed at two levels (Fig. 1): (a) the analysis of the variables themselves, which gives an insight on the relationships among the different indicators and their relevance in the water sector; and (b) the analysis of the observations (countries) which offers a classification of the countries looking at the water sector. This geographical analysis is also used to observe country status regarding the selected variables, and therefore to highlight countries where efforts should be put in priority. Several countries' profiles are proposed as a tool to identify

weaknesses restricting country development and population well-being.

Therefore, within this framework, this paper presents the data sources and variable selection criteria that are included in the WatSan4Dev database (Sect. 2) and the methods and analysis performed to establish the coherence of the dataset (Sect. 3). It finally presents the results of the coherency verification process of WatSan4dev for 2004 (Sect. 4). This year is taken as reference because of its being the last release of the MDG's variables regarding water and sanitation when this work started. Key variables impacted or being influenced by the level of WSS are presented in Sect. 4. A deep analysis of countries behaviours is also described through the definition of five profiles.

## 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 the MDGs are under the responsibility of the United Nations. These data are freely available, accessible through web online databases and considered by the International Community as reference data (AQUASTAT, CPI , Earth trends, GEO portal, JMP, OECD and World Bank databases). However data collection and processing methodologies are heterogeneous calling for caution and requiring additional analyses before use. Better data and methods harmonization between international organisations is the main mean to reduce errors in analyses.

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, and 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 these indicators, in terms of geographical and temporal scales, are major constraints 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 indicators 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: *Water Resources* (water resources and environmental variables), *Human activity pressure on water resources* (human activity pressure variables), *Governance* (variables measuring various aspects of governance including environmental aspects), *Human development* (social and health variables).

Complementary to these four thematic 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.

Initially, 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 WSS level; (ii) data availability: the variable collected has enough observations (less than 50 % of data in the indicator are missing); (iii) reliability: the variable is produced by trustful-official providers and the method is fully described; and, (iv) inter-correlation: variables with high correlations with other variables (above 0.9) are directly removed. After applying these different criteria to the whole set of variables, 42 variables are pre-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 variables collected (the relationships between the variables and magnitudes of the values) are tracked through Principal Component Analysis (PCA).

Concerning uncertainty, standards errors and boundaries are poorly available in particular regarding socio-economic indicators. However variables are collected from official international providers that do a substantial work of verification and consolidation of datasets. For instance, JMP holds "data reconciliation processes" to reduce discrepancies among data national sources and between these sources and international estimates generated by the JMP. These processes aim at ensuring the quality of the data being collected in a country. In addition to validation through human interaction, FAO performs cross-comparison with similar countries as well as historic data for the country in question and mathematically checks for consistency and correctness.

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.

Although WatSan4Dev database includes 192 countries, only 101 countries are considered in this study: 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 considered in this paper. The World Bank set the threshold at 12 276 current $ per capita, below which a

**Table 1.** Indicators-Variables integrated in the WatSan4Dev dataset.

| | | Variables to be explained | Data provider |
|---|---|---|---|
| Access to Water Supply and Sanitation (WSS) services | | Access to improved Water source (WS)* Access to improved Sanitation ($S$)* | 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)* | UNDP |
| | | Poverty Rate (Poverty)* | ILO |
| | | Female economic activity rate (femal eco)* | World Bank World Bank |
| | Urbanization | Urban population (Urban Pop)* | UN-HABITAT |
| | | Urban Slum population (% slums)* | HABITAT |
| | | Rural population growth (rural growth) | UN-UN-HABITAT |
| | | Urban population growth (urban growth) | UN-HABITAT |
| | Health | Malaria cases (Malaria)* | WHO/UNICEF |
| | | Fertility Rate (fertility) | UN HABITAT |
| | | Mortality Rate for children under 5 (Child Mortal-5)* | UNICEF |
| | | Life expectancy at birth (life Expect Birth) | UN DESA Population |
| | | Health expenditure (Health exp) | WHO |
| | Education | Ratio Girls to boys in preliminary school (School G/B)* | World Bank |
| | | Gross enrolment at primary, secondary and tertiary school (School enrol)* | UNESCO |
| Water Resources | Environmental general characteristics | Environmental Sustainability Index (ESI) | Yale and Columbia Universities |
| | | Water Poverty Index (WPI) | Wallingford, Centre for Ecology and Hydrology. |
| | | Percentage of land under risk of desertification (Desert risk)* | FAO |
| | | Precipitations (Precipit)* | |
| | | National Biodiversity Index (NBI) | FAO Convention on Biological Diversity. |
| | | Surface of Water bodies (WB) | FAO |
| | Water resources availability | Total renewable water resources (TWRR)* | FAO |
| Human activity pressure on WR | Water demand | Total Withdrawals (Total withdrawal)* | FAO |
| | | Withdrawal by sectors of activities: (Withdrawal municipal, Withdrawal industrial)* | FAO |
| FAO | Agricultural | Agriculture area (% Agri area) | FAO |
| | pressure | Total surface in irrigation (% irrigated areas)* | FAO |
| | | Water use intensity in Agriculture (Water Use Int Agri)* | FAO |
| | | Dam capacity (DAM) | |

**Table 1.** Continued.

| | | Variables to be explained | Data provider |
|---|---|---|---|
| Access to Water Supply and Sanitation (WSS) services | | Access to improved Water source (WS)* | Joint Monitoring Programme |
| | | Access to improved Sanitation (S)* | Joint Monitoring Programme |
| Governance | Governance efficiency | Voice and accountability (WGI AV)* | 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 to International Environmental agreements (Particip)* | International conventions and agreements. |
| Financial flow | Aid flow | Official Development Aid (ODA)* Disbursement for the water sector, breakdown by subsectors (ODA-WSS)* | OECD |

* Variables selected for multi-variate analyses in this paper (total variables selected = 25).

country belongs to low income country category[2]. Bahrain, Kuwait Qatar and the United Emirates are included here because they are considered emerging and developing countries by the FMI. Countries from Eastern Europe and Central Asia are excluded from this analysis because they are out of our target countries and second, they are a group with particular behaviour that imply a specific analysis.

# 3 Methods

Taking into account the strong heterogeneity and origin of the indicators, not only from a point of view of the data collection-gathering methods but also from the geographical origin of the data, several data pre-process methods are used.

First, normalisation and multi-variate data imputation are applied to impute missing values. Then principal component analysis and factor analysis are used in two different ways: (a) to analyse the relationships between variables and track error or inconsistency within the dataset; (b) to reduce the number of variables but maximising the information coming from the variables.

This reduction of variables leads to the construction of relevant factors that allow: (a) clustering the variables to identify the relationships among them and detect collinearity; (b) clustering observations-countries to build country profiles

with homogenous behaviours. From the first group of factors (variable analysis), linear regression is chosen as a robust method to identify the keys indicators being influenced or influencing WSS variables. From the second group of factors (observations-countries), the construction of these profiles implies the application of k-means clustering methods to identify similar behaviour among countries.

## 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 42 variables. 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 affected by the normalisation process.

---

[2]World Bank method for country classification http://data.worldbank.org/about/country-classifications

**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 | 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 |
| 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).

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 ($m$) 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 note that quantitative interpretations are 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.

The EMB algorithm (Expectation Maximization Bootstrap)[3] is used in this paper for imputation. 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 combined under the analyst control. The assumptions are: (1) the imputation model assumes that the complete data (that is, both observed and unobserved) are multi-variate normal; and, (2) the data are randomly missing (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 applied to 156 observations – countries. The imputation process includes manual verifications and specific case-by-case corrections.

## 3.2　Analysis methods

Both PCA and FA are variable reduction methods. Effectively, by analysing the correlation between variables in a dataset, the variables can be reduced to a smaller amount of Factors (FA) or Principal Components (PCA). Both methods provide a set of loadings (correlations between original variables and extracted factors or components) and a set of scores (values each data item gets on the extracted factors/components after the variable reduction).

The most important differences between PCA and FA are:

1. PCA analyses all the variance present in the dataset. PCA is used to find optimal ways of combining variables by establishing potential relationships among the dataset variables for obtaining an empirical summary of the data.

2. FA analyses only the common variance that means uncontaminated by unique and error variability. Therefore, FA is less sensitive to noise in the dataset as based on assumed underlying factors. A thematic knowledge (a set of hypotheses that form the conceptual basis of the study) of the dataset is required for applying the FA.

---

[3]Imputation have been made using Amelia II software (Honaker et al., 2011).

FA analyses are usually jointly combined with PCA for confirming the variable reduction proposed by the PCA through the thematic knowledge (FA) of the dataset.

### 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 advantages 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 (called Principal Components). The PCA technique is widely used in the multi-variate 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 high a 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 of only the three main PCA factors in Fig. 2 because of 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 dataset 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 observations: the behaviour of 25 variables-indicators for 101 observations-countries. PCA provides a global view of the correlations among variables and/or countries.

### 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 produce 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). Significant factors highlighted by the FA will be the basis for the country analyses in this paper.

### 3.2.3 K-mean clustering

The K-mean clustering algorithm was developed by Mac-Queen (1967). K-mean clustering is an algorithm to classify or to group your objects based on attributes/features into $K$ number of groups. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. This clustering method is used to analyse countries' characteristics to define different country profiles.

