

## ***Interactive comment on “Hydrochemical assessment of Semarang area using multivariate statistics: A sample based dataset” by Dasapta Erwin Irawan and Thomas Triadi Putranto***

**Anonymous Referee #1**

Received and published: 5 August 2016

The MS addresses the current and interesting topic of groundwater quality data handling. It presents a dataset from Indonesia consisting of 58 samples of 20 parameters. This amount was proven to be sufficient for multivariate data analysis such as the ones presented in the MS.

Unfortunately, despite the importance and interesting thoughts provoked in the introduction, we only gain a small knowledge about the dataset, in addition I have found numerous flaws and information missing from the MS. Without these, the MS is just a good draft, a skeleton if I may, which – if filled up with the necessary information - may later on serve as a good paper. Nevertheless, if these issues are handled to a sufficient extent, I urge the authors to resubmit their MS to ESSD.

C1

Major concerns about the MS:

-The Authors state the aim of their research project in the Introduction, but not the aim of the MS itself. In addition, an extra aim is mentioned in Section 2.2: underlining the policy of the respective ministry i.e. letting them use and publish the data. This is truly important especially in the view of the data handling policy of Eastern European countries. Is this the first openly available groundwater dataset from Indonesia for example?

-Section 2.1.: Why were not the years between 1993-2003, and between 2003-2006 included?

-Such as in Section 3, the importance of the temporal coverage has to be discussed as well, why were these years chosen, e.g. in E Eurpe covering the turn of the 1990s is highly important with respect to the restructuring of agriculture and industry due to the collapse of the Soviet Union.

-Section 4.: The discussion is too brief. Consider e.g. <http://link.springer.com/article/10.1007/s10040-013-1093-x/fulltext.html> and try to place your methodology in that ideology or any other related to groundwater and explain why these methods were chosen, what is the novelty in choosing especially these, why did not you use e.g. CCDA (<https://cran.r-project.org/web/packages/ccda/ccda.pdf>) to be objective during the grouping

-Related to this section the specific aims are missing, e.g. why did you want to group the data etc. My main concern here is that the script describes the detailed steps of the data analysis but not its preparation, errors drawbacks etc. The results of the analysis described in the script are valuable in a “classic research paper” with proper discussion. However, without their interpretation - please do not misunderstand me – the results are useless for the readers of ESSD. The results without the knowledge of the Authors are hard to interpret, especially since only little is known of the background of the dataset, which should have been the main point in the first place.

C2

-Section 5.: In order to see the novelty of the MS at least the major results and pre-processing issues should be lined up point-by-point and more discussion is needed on the data handling policy of the ministry, prior to this publication. If this MS is a flagship in this sense, why?

Minor comments:

-Regarding the missing data: was not there an opportunity to interpolate the missing values? That could have been a big additional value of the paper? I would sum NAs as well: `sum(is.na(df))`

-Is there a spatial/temporal trend in the data, why and how were the sites chosen, there are more densely sampled areas in the western areas (Fig. 1), why?

-Use upper/lowercase letters for chemical formulae

-Instead of groundwater use the terminology of Tóth, J., 2009. Gravitational systems of groundwater flow: theory, evaluation, utilization. Cambridge University Press.

-Use "in-depth" instead of "in depth".

-Please state in the MS that the dataset is available as a supplement and the script itself as well.

In fine:

If you want to publish your MS and dataset in ESSD, please restructure the whole MS and move the emphasis from the analysis to the data preparation. If you are able to provide a script with justification about the data-preparation and fill in the gaps outlined in the review, I am sure your dataset and approach will be an excellent material and fill in an important gap in literature.

Yours sincerely,

Reviewer #1

---

C3

Interactive comment on Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-29, 2016.

C4