

# Preliminary study of soil permeability properties using principal component analysis

M Yulianti<sup>1</sup>, Y Sudriani<sup>1</sup> and H A Rustini<sup>1</sup>

<sup>1</sup> Research Center for Limnology, Indonesian Institute of Sciences (LIPI), Cibinong Science Center-Botanical Garden, Cibinong 16911, Indonesia  
E-mail: meti@limnologi.lipi.go.id

**Abstract.** Soil permeability measurement is undoubtedly important in carrying out soil-water research such as rainfall-runoff modelling, irrigation water distribution systems, etc. It is also known that acquiring reliable soil permeability data is rather laborious, time-consuming, and costly. Therefore, it is desirable to develop the prediction model. Several studies of empirical equations for predicting permeability have been undertaken by many researchers. These studies derived the models from areas which soil characteristics are different from Indonesian soil, which suggest a possibility that these permeability models are site-specific. The purpose of this study is to identify which soil parameters correspond strongly to soil permeability and propose a preliminary model for permeability prediction. Principal component analysis (PCA) was applied to 16 parameters analysed from 37 sites consist of 91 samples obtained from Batanghari Watershed. Findings indicated five variables that have strong correlation with soil permeability, and we recommend a preliminary permeability model, which is potential for further development.

## 1. Introduction

Knowledge of soil permeability and its variations is needed in many environmental applications such as for rainfall-runoff modelling, irrigation systems, design of the earth dams [1] and flood management. The role of soil permeability capacity to flood is more significant than most people expect [2,3]. The higher the permeability of a soil layer, the faster water can infiltrate through to avoid flooding.

Variability in permeability properties is very high considering the directions into horizontal and vertical inside soil [4]. The permeability is normally determined from laboratory test that is laborious, time-consuming, and costly. Moreover, laboratory test could not represent the heterogeneity of soil permeability in the field [5]. Therefore, it is desirable to develop the prediction model. Several studies of empirical equations for predicting permeability have been undertaken by many researcher including [6] and [7]. These studies derived the models from areas which soil characteristics are different from Indonesian soil, which suggest a possibility that these permeability models are site-specific.

Since it is not practical to take into consideration every possible influential factor, we need to know what variables are closely related to permeability in the early stage. Such information can then be used to acquire reliable permeability prediction model. The purpose of this study is to identify which soil parameters corresponds strongly with soil permeability and propose a preliminary model for permeability prediction.

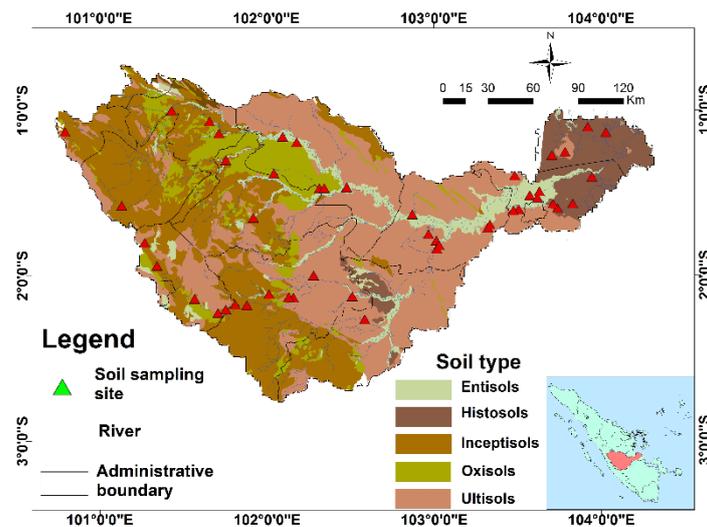
## 2. Materials and Methods

### 2.1. Soil sample collection

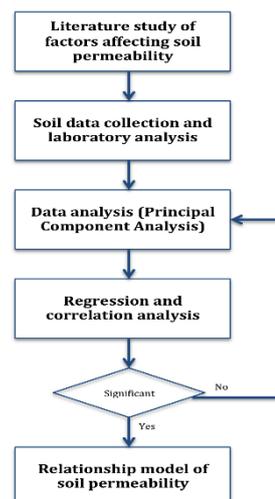
The study utilized a soil data set sampled from May 2015 to August 2016. The soil samples were collected from 37 sites consisting 91 samples in Batanghari Watershed, situated in Jambi and West



Sumatera Province. The Batanghari Watershed is a large river basin area (47.479 km<sup>2</sup>) dominated by forests, plantations, and agricultural land uses. An intensive conversion of land use from forest to agriculture encountered in this basin resulted in various water management related issues such as flooding and water availability. Therefore, the study of soil's ability to pass and store water indicated by the nature of soil permeability is essential in this area. Collected samples represented soil types and land uses from the upper, middle and lower part of the catchment area (Figure 1). The soil samples were examined and several parameters of soil properties were measured, which are water content (WC), bulk density (BD), particle density (PD), percentage of sand, percentage of silt, percentage of clay, percentage of fine clay, porosity, quick drainage, slow drainage, soil moisture retention (pF 1, pF 2, pF 2.54, pF 4.2), organic matter, water availability, and soil permeability itself.



