INTERPOLATION METHODS FOR GROUNDWATER QUALITY ASSESSMENT IN TANK CASCADE LANDSCAPE: A STUDY OF ULAGALLA CASCADE, SRI LANKA

KUMARI, M. K. N.[1,2*] – SAKAI, K.[3*] – KIMURA, S.[3] – NAKAMURA, S.[3] – YUGE, K.[4] – GUNARATHNA, M. H. J. P.[1,2] – RANAGALAGE, M.[5] – DUMINDA, D. M. S.[2]

[1]*United Graduate School of Agricultural Sciences, Kagoshima University, 1-21-24 Korimoto, Kagoshima-shi, Kagoshima 890-0065, Japan*

[2]*Faculty of Agriculture, Rajarata University of Sri Lanka, Puliyankulama, Anuradhapura 50000, Sri Lanka*

[3]*Faculty of Agriculture, University of the Ryukyus, 1 Senbaru, Nishihara-cho, Okinawa 903-0213, Japan*

[4]*Faculty of Agriculture, Saga University, 1 Honjo-machi, Saga 840-8502, Japan*

[5]*Faculty of Social Science and Humanities, Rajarata University of Sri Lanka, Mihintale, Anuradhapura 50000, Sri Lanka*

**Corresponding authors*
e-mail/phone/fax: nadeeta@gmail.com, +81-80-1787-5910 (M. K. N. Kumari); ksakai@agr.u-ryukyu.ac.jp, +81-89-895-8783, +81-89-895-8734 (K. Sakai)

(Received 22nd May 2018; accepted 11th Jul 2018)

Abstract. Interpolation methods are extensively used to map the spatial distribution of water quality parameters. However, the selection of the most appropriate method is a critical issue in environmental studies. The relative performances of deterministic and geostatistical methods in explaining the spatiotemporal variation of water quality parameters/indices in a tank cascade landscape were assessed. Inverse distance weighted (IDW), global polynomial interpolation (GPI), local polynomial interpolation (LPI), radial basis function (RBF), kriging (KR), and empirical Bayesian kriging (EBK) methods were evaluated using root mean square error (RMSE) in a leave-one-out cross-validation. Coefficient of variance, normality, level of autocorrelation, and extreme values near boundaries of the dataset showed a clear relationship with the relative performances of the different interpolation methods. Therefore, a clear understanding of the quality of the dataset is required in order to select the appropriate method to interpolate water quality parameters. EBK performed well for most parameters throughout the study period and is recommended as the best method to interpolate water quality parameters/indices in the Ulagalla cascade and other tank cascade landscapes in Sri Lanka and similar environments.
Keywords: *deterministic interpolation, empirical Bayesian kriging (EBK), geostatistical methods, root mean square error (RMSE), spatial variability*

Introduction

The tank cascade system (TCS) is a unique water storage and supply system used in the intermediate and dry zones of Sri Lanka. The system has been in use since the third century BCE (Madduma Bandara, 1985), mainly for irrigation and domestic water use. The main principle behind the TCS is re-use and recycling of water through a connected series of tanks. Hence, the TCS is defined as a connected series of tanks arranged within a micro- (or meso-) catchment of the dry zone landscape for storing, conveying and utilizing of water from an ephemeral rivulet (Madduma Bandara, 1985). As can be seen

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

in *Figure 1,* the major elements of TCS are categorized as meso-catchment, micro-catchment (catchment area of the individual tanks within the cascade), the main valley, side valleys and irrigated paddy lands. It has been recognized as globally important agricultural heritage site by Food and Agriculture Organization of the United Nations (FAO).

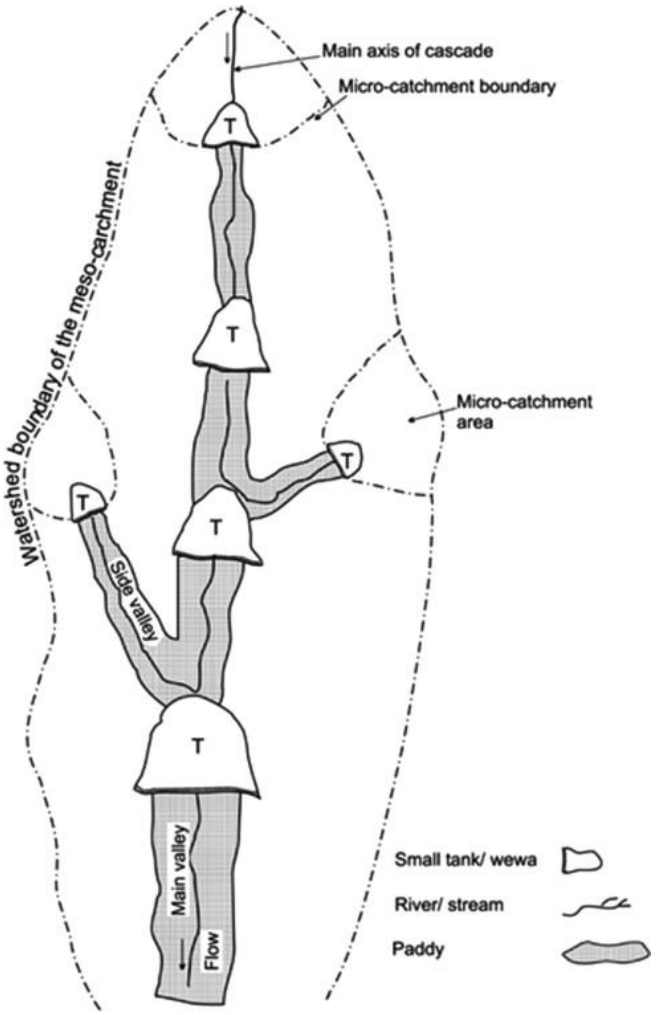

Figure 1. *Schematic representation of tank cascade system in dry zone of Sri Lanka (adapted from Panabokke et al., 2002)*

The dry zone of Sri Lanka receives an annual rainfall less than 1750 mm whereas, annual evaporation ranges from 1700-1900 mm, which implies the water stress condition during dry periods (Panabokke et al., 2002). This area is characterized with short rainy period (from September to January) which receives 80% of the total rainfall and long dry period (from February to October). As this area is dominated with reddish brown earth with low water retention capacity, the water scarcity problem is intensified in this area (Panabokke et al., 2002). This spatial and temporal variation of rainfall has led the ancient farming communities to invent TCS which can act as a sustainable water management system. TCS provides cooler micro-climate which enhances the plant and animal biodiversity while providing habitat for endangered elephants, resident and migrant water birds. Though it is not totally similar to Sri Lankan tank cascade systems,

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

comparable environments are in use for paddy irrigation in India (Bitterman et al., 2016; Van Meter et al., 2016; Bebermeier et al., 2017).

As a convenient substitute for insufficient surface water resources in the dry zone of Sri Lanka, groundwater use has dramatically increased during the last three decades, coinciding with changes in agriculture and livelihoods (Jayakody, 2006; Kumari et al., 2013). Moreover, a close relationship between the groundwater and surface water has been identified in tank cascade landscape (Bebermeier et al., 2017). Hence, the sustainability of the tank cascade landscape is endangered along with the overexploitation and quality deterioration of groundwater. Thus, an increasing amount of attention has been given to sustainably managing water resources in tank cascade landscapes, and several monitoring studies of groundwater in this landscape have been conducted (Wijesundara et al., 2012; Gunarathna et al., 2016a, b; Kumari et al., 2016). However, no appropriate continuous monitoring system has been put in place, because continuous monitoring of groundwater over a large area for an extended duration is expensive and labor intensive. Therefore, a suitable method for estimating groundwater availability and quality is needed that requires a minimum number of sampling sites in order to better manage the water resources in tank cascade landscapes.

