

## Referee comment 1

This paper presents an analysis of possible inter-relationships among water, governance, human development variables in developing countries, based on a large dataset collected (WatSan4Dev). I must say that I find this paper an introduction to the upcoming research (as stated by authors) on data modelling, rather than a standing alone research work. It is a well presented description of the use of different sources to collect a large number of data to create the WatSan4Dev dataset, but I can't see the immediate use of the dataset in real world practice nor can I say that the information provided is scientifically interesting for a journal such as HESS. Furthermore, I believe that some information is missing on the description of data availability: it is not clear to me what is the temporal coverage of the available data? is it only the year 2004? Authors claim the dataset to be coherent and consistent, but there is no mention of the uncertainty in the collected data. In terms of quality of the presentation, I find the paper well structured and written. The quality of the Figure could be improved.

## Response

**Comment 1:** I must say that I find this paper an introduction to the upcoming research (as stated by authors) on data modelling, rather than a standing alone research work. I can't see the immediate use of the dataset in real world practice ...

**Answer 1:** In order to answer and monitor the objectives of the Millennium Development Goals (MDGs), the international community and the states decided to develop several monitoring processes. For this purpose, several international institutions in charge of monitoring the Access to Water Supply and Sanitation (WSS) developed two indicators. Although the WSS indicators are proving that the access to water and sanitation are improving, today it is difficult to say what are the main factors involved in this improvement and what are the relationships between the different factors. This is the main objective of this research.

The work proposed in this paper is one of the first steps to answer these questions involving the MDG targets. An important research effort has been done to structure and analyse all the variables that could have a direct and/or an indirect influence on WSS indicators.

To build the dataset was a critical part of this research as it's the result of the analysis of the variables and the normalisation of a huge amount of data. The data processing and methodologies used will be of interest for future researchers in the domain, not only because of future analysis of the dataset but also because the methodologies as proposed in the paper can be applied to other data.

The preliminary analyses of this dataset have allowed two different things:

- 1) To verify that the relationships among the variables are coherent with literature and common knowledge in the domain. This allows us to say that a strong base has been developed to go further in the analysis.
- 2) It also allows the creation of country profiles based on these indicators opening new interpretation possibilities in developing countries.

Based on this exploration between variables and countries, we will build a tool to create probabilistic scenarios. The latter will be the subject of the next paper as we considered that this submitted paper is coherent as such, separated from the second part.

**Comment 2:** I cannot say that the information provided is scientifically interesting for a journal such as HESS.

**Answer 2:** We asked directly the editorial team before submitting the article to see if our work was in the scope of HESS. The answer was positive, it fits with the interdisciplinary view of HESS. Looking at Hess Aims and Scope ; HESS encourages and supports fundamental and applied research that seeks to understand the interactions between water, earth, ecosystems and man. A multi-disciplinary approach is encouraged that enables a broadening of the hydrologic perspective and the advancement of hydrologic science through the integration with other cognate sciences, and the cross-fertilization across disciplinary boundaries. HESS, therefore, has the ambition to serve not only the community of hydrologists, but all earth and life scientists, water engineers and water managers, who wish to publish original findings on the interactions between hydrological processes and other physical, chemical, biological and societal processes within the earth system, and the utilization of this holistic understanding towards sustainable management of water resources, water quality and water-related natural hazards.

**Comment 3:** the temporal coverage of the available data? is it only the year 2004?

**Answer 3:** Responding to the point on data availability, the coherency of data is validated for 2004 (as mentioned p 489 line15-19) for several reasons: i) it was the last release of the two variables measuring the percentage of access one to water supply, the other to sanitation services provided by the Joint Monitoring Programme (JMP). JMP updates these indicators on a five year basis ii) Considering a large number of variables, 2002-2004 was the time period for which the datasets was the most completed and representative of developing countries. Indeed, we excluded quite a number of variables because of too many missing values. Broadening the discussion on this point, the global research approach was to first validate the coherency and the modelling of data for 2004 to test methods before looking for potential time series (requiring a consequent amount of work).

**Comment4** The quality of the Figure could be improved.

**Answer 4:** OK

**Comment5** no mention of the uncertainty in the collected data

**Answer 5:** Concerning the point on data uncertainty, standards errors /margins are not often available in particular regarding socio- economic indicators but data are collected from official international providers that do an important work of verification and consolidation of datasets. For instance FAO performed several steps of validation as well as adjustments: “the first two levels of validation include human interaction, and are done based on cross-comparison with similar countries as well as historic data for the country in question. The last validation step, prior to saving of new results, is an automated

validation step that mathematically checks updated data for consistency and correctness. Almost 200 validation rules are used by this automated validation routine.” Such information (margin error) is only available for Worldwide Governance Indicators (provided by the World bank) and Corruption Perception Index (provided by Transparency International). This is one of the reasons why we closely checked the series with the literature, experience from the field to go beyond the uncertainty of data.