### 3.2.4 Ordinary Least Square (OLS) linear regression analysis

Although the modelling phase is not part of this paper, 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.

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 follows:

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

where $Y_i$ are the dependent variables, $\beta_{0,i}$ are the intercepts of the model, $X_{j,i}$ corresponds to the $j$-th explanatory variable of the model ($j = 1$ to $p$), and $\varepsilon_i$ is the random error with expectation 0 and variance $\sigma^2$. The $\beta_{0,i}$ and $\beta_{j,i}$ parameters and $\varepsilon_i$ 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

As described in the following paragraph, only 25 variables are selected using PCA and FA; finding those sufficient and relevant to perform robust analyses in accordance with the
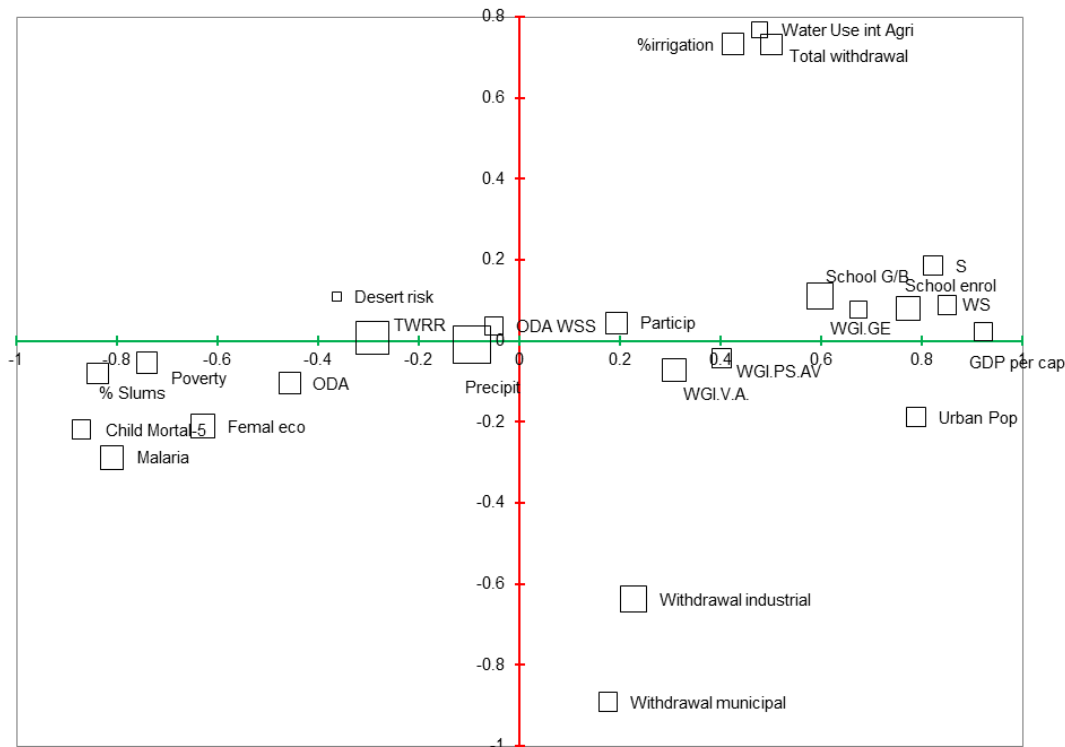
**Fig. 2.** Graphical representation of the WatSan4Dev variables in the three first components of the PCA (57 %). The first component (HDP) is represented by the green axis; in red, the second component (AP); and, the size of the point represents the factor loading in the third component (WR).

approach described by Fig. 1. These variables are considered as coherent if, other than showing statistical coherency and robustness, they also comply with field experiences, case studies and scientific literature. The following analysis of variables aims at identifying their relationships. PCA and FA are also used to synthesize the information through a smaller representative numbers of variables-components. The components are then used to analyse the observations and build country profiles. Two other main results are obtained: (i) the understanding of relationships between variables and (ii) the proposition of five profiles to support a geographical analysis of data aiming at identifying less favourable country situations that will help in orienting efforts.

### 4.1 Verifying the coherence of the variables before analyses

#### 4.1.1 Selection of variables

In this work, only 25 representative variables out of the 42 WatSan4Dev variables are used for the statistical analyses; excluded and selected variables are listed hereafter. Pearson correlation coefficient is used for multi-variate analyses.

Governance variables, namely the Worldwide Governance Indicators (WGI), Corruption Perception Index (CPI) and Environmental governance (Env Gov) are highly correlated

together (correlation values range from 0.645 to 0.852) showing in this way coherency despite different gathering methods and data sources. Their number is reduced to avoid multi-collinearity. The Governance Effectiveness (WGI-GE) is taken as representative of several other governance indicators because of high correlation with corruption perception index (0.881 correlation – CPI), Rule of law (0.906 correlation – WGI-RofL), Regulatory Quality (0.876 correlation – WGI-RQ) and environmental governance (0.807 – Env gov). In the same way, Child mortality rate (Child Mort-5) represents the life expectancy at birth (life expect) and the fertility rate (fertility): a high child mortality rate means a low life expectancy at birth and therefore in developing countries is correlated with high fertility rate. Health expenditure (Health expect) has a high correlation ($> 0.9$) with the income per capita (GDP per capita).

Water bodies surface (WB) and National Biologic Index (NBI) are correlated with few variables, mainly the amount of precipitation (respectively 0.439 and 0.732 correlation – Precipit) and the amount of water resources available per capita (0.569 correlation value with NBI-TWRR) but not correlated with WSS. In the same way, Dam capacity (DAM) shows no significant correlations with any other variable, indicating that the equipment in dam-reservoirs neither impacts directly nor indirectly on the level of access to WSS.

Therefore, these variables are excluded from the analysis for not contributing to the explanation of water supply (WS) or sanitation (S) variables.

Rural and urban population growths (Growth Urb pop and Growth Rural Pop) variables represent complex processes that are not targeted in this paper. Effectively, the evolution of rural/urban population requires considering both the repartition of the population (Urban pop) and their own dynamic (Growth Urb pop and Growth Rural Pop). The interaction between these variables involves demographic, economic and urban transition processes[4]. Indeed, developing countries are at the beginning of this multi-dimensional transition. Being a long and complex process, countries can be at a diverse stage of this transition 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).

The Environmental Sustainability Index (ESI), the Water Poverty index (WPI), Human development Index (HDI) are also excluded because of being composite indicators overlapping simple variables. They were only used as reference indicators projected into the space of the PCA to support the factors interpretation.

The PCA analysis shows that the proportion of country land dedicated to Agriculture (% agri area) is not significant in this study and therefore removed.

### 4.1.2 Understanding relations between variables

Multi-variate analysis, successively PCA and FA, are used with a twofold objective (1) verifying the consistence of selected variables through statistical and theory/field experience compliance (2) analyzing their relationships.

The first five components explain more than 73 % of the variability with an eigenvalue above 1 (Table 3 and Fig. 2). These components are the basis for the following analysis of relationships between variables.

**Factor 1 (HDP)**

Factor 1 (HDP) represents the human development of the country taking into account health (Child Mortal-5, Malaria), education (School enrol, School G/B), living conditions (Urban Pop, % Slums, WS, S), the income (GDP per cap, femal eco, poverty) and governance (WGI-GE). Following the effort of synthesis of the HDI, the HDP factor represents the level of human development in a more multi-dimensional way by including health, gender, urbanization and governance aspects. The HDI is, indeed, based on four indicators: life expectancy at birth, mean years of schooling, expected years of schooling and Gross National Income per capita.

---

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

**Table 3.** Rotated PCA factors matrix after varimax rotation and kaiser normalization. The five components explain 73 % of the variability. In bold, the highest factor loading.

| | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| Eigenvalue | 9.625 | 3.030 | 2.262 | 2.051 | 1.448 |
| Variability (%) | 38.502 | 12.121 | 9.048 | 8.204 | 5.791 |
| Cumulative (%) | 38.502 | 50.623 | 59.671 | 67.875 | 73.666 |
| Variables | HDP | AP | WR | ODA | CEC |
| TWRR | −0.293 | 0.008 | **0.770** | 0.027 | 0.184 |
| S | **0.823** | 0.186 | −0.009 | −0.051 | 0.065 |
| Child Mortal-5 | **−0.871** | −0.219 | −0.124 | 0.075 | −0.120 |
| WS | **0.850** | 0.089 | −0.051 | −0.075 | 0.201 |
| School G/B | **0.597** | 0.111 | 0.268 | 0.111 | 0.226 |
| Particip | 0.193 | 0.045 | 0.006 | −0.187 | **0.815** |
| GDP per cap | **0.922** | 0.023 | −0.011 | 0.093 | 0.078 |
| % irrigation | 0.425 | **0.733** | 0.121 | −0.114 | 0.037 |
| ODA WSS | −0.051 | 0.037 | −0.047 | **0.840** | −0.036 |
| ODA | −0.457 | −0.104 | 0.064 | **0.708** | −0.055 |
| Malaria | **−0.811** | −0.287 | 0.183 | 0.127 | 0.001 |
| Water Use int Agri | 0.477 | **0.768** | −0.163 | 0.004 | 0.047 |
| Precipit | −0.095 | −0.010 | **0.929** | −0.054 | 0.057 |
| Desert risk | −0.363 | 0.110 | **−0.667** | −0.059 | 0.344 |
| Total withdrawal | 0.501 | **0.732** | 0.025 | −0.073 | 0.083 |
| WGI.V.A. | 0.307 | −0.072 | 0.155 | 0.289 | **0.762** |
| WGI.PS.AV | 0.401 | −0.042 | −0.023 | **0.657** | 0.320 |
| WGI.GE | **0.673** | 0.078 | −0.160 | 0.324 | 0.483 |
| School enrol | **0.773** | 0.081 | 0.223 | −0.053 | 0.229 |
| Femal eco | **−0.629** | −0.210 | 0.218 | −0.082 | 0.261 |
| % Slums | **−0.839** | −0.080 | 0.091 | 0.017 | −0.234 |
| Withdrawal municipal | 0.176 | **−0.892** | −0.050 | −0.051 | 0.055 |
| Withdrawal industrial | 0.226 | **−0.638** | 0.379 | −0.006 | 0.069 |
| Poverty | **−0.741** | −0.053 | 0.126 | −0.017 | −0.072 |
| Urban Pop | **0.788** | −0.188 | −0.107 | −0.073 | 0.045 |

As expected, Fig. 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 (the income per capita – "GDP per cap", sanitation and water supply level – "WS" and "S") .

The participation of women in economic activities is correlated with malaria (0.589 correlation), urban slums proportion (0.511 correlation) water supply access (−0.498 correlation), the mortality of children under 5 yr old (0.477 correlation–Child Mortal-5), and other variables expressing the poverty of the country. 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 (see Appendix B). Developing countries in the WatSan4Dev dataset are clearly in line with the first part of this behaviour: a high female economic activity rate is therefore an expression of poverty.

Governance (WGI-GE) is correlated mainly with the development of the country: 0.724 correlation with Gross

Domestic Product per capita (GDP per cap), 0.647 correlation with Water Supply access (WS), 0.559 with basic Sanitation access (*S*), and 0.518 correlation with gross school enrolment (School enrol).

## Factor 2 (AP)

Factor 2 (AP) shows high factor loadings for variables related to water consumption according to their uses (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 as a measure of how economic activities put pressure on water resources in a country. As is well known, water demand in developing countries is dedicated in majority for agricultural purposes (Hinrichsen et al., 1997). Almost 82 % of developing countries considered (83 out of 101) dedicate water to agriculture at more than 50 %. Therefore, total water withdrawal is negatively correlated to municipal ($-0.553$ correlation) and industrial ($-0.251$ correlation) withdrawals. The more the factor loading is positive, the more agriculture is predominant. On the contrary, the more the factor loading is close to 0 or negative, the more the share of municipal and industrial activities in water consumption growths.

In addition to main high correlations with withdrawals and agriculture intensity (Water Use intensity in agri), the level of agricultural land equipped in irrigation system (% irrigation) is more correlated with to the level of development of the country than to the water resources variables (TWRR and Precipit). We advance the hypothesis that irrigated perimeters captured by this variable are medium or large scale schemes; very small irrigation 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. Water resources availability accounts for little in the choice of irrigation development in this case in developing countries.

## Factor 3 (WR)

Factor 3 (WR) gathers high factor loadings for variables expressing the amount of water resources available through the amount of precipitations (precipit), the estimated amount of water renewable resources available per capita (TWRR) and the proportion of land under desertification risk (desert risk). The more the value is negative, the more the country presents arid areas, either directly desert like in the Middle East or areas receiving low water resources that can be under threat like Sahel countries.

## Factor 4 (ODA)

Factor 4 (ODA) represents the Official Development Aid, both the global flow and the specific flow dedicated to Water Supply and Sanitation. The governance indicator related to the political stability and absence of violence (WGI PS AV) is included in this component. ODA are funds dedicated to development not to emergency purpose. This component shows that donors request beneficiary countries for a minimum of political stability before committing and transferring funds. Other than the stability and absence of violence (WGI PS AV), the global official development Aid (ODA) is also correlated with health (0.460 correlation – Child mortal-5), access to sanitation and water facilities ($-0.450$ and $-0.412$ correlation, respectively) and poverty variables (0.408 correlation – Poverty, 0.405 correlation – % slums). These are therefore the main indicators being considered by donors for fund commitment.

## Factor 5 (CEC)

Factor 5 (CEC) expresses the degree of concern of a country for the environment empowered by government accountability towards its population (WGI-VA). WGI Voice and accountability (WGI-VA) measures the ability of citizens to participate in selecting their government and examining its actions. The indicator also includes the level of freedom of expression/media, freedom of association that constitute the necessary conditions for the civil society to be an opposition or controlling force to government. From the perspective of this analysis, this is an expression of the essential role of the civil society in pushing a global environmental policy in the country; for instance, regarding climate change mitigation, protection of biodiversity or hazardous waste management. The economic-industrial interests are often an obstacle to environmental protection and related policy making. However, citizens are key actors for pushing such protection processes because being directly affected by an environmental degradation (health, household budget, food security, future generations, ...). In a more direct way, many studies (e.g. Barrett and Graddy, 2000) found that an increase in civil and political freedoms significantly reduces some measures of pollution and hence improves the quality of the environment.

FA confirms the PCA results by showing the same factor loading matrix (Table 4) therefore should be interpreted in the same way.

## 4.2 Identifying key variables

This section describes the analyses performed to identify the key independent variables involved in the WSS variables when considering all developing countries. The relations between the independent variables that are common to developing countries are analysed. For each dependent variable (water supply – WS, and sanitation – *S*), two linear regression analyses are performed. The objective is to identify the main key explanatory – independent indicators and not the associated models.

The first linear regressions for both dependent variables allow the identification of the proportion of the urban slums,

**Table 4.** Rotated FA factors matrix with principal component as extraction method, varimax rotation and Kaiser normalization. The five components explain more than 73 % of the variability. In bold, the highest factor loading.

|                      | D1      | D2      | D3      | D4      | D5      |
|----------------------|---------|---------|---------|---------|---------|
| Variability (%)      | 35.480  | 12.515  | 9.533   | 7.873   | 8.265   |
| Cumulative (%)       | 35.480  | 47.994  | 57.527  | 65.400  | 73.666  |
| Variable             | HDP     | AP      | WR      | ODA     | CEC     |
| TWRR                 | −0.307  | 0.004   | **0.763** | 0.024 | 0.191   |
| S                    | **0.826** | 0.170 | 0.005   | −0.053  | 0.071   |
| Child Mortal-5       | **−0.872** | −0.199 | −0.139 | 0.077  | −0.128  |
| WS                   | **0.851** | 0.072 | −0.039  | −0.074  | 0.207   |
| School G/B           | **0.594** | 0.096 | 0.278   | 0.111   | 0.230   |
| Particip             | 0.189   | 0.038   | −0.002  | −0.171  | **0.820** |
| GDP per cap          | **0.922** | 0.004 | 0.004   | 0.092   | 0.080   |
| % irrigation         | 0.436   | **0.723** | 0.135 | −0.116  | 0.045   |
| ODA WSS              | −0.047  | 0.039   | −0.038  | **0.840** | −0.053 |
| ODA                  | −0.457  | −0.096  | 0.063   | **0.708** | −0.072 |
| Malaria              | **−0.819** | −0.273 | 0.168  | 0.128   | −0.004  |
| Water Use int Agri   | 0.494   | **0.760** | −0.146 | 0.005  | 0.050   |
| Precipit             | −0.111  | −0.019  | **0.926** | −0.061 | 0.068  |
| Desert risk          | −0.351  | 0.124   | **−0.676** | −0.044 | 0.336 |
| Total withdrawal     | 0.514   | **0.721** | 0.040 | −0.074  | 0.090   |
| WGI.V.A.             | 0.300   | −0.082  | 0.153   | 0.301   | **0.759** |
| WGI.PS.AV            | 0.402   | −0.051  | −0.014  | **0.662** | 0.309 |
| WGI.GE               | **0.676** | 0.065 | −0.150  | 0.333   | 0.478   |
| School enrol         | **0.770** | 0.063 | 0.233   | −0.053  | 0.236   |
| Femal eco            | **−0.638** | −0.201 | 0.201 | −0.077  | 0.261   |
| % Slums              | **−0.841** | −0.064 | 0.080 | 0.014   | −0.237  |
| Withdrawal municipal | 0.159   | **−0.895** | −0.059 | −0.049 | 0.053  |
| Withdrawal industrial| 0.206   | **−0.647** | 0.374 | −0.009  | 0.073   |
| Poverty              | **−0.744** | −0.040 | 0.114 | −0.017  | −0.074  |
| Urban Pop            | **0.786** | −0.203 | −0.098 | −0.074  | 0.049   |

the children under five years mortality rate and malaria as the key explanatory variables with an adjusted $R^2$ of 0.711 and 0.680 for the WS and $S$, respectively. It is accepted that improving Water Supply and Sanitation services reduces mortality due to waterborne diseases, mainly diarrhoea in particular for children under 5 yr, malaria, and cholera (UNICEF and WHO, 2009). However, the existence of such serious health issues is also an important incentive for policy and commitments towards WSS. Therefore the variables children under five years' mortality rate and malaria being rather consequences than causes of WSS, we proceed to a second linear regression but removing those variables from the analysis. This allows the identification of complementary key variables masked by the first results. This time the adjusted $R^2$ are equal to 0.686 and 0.635 for WS and $S$, respectively. As the objective of this study is rather about the identification of the key variables impacting WSS than the consequences of a good WSS, we will describe in this section the results of the second analysis and not those of the first. The first analyses confirm the impact of access to improved WSS on health, represented here by child mortality and the prevalence of malaria.

