Groundwater and river water interaction to solve water shortage: a case from Tasikmalaya, Indonesia

Dasapta E. Irawan[a*)], Cut N. Rachmi[b)], Aditya Pratama[a)], Gusti P. Tulak[a)], Ahmad D. Rochman[a)]

[a)] Faculty of Earth Sciences and Technology, Bandung Institute of Technology, Institut Teknologi Bandung, Jl. Ganesa No. 10, Bandung, 40132

[b)] Department of Paediatrics and Child Health, University of Sydney Clinical School, Children's Hospital at Westmead, Sydney, Australia

[*)]Corresponding author: Dasapta E. Irawan, address: Faculty of Earth Sciences and Technology, Bandung Institute of Technology, Institut Teknologi Bandung, Jl. Ganesa No. 10, Bandung, 40132, Phone: +6282121424210, Fax: +62 22 2514837, email: dasaptaerwin@outlook.co.id

Abstract.

Background. Water shortage is a common problem in the high density settlement along the riverbank of Ciromban and Cibeureum River, Tasikmalaya, as the quality of the water also decreases. One of the solution is to maximize the use of river water. This study aims to investigate the interaction between river and groundwater along the riverbank as a function of land use impact.

Methods. A river water and unconfined groundwater level mapping has been conducted to make water flow map, assuming both waters are in the same flow system. Physical parameters, temperature, TDS, and pH were measured at each stations to understand water characteristics.

Results and discussions. Based on observations at 50 dug wells and 12 river stations on July-August 2014, a close interaction between both water bodies has been identified with two flow systems: effluent flow (or gaining stream) at Cibereum river segment and influent flow (losing stream) at Ciromban river segment. Physical parameters show a high correlation in temperature, pH, and TDS. Hence, further evaluation from health point of view should be taken before using river water as raw water supply in Tasikmalaya area.

INTRODUCTION

Tasikmalaya as one of the growing city in West Java Province, is also sufferring from drought. The city which one was a full agriculture-based area, now have grown to a industrial and commercial city. This condition has forced the local government and water supply company to look for alternative source of water. Currently the source of water is exclusively supplied by several deep water wells and groundwater springs at the foot slope of Mount (Mt) Galunggung, a 2167 meter above mean sea level (masl) strato-volcano located at north-western area from the city. The volcano is the upstream of several rivers, including Ciromban and Cibeureum river. In drought period, the available water sources can maintain their supply to the city (see Fig. 1). Therefore the local authority need to look for alternative sources. One of the thought was to use the water from the Ciromban and Cibeureum river. However, the interaction of both rivers with the groundwater system has not been closely observed. This research is placed as a preliminary study of such interaction based on water quality analysis. The question we want to answer here are whether we can separate the groundwater and river water and the mixin between both waters from the physical parameters that will be discussed in the Method section. The interactions between groundwater and surface water represent important issues in water resources management. The exchange of water between both water bodies is an important pheomenon that arises form infiltration, spring exposure, water flow, and geological layers (Cavazza & Pagliara, 2000).

MATERIALS AND METHODS

The dataset was composed of 62 points: 50 points of groundwater measurements at dug wells along the riverbank and 12 points of river water measurements near the dug well measurement. In the field survey, we used handheld instruments from Hanna Instrument, consists of water level detector (WLD) for water level depth, total dissolved solids (TDS) meter for temperature and TDS measurement, and Dissolved Oxygen (DO) meter to measure DO concentration. The precision of the instruments were: 0.1 cm for the WLD, 0.1°C for temperature, 0.1 µSiemens/cm, 0.01 ppm or mg/L for the TDS and DO. All tools were calibrated prior of use. Figure 1b shows station locations which were taken in the period from July to August 2014.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.720v1 | CC-BY 4.0 Open Access | rec: 18 Dec 2014, publ: 18 Dec 2014

Here we introduce an statistical analysis to explore the dataset, consists of: basic exploratory analysis and decision tree. Both methods are used to see the possible separation between both water bodies and mixing of water quality. We used R version 3.1.1 (R Core Team, 2014) and Rstudio version 0.98.1028 (RStudio Team, 2014). R is a widely used open source statistical programming, available as open source software. It compiles and runs on UNIX platforms (including Linux), MacOS and Windows. RStudio is an integrated development environment (IDE) for R and it is available in open source and commercial editions and also runs on Linux, Mac, and Windows. We used several add-on packages of R to do the analysis: (a) graphical plotting using "lattice" (Sarkar, 2008), (b) data manipulation with "dplyr" (Wickham & Francois, 2014), (c) spatial analysis using "sp" (Roger S. Bivand, 2013), and (d) principal component analysis using "pcaMethods" (Stacklies et al., 2007) and "randomForest" (Liaw & Wiener, 2002a).