**Figure 1.** Map of soil sampling sites in Batanghari Watershed, Sumatera Island.



**Figure 2.** Flow diagram for permeability analysis

## 2.2. Methodology

In general, the stages of this study are to identify and determine the factors related to the soil permeability value as shown in Figure 2. The analysis is done using Principal Component Analysis (PCA). After PCA Process, the next step is to perform regression analysis to search the equation for correlation between parameters.

Principal component analysis (PCA) is one of the multivariable analysis techniques performed to reduce a set of raw data into a number of principal components, which retain the most variance within the original data in order to identify possible patterns or clusters between objects and variables [8]. PCA estimates the correlation structure of the variables. The PCA will produce the value for factor loading and factor score. The number of these factors limited to the root of the eigenvalue = 1.

Correlations between selected variables were also obtained allowing identification of the most important parameters when characterizing soil permeability indicator most influence. The analysis was performed using Python 3.4 and the variables investigated were 16 parameters of soil properties. Steps on how to execute PCA on data set [8]:

- Gather the  $n$  samples of  $m$ -dimensional data  $x_1, \dots, x_n$ . Compute the mean  $\mu$  (Equation 1), Build the matrix  $B$  (Equation 2), and compute  $S$  (Equation 3).

$$\mu = \frac{1}{n}(x_1 + \dots + x_n) \quad (1)$$

$$B = |x_1 - \mu| \dots |x_n - \mu| \quad (2)$$

$$S = \frac{1}{n-1} BB^T \quad (3)$$

- Find the eigenvalue  $\lambda_1 \dots \lambda_m$  of  $S$  (arranged in decreasing order), as well as an orthogonal set of eigenvectors  $U_1 \dots U_m$
- Interpret the results:
  - If a small number of the  $\lambda_i$  is larger than all the others, so this indicates a dimension reduction is possible;
  - define  $n$  variables that are the most important in the first, second, etc of principal components; and
  - define factors that appear with the same or opposite value as others factors.

Regression analysis is used to model and analyze numerical data consisting of values of an independent variable  $X$  (the variable that we fix or choose deliberately) and dependent variable  $Y$ . Regression can be used to describe the nature and strength of the relationship between two continuous variables. Step to perform regression analysis [9]:

- At any point  $x_i$  the corresponding point on the line is given by

$$y_i = a + bx_i$$

Residual errors =

$$e_i = y_i - (a + bx_i)$$

Linear model =

$$y_i = a + bx_i + e_i, e_i \sim \text{normal}(0, \sigma^2)$$

- If the errors/residual are correlated or have unequal variance then least squares is not the best way to estimate the regression coefficient. Minimize the sum of squares ( $S$ ) of the vertical distances of the observations from the fitted line (residuals).

$$S = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n [y_i - (a + bx_i)]^2$$

- Find the intercept and regression coefficient that minimize:

- Partially differentiate with respect to the intercept and the regression coefficient, giving two equations
- Set these two equations equal to zero
- Solve the two simultaneous equations (sometimes referred to as normal equations) to obtain the estimates of the intercept and regression coefficients that minimize the sum of squares

Nash-Sutcliffe efficiency ( $E_{NS}$ ) was employed to evaluate the performances of model.  $E_{NS}$  is used to compare an estimated permeability to an observed one and represent the accuracy of prediction models [10].

$$E_{NS} = 1 - \frac{\sum_{i=1}^n (P_{sim}^i - P_{obs}^i)^2}{\sum_{i=1}^n (P_{sim}^i - \bar{P})^2} \quad (4)$$

Where  $P_{sim}$  is the predicted (cm/h),  $P_{obs}$  is the observed permeability (cm/h), and  $\bar{P}$  is the mean daily observed permeability. Ideally, the  $E_{NS}$  should be close to 1 to be considered as an efficient model with acceptable prediction accuracy.

### 3. Results and Discussions

Permeability expresses the ability of soil to pass water either laterally or horizontally. The rate of permeability, commonly expressed in unit of cm/h, is a function of various physical properties of the soil. PCA has generated five main factors out of sixteen permeability predictor variables, which are water content, bulk density, particle density, porosity, and pF 1 (soil water retention). PCA result is shown in Table 1, where the variable with eigenvalue  $> 1$  is identified as primary factors for soil permeability predictor. These five components explain 76.1% of data variance.