Although several variable introduction methods have been tested, such as stepwise and backward, these provide similar

results which show once again the stability of the variables and their relationships.

### 4.2.1 Water Supply (WS) level

Table 5 presents the parameters of the second linear regression with stepwise method and Adjusted $R^2 = 0.686$. The income per capita (Gdp per cap), the proportion of irrigation (% irrigation), urban population (Urban pop) and, the percentage of informal settlement (% slums) are found as the most significant in explaining the access level to water supply (WS).

Developing countries, being in an important and often informal urbanisation process, encounter difficulties in providing basic services in slums, explaining the negative coefficient $\beta$ (−0.333). More generally, urbanisation takes an important role in the access to water supply and sanitation facilities and more concretely to 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 population flows from rural areas. Slum development is expected to rise and by 2050, urban dwellers will likely account for 66 % of that in the less developed regions (UN DESA, 2010). Therefore, extending the access to sanitation and water supply in these conditions can be hindered or slowed 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 medium income (GDP per cap) understood as an indicator of living standard is an important condition for households to access improved infrastructure. Finally, the development of irrigation appears as a positive factor. The reason may stand in the multi-purpose of the irrigation infrastructure: water supply source used for crops is also used for water delivery for livestock and population.

### 4.2.2 Sanitation (S) level

Table 6 presents the parameters of the second linear regression with stepwise method and an Adjusted $R^2 = 0.635$. The total water withdrawal per capita (total withdrawal), the proportion of urban population (Urban pop) and within this population, the percentage of informal settlement (% slums), are found as the most significant in explaining the access level of sanitation ($S$).

The type of urban development is crucial in explaining the observed access to sanitation for the same reasons mentioned in Sect. 4.2.1. Total water withdrawal represents the capacity of the society to mobilize water resources available without distinction of uses. It is generally accepted that the total water consumption per inhabitant (total water withdrawal) increases with development of the country because of the extension of household consumption (Gleick, 1996) and/or

**Table 5.** Model parameters with water supply as dependent variable – method stepwise (adjusted $R^2 = 0.686$).

| Variables | Standardized Coefficients | | $t^*$ | Sig.* | 95 % confidence Interval for $\beta$ | |
| --- | --- | --- | --- | --- | --- | --- |
| | $\beta^*$ | 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 $\beta$ is the constant, Std error is the standard error of $\beta$, $t$ is $\beta$ divided by the std error and Sig. is the p-value.
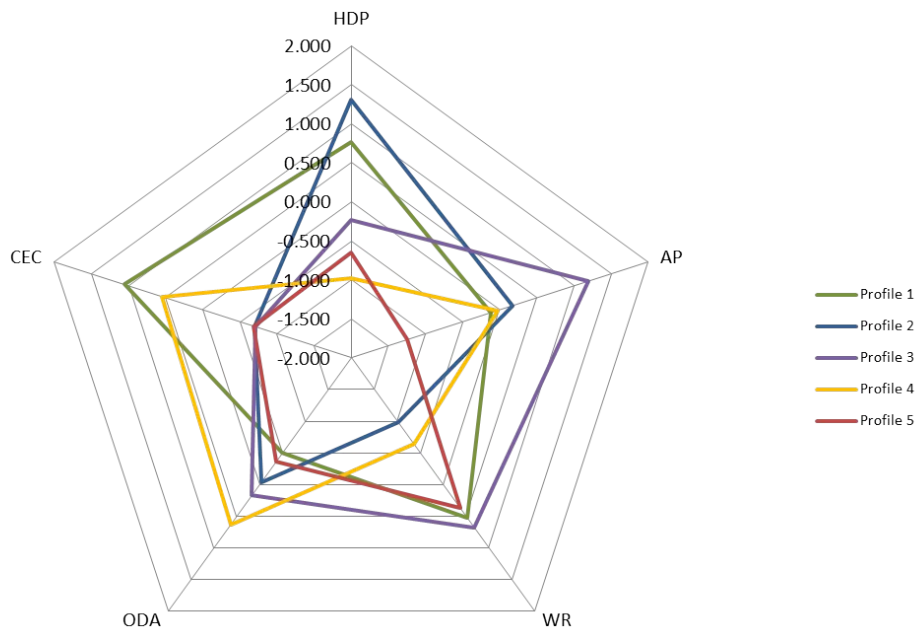


**Fig. 3.** The five country profiles. Each color shows the centroid of the class for each profile. HDP (D1) stands for Human Development against Poverty in the country, AP (D2) for Human Activity Pressure, WR (D3) for Water Resources, ODA (D4) for Official development Aid, and CEC for Country Environmental Concern.

increase of water volume needed due to economic activities development (industrial, commercial and agriculture). The link with sanitation in this context is indirect; the development of the country leads to development of both offer and demand for sanitation services.

The development of water supply capacities (e.g. technological aspects, operator and financial capacities) may also overflow to sanitation services. Stakeholders involved in the water supply production start investing at one point on the end of the water cycle, meaning wastewater treatment and sanitation. In that case, the link with the amount water withdrawals is direct.

### 4.3 Analysing country observations

#### 4.3.1 Clustering countries with similar profile

The K-mean clustering method is used to identify groups of countries with a similar profile (Fig. 3). The interpretation is relative to the 101 country observations. Hierarchical Agglomerative Clustering method was tested but the K-mean method provided more relevant clusters. The analysis was performed testing different numbers of clusters and five clusters was found to be the most optimised choice and coherent with the factors obtained from the previous analysis (HDP – Human Development against Poverty, AP – Human Activity Pressure, WR – Water resources, ODA – Official development Aid, CEC – Country Environmental Concern).

**Table 6.** Model parameters with Sanitation as dependent variable, stepwise method (adjusted $R^2 = 0.635$).

| Variables | Standardized Coefficients | | $t^*$ | Sig.* | 95 % confidence Interval for $\beta$ | |
|---|---|---|---|---|---|---|
| | $\beta^*$ | 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 $\beta$ is the constant, Std error is the standard error of $\beta$, $t$ is $\beta$ divided by the std error and Sig. is the p-value.
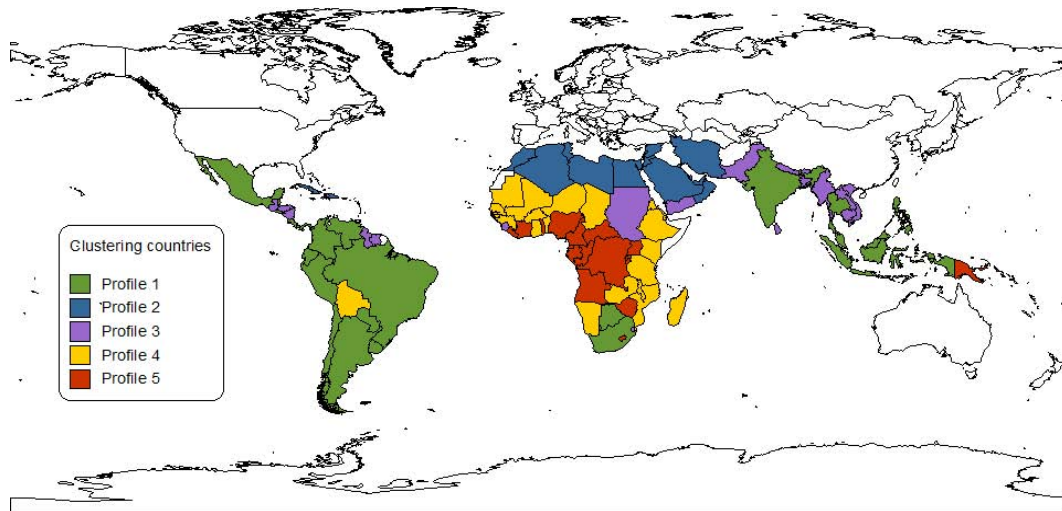


**Fig. 4.** Geographical distribution of the five country profiles.

### 4.3.2 Description of the five profiles

Country profiles (Fig. 3) are ordered from most favourable (profile 1) to the less favourable country situation (profile 5). Profile 1 presents high values on WR, HDP and CEC implying little need for external support (ODA). The water demand reflects a relative balance between municipal/industrial activities and agriculture, leaving the latter as the dominant activity. Profile 2 shows weaknesses in terms of accountability and population's freedom, leading to a low commitment towards environment. Profile 3 indicates that the economy is mainly driven by agricultural activities facilitated by a context of abundant natural resources. Profile 4 and 5 are the less favourable profiles when considering human development. Although profile 4 benefits from a higher level of freedom (CEC), political stability allows a high commitment of ODA and a more balanced economy, which corresponds with value around 0 in the AP component. Profile 5 shows an economy mainly based on primary resources exploitation (negative value in AP) with little consideration on environmental aspects.