The principal component analysis (PCA) is a classic multivariate statistics, used to reduce the size of data dimension. It basically classifies the variables in to principal components (PC). It succesfully used as tools to analyse hydrogeological nature based on water quality (Irawan et al., 2009, 2014; Irawan, Puradimaja & Silaen, 2012). The random forest (Breiman, 2001; Hastie, Tibshirani & Friedman, 2001; Liaw & Wiener, 2002b) is an ensemble approach that can also be thought of as a form of nearest neighbor predictor. The random forest starts with a standard machine learning technique called a "decision tree".

RESULTS AND DISCUSSIONS

The following analysis consists of basic exploratory analysis using "summary()" function and "pairs()" plot functions (see Table 1 and Fig. 2a). The pairs plot shows the interaction between variables. According to the summary, the sample locations are located at the elevation below 400 masl with TDS value fall in the range of fresh water. However we can see some river water spots with high TDS 800 ppm. This condition indicates a possible light contamination in the water. The pH, resisitivity, and DO values show normal range, with some groundwater water points having DO more than 2 ppm. One of the reason is due to the high population of organic substances in the water. Moreover, the mixing of groundwater (blue) and river water (red) points in the pairs plot reflects the possible mixing of water.

PCA calculates variable and case loadings to the PC as seen in the screeplot and biplot (see Fig. 2b and 2c respectively). The first two PC explains 70% of the total variance. This is good enough considering the complex nature of interaction between shallow groundwater and river water. Both water types are separated in the biplot: groundwater samples are grouped in the left side while river water samples are located in the right side. Groundwater samples are controlled largely by y coordinate trend, DO, and resistivity. The y-trend reflects more dominant north-south groundwater flow direction than east-west flow direction. Major DO and resisitivity role are explained by possibility light organic contaminant in the water. Currently we haven't had more clear evidence whether the source is from river water seepage to the aquifer of a direct urban contamination from the surface.

On the river water side, x coordinate, temperature, pH, and TDS are the most dominant variable. The x-trend shows the dominant of east-west flow direction than north-south direction. Aside to that, the temperature, pH, and TDS indicates the strong influence of surficial activity. It also supports groundwater discharge to the river, from water level mapping. Both water, however, have a fair share to level variable. The variable importance using "varImp()" function in Random Forest also supports the interaction between both water (see Fig. 3). It shows the high importance of resistivity and y-trend from groundwater side and x-trend, pH, DO, and temperature from river side.

CONCLUSIONS

From the analysis we can draw several remarks: (1) A close interaction occurs between groundwater and river water, as shown by the water level maps. In the period of dry season (July-August 2014), we identified a low gradient groundwater flow to river in the Cibeureum segment (effluent groundwater flow or gaining stream), where as a similar low gradient seepage occurs in the Ciromban river segment (influent groundwater flow or losing stream). The close interaction is also shown by pairs plot; (2) A clear separation between both waters can be drawn form the PCA. Moreover, Random Forest technique has shown stronger role of some variables than the others;

and (3) Based on the result, we recommend that the authority should not use the river water directly without proper treatment. The pumping of river water should also be carefully design as it will influence shallow groundwater level.

ACKNOWLEDGEMENTS

We wish to express our gratitude to undergraduate students that give full support in the field survey stage. We wish to thank Dr. Wouter Buytaert of Imperial College London and Dr. Reik Donner of PIK Postdam for introducing R to us in 2011, and Dr. Willem Vervoort University of Sydney for his invaluable guidance and support. We also would like to thank Prof.Dr. Lambok Hutasoit as Head of Applied Geology Group, Prof.Dr. Deny Juanda, Prof.Dr. Sudarto Notosiswoyo, Dr. Prihadi Sumintadireja, and Dr. Suyatno Yuwono who have built a strong research to us.

REFERENCES

Breiman L. 2001. Random Forests. In: *Machine Learning*. 5–32.

Cavazza S, Pagliara S. 2000. Groundwater and surface water interactions. *EOLSS Encyclopedia of Life Support System* 1.

Hastie T, Tibshirani R, Friedman J. 2001. *The Elements of Statistical Learning*. New York, NY, USA: Springer New York Inc.

Irawan DE, Puradimaja DJ, Notosiswoyo S, Soemintadiredja P. 2009. Hydrogeochemistry of volcanic hydrogeology based on cluster analysis of Mount Ciremai, West Java, Indonesia. *Journal of Hydrology* 376:221 – 234.