Spatial interpolation, including deterministic and geostatistical interpolation techniques in ArcGIS, has been used to understand the spatial and temporal variation of natural resources, including groundwater, and related environmental concerns (Chai et al., 2011; Gunaalan et al., 2018). Deterministic interpolation techniques include inverse distance weighted (IDW), radial basis functions (RBFs), global polynomial interpolation (GPI), and local polynomial interpolation (LPI) methods; geostatistical interpolation techniques include kriging/co-kriging (ordinary kriging [OK], simple kriging [SK], universal kriging [UK], etc.), areal interpolation, and empirical Bayesian kriging (EBK). The ArcGIS Geostatistical Analyst extension can fill the gap between geostatistics and GIS analysis and has been used to characterize the spatial variability of variables in detail (Kumar et al., 2007; Uyan and Cay, 2013; Bao et al., 2014).

Interpolation accuracy is sensitive to the precise demarcation of boundaries and areas (Mirzaei and Sakizadeh, 2016; Gunaalan et al., 2018), the effectiveness of predicting parameters of unknown locations using known values, sample size (Stahl et al., 2006), spatial distribution of sampling sites (Güler, 2014), normality of the dataset (Wu et al., 2016), grid size or resolution (Hengl, 2007), and interpolation method (Luo et al., 2008; Xie et al., 2011). Moreover, if the distribution of sampling locations or wells does not appropriately represent the spatial variation of water quality parameters, any biases will be intensified (Heistermann and Kneis, 2011; Wagner et al., 2012). However, different interpolation methods tend to provide similar predictions at low (Mirzaei and Sakizadeh, 2016) and very high sampling densities (Gunnink and Burrough, 1996). In most cases, interpolation methods have been used without proper assessment of their accuracy. Only a few assessments of accuracy have been conducted. Mirzaei and Sakizadeh (2016) evaluated three interpolation methods to estimate a water quality index and found EBK to be the best method. Xie et al. (2011) stated that the best interpolation method to explain the spatial variation of heavy metals in soil varied with the size of the polluted area. Seyedmohammadi et al. (2016) compared five interpolation methods to estimate the spatial variation of electrical conductivity (EC) in groundwater and reported that OK was superior to the others. Based on the relative performance of four interpolation methods to interpolate EC, total dissolved solids (TDS), and pH, EBK was found as the best method (Gunarathna et al., 2016a, b).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

To date, no study has evaluated these interpolation methods with an extensive number of parameters covering all contaminant groups (anions, cations, nutrients) and water quality indices along with temporal effects. Because the assessment of spatial and temporal variation of groundwater is essential in sustainable management of water resources, the objective of this study was to describe and predict the relative performance of deterministic (IDW, LPI, GPI, and RBFs) and geostatistical (UK, OK, and EBK) interpolation methods and to select the best interpolation method to explain the spatial and temporal variation of groundwater quality in the Ulagalla cascade, Sri Lanka. Many physicochemical parameters were studied, including anions, cations, nutrients, and other water quality indices, as well as temporal effects. The relationships between characteristics of datasets and those of different interpolation methods were also examined.

Materials and methods

Study area

Ulagalla cascade covers approximately 51 km^2 in the Anuradhapura district of Sri Lanka (8°5'–8°14'N; 80°31'–80°34'E). The economy of this area is based on agriculture, which comprises tank-based paddy cultivation and rainfed or irrigated upland crop cultivation using groundwater. Mean annual rainfall in Anuradhapura is 1255 mm, and there is a distinct dry period from May to September. The monthly average maximum and minimum temperatures in the dry zone range from 25.0 to 37.7 °C and 17.4 to 26.8 °C, respectively. (Gunarathna and Kumari, 2013; Abeysekara and Punyawardena, 2016). A shallow regolith aquifer of the hard rock region is the main aquifer type in the study area. Groundwater potential is comparatively limited because of the low groundwater storage capacity and transmissivity of the underlying crystalline basement (Sirimanne, 1952).

Data collection and data preparation

The total cascade area was divided into 1-km^2 cells, and one agro-well was purposely selected to represent each cell so as to evaluate the quality of groundwater in the study area. Within the existence of agro-wells and the availability of water in the agro-wells throughout the study period, a total of 29 wells were selected (*Fig. 2)*. Three replicates from aforesaid 29 wells were collected monthly from April 2016 to March 2017 to measure the water quality parameters using standard procedures (APHA, 2005). All the chemical analyses were carried out at the laboratory of soil and water sciences, Department of Agricultural Engineering and Soil Science, Faculty of Agriculture, Rajarata University of Sri Lanka. Sodium adsorption ratio (SAR) (Wilcox, 1955) and total hardness (TH) (Todd and Mays, 2005) were calculated from measured data. The following 12 water quality parameters or indices were used: electrical conductivity (EC); pH; concentrations of sodium (Na^+), potassium (K^+), calcium (Ca^{2+}), magnesium (Mg^{2+}), chloride (Cl^-), nitrate (NO_3^--N), phosphate (PO_4^{3-}), and bicarbonate (HCO_3^-); sodium adsorption ratio (SAR); and total hardness (TH).

Attribute data containing information about the physicochemical parameters/indices were joined with geographic coordinates obtained with a handheld global positioning system (GPS) receiver (eXplorist 510, Magellan, USA) of each sampling point (*Table A1* in the *Appendix*). ArcGIS 10.2 (ESRI, California, USA) and R statistical

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

software (R Foundation for Statistical Computing, Vienna, Austria) (Team R, 2016) were used for the interpolation analysis and statistical analysis, respectively.

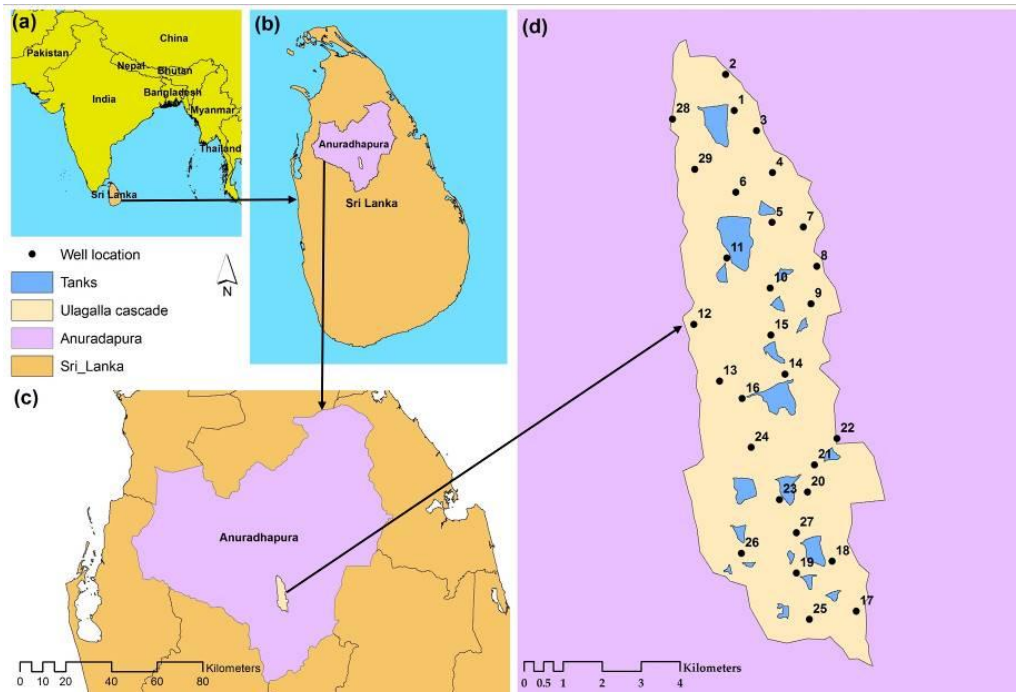

Figure 2. *Location of the study areas: (a) South Asia; (b) Sri Lanka; (c) Anuradhapura district; (d) groundwater sampling locations in Ulagalla cascade*

Interpolation methods

Inverse distance weighted (IDW)

In IDW, the interpolation weights are calculated as a function of the observed sampling point and the prediction point (Gunnink and Burrough, 1996). The accuracy of IDW depends on the number of closest neighboring sampling points (Yao et al., 2013). The values for unknown points are estimated with *Equation 1:*

$$Z(x_0) = \frac{\sum_{i=1}^{n} \frac{x_i}{h_{ij}^{\beta}}}{\sum_{i=1}^{n} \frac{1}{h_{ij}^{\beta}}} \qquad \text{(Eq.1)}$$

where $Z(x_0)$ is the interpolated value, x_i is the i^{th} data value, β is the user-defined exponent for weighting, n is the total number of sampling data values and h_{ij} is the distance between the known point and the unknown point (Seyedmohammadi et al., 2016).