In Fig. 4, the African continent shows the highest situation diversity, having countries belonging to all five profiles. Latin America concentrates a majority of "best profiles", except Bolivia, while South East Asia presents a mixed distribution with countries mainly ranging between advanced development (profile 1) and agricultural-oriented countries (profile 3).

**Profile 1: towards well-being (Fig. 5, Table 7)**

Profile 1 is characterised by good development (HDP), a high level of freedom, together with an environmental engagement (CEC). Although water resources are still mainly dedicated to agriculture, these resources are also devoted to domestic, commercial and/or industrial uses as shown by the average value ranging around 0 in AP axe. These countries benefit from a positive context of water resources availability (WR). The countries in this profile benefit from the lowest level of external aid (ODA). Therefore, the profile is considered the most advantageous, showing high human development, environmental concerns, degree of freedom and a certain balance between water uses (agriculture versus industry/municipal uses), indicating a diversified economic
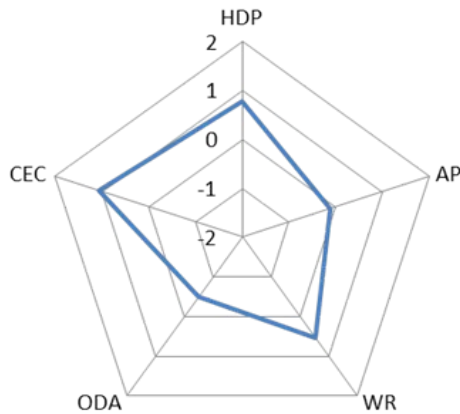
**Fig. 5.** Profile 1 – this profile corresponds to the most advanced countries having a relatively good human development (HDP) based on a diversified economy (AP) and political/democratic organization (CEC). WSS level is, therefore, relatively high compared to other developing countries; consequently, these countries benefit from low ODA from donors.
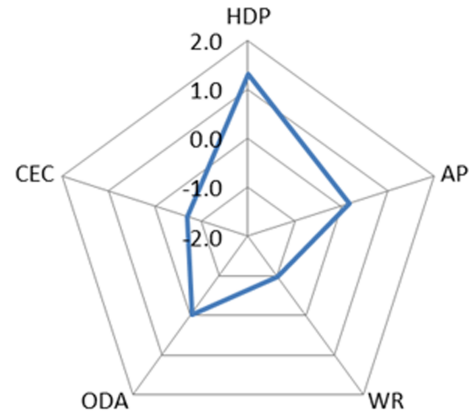


**Fig. 6.** Profile 2 – besides good human development (HDP), this profile presents very low values in terms of level of accountability/democracy, together with low interest in environmental matters (CEC). The relatively wealthy situation allows stepping over the scarce water resources context (WR) by showing a high level of WSS services.

development. The context implies relatively high access levels to WSS (around 80 % for WS) even if access to basic sanitation always falls behind (around 70 %).

The Philippines is the most representative country in this cluster (the closest to the centroid, Table 7). Belize shows a similar shape to that of profile 5, but the amplitude of values in particular regarding ODA, CEC and HDP justifies its inclusion in profile 1.

Botswana has a major share of water withdrawals due to municipal and industrial activities (59 %), the agriculture is not dominant. Its economy is based on natural resources exploitation but it has succeeded in maintaining other sectors such as tourism, financial services, subsistence farming, and cattle breeding[5]. This tends to the ideal of economic diversification of this profile.

Jamaica, Malaysia and Mauritius show similar values toward ODA, which is superior to the average. Thailand, as well as India, shows a maximum in AP indicating the dominance of agricultural activities (respectively consuming 96 and 87 % of water resources). Thailand has a profile close to the average, except its AP value. India also shows a significant low human development index. Being in the periphery of this cluster, India could be an intermediary country between class 1 and 3. The clustering algorithm has included it in this cluster because having three components (CEC, ODA, WR) out of the five match with the average values in this profile.

### Profile 2: freedom/democracy black spot (Fig. 6, Table 8)

This profile presents a high HDP value close to profile 1, but with a low-medium level of ODA. Agriculture is consuming

the majority of the drawn water resources but enough resources remain for satisfying municipal needs, industries being marginal. The main differences with profile 1 stand in the water scarcity and the poor level in CEC. A majority of these countries leave little space for citizen expression, being monarchies or authoritarian regimes such as in the Arabian Peninsula. In addition, they are scarcely exposed to global environmental matters (e.g. climate change, protected species trade, hazardous waste transport) coherent with the low level of CEC. Regarding WSS, these countries present the best levels of access (close to 100 %).

Lebanon is the referenced country of this cluster (Table 8). Cuba and the Dominican Republic diverge from the average because they benefit from better climate conditions and more water resource availability. Saudi Arabia and Libya have the poorest environmental involvement at international level together with little accountability to citizens, implying a high divergence on CEC.

### Profile 3: agricultural economy (Fig. 7, Table 9)

These countries show a medium human development level sustained by a rather important external Aid. The economy of these countries is almost oriented, in some cases exclusively, towards agricultural activities. This agriculture predominance is often accompanied by an important practice of irrigated agriculture. This intensity of agriculture is facilitated by the abundance of water resources. Environmental concerns and government accountability come last in importance. These countries are also defined by an important gap in term of access to basic sanitation (S) compared to access to an improved water supply (WS). Sanitation access is clearly neglected, falling behind with at least 25–30 % less than WS access.

---

[5]CIA factbook Botswana
https://www.cia.gov/library/publications/the-world-factbook/geos/bc.html

**Table 7.** Factors loadings and distances to the centroid of the profile 1: the last 7 countries of the list are considered outliers with divergence behavior from the average value.

| Observation | HDP | AP | WR | ODA | 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 8.** Factors loadings and distances to the centroid of the profile 2.

| Observation | HDP | AP | WR | ODA | 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 |

Nepal is a good example of this cluster with a WS access at around 90 % and a sanitation access level at 35 %. The fact that 81 % of the population lives in rural areas[6] pushes away the efforts to improve sanitation by keeping cheaper and easier alternative sanitation options, such as open defecation, acceptable.

Vietnam is the closest country to the centroid of this profile with agriculture accounting for 68 % of WR, and 31 % of agricultural land equipped with irrigation. Table 9 indicates distances of the different countries to the centroid.

---

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

**Table 9.** Factors loadings and distances to average for profile 3. The last 5 countries in the list are considered outliers of the cluster, showing divergences with the average value of the profile.

| Observation | HDP | AP | WR | ODA | CEC | Distance |
|---|---|---|---|---|---|---|
| Vietnam | 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 |
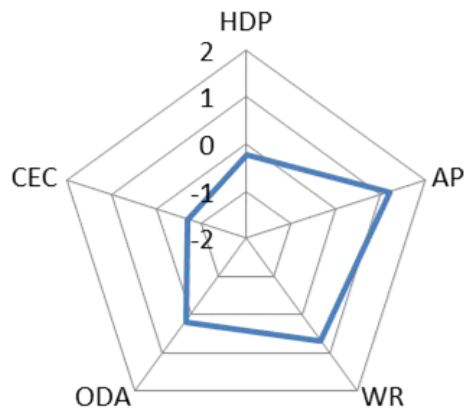


**Fig. 7.** Profile 3 – mainly organised around agriculture (AP), often irrigation, the countries in this profile show limited human development (HDP) with low accountability and concern for the environment (CEC). An important gap in basic sanitation access is observed that steps generally far behind water supply access.



**Fig. 8.** Profile 4 – mean feature of this profile is the important level of External Aid (ODA) facilitated by a relatively good political accountability and stability (CEC) despite a low human development (HDP). In line with this general context, access to WSS is still below the MDG's targets.

Several countries diverge from the average profile: (i) Honduras, Nicaragua and Guatemala have developed little irrigated/intensive agriculture compared to Southern East Asia countries with less than 3 % irrigated areas; however, in these countries, agriculture demands around 80 % of the water consumption; (ii) Yemen's water resources are low, however agricultural represents almost all the amount of water consumption (at 95 %); (iii) Bhutan receives more ODA than average (> 125 $ per cap, 5th position); (iv) Bangladesh shows a higher concern for environmental issues than average. One hypothesis is that Bangladesh is concerned by
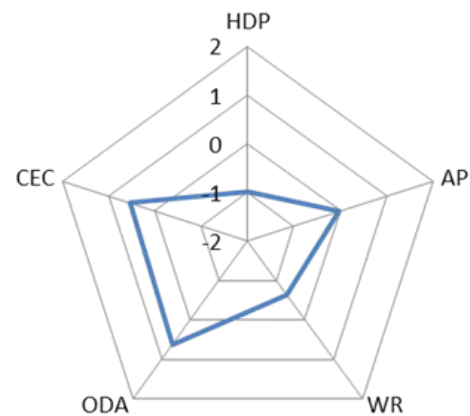
consequences of environmental issues like floods from sea rising due to monsoons.