Irawan DE, Puradimaja DJ, Silaen H. 2012. Hydrodynamic Relationship between Man-Made Lake and Surrounding Aquifer, Cimahi, Bandung, Indonesia. *International Groundwater Engineering Conference, WASET*.

Irawan DE, Silaen H, Sumintadireja P, Lubis RF, Brahmantyo B, Puradimaja DJ. 2014. Groundwater–surface water interactions of Ciliwung River streams, segment Bogor–Jakarta, Indonesia. *Environmental Earth Sciences*:1–8.

Liaw A, Wiener M. 2002a. Classification and Regression by randomForest. *R News* 2:18–22.

Liaw A, Wiener M. 2002b. Classification and Regression by randomForest. *R News* 2:18–22.

R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Roger S. Bivand VG-R, Edzer Pebesma. 2013. *Applied spatial data analysis with R, Second edition*. New York: Springer.

RStudio Team. 2014. *RStudio: An Integrated Development Environment to R*. Vienna, Austria: RStudio.

Sarkar D. 2008. *Lattice: Multivariate Data Visualization with R*. New York: Springer.

Stacklies W, Redestig H, Scholz M, Walther D, Selbig J. 2007. pcaMethods – a Bioconductor package providing PCA methods for incomplete data. *Bioinformatics* 23:1164–1167.

Wickham H, Francois R. 2014. *dplyr: A Grammar of Data Manipulation*.

FIGURE 1. Maps of the study area, as part of Tasikmalaya and Mt. Galunggung watershed and sample points along the Ciromban and Cibeureum riverbank (shaded area).

6

TABLE 1. Statistical summary of the dataset and pairs plot showing the interaction between variables

	Level (masl)	TDS (ppm)	Temp (oC)	pH	Resistivity (mS/cm)	DO (ppm)
min	363	276	25	6.7	-136.6	0.53
1st quantile	375	512	26.1	6.89	-45.2	1.51
median	381	558	26.8	7	-37	1.75
mean	384	579	26.9	7.09	-41.1	1.81
3rd quantile	390	647	27.3	7.3	-30.3	2.02
max	430	822	29.3	7.68	-14.8	4.93

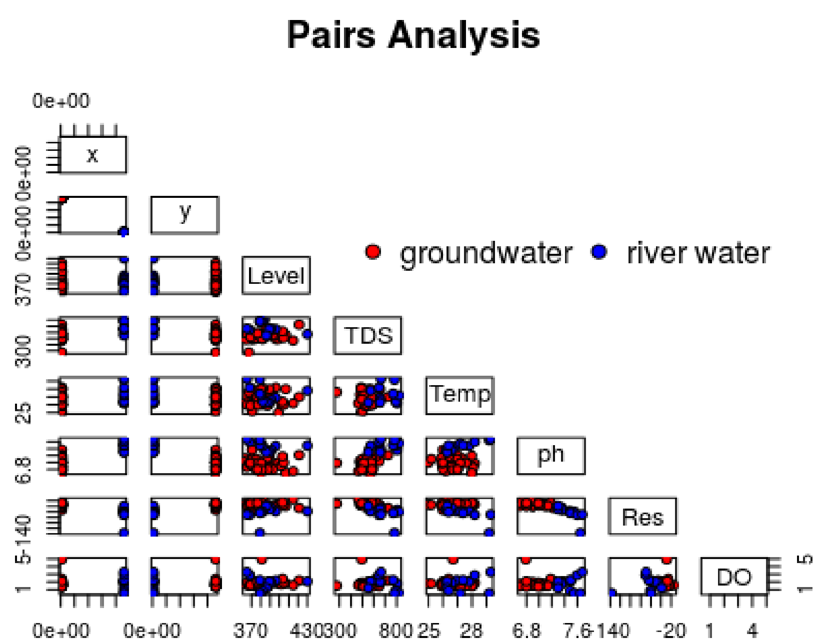

Fig. 2a

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.720v1 | CC-BY 4.0 Open Access | rec: 18 Dec 2014, publ: 18 Dec 2014