Global polynomial interpolation (GPI)

The GPI method positions a plane between sample points by fitting a polynomial formula to the points. Using a value on the plane that relates to the prediction location,

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

the unknown point is determined by minimizing the errors (Webster and Oliver, 2008) With the use of low order polynomials GPI creates slowly while describing the physical processes. However, with complex polynomials, it is difficult to ascribe physical meaning to GPI (Johnston et al., 2003).

Local polynomial interpolation (LPI)

Whereas GPI fits one polynomial to the entire surface, LPI fits many polynomials, each within specified overlapping local neighborhoods. Although this method produces smooth surfaces, it is best suited for use only with data that have a narrow range of variation. LPI creates a surface from many different polynomial formulas, each of which is optimized for a specified neighborhood, neighborhood shape, and maximum and minimum number of points. LPI is sensitive to the neighborhood distance, and the sample points in a neighborhood can be weighted by their distance from the prediction location. Because LPI is sensitive to neighborhood distance and a small search neighborhood may create empty areas in the prediction surface, the method shows better results with grid-based sampling data than with random point sampling (Johnston et al., 2003; Hani et al., 2011).

Radial basis functions (RBFs)

RBFs are a form of artificial neural networks with a series of exact interpolation techniques. They use an equation derived from the distance between an interpolated point and the sampling points (Lin and Chen, 2004; Aguilar et al., 2005). The method consists of five deterministic interpolation techniques: thin plate spline, spline with tension, completely regularized spline, multi-quadratic function, and inverse multi-quadratic function. The RBF method is used mainly to create smooth surfaces from a large number of data points. Although RBFs give good results for areas with gently varying surfaces, the method will not provide accurate results if there are any large variations in the surface within a short horizontal distance (Johnston et al., 2003). The most commonly used RBF technique, completely regularized spline was used for this analysis.

Kriging

Kriging is a linear interpolation method that assumes that the parameter to be interpolated can be modeled by random processes with spatial autocorrelation. Hence, kriging techniques are widely used to describe and model spatial patterns and predict values at unmeasured locations. Three types of kriging were evaluated in this study: ordinary kriging, universal kriging and empirical Bayesian kriging.

Ordinary kriging (OK)

OK is the most widely used kriging method. It uses an average of a subset of neighboring points to produce a particular interpolation point. OK can use either semivariograms or covariances to explain the autocorrelation and can use transformations to avoid trends (Johnston et al., 2003), but the semivariance function plays a major role in deriving weights of OK (Johnston et al., 2003). The empirical semivariance function can be used to estimate the parameters of the semivariogram function and the nugget effect as expressed in *Equation 2:*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad \text{(Eq.2)}$$

where γ is the semivariance, $N(h)$ is the number of data pairs within a given class of distance and direction, h is the lag distance, and $Z(x_i)$ and $Z(x_i + h)$ are the sample values at two points separated by the distance interval h (Xie et al., 2011).

Universal kriging (UK)

UK can be used to produce prediction, quantile, probability, or standard error maps. The method is used to estimate the spatial means when the data have a strong trend, and the trend is modeled using simple functions. The use of UK is limited to large surfaces, such as a large country, because it is difficult to follow a trend along the direction of spreading (Kis, 2016).

Empirical bayesian kriging (EBK)

EBK is different from other classical kriging methods because the parameters are automatically optimized using a number of semivariogram models instead of a single semivariogram. The following steps are used in EBK: (1) A semivariogram model is estimated using available data. (2) A new value is simulated for each input data location using the semivariogram model. (3) Based on the simulated data, a new semivariogram model is estimated. Bayes's rule is then used to calculate the weight of the new semivariogram model. By repeating the steps 2 and 3, the semivariogram estimated in step 1 is used to simulate a new set of values at the input locations (Krivoruchko, 2012).

Data preprocessing

Because the Kriging methods require the sample distribution to be normal, Shapiro Wilk test was performed for all 144 datasets (12 water quality parameters/indices × 12 months) to check the goodness-of-fit of the data ($P < 0.05$). The results showed that K^+, Mg^{2+}, NO_3^--N, Cl^- and EC were not normally distributed at any time, and the other parameters and indices were normally distributed only during several months. Hence, datasets that were not normally distributed were log-transformed and thereafter except very few, all the other data sets were normally distributed.

Validation and model evaluation

Cross-validation and validation with an independent dataset are the most common methods used to compare different interpolation methods, whereby the data are divided into a training set and a validation set. The validation set is used to test the model acquired from the training set. Those allowed us to assess the goodness-of-fit of interpolation methods and the appropriateness of the neighborhood (Dashtpagerdi et al., 2013; Gunarathna et al., 2016a, b). Because the number of sampling points was limited, we used leave-one-out cross-validation (Gunarathna et al., 2016a, b) to estimate the spatial variation of water quality parameters/indices in the study area, removing one data point from the known dataset and estimating its value from the other known values. If a model has a standardized mean error close to 0, the RMSE and average standard error are as small as possible as compared with other models, which means the model

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

provides the most accurate predictions. Hence, we used RMSE (*Eq. 3*) to compare the models:

$$RMSE = \sqrt{\frac{1}{n}\left(Z_i - \hat{Z}_i\right)^2} \quad \text{(Eq.3)}$$

where $\hat{Z}_i$ is the estimated value, Z_i is the measured value at sampling point i (i = 1,.., n), and n is the total number of observations.

The coefficient of variance (CV), which is the ratio of standard deviation to the mean of each parameter/index, was used to study the relative variability of the dataset. We used the local Moran's Index (MI), one of the most commonly used criteria for spatial autocorrelation of quantitative data (Moran, 1950), to estimate the level of spatial autocorrelation of water quality parameters/indices in the Ulagalla cascade.

Results and discussion

Relative performance of deterministic and geostatistical interpolation methods

The RMSE values of cross-validation for the 12 water quality parameters/indices during 12 consecutive months are summarized in *Tables A2–A13* in the *Appendix*. Note that the RMSE values of the OK and UK interpolation methods were similar to each other for all the parameters/indices, and are considered together as kriging (KR). EBK was superior to all other interpolation methods in estimating spatial variation of K^+, Mg^{2+}, NO_3^--N, and EC in all 12 months (*Tables 1, A4, A6, A7* and *A11*); of Na^+, HCO_3^-, Cl^-, and TH in 11 months (*Tables 1, A3, A9, A10* and *A13*); of SAR in 10 months and of pH and PO_4^{3-} in 7 months (*Tables 1, A2, A12, A8*). EBK was outperformed in only 6 months in the interpolation of Ca^{2+} (*Tables 1* and *A5*). Overall, EBK was the best method for interpolating groundwater quality parameters/indices in 121 out of the 144 incidences.