**Profile 4: essential external support (Fig. 8, Table 10)**

This profile is mainly defined by the high level of ODA being justified by the low level of human development (HDP). Despite this fact, these countries show an important commitment toward environmental and international issues and have significantly developed freedom for citizens (high values in CEC). Agriculture monopolises a majority of the drawn

**Table 10.** Factors loadings and distances to average for profile 4. The last 5 countries of the list are considered outliers of the cluster, showing divergence with the average value of the profile.

| Observation | HDP | AP | WR | ODA | 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 |

water resources but remains enough for the development of other activities (domestic, commercial and/or industrial) in the context of water resources stress. Indeed, most of the countries belonging to this profile are facing stress on water resources mainly located in Sub-Sahel and South East Africa. The consequences of this situation is poor access to Water services and an even worse sanitation service. This also includes some very low levels observed such as Ethiopia with 22 % for water supply and 13 % for sanitation.

Table 10 shows the distances of the different countries to the centroid. Zambia is the most representative country, having a dominant agriculture activity (76 %) but with 86 % of its land under desertification risk. Zambia also benefits from almost 100 $ per cap in ODA.

Bolivia is the only country from Latin America in this profile. Bolivia is the second target country for external Aid (ODA) (after Honduras) in Latin America, having one of the lowest human development index of the region (HDI). Its economy is dominated by agriculture but also exploits natural resources as indicated by the average value in AP. This country is also concerned by environmental issues by facing, for instance, desertification processes. With the existence of a democratic regime since 1982, this leads to a high value in CEC which is characteristic of this profile.

Madagascar and Eritrea almost exclusively dedicate drawn water resources to agriculture (at 95 % like in profile 3);

being superior to the average value, the other four components still correspond to this profile.

Cape Verde has a development superior to the other countries but keeps a similar shape for the other factors in profile 4. Djibouti distinguishes itself by having scarce water resources. Being an already existing desert area, this country is no more considered "under desertification risk" by definition.

**Profile 5: primary material consumption (Fig. 9, Table 11)**

This profile is clearly characterised by the abundance of water resources. However these resources are mainly dedicated to industrial and municipal uses, in contrast with an underdeveloped agriculture. In fact, countries in this profile have an economy based on mining, or primary material exportation, being the main consumer of water resources. This is typically the case of Gabon, Congo, Angola Liberia, Papua New Guinea, Togo and Nigeria that exploit subsoil resources such as oil, gold, copper, phosphates, diamonds, or other natural resources such as timber.

This profile cumulates both non-favourable conditions for a sustainable development and an economy focused on primary resources exploitation consuming water resources in a non-sustainable way. These countries also have poor human development in a context of political instability. The

**Table 11.** Factors loadings and distance to centroid of the profile 5. Equatorial Guinea is considered outlier of the cluster, showing divergence with the average value of the profile.

| Observation | HDP | AP | WR | ODA | 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 |

consequences are that they benefit from limited external financial commitments from international donors. Although these countries have a low level of human development, they benefit from a rather low level of ODA. One hypothesis is because of the political instability and context of violence predominant in these countries. Effectively, as already mentioned, ODA is dedicated to development and not to humanitarian emergencies. International donors seem to consider that a minimum security and stability is necessary to engage and implement this medium- to long-term vision. From a point of view of WSS, the situation is similar to profile 4, with low level of access for the population either to water supply or sanitation.

Comoros stands as the reference country of profile 5. Although Gabon has a high value of HDP, the other components still fit well in the profile. Zimbabwe and Burundi diverge from the average profile by keeping agriculture as central economic activities. These two countries can be seen as intermediate cases between profile 3 and 5. They belong to cluster 5 because three components out of five (HDP, ODA, WR) fit with the average behaviour of this profile. Equatorial Guinea has had an economy based on oil exploitation since 1996. They are in the group of least developing countries[7] (low HDP). The main divergence stands in the very negative value in CEC due to authoritarian regimes less oriented to accountability and freedom of its citizen and low concern for the environment (considered as extra costs and economic constraints). However its "political stability" since 1979 allows the delivery of ODA at a significant level.

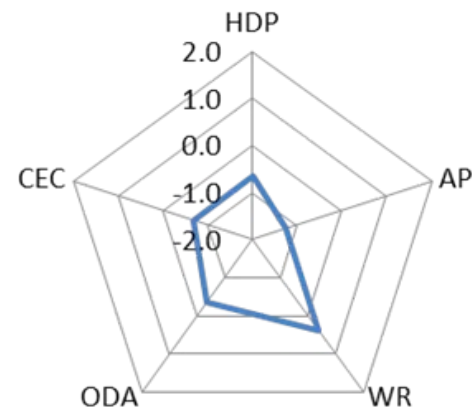[7]Least developed countries list http://www.unohrlls.org/en/ldc/25/



**Fig. 9.** Profile 5 – this profile shows low values either in human development (HDP) and accountability toward population and environment (CEC). Because of the political instability or violence within the country, external aid remains low. The environmental concern may be undermined by the type of economy, mainly based on raw material or natural resources exploitation.

Lesotho enjoys a political stability and benefits from important external financial flows (48 $ per capita). It also has a high level of CEC, indicating a rather high level of freedom (similar to profile 4).

## 5 Discussion

Environmental variables, namely the amount of water resources available per capita (TWRR), land under desertification threat (Desert Risk), precipitation, National Biological

Index (NBI), water bodies surface (WB), and Environmental Sustainability Index (ESI); they are coherent and correlated among themselves but they show little correlation to the percentages of the population having access to water supply and sanitation. On another side, the Environmental Sustainability Index (ESI) is a cross-cutting and very complex index (the list of variables being part of this index are detailed in Appendix A) showing the capacity of a country to manage its environment in a sustainable way. Therefore, it is highly correlated with the water resources availability (0.528 correlation – TWRR) and governance (0.420 correlation – WGI-VA) but little correlated with the WSS. In addition, the water quality sub-indicators included in ESI computation often have 86 % of missing values. This undermines the reliability of these sub-indicators and can contribute to explaining this absence of correlation. Water bodies surface (WB) and National Biological Index (NBI) are also correlated to the amount of precipitations (respectively 0.439 and 0.732 correlation – Precipit) but are not linked to WSS.

Therefore, although this paper shows no significant correlations between environmental variables and WSS services, it is evident that WSS has an important impact on the environment. For example, wastewater treatment or sanitation excreta, when not adequately treated, have a direct impact on surface or groundwater quality and vice versa.

This proves that considering only the quantity of water resources is not sufficient for establishing links with WSS but should be combined with qualitative water indicators. The main issue is that at national scale, indicators on quality, accessibility, and management capacities of water resources are not available in most developing countries. Downscaling analyses may permit the availability of such indicators, for instance at basin level. Garriga et al. (2009) when modelling data on the Turkana district in Kenya obtained information on water quality (water point protection, pollutant agents from livestock or waste, water salinity) through qualitative questionnaires.

The representativeness of water resources at national scale still requires reinforcement data collection and reliability of current indicators, finding proxies or designing appropriate new indicators at national scale. The FAO and UNEP/GEMS-Water mainly engaged in these efforts provide data on biochemical, nitrogen concentrations at national scale and water quality surface waters at watershed scale. The JMP performed pilot country projects (Ethiopia, Jordan, Nicaragua, Nigeria, Tajikistan) in 2004–2005 to investigate new methods to monitor quality of drinking water supplies. The conclusions of these pilots have been available since 2010 but the generalisation of data collection and methods worldwide has not been initiated yet. To summarize, further research on country indicators defining quality (accuracy and uncertainty), accessibility, management capacities of resources, and beyond environment state are needed to refine analyses of the relationship between environment, human development and WSS.

Despite this constraint on data availability and quality, this paper proposes a common analytical framework to developing countries and using multi-variate analyses in combination with OLS methods. It provides a general view of the water sector in developing countries, including other sectors on which efforts should be focussed in priority to facilitate the WSS sustainable development. This can be used either by international donors or national authorities that have to face and handle these issues. Using robust and simple methodologies, these analyses can be replicated easily.

Commitment towards waterborne diseases reduction and formalisation of the ongoing urbanisation process came out as main leverages to increase WSS sustainable access. The rapid urban growth throughout the developing world observed over the last 20 yr is seriously "outstripping the capacity of most cities to provide adequate services for their citizens" (Cohen, 2006). Cities are an opportunity for investment in water supply and sanitation because the critical consumer mass to create new infrastructures and management capacities are easily reachable. Breakdown statistics between rural and urban areas can illustrate this difference by often showing a gap between levels of WSS access. This concentration of population is therefore an a priori chance to foster WSS access. However, as highlighted by UN DESA (2010), urbanisation should be organised to avoid the development of slums (66 to 86 % of the developing countries' population will live in slums by 2050). Urbanisation is expected to continue in the next 30 years, mainly in medium and small cities. Cohen (2006) mentioned that these latter city categories are facing higher health issues and lack of capacities more than large cities. Fighting and finding solutions to organise urbanisation processes, in particular in medium and small cities, should be therefore a priority.

