Hydrology and Earth System Sciences Discussions
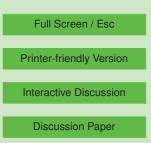
Interactive
Comment

Interactive comment on "Analysing inter-relationships among water, governance, human development variables in developing countries: WatSan4Dev database coherency analysis" *by* C. Dondeynaz et al.

Anonymous Referee #2

Received and published: 15 April 2012

This manuscript describes the development and preliminary analysis of a dataset combining data for multiple variables from a wide range of sources for developing countries. This is a challenging task and the authors should be recognised for attempting it. The paper is well organized. It could use some work by an editor to address consistent grammatical issues that probably derive from English as a second language. I have two general concerns about this manuscript: thematic and technical.

Thematic: This paper is organized around a technical task that assumes some future useful application, rather than a research question. If pretty much stays on this tech-

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

nical task. Those results that it produces simply confirm what is already known about the relationships among variables and indicators of development. As such this interest of this paper is limited. These is potential, however, to further develop some of the outcomes of this work so as to inform development efforts, to add to theory about development, to direct international development or aid efforts at the macro scale, etc.. For example the categorization of counties briefly explored in this manuscript could be further developed to further inform models such as the demographic transition, the mobility revolution, etc..

Technical:

âĂć Principle components analysis is related to factor analysis (it is basically factor analysis with commonalities equal to 0). Why do both?

âĂć For a PCA to be considered useful it should explain around 70% or more of the variance in the data set. For subset of African countries only about 50% of variance was explained with 3 components. This is rather low. Were there other components that explained enough variance to include? How were the number of components to include determined (eigenvalues alone? Scree test?).

âĂć Nations are not consistent in such things as definitions of key concepts and terms, methods of data collection, etc. For example what is considered urban in some countries would be considered rural in others. This will be a profound problem in using measures that are not standardized across countries. How was this issue addressed in this work?

âĂć I don't see how BOD could be used in this dataset. BOD is not reported as a aggregate variable at a country scale, and if it was it would be meaningless. BOD will vary within and across water bodies and streams. Perhaps there is an explanation that is missing from this paper. Otherwise this should be removed from the analysis.

âĂć If I read this correctly some of the variables used in the analysis are aggregates

(e.g., indices) of other variables that are also used in this analysis. This is like using the same variable, same measurements in the same analysis, leading to a problem of multicollinearity. On or the other should be removed.

âĂć Why was hierarchical clustering chosen? Is it better for this application that something like K-means clustering?

âĂć What is meant by "coherency" and "robustness" of the data is not well enough explained.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 485, 2012.