Table 1. *Summary of selected best interpolation methods for different parameters/indices during the study period*

	pH	Na^+	K^+	Ca^{2+}	Mg^{2+}	NO_3^--N	PO_4^{3-}	HCO_3^-	Cl^-	EC	SAR	TH
Apr	EBK	EBK	EBK	GPI	EBK	EBK	EBK	EBK	KR	EBK	RBF	EBK
May	EBK	EBK	EBK	EBK	EBK	EBK	KR	EBK	EBK	EBK	KR	EBK
Jun	EBK	EBK	EBK	EBK	EBK	EBK	EBK	EBK	EBK	EBK	EBK	EBK
Jul	IDW	EBK	EBK	EBK	EBK	EBK	RBF	EBK	EBK	EBK	EBK	EBK
Aug	GPI	EBK	EBK	EBK	EBK	EBK	RBF	EBK	EBK	EBK	EBK	EBK
Sep	EBK	EBK	EBK	GPI	EBK	EBK	KR	EBK	EBK	EBK	EBK	EBK
Oct	EBK	EBK	EBK	LPI	EBK	EBK	EBK	EBK	EBK	EBK	EBK	LPI
Nov	EBK	IDW	EBK	GPI	EBK	EBK	RBF	EBK	EBK	EBK	EBK	EBK
Dec	GPI	EBK	EBK	EBK	EBK	EBK	RBF	KR	EBK	EBK	EBK	EBK
Jan	LPI	EBK	EBK	GPI	EBK	EBK	EBK	EBK	EBK	EBK	GPI	EBK
Feb	LPI	EBK	EBK	LPI	EBK	EBK	EBK	EBK	EBK	EBK	RBF	EBK
Mar	EBK	EBK	EBK	EBK	EBK	EBK	RBF	EBK	EBK	EBK	RBF	EBK

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

Based on the number of success incidences obtained from the cross-validation results, the interpolation methods can be sorted, as EBK > LPI > GPI > IDW > KR > RBF (*Table 1*). This ranking supports the findings of Gunarathna et al. (2016a, b) and Mirzaei and Sakizadeh (2016), who also found EBK to be superior to other interpolation methods with the use of a limited number of variables. Even though EBK recorded the lowest RMSE value for most of the parameters/indices that was not the case with Ca^{2+} and PO_4^{3-}, for which several methods had the lowest RMSE during different months (*Table 1*). Hence, we selected Ca^{2+} (*Table 2*) and PO_4^{3-} (*Table 3*) to elucidate differences in the methods.

As inexact interpolators among the selected methods, GPI and LPI showed quite similar results compared to other methods used in the present study and this was supported by Wang et al. (2014). Xiao et al. (2016) also confirmed that GPI is suitable only when the variability of the dataset is relatively small. Although GPI can be used to analyze the surface trend of regionalized variables, it is not accurate when extreme values are present (Mutuna and Kurima, 2012; Wang et al., 2014). LPI is also capable of simulating a narrow range of variability with high accuracy (Xiao et al., 2016). GPI was ranked in the top three methods during 10 months of the study period, and the CV was relatively small in 6 of those 10 months (*Table 2*). Moreover, no extreme values were recorded near the boundaries in those 6 months. In the other 4 months, the variation of the dataset was moderate, and no extreme values were recorded near the boundaries (*Table 2*). According to the summary statistics of phosphate concentration (*Table 3*), the conditions of low variation and no extreme values near boundaries were met in only 3 months, and GPI was ranked among the top three only once.

Table 2. *Summary statistics of calcium concentration in Ulagalla cascade*

	Apr	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**	**Jan**	**Feb**	**Mar**
Mean	*182.9*	*38.2*	*63.6*	*70.4*	*35.3*	*49.7*	*30.1*	*35.0*	*92.2*	*86.4*	*84.5*	*37.3*
Standard deviation	*89.7*	*28.4*	*30.3*	*47.2*	*19.5*	*25.1*	*20.6*	*17.5*	*44.0*	*36.5*	*53.8*	*23.1*
CV	49	74	48	67	55	50	69	50	48	42	64	62
Moran's Index (MI)	−0.19	−0.12	−0.39	0.04	−0.18	−0.30	0.13	−0.26	−0.26	−0.22	−0.03	−0.07
P-value of MI	0.31	0.57	0.02	0.60	0.32	0.08	0.22	0.13	0.12	0.22	0.96	0.79
Skewness (original data)	*0.40*	*1.28*	*0.62*	*0.91*	*1.45*	*0.84*	*1.99*	*0.81*	*1.22*	*0.22*	*1.62*	*1.66*
Skewness (after log transformation)	0.4	0.07	0.62	−0.18	0.33	0.84	−0.19	0.81	−1.04	0.22	−0.43	−0.07
Range	*350*	*114*	*117*	*161*	*85*	*106*	*100*	*80*	*226*	*129*	*264*	*100*
Lowest RMSE	**GPI**	**EBK**	**EBK**	**EBK**	**EBK**	**GPI**	**LPI**	**GPI**	**EBK**	**GPI**	**LPI**	**EBK**
2nd lowest RMSE	**LPI**	**KR**	**KR**	**GPI**	**KR**	**EBK**	**EBK**	**LPI**	**GPI**	**EBK**	**EBK**	**KR**
3rd lowest RMSE	**KR**	**LPI**	**GPI**	**LPI**	**GPI**	**KR**	**RBF**	**EBK**	**LPI**	**LPI**	**GPI**	**GPI**
4th lowest RMSE	EBK	GPI	LPI	KR	LPI	LPI	IDW	KR	KR	KR	KR	LPI
5th lowest RMSE	IDW	RBF	IDW	IDW	RBF	IDW	KR	IDW	IDW	IDW	IDW	RBF
6th lowest RMSE	RBF	IDW	RBF	RBF	IDW	RBF	GPI	RBF	RBF	RBF	RBF	IDW
Well numbers with extreme values		1		10, 14	1	10	27, 28	10	10		10	1, 10

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

Table 3. *Summary statistics of phosphate concentration in Ulagalla cascade*

	Apr	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**	**Jan**	**Feb**	**Mar**
Mean	0.04	0.05	0.06	0.57	0.58	0.06	0.11	0.16	0.09	0.04	0.32	0.70
Standard deviation	0.04	0.06	0.05	0.24	0.27	0.08	0.16	0.15	0.12	0.05	0.06	0.78
CV	97	119	87	42	47	135	139	91	134	106	20	111
Moran's Index (MI)	0.33	0.26	0.37	0.43	0.36	0.25	0.17	0.26	0.25	0.23	0.03	0.20
P-value of MI	0.00	0.02	0.00	0.00	0.00	0.03	0.14	0.04	0.04	0.06	0.61	0.09
Skewness (original data)	*2.24*	*2.88*	*2.42*	*1.93*	*1.89*	*2.42*	*2.20*	*2.01*	*2.05*	*2.01*	*−2.06*	*2.72*
Skewness (after log transformation)	0.80	0.78	0.23	0.97	0.49	0.80	0.53	1.24	0.49	0.67	−2.55	1.00
Range	0.15	0.26	0.26	0.99	1.30	0.34	0.63	0.57	0.44	0.17	0.26	3.69
Lowest RMSE	**IDW**	**RBF**	**RBF**	**EBK**	**RBF**	**RBF**	**EBK**	**EBK**	**EBK**	**EBK**	**EBK**	**EBK**
2nd lowest RMSE	**EBK**	**KR**	**EBK**	**KR**	**EBK**	**KR**	**RBF**	**GPI**	**RBF**	**RBF**	**LPI**	**LPI**
3rd lowest RMSE	**RBF**	**LPI**	**IDW**	**RBF**	**KR**	**LPI**	**IDW**	**KR**	**KR**	**KR**	**GPI**	**GPI**
4th lowest RMSE	LPI	EBK	KR	IDW	IDW	EBK	LPI	LPI	IDW	IDW	IDW	KR
5th lowest RMSE	GPI	IDW	GPI	LPI	LPI	IDW	GPI	RBF	LPI	GPI	RBF	RBF
6th lowest RMSE	KR	GPI	LPI	GPI	GPI	GPI	KR	IDW	GPI	LPI	KR	IDW
Well numbers with extreme values	14, 16, 12	14, 16	2, 14	14, 16	1, 2, 14, 15, 16	1, 2, 11, 14, 15, 16	14, 16, 11	11, 14, 16, 25	14, 16, 11	14, 11	14, 16, 11	16, 15, 11

IDW is widely used in the field of environmental sciences, but it is rarely recommended as the best interpolation method in comparison studies (Li and Heap, 2011). In their review, Li and Heap (2011) reported that IDW is highly sensitive to sample density and data variation (CV). The poor performance of IDW in our study could be attributed to the high spatial data variation and relatively low sample size. In the classification using Ca^{2+} and PO_4^{3-}, IDW was never the best-fit model when the CV was high. Therefore, we do not recommend the use of IDW to interpolate spatial variation of groundwater quality parameters/indices in tank cascade landscapes unless the data show low spatial variation and have a higher sample density with an evenly distributed sampling pattern.