The new classification of countries around country profiles proposed in this paper helps in targeting countries in difficulty but also identifying weaknesses restraining country development and well-being. The approach of evaluation of country performance on several dimensions appears as necessary to best adapt the state/international donor actions to the various situations observable among developing countries. The UN institutions in charge of monitoring the progress toward MDG's targets, among those the WSS'ones, analyse data with a geographical approach: at global level or by subregions. This geographical approach masks disparities among the different countries but also the countries' specific needs within the analysed area. The approach proposed in this paper considers a different angle by analysing groups of countries which have similar profiles to avoid this "bias". International donors and policy makers can use the different clusters of countries to establish joint strategies but also for monitoring and assessing the impact of their development and cooperation policies in a more strategic way.

## 6    Conclusions

The WatSan4Dev dataset proposed in this paper is a new database of development (socio-economic, environmental and governance indicators) collecting the indices widely dispersed in the different international institutions in charge of their management. Because of the different origins and gathering methods, raw data have been cleaned, and robust missing data processing methods applied for filling the gaps. As a result of the statistical processing, since the coherence and robustness among the different indicators has been increased and proved that these indicators can be used for further and deeper studies, the WatSan4Dev dataset should only be used for qualitative estimations and analysis.

The main objective of this paper was to validate the coherency of relationships among the different indicators and establish cross-sector links with WSS. From a first analysis of the variables in WatSan4Dev (Table 1), it has been proved that 25 variables out of the 42 variables considered are actually independent and coherent with the literature and experience in the field.

The variables organise themselves around five factors:

– Factor 1 (HDP) represents the human development of a country not only as a function of income, life expectancy and education (as this is the case of the HDI index) but also takes into account population living conditions (Urban Pop, % Slums, WS, $S$), and governance aspects (WGI-GE).

– Factor 2 (AP) expresses how economic activities pressure water resources in the country.

– Factor 3 (WR) gathers high factors loadings for variables expressing the amount of water resources available.

– Factor 4 (ODA) represents the external financial flows ODA and ODA WSS conditioned by the political stability and absence of violence in the country.

– Factor 5 (CEC) links the involvement of the country in International Environmental Agreements with the governance (WGI-VA). WGI Voice and Accountability (WGI-VA) measures the ability of citizens to participate in electing their government (democracy) and to be a government controlling force in the country.

The identification of the key variables in an improved access to WSS showed that urbanisation takes an essential role in the access to improved water supply and sanitation facilities, together with 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 flows from rural areas. The slums' development is expected to rise and, by 2050, urban dwellers will likely account for 66 % in the less developed regions (UN, 2010). Therefore, extending the access to sanitation and water supply in these conditions can be hindered or slowed down. However, cities are an opportunity for investment in water supply and sanitation because the critical mass to create new infrastructures and maintain management capacities is easily reachable.

The relationship between Child mortality under 5 yr (Child Mortality-5) and access to water supply and sanitation (WSS) is clearly depicted with an important impact on children's health. Indeed, more generally, WSS is clearly correlated with health, which is an essential leverage for improving living conditions.

Irrigation also pushes towards improvement of water supply thanks to multi-purpose infrastructure. Environmental conditions referring to water resources (total water renewable resources, the precipitation, desert risk) are secondary factors in explaining the WSS level.

Built around the five factors, five different country profiles have also been established (Fig. 3). Profile 1 represents the equilibrium among the different development indicators with countries benefiting from access to Water Supply around 80 %. These countries are close to reaching MDGs, however, access to basic sanitation still falls behind (around 70 %). Profile 2 can also be considered as a good profile with high WSS access but authoritarian political systems and deficit accountability is the major black spot of this profile. Profile 3 illustrates the fact that Sanitation is clearly neglected, falling behind with at least 25–30 % less than WS. Profile 4 has the lowest human development expressed in terms of poor access to Water services and even worth sanitation services. Profile 5 cumulates both non-favourable conditions for a sustainable development and an economy focused on primary resources exploitation and consuming water resources in a non-sustainable way. These countries also have poor human development in a context of political instability.

This new classification based on development indicators is a new way to identify countries in difficulty but also indicates weaknesses restraining countries' development and well-being. International donors can use these profiles for establishing more adapted development strategies and better monitoring to assess the impact of their policies.

Countries in profiles 4 and 5 are less advanced but efforts should be put on different domains: (a) profile 4 countries are facing water stress from scarcity in a context of low development. Therefore, efficient resources management, mitigation measures and adapted solutions appear crucial to improve WSS; (b) in the case of profile 5, efforts should be concentrated to ensure political stability and reduction of violence to get out of the "dog eat dog organisation" implied by such a context. This appears as the a priori condition to ensure the existence of individual/collective initiatives towards sustainable development.

For countries in profile 3, sanitation should be a priority. Within this agricultural irrigation context, water supply

services are often ensured, while basic sanitation development is neglected. Rural areas keep, indeed, feasible/acceptable non-improved practices such as open defecation, while investments in basics sanitation facilities may be seen as unnecessary, in particular for poor households.

It is to be noted that countries from profile 2 benefit from a significant level of ODA, while WSS level is almost at 100 %. A detailed analysis of the external aid provided to these countries should be considered to better understand in what kind of activities the aid is engaged.

As shown in this paper, the biophysical variables included to characterize the environmental state are not sufficient or appropriate for exploring the links among water resources, WSS and human development as they describe only the quantity of water resources. Further research on country indicators defining quality, accessibility, management capacities of resources, and beyond environment state is needed to refine analysis of the relationships among the environment, human development and WSS at national country scale. This paper finally proposes that the international organisations in charge of monitoring MDG's targets should collect qualitative water indicators at national level to establish links between the environment and WSS.

On 6 March 2012[8], the UN announced that the Millennium targets for safe drinking water were reached, although the one for sanitation is still out of reach. Beyond this positive quantitative improvement, access to safe water and basic sanitation is not ensured yet. Quality of water resources still remains a challenge in a majority of developing countries, for instance, in Latin America. Disparities also appear across regions and according to urban/rural areas. Progress towards WSS may require even more efforts in the future because of these potential rising constraints, like the increase of population, urbanisation, and/or climate change and variability consequences.

The next part of this research will introduce the different data models 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.

---

[8]UN Press release: http://www.who.int/mediacentre/news/releases/2012/drinking_water_20120306/en/

## Appendix A

### WatSan4Dev indicators: description (in alphabetic order: short name, full name)

#### 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.

#### 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.

#### DAM, Dam capacity

The capacity of the dams indicated is the theoretical initial capacity; it does not change with time. The current dam capacity is the state of the dams at a given time that can be decreased by stilting.

#### Desert risk, dryland 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. Esty et al. (2005) describes the ESI and methods used.

### Femal eco, female economic activity rate

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 US. 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.

### 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 designed for heath, but excludes the provision of water and sanitation.

### Malaria, Malaria cases

Number of reported cases per 1000 persons in country.

### 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, participation to international environmental agreements

It is calculated taking into account the participation 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.

### 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.

### % 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).

## Urban-rural growth, population growth

Variation of the population respectively in urban and rural areas between 2000–2005.

## % 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.)

## 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.

## 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.

## 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.

## 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 higher this index is, the lower the water constraint is.

## TOT-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
$+$Domestic Water Use $+$ Industrial Water Use.

## Withdrawal-municipal, water withdrawal for municipal purpose

Annual quantity of water used 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.

## Withdrawal-Industrial, water withdrawal for industrial purpose

Annual quantity of water used 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
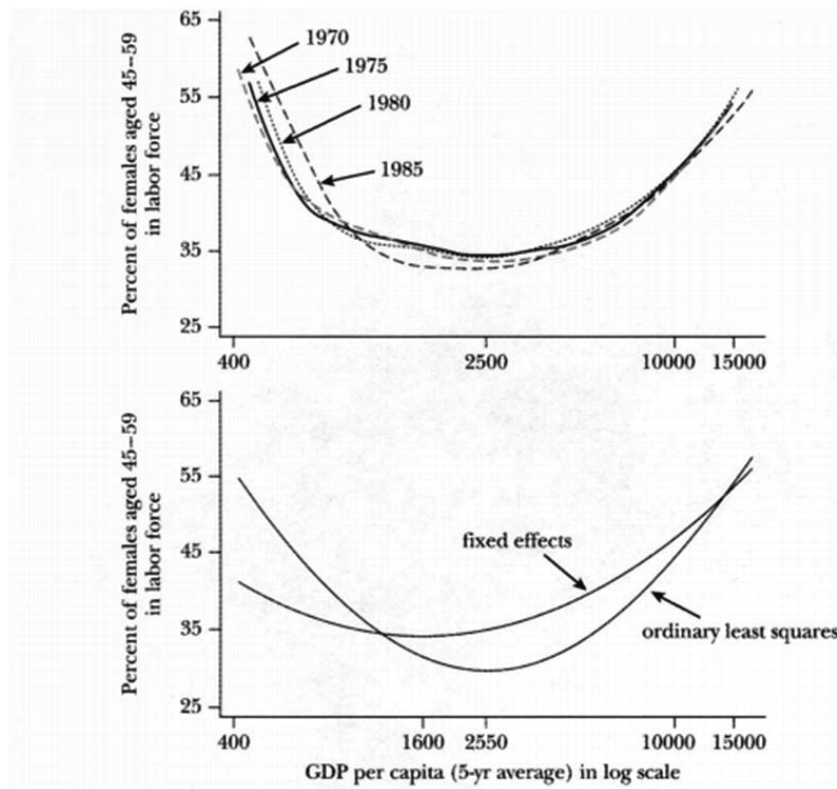
**Female Labor Force Participation**