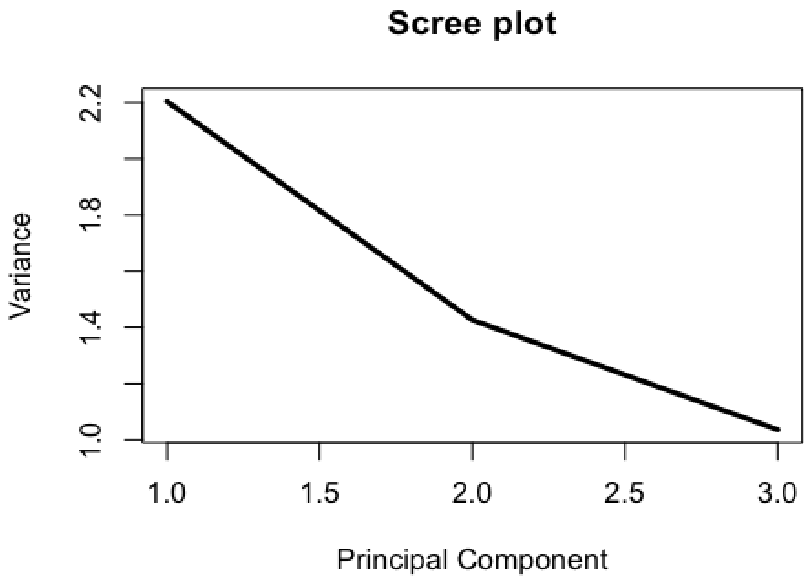

Fig. 2b

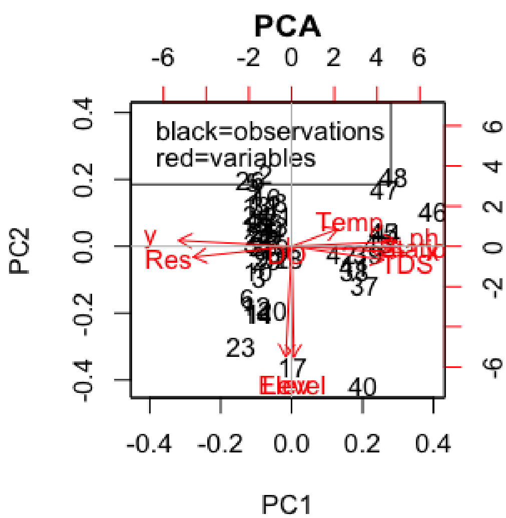

Fig. 2c

FIGURE 2. (a) The Pairs plot of the variables; (b) The screeplot shows the variance of each principal components; (c) The Biplot of PCA shows the separation of the groundwater points (on the left) and river water point (on the right), and the dominating variables on each group.

8

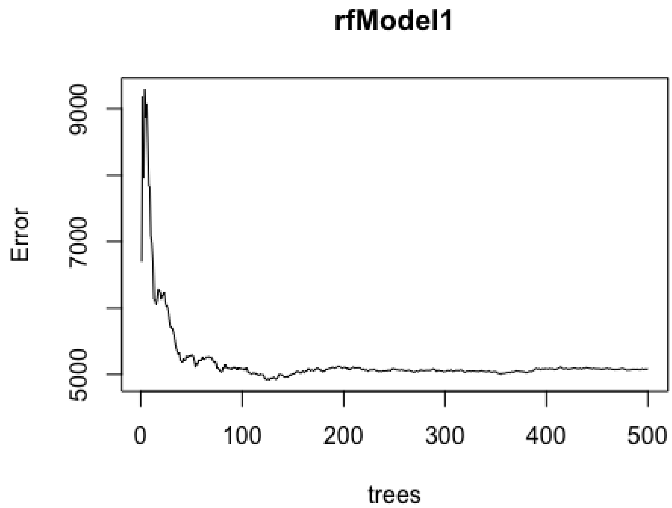

Fig. 3a

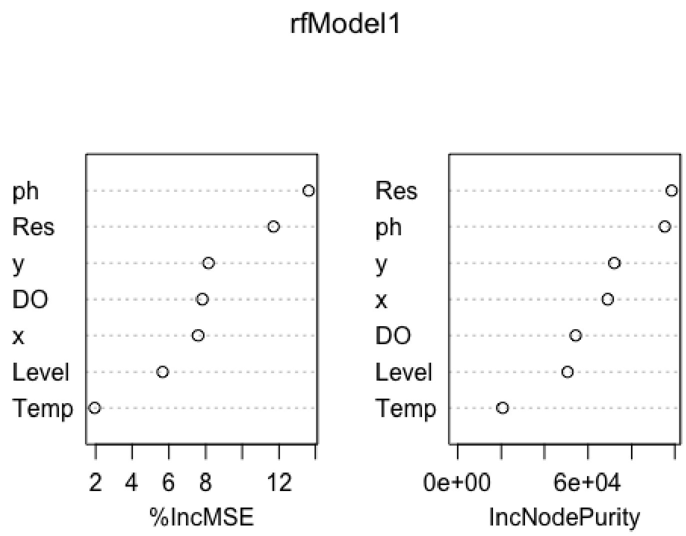

Fig. 3b

FIGURE 3. RandomForest plot shows the important set of variables. (a) The trees and error plot shows the variance of each regression tree (from the 1st iteration to the 500th); (b) The variable importance plot shows the strong variables.