The variable that has the most significant effect on the permeability value is water content. It is commonly known that permeability changes significantly as water content changes. This is in line with a similar study conducted by [11] which states that changes in water content are very influential in the process of water movement in soil. The effect of degree of compaction in permeability that is shown by density value of soil is obvious from Table 1. Soil properties associated with this are bulk density (BD) and particle density (PD) that have a considerable eigenvalue and contribute to the variability of permeability value. Another factor that is closely related as a determinant of the permeability value is porosity, which indicates the presence of pore or space in the soil. It has been proven that porosity as one of the factors that needs to be involved in the function of permeability [12]. The fifth factor based on PCA which is the predictor variable for the soil permeability model is the pF 1 value. This property shows the condition of moisture content in the soil at 1 atm pressure. Moreover, by using five selected variables the model of soil permeability predictor equation is then obtained as follows:

$$\begin{aligned} \text{Soil permeability} & \quad (5) \\ & = 29.45 - 0.1877 \text{ water content} - 18.81 \text{ bulk density} \\ & + 2.03 \text{ particle density} + 0.019 \text{ porosity} - 0.1039 \text{ pF1} \end{aligned}$$

Predicted versus observed values of permeability were plotted to evaluate reliability of the equation. Plot showing the comparison is depicted in Figure 3.

The finding of this study to some extent different from many previous models. Most of prediction models show a strong relationship between grain size distribution and soil texture with permeability [13-16], while other studies show that permeability determined by soil compaction [17]. Among the five primary factors resulted from PCA, bulk density is the only similar parameter taken into account in soil characteristics model proposed by [18]. While our result stated that organic matter (OM) is insignificant for permeability estimation, model intended in [18] took into account OM as one of the variables used for determining moisture factor. Considering the heterogeneity of soil characteristics, no single model

would fit for all purposes. It becomes challenging to construct equations that can explain 100% variability of permeability values among different types of soil types [14]. Accordingly, more research is needed to find a reliable mathematical model for soil permeability.

### Acknowledgements

This study was part of the Priority Research Program of the Indonesian Institute of Sciences (LIPI), Sub-Program IV, entitled "Evaluation and Projection of Climate Change Impact on Flood Risk with High Precision for the Formulation of Flood Disaster Mitigation Strategy" (PI: Dr. Apip, M.Eng), funded by LIPI through the fiscal year of 2015-2017. We thanked the reviewers and editor of GCGE2017 for their comments and suggestions.

### References

- [1] Das B M 2008 *Advanced soil mechanics* (New York: Taylor & Francis)
- [2] Bonell M, Coles N, MacDonald A M, Auton, C A and Stevenson R 2013 Soil characteristics and landcover relationships on soil hydraulic conductivity at a hillslope scale: A view towards local flood management *J. Hydro* **497** 208-222.
- [3] Sampson S E The correlation between soil permeability and flooding in the northeast sector of the dog river watershed
- [4] Jabro J D 1992 Estimation of saturated hydraulic conductivity of soils from particle size distribution and bulk density data *Trans. ASAE* **35:2** 557-560
- [5] Holtz R D, Kovacks, W D and Sheahan T C 2011 *An introduction to Geotechnical Engineering* (Upper Saddle River, NJ: Prentice-Hall) p 853
- [6] Kenney T C, Lau D and Ofoegbu G I 1984 Permeability of compacted granular materials *Canadian Geo. J* **21** 726-729
- [7] Alyamani, M S and Sen Z 1993 Determination of hydraulic conductivity from grain size distribution curves *Gro. Wat.* **31** 551-555
- [8] Wold S, Esbensen, K and Geladi P, 1987 Principal component analysis *Chemo. Int. Lab. Sys.* **2(1-3)** 37-52
- [9] Rawlings, J O, Pantula, S G and Dickey, D A, 2001 *Applied regression analysis: a research tool* Springer Science & Business Media
- [10] Nash J and Sutcliffe J V 1970 River flow forecasting through conceptual models part I—A discussion of principles *J. Hydro* **10(3)** 282-290
- [11] Nishiyama N, Yokoyama, T and Takeuchi S, 2009 *AGU Fall Meet.* Relationship between water content and permeability of a rock: an example of Fontainebleau sandstone.
- [12] Urumović K and Urumović Sr K 2014 The effective porosity and grain size relations in permeability functions *Hydro. Earth. Sys. Sci. Disc.* **11(6)** 6675-6714
- [13] Horn M E 1971 Estimating soil permeability rates *J. Irrig. Dra. Div. ASCE* **97(IR2)** 263-274
- [14] Chen C Y, Bullen A G R, Elnaggar H A 1977 Permeability and related principles of coal refuse *Trans. Res. Rec.* **640** 49-52
- [15] Sinha S K and Wang M C 2008 Artificial neural network prediction models for soil compaction and permeability *Geotech. Geol. Eng.* **26(1)** 47-64
- [16] Onur E M 2014 Predicting the permeability of sandy soils from grain size distributions *Doctoral dissertation* Kent State University
- [17] Wang M C and Huang C C 1984 Soil compaction and permeability prediction models *J. Env. Eng.* **110(6)** 1063-1083
- [18] Saxton K E, Rawls W, Romberger J S and Papendick R I 1986 Estimating generalized soil-water characteristics from texture *SSSA. J.* **50(4)** 1031-1036