Kriging (KR) is one of the most widely used interpolation methods in the field of environmental sciences, and it has been recommended in comparison studies (Li and Heap, 2011). Basic assumptions of KR are spatially autocorrelated observations (a function of distance between observations) and normally distributed data (Zimmerman et al., 1999). Therefore, KR has a strong ability to predict the overall trend of groundwater contamination when a dataset is autocorrelated (Ahmed, 2002; Nas, 2009; Xie et al., 2011; Mutuna and Kurima, 2012; Zehtabian et al., 2013). As shown in *Table 3*, KR was ranked on first three during 6 months when the data was significantly autocorrelated in 8 months and was in the top three in 5 of those 8 months. This characteristic was also observed in the Ca^{2+} dataset where KR performed well in June when the data were autocorrelated.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

Because EBK divides the dataset into subsets and simulates a semivariogram for each subset to reduce the uncertainty relative to that in KR, it provides relatively better interpolation accuracy on small datasets and non-stationary datasets than KR (ESRI, 2015). In *Tables 2* and *3*, EBK was listed in the top three in 21 of 24 incidences. This shows that EBK performs well irrespective of CV, MI, extreme values near boundaries (EVnB), and skewness (SK). However, when the dataset was less variable, showed autocorrelation, and had no EVnB, other interpolation methods performed better than EBK. Hence, we recommend using EBK when data show high variability and no autocorrelation, and extreme values near boundaries and are not normally distributed.

As the quality of the dataset determines the accuracy of different interpolation methods, we assessed the importance of CV, MI, SK, and EVnB (characteristics that show the quality of a dataset) to the success of each method using an attribute evaluation option available in the CORElearn package (Robnik-Sikonja and Savicky, 2017) of R software (*Table 4*). The relative importance of the characteristics can be sorted as EVnB > CV > SK > MI for GPI and EVnB > SK > CV > MI for LPI. These results confirm the sensitivity of GPI and LPI to CV and EVnB. The relative importance for RBF was CV > EVnB > SK > MI, demonstrating that RBF is sensitive to the variability of the dataset and EVnB. The relative failure of RBF in this study can be explained by the high spatial variation of the water quality parameters/indices. The relative importance for IDW was EVnB > CV > MI > SK, confirming its sensitivity to dataset variability and extreme values. The relative importance for KR was EVnB > MI > CV > SK, confirming that KR can be successfully used when the dataset is autocorrelated with low variability while lacking extreme values near boundaries.

Table 4. *Relative importance of attributes on different interpolation methods*

	Relative importance					
Attributes	**GPI**	**LPI**	**RBF**	**IDW**	**KR**	**EBK**
CV	0.013	0.036	0.181	0.274	0.013	0.025
MI	−0.031	0.010	0.010	0.243	0.052	0.009
SK	−0.02	0.081	0.050	0.730	0.001	−0.013
EVnB	0.262	0.093	0.077	0.422	0.218	0.154

CV – coefficient of variance, MI – Moran's Index, SK – Skeweness, EVnB – extreme values near boundary

Distribution pattern of observed and simulated data using EBK method

Figure 3 shows boxplot diagrams of the observed and EBK-predicted values of the 12 parameters/indices. The measured and predicted values of almost all the parameters/indices were right skewed except pH, showing that the majority of the values are clustered below the median and the means are greater than the median (*Table A14*). Further, it can be noted that EBK was unable to properly predict extreme values, but there were no significant differences between observed and predicted values. The final spatial distribution maps of water quality parameters/indices prepared using the EBK interpolation method for the mean values of the 12 parameters/indices are shown in *Figure 4*. It was observed that the concentration of most parameters/indices were comparatively low at the upper part of the cascade and it has increased at the lower part of the cascade due to the accumulation effect.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

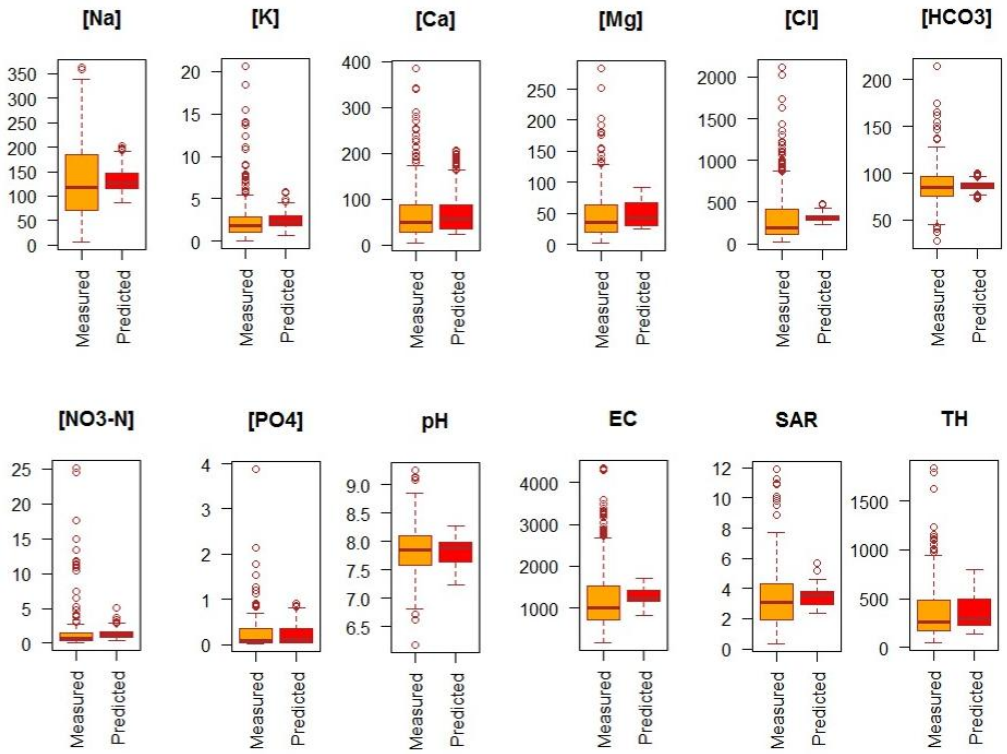

Figure 3. *Boxplot diagrams of observed and EBK-predicted values of (a) Na^+ (mg/l) (b) K^+ mg/l), (c) Ca^{2+}(mg/l), (d) Mg^{2+} (mg/l), (e) Cl^- (mg/l), (f)HCO_3^- (mg/l), (g) NO_3^--N (mg/l), (h) PO_4^{3-} (mg/l), (i) pH, (j) EC (µS/cm), (k) SAR and (l) TH (mg/l) in Ulagalla cascade*