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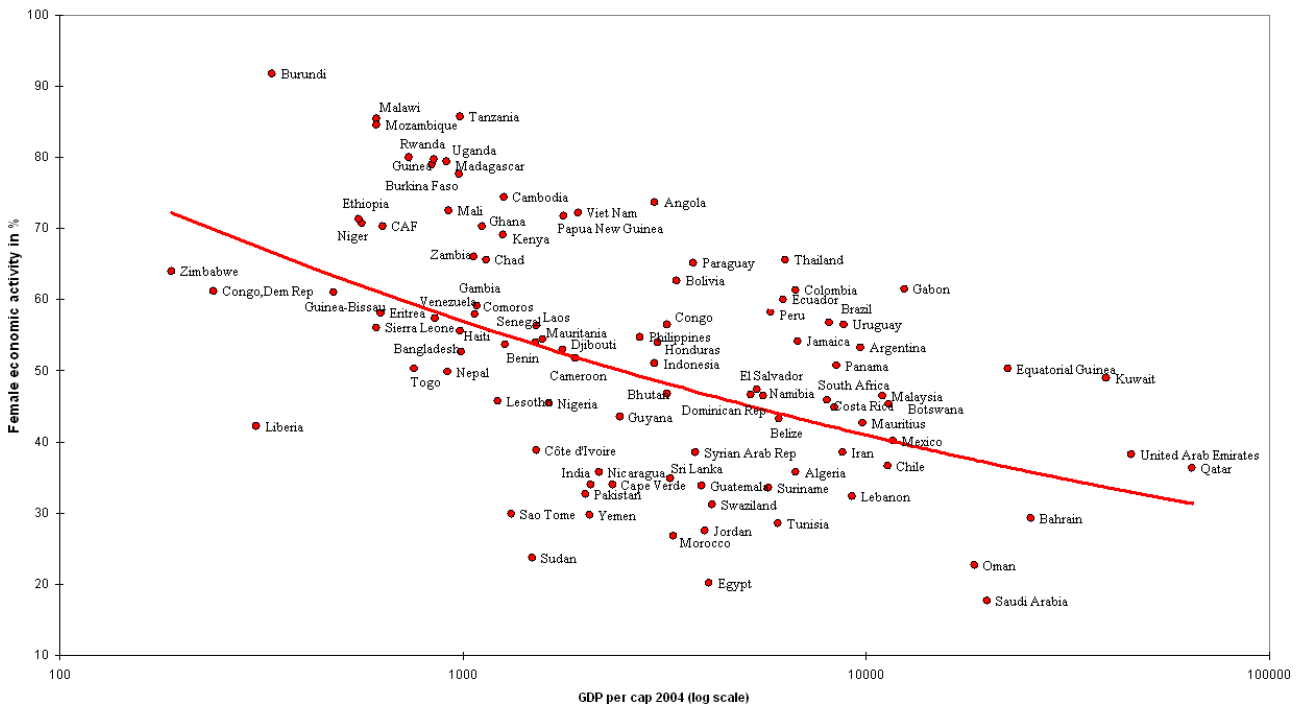


**Fig. B2.** Female economic participation versus income per capita, distribution and trendline for 2004.

crossed with the income level per capita for the 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) has risen in absolute terms since 1985, at least because of inflation, 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 proof 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, 1999; Boserup, 1989; Beguy, 2009).

# 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, 2010.

Ahn,N. and Mira, P.: A Note on the Changing Relationship Between Fertility and Female Employment Rates in Developed Countries, FEDEA Working paper, http://www.fedea.es/pub/Papers/1999/DT99-09.pdf (last access: 23 October 2012), 1999.

AQUASTAT: A global information system on water and agriculture, http://www.fao.org/nr/water/aquastat/main/index.stm (last access: 23 October 2012), 2011.

Barrett, S. and Graddy, K.: Freedom, Growth, and the Environment, Environ. Develop. Econ., 5, 433–456, 2000.

Bartholomew, D. J., Steele, F., Galbraith, J., and Moustaki, I.: Analysis of Multivariate Social Science Data, 2nd Edn., Chapman & Hall/Crc, New York, 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, Population and development review, 15, Supplement: rural development and population: institutions and policy, Population council publisher, http://www.jstor.org/stable/2807921, last access: 23 October 2012, 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, Globalization Health, 6, 1–8, doi:10.1186/1744-8603-6-12, 2010.

Cohen, B.: Urbanization in developing countries: Current trends,future projections, and key challenges for sustainability, Technol. Soc., 28, 63–80, doi:10.1016/j.techsoc.2005.10.005, 2006.

CPI: database from Transparency International, http://www.transparency.org/policy_research/surveys_indices/cpi/2004, last access: 23 October 2012.

Dublin statement on water and sustainable development 1992, http://www.wmo.int/pages/prog/hwrp/documents/english/icwedece.html (last access: 23 October 2012), 1992.

Earth trends: the environmental information portal, http://earthtrends.wri.org, last access: 23 October 2012, database in maintenance since 2012, 2011.

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, http://www.yale.edu/esi/ (last access: 23 October 2012), 2005.

Europe Aid: Water Sector Development and Governance: Complementarities and synergies between sector-wide Approach and Integrated Water Resource Management, European Commission publication Office, 44 pp., http://ec.europa.eu/europeaid/infopoint/publications/europeaid/176a_en.htm (last access: 23 October 2012), 2009.

FMI list of Emerging and developing countries for 2004: http://www.imf.org/external/pubs/ft/weo/2004/02/data/groups.htm (last access: 23 October 2012), 2004.

Fonseca, C., Diaz, C., Verdeil, V., Cotton, A., Zipper, T., Van Koppen, P., and Vink, M.: Working together to improve Aid effectiveness in the water sector, EUWI and IRC international water and sanitation center, 36 pp., http://www.euwi.net/files/EUWI-AWG_finalreport_Eng_Dec_20081.pdf (last access: 23 October 2012), 2008.

GEO Portal: http://geodata.grid.unep.ch/# (last access: 23 October 2012), 2011.

Giné Garriga, R., Pérez Foguet, A., Molina, J. L., Bromley, J., and Sullivan, C.: Application of Bayesian networks to assess water poverty, II International Conference on Sustainability Measurement and Modelling, CIMNE, Barcelona, 24 pp., 2009.

Gleick, P. H.: Basic water requirements for human activities: Meeting basic needs, http://www.bvsde.paho.org/bvsacd/cd17/basicwate.pdf, last access: 23 October 2012, Water Int., 21, 83–92, 1996.

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, http://www.k4health.org/pr/m14/m14chap2_2.shtml (last access: 23 October 2012), 1997.

Honaker, J. and King, G.: What to Do about Missing Values in Time-Series Cross-Section Data, http://gking.harvard.edu/files/pr.pdf, last access: 23 October 2012, Am. J. Polit. Sci., 54, 561–581, 2010.

Honaker, J., King, G., and Blackwell, M.: Amelia II: A Program for Missing Data, Version 1.2–17, http://r.iq.harvard.edu/docs/amelia/amelia.pdf (last access: 23 October 2012), 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. Ed. Phych., 26, 139-142, 1935.

JMP – Joint Monitoring Programme: Portal, http://www.wssinfo. org/ (last access: 23 October 2012), 2011.

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.

MacQueen, J. B.: Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1, 281–297, 1967.

Mammen, K. and Paxson, C.: Women's Work and economic Development, J. Econ. Persp., 14, 141–164, 2000.

OECD: http://stats.oecd.org/qwids/ (last access: 23 October 2012), 2011.

Pearson, E. S.: On lines and planes of closest fit to systems of points in space, Phil. Mag., 2, 559–572, 1902.

Thurstone, L. L.: Multiple-Factor Analysis, University of Chicago Press, 1947.

UN Department of Economic and Social Affairs, Population Division: World Urbanization Prospects: The 2005 Revision, Working Paper No. ESA/P/WP/200, UN publications, New York, 2006.

UNICEF, WHO: Diarhhoea: Why children are still dying and what can be done?, WHO publications, Geneva, 57 pp., http://whqlibdoc.who.int/publications/2009/9789241598415_eng.pdf (last access: 23 October 2012), 2009.

UN DESA – United Nations, Department of Economic and Social Affairs, Population Division: World Urbanization Prospects, the 2009 Revision: Highlights, UN publications, New York, http://esa.un.org/unpd/wup/Documents/WUP2009_Highlights_Final.pdf (last access: 23 October 2012), 2010.

UNFPA: State of the world population 2007: Unleashing the Potential of Urban Growth, United Nations Population Fund publications, 100 pp., http://www.unfpa.org/swp/swpmain.htm (last access: 23 October 2012), 2007.

World Bank: http://data.worldbank.org/ (last access: 23 October 2012), 2011.