Dear Referee #1

Thank you for your interactive comment on our paper "Hydrochemical assessment of Semarang area using multivariate statistics: A sample based dataset". The following document contains our clarifications and answers.

About the MS:

- **Comment:** 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?
    - **Author's response:** Regarding the dataset, the dataset were conducted from several reports from the ministry of energy and mineral resources. Due to elaborate with local government (Provincial level), the reports were open data set and freely for researchers at university, and NGO who want to use the dataset for further analyzing as long as they mentioned the data source. This data set is not the first openly available groundwater dataset from Indonesia for example. A lot of previous (older than our hydrochemistry data) were available in the repository of environmental geology library at the Indonesia Geological Agency (Ministry of Energy and Mineral Reources). Unfortunately, the reports were mainly in Bahasa. However, based on our observation, we (Indonesian researchers) don't have the habit and knowledge of submitting dataset to journals. Therefore, this MS is one of the first dataset to be published in journals.
- **Comment:** Section 2.1: Why were not the years between 1993-2003, and between 2003-2006 included?
    - **Author's response:** The reports were mainly based on the internal project funded by the ministry. Thus, annual monitoring was difficult to apply due to our limited budget to handle 432 groundwater basins in total. Moreover the wells were not setting in automatic data recording, and we have invite the participation of private parties to install a cheap monitoring equipment in their wells.
- **Comment:** 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.
    - **Author's response:** Thank you for your comment and we will add it with some relevant informations, as follows.
    - Due to increasing groundwater used starting in the early 1990s, the ministry allocated project funding in Semarang groundwater basin. In 1992-1993, the project was started with the topic of mapping of groundwater reserve and then continued with the study of

groundwater conservation. To evaluate the groundwater condition in the new millennium, the ministry then continued the assessment of groundwater conservation in the early 2000's. Regarding the changing of political policy in the mid 2000's that local governments (both in Provincial and City) have the authority to manage natural resource including groundwater, but withing the Ministry supervision. Unfortunately, projects on basin monitoring and assessment were not granted regularly.

- Comment: 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
    - Author's response: We will revise the discussion section. Our main concern in choosing the R package was the visualization part. We use the ones that can produce the easiest plots. However, thank you for introducing the CCDA package, and we will examine it.
- Comment: 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.
    - Author's response: We will revise that part in the script accordingly. If we're not mistaken, this paper should only include explanations in data sampling, preparation, and handling, therefore the scripts must also follow that direction. Then we will delete the analysis part in the script.
- Comment: Section 5.: In order to see the novelty of the MS at least the major results and preprocessing 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?
    - Author's response: We will revise the data preprocessing and data handling section as mentioned in the comment. We don't carry any flagship with this MS.

Minor comments:

- Comment: 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)).
    - **Author's response:** Yes we can do the extrapolation.
- **Comment:** 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?
    - **Author's response:** We did not follow any trend during data sampling. The sites were chosen based on the availability of well data. Given on the situation, the western areas has more densely sites than the other part of the area, as it is the center of industrial and other economic activities in Semarang City.
- **Comment:** Use upper/lowercase letters for chemical formulae.
    - **Author's response:** we will correct the typos.
- **Comment:** Instead of groundwater use the terminology of Tóth, J., 2009. Gravitational systems of groundwater flow: theory, evaluation, utilization. Cambridge University Press.
    - **Author's response:** we will include it in the revision.
- **Comment:** Use "in-depth" instead of "in depth".
    - **Author's response:** we will correct the typos.
- **Comment:** Please state in the MS that the dataset is available as a supplement and the script itself as well.
    - **Author's response:** The dataset and scripts are available as supplementary files as they were uploaded in our Zenodo repository.

In fine:

- **Comment:** 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.
    - **Author's response:** we will revise the MS following the given the constructive comments.