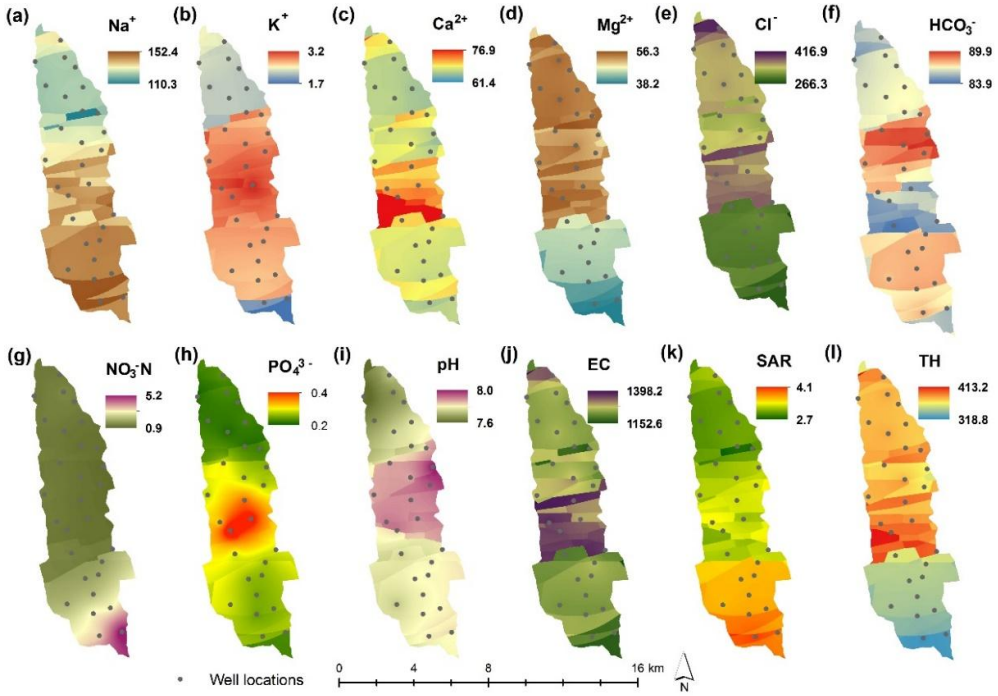

Figure 4. *Spatial distribution of (a) Na^+ (mg/l), (b) K^+ (mg/l), (c) Ca^{2+}(mg/l), (d) Mg^{2+} (mg/l), (e) Cl^- (mg/l), (f)HCO_3^- (mg/l), (g) NO_3^--N (mg/l), (h) PO_4^{3-} (mg/l), (i) pH, (j) EC (µS/cm), (k) SAR, and (l) TH (mg/l) in Ulagalla cascade*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

Conclusions

A clear understanding of spatial and temporal variation in water quality parameters/indices is a key issue in agriculture as well as in environmental studies. At present, many algorithms are used with the aim of selecting the best interpolation method for delineation of the spatial distribution of water quality parameters/indices. We investigated the interpolation accuracy of a variety of methods in a tank cascade landscape.

The EBK method proved superior to deterministic and other geostatistical methods in interpolating groundwater quality parameters (anions, cations, and nutrients) and indices associated with tank cascade landscapes. Kriging interpolation was successful when the dataset was autocorrelated with low variability. IDW had the worst results in estimating the spatial distribution of water quality parameters/indices. Better performance was obtained with the GPI and LPI methods when the dataset was less variable and had no extreme values near boundaries.

Because groundwater monitoring is labor intensive and expensive, it is important to use optimum sampling density and to choose the design in a methodical way. Furthermore, it is advisable to decide on the interpolation method before sampling and then schedule sample density and design accordingly. This study can be used as a guide for such decision making for groundwater monitoring in a tank cascade landscape.

In general, the preparation of composite water quality zonation map for the Ulagalla cascade with the integration of EBK method and water quality indices/parameters can be carried out. Future research can be conducted to find out the optimum number of sampling points to obtain precise estimation of water quality in tank cascade landscape.

Acknowledgements. This work was financially supported by Research, Publication and Higher Degrees Committee, Rajarata University of Sri Lanka, Mihintale, Sri Lanka, under the grant number RJT/RP&HDC/2016/Agri/R/3.

REFERENCES

[1] Abeysekara, A. B., Punyawardena, B. V. R. (2016): Potential and Constraints of Climate for Groundwater Management in the Dry Zone of Sri Lanka. – In: Pathmarajah, S. (ed.) Groundwater Availability and Use in the dry Zone of Sri Lanka. Cap-Net Lanaka, Postgraduate Institute of Agriculture, University of Peradeniya, Sri Lanka, pp.1–32.

[2] Aguilar, F. J., Aguera, F., Aguilar, M. A., Carvajal, F. (2005): Effects of terrain morphology, sampling density, and interpolation methods on grid DEM accuracy. – Photogrammetric Engineering & Remote Sensing 71: 805–816. doi: 10.14358/PERS.71.7.805.

[3] Ahmed, S. (2002): Groundwater Monitoring Network Design: Application of Geostatistics with a Few Case Studies from a Granitic Aquifer in a Semiarid Region. – In: Sherif, M. M., Singh, V. P., Al-Rashed, M. (ed.) Groundwater Hydrology, vol. 2, pp. 37–57. Balkema, Rotterdam.

[4] APHA (2005): Standard Methods for the Examination of Water and Wastewater, Standard Methods. – American Public Health Association/American Water Works Association/Water Environment Federation, Washington D. C.

[5] Bao, Z., Wu, W., Liu, H., Chen, H., Yin, S. (2014): Impact of long-term irrigation with sewage on heavy metals in soils, crops and groundwater. A case study in Beijing. – Polish Journal of Environmental Studies 23(2): 309–318.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

[6] Bebermeier, W., Meister, J., Withanachchi, C. R., Middelhaufe, I., Middlehaufe, B. (2017): Tank cascade systems as a sustainable measure of watershed management in South Asia. – Water (Switzerland) 9(3): 1–16. DOI: 10.3390/w9030231.
[7] Bitterman, P., Tate, E., Van Meter, K. J., Basu, M. B. (2016): Water security and rainwater harvesting: A conceptual framework and candidate indicators – Applied Geology 76: 75–84. DOI: 10.1016/j.apgeog.2016.09.013.
[8] Chai, H., Cheng, W., Zhou, C., Chen, X., Ma, X., Zhao, S. (2011): Analysis and comparison of spatial interpolation methods for temperature data in Xinjiang Uygur Autonomous Region, China. – Natural Science 3(12): 999–1010. doi: 10.4236/ns.2011.312125.
[9] Dashtpagerdi, M. M., Vagharfard, H., Honarbakhsh, A. (2013): Application of cross-validation technique for zoning of groundwater levels in Shahrekord plain. – Agricultural Sciences 4(7): 329–333.
[10] ESRI (2015): What is empirical bayesian kriging. – http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/geostatistical-analyst/what-is-empirical-bayesian-kriging-.htm. Accessed on 12.12.2017.
[11] Güler, M. (2014): Comparison of different interpolation techniques for modelling temperatures in Middle Black Sea region. – Journal of Agricultural Faculty of Gaziosmanpasa University 31(2): 61–71. DOI: 10.13002/jafag714.
[12] Gunaalan, K., Ranagalage, M., Gunarathna, M. H. J. P., Kumari, M. K. N., Vithanage, M., Saravanan, S., Warnasuriya, T. W. S. (2018): Application of geospatial techniques for groundwater quality and availability assessment : A case study in Jaffna Peninsula, Sri Lanka. – International Journal of Geo-Information 7: 20. DOI: 10.3390/ijgi7010020.
[13] Gunarathna, M. H. J. P., Kumari, M. K. N. (2013): Rainfall trends in Anuradhapura : Rainfall analysis for agricultural planning. – Rajarata University Journal 1: 38–44.
[14] Gunarathna, M. H. J. P., Kumari, M. K. N., Nirmanee, K. G. S. (2016a): Evaluation of interpolation methods for mapping pH of groundwater. – International Journal of Latest Technology in Engineering, Management and Applied Science 5(3): 1–5.
[15] Gunarathna, M. H. J. P., Nirmanee, K. G. S., Kumari, M. K. N. (2016b): Are geostatistical interpolation methods better than deterministic interpolation methods in mapping salinity of groundwater? – International Journal of Research and Innovations in Earth Sciences 3(3): 59–64.
[16] Gunnink, J. L., Burrough, P. A. (1996): Interactive Spatial Analysis of Soil Attribute Patterns Using Exploratory Data Analysis (EDA) and GIS. – In: Masser, I., Salge, F. (ed.) Spatial Analytical Perspectives on GIS. Taylor & Francis, New York.
[17] Hani, A., Ali, S., Abari, H. (2011): Determination of Cd, Zn, K, pH, TNV, organic material and electrical conductivity (EC) distribution in agricultural soils using geostatistics and GIS (Case Study : South- Western of Natanz-Iran). – World Academy of Science, Engineering and Technology 5(12): 852–855.
[18] Heistermann, M., Kneis, D. (2011): Benchmarking quantitative precipitation estimation by conceptual rainfall-runoff modeling. – Water Resources Research 47(6): 1–23. DOI: 10.1029/2010WR009153.
[19] Hengl, T. (2007): A Practical Guide to Geostatistical Mapping of Environmental Variables. – Office for Official Publication of the European Communities, Luxembourg.
[20] Jayakody, A. N. (2006): Large diameter shallow agro-wells. A national asset or a burden for the nation? – The Journal of Agricultural Science 2(1): 1–10.
[21] Johnston, K., Ver Hoef, J., Krivoruchko, K., Lucas, N. (2003): Using ArcGIS Geostatistical Analyst. – ESRI press, Redlands, California.
[22] Kis, I. M. (2016): Comparison of ordinary and universal kriging interpolation techniques on a depth variable (a case study of the Šandrovac Field). – The Mining-Geology-Petroleum Engineering Bulletin: 41–58. DOI: 10.17794/rgn.2016.2.4.
[23] Krivoruchko, K. (2012): Empirical Bayesian Kriging - Implemented in ArcGIS Geostatistical Analyst. – ESRI Press, Redlands, CA.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

[24] Kumar, A., Maroju, S., Bhat, A. (2007): Application of ArcGIS geostatistical analyst for interpolating environmental data from observations. – Environmental Progress 26(3): 220–225. DOI: 10.1002/ep. 10223.
[25] Kumari, M. K. N., Pathmarajah, S., Dayawansa, N. D. K. (2013): Characterization of agro-well water in Malwathu Oya cascade-I in Anuradhapura district of Sri Lanka. – Tropical Agricultural Research 25(1): 46–55.
[26] Kumari, M. K. N., Pathmarajah, S., Dayawansa, N. D. K., Nirmanee, K. G. S. (2016): Evaluation of groundwater quality for irrigation in Malwathu Oya cascade-I in Anuradhapura district of Sri Lanka. – Tropical Agricultural Research 27(4): 310–324.
[27] Li, J., Heap, A. D. (2011): A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. – Ecological Informatics 6: 228–241. DOI: 10.1016/j.ecoinf.2010.12.003.
[28] Lin, G., Chen, L. (2004): A spatial interpolation method based on radial basis function networks incorporating a semivariogram model. – Journal of Hydrology 288: 288–298. DOI: 10.1016/j.jhydrol.2003.10.008.
[29] Luo, W., Taylor, M., Paker, S. (2008): A comparison of spatial interpolation methods to estimate continous wind speed surfaces using irregularly distributed data from England and Wales. – International Journal of Climatology 28: 947–956. DOI: 10.1002/joc.1583.
[30] Madduma Bandara, C. M. (1985): Catchment Ecosystem and Village Tank Cascade in the Dry Zone of Sri Lanka: A time-Tested System of Land and Water Resources Management. – In: Lundqvist, J., Lohm, U., Falkenmark, M (ed.) Strategies for River Basin Management. Linkoping, Sweden.
[31] Mirzaei, R., Sakizadeh, M. (2016): Comparison of interpolation methods for the estimation of groundwater contamination in Andimeshk-Shush Plain, Southwest of Iran. – Environmental Science and Pollution Research 23(3): 2758–2769. DOI: 10.1007/s11356-015-5507-2.
[32] Moran, P. A. P. (1950): Notes on Continuous Stochastic Phenomena. – Biometrika 37: 17–23. DOI: 10.2307/2332142.
[33] Mutuna, F., Kurima, D. (2012): A comparison of spatial rainfall estimation techniques: A case study of Nyando river basin Kenya. – Journal of Agriculture, Science and Technology 14(2): 149–165.
[34] Nas, B. (2009): Geostatistical approach to assessment of spatial distribution of groundwater quality. – Polish J. of Environ. Stud. 18(6): 1073–1082.
[35] Panabokke, C. R., Sakthivadivel, R., Weerasinghe, A. D. (2002): Small Tanks in Sri Lanka: Evolution, Present Status and Issues. – International Water Management Institute, Colombo, Sri Lanka.
[36] Robnik-Sikonja, M., Savicky, P. (2017): CORElearn: Classification, Regression and Feature Evaluation. R package version 1.50.3. – https://cran.r-project.org/package=CORElearn.
[37] Seyedmohammadi, J., Esmaeelnejad, L., Shabanpour, M. (2016): Spatial variation modeling of groundwater electrical conductivity using geostatistics and GIS. – Modeling Earth Systems and Environment 2(169): 1–10. DOI: 10.1007/s40808-016-0226-3.
[38] Sirimanne, C. H. L. (1952): Geology for water supply Ceylon. – As. Advmt. Sci. 8th Annual Session, pp. 87–118.
[39] Stahl, K., Moore, R. D., Floyer, J. A., Asplin, M. G., Mckendry, I. G. (2006): Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. – Agricultural and Forest Meteorology 139: 224–236. DOI: 10.1016/j.agrformet.2006.07.004.
[40] Team R, R. C. (2016): A language and environment for statistical computing. R foundation for statistical computing: Vienna, Austria. – http://www.r-project.org/.
[41] Todd, D. K., Mays, L. (2005): Groundwater Hydrology. – John Wiley & Sons, USA, New York.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

[42] Uyan, M., Cay, T. (2013): Spatial analyses of groundwater level differences using geostatistical modeling. – Environmental and Ecological Statistics 20: 633–646. DOI: 10.1007/s10651-013-0238-3.
[43] Van Meter, K. J., Steiff, M., McLaughlin, D. L., Basu, N. B. (2016): The socioecohydrology of rainwater harvesting in India: Understanding water storage and release dynamics across spatial scales. – Hydrology and Earth Sciences 20(7): 2629–2647. DOI: 10.5194/hess-20-2629-2016.
[44] Wagner, P. D., Fiener, P., Wilken, F., Kumar, S., Schneider, K. (2012): Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. – Journal of Hydrology 464–465: 388–400. DOI: 10.1016/j.jhydrol.2012.07.026.
[45] Wang, S., Huang, G. H., Lin, Q. G., Li, Z., Zhang, H., Fan, Y. R. (2014): Comparison of interpolation methods for estimating spatial distribution of precipitation in Ontario, Canada. – International Journal of Climatology 34: 3745–3751. DOI: 10.1002/joc.3941.
[46] Webster, R., Oliver, M. (2008): Geostatistics for Environmental Scientists. – John Wiley & Sons, Ltd, New York.
[47] Wijesundara, W. M. G. D., Nandasena, K. A., Jayakody, A. N. (2012): Spatial and temporal changes in nitrogen, phosphorus and potassium concentration in water in the Thirappane tank cascade in dry zone of Sri Lanka. – Journal of Environmental Professionals Sri Lanka 1(1): 70–81.
[48] Wilcox, L. V. (1955): Classification and Use of Irrigation Waters. – SDA Circular No. 969, United States Department of Agriculture. Washington, D. C.
[49] Wu, W., Tang, X., Ma, X., Liu, H. A. (2016): A comparison of spatial interpolation methods for soil temperature over a complex topographical region. – Theoretical and Applied Climatology 125(3–4): 657–667.
[50] Xiao, Y., Gu, X., Yin, S., Shao, J., Cui, Y., Zhang, Q., Niu, Y. (2016): Geostatistical interpolation model selection based on ArcGIS and spatio-temporal variability analysis of groundwater level in piedmont plains, northwest China. – SpringerPlus 5: 425. DOI: 10.1186/s40064-016-2073-0.
[51] Xie, Y., Chen, T. B., Lei, M., Yang, J., Guo, Q. J., Song, B., Zhou, X. Y. (2011): Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: Accuracy and uncertainty analysis. – Chemosphere 82: 468–476. DOI: 10.1016/j.chemosphere.2010.09.053.
[52] Yao, X., Fu, B., Lu, Y., Sun, F., Wang, S., Liu, M. (2013): Comparison of four spatial interpolation methods for estimating soil moisture in a complex terrain catchment. – PLoS ONE 8(1): 1–13. DOI: 10.1371/journal.pone.0054660.
[53] Zehtabian, G., Asgari, H., Tahmouresc, M. (2013): Assessment of spatial structure of groundwater quality variables based on the geostatistical simulation. – Desert 17: 215–224.
[54] Zimmerman, D., Pavlik, C., Ruggles, A. (1999): An experimental comparison of ordinary and universal kriging and distance weighting. – Mathematical Geology 31(4): 375–390.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

APPENDIX

Table A1. *Longitude and latitude of the sampling locations*

Well no.	Longitude	Latitude	Well no.	Longitude	Latitude
1	80° 32' 28" E	8° 13' 35" N	16	80° 32' 35" E	8° 9' 34" N
2	80° 32' 21" E	8° 14' 5" N	17	80° 34' 11" E	8° 6' 36" N
3	80° 32' 47" E	8° 13' 18" N	18	80° 33' 51" E	8° 7' 18" N
4	80° 33' 1" E	8° 12' 43" N	19	80° 33' 21" E	8° 7' 8' N
5	80° 33' 0" E	8° 12' 1" N	20	80° 33' 30" E	8° 8' 16" N
6	80° 32' 30" E	8° 12' 27" N	21	80° 33' 36" E	8° 8' 38" N
7	80° 33' 27" E	8° 11' 58" N	22	80° 33' 55" E	8° 9' 0" N
8	80° 33' 38' E	8° 11' 25" N	23	80° 33' 7" E	8° 8' 9" N
9	80° 33' 33" E	8° 10' 53" N	24	80° 32' 43" E	8° 8' 53" N
10	80° 32' 59" E	8° 11' 6" N	25	80° 33' 32" E	8° 6' 29" N
11	80° 32' 23" E	8° 11' 32" N	26	80° 32' 35" E	8° 7' 24" N
12	80° 31' 55" E	8° 10' 36" N	27	80° 33' 21" E	8° 7' 42" N
13	80° 32' 16" E	8° 9' 49" N	28	80° 31' 37" E	8° 13' 28" N
14	80° 33' 11" E	8° 9' 54" N	29	80° 31' 56" E	8° 12' 46" N
15	80° 32' 60" E	8° 10' 27" N			

Table A2. *Accuracy of different methods at predicting pH*

	RMSE						
	IDW	**GPI**	**RBF**	**LPI**	**OK**	**UK**	**EBK**
April	0.422	0.421	0.442	0.440	0.482	0.482	*0.393*
May	0.294	0.269	0.294	0.276	0.270	0.270	*0.266*
June	0.406	0.430	0.425	0.438	0.409	0.409	*0.402*
July	*0.303*	0.314	0.319	0.310	0.313	0.313	0.306
August	0.401	*0.392*	0.412	0.400	0.413	0.413	0.404
September	0.468	0.456	0.486	0.469	0.499	0.499	*0.451*
October	0.417	0.395	0.426	0.406	0.440	0.440	*0.391*
November	0.397	0.398	0.408	0.405	0.398	0.398	*0.370*
December	0.394	*0.356*	0.397	0.357	0.395	0.395	0.377
January	0.373	0.358	0.375	*0.352*	0.367	0.367	0.359
February	0.355	0.373	0.359	*0.344*	0.351	0.351	0.347
March	0.257	0.249	0.252	0.246	0.243	0.243	*0.236*

Table A3. *Accuracy of different methods at predicting sodium concentration*

	RMSE						
	IDW	**GPI**	**RBF**	**LPI**	**OK**	**UK**	**EBK**
April	92.45	92.48	95.10	94.76	96.52	96.52	*90.46*
May	78.64	79.38	79.95	81.00	81.91	81.91	*77.20*
June	72.20	64.76	74.33	67.48	63.97	63.97	*63.13*
July	69.93	64.29	70.40	67.04	64.68	64.68	*62.25*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

August	76.94	71.68	80.03	74.68	77.974	77.974	*70.36*
September	78.63	76.92	78.91	79.49	82.32	82.32	*70.71*
October	59.37	57.47	60.22	59.78	62.48	62.48	*52.50*
November	*59.40*	61.57	60.22	59.60	60.392	60.392	60.15
December	101.53	101.44	103.13	104.55	104.267	104.267	*98.98*
January	66.72	66.66	65.96	68.75	67.62	67.62	*63.51*
February	91.87	94.39	93.75	97.04	94.024	94.024	*91.52*
March	77.68	77.91	79.63	78.36	82.09	82.09	*71.69*

Table A4. *Accuracy of different methods at predicting potassium concentration*

	RMSE						
	IDW	**GPI**	**RBF**	**LPI**	**OK**	**UK**	**EBK**
April	2.497	2.512	2.509	2.461	2.435	2.435	*2.413*
May	2.094	2.149	2.099	2.131	2.095	2.095	*2.040*
June	3.008	2.922	2.954	2.847	2.886	2.886	*2.841*
July	2.853	2.805	2.872	2.783	2.818	2.818	*2.693*
August	2.272	2.311	2.297	2.256	2.187	2.187	*2.190*
September	1.873	1.907	1.881	1.842	1.810	1.810	*1.795*
October	1.509	1.566	1.528	1.553	1.486	1.486	*1.453*
November	1.710	1.712	1.707	1.703	1.707	1.707	*1.652*
December	3.669	3.686	3.694	3.674	3.659	3.659	*3.500*
January	2.753	2.719	2.777	2.710	2.661	2.661	*2.558*
February	3.300	3.482	3.324	3.307	3.249	3.249	*3.198*
March	1.043	1.041	1.040	1.016	1.075	1.075	*1.011*

Table A5. *Accuracy of different methods at predicting calcium concentration*

	RMSE						
	IDW	**GPI**	**RBF**	**LPI**	**OK**	**UK**	**EBK**
April	99.562	*90.209*	100.446	91.437	93.295	93.295	94.296
May	32.981	30.199	32.803	30.184	29.581	29.581	*28.213*
June	36.401	32.650	37.321	33.056	32.387	32.387	*31.622*
July	49.471	47.518	50.900	47.606	49.863	49.863	*45.759*
August	23.423	20.558	23.355	20.890	20.212	20.212	*19.383*
September	29.037	*26.220*	29.319	26.744	26.700	26.700	26.594
October	20.529	21.708	20.212	*17.601*	20.772	20.772	20.183
November	19.873	*17.981*	20.212	18.285	18.807	18.807	18.682
December	49.393	45.419	49.848	45.836	46.767	46.767	*43.777*
January	41.396	*38.672*	42.099	38.988	39.115	39.115	38.914
February	55.709	53.134	56.238	*51.639*	54.282	54.282	52.781
March	25.096	23.094	25.008	23.289	22.864	22.864	*22.832*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

Table A6. *Accuracy of different methods at predicting magnesium concentration*

	RMSE						
	IDW	**GPI**	**RBF**	**LPI**	**OK**	**UK**	**EBK**
April	52.09	48.59	51.09	49.33	51.61	51.61	*46.22*
May	38.74	35.86	38.09	36.70	36.62	36.62	*33.58*
June	33.33	30.96	33.06	31.05	30.94	30.94	*29.02*
July	48.02	48.26	46.85	46.06	45.52	45.52	*43.98*
August	21.45	19.44	21.72	19.79	22.28	22.28	*18.13*
September	20.26	18.93	19.76	19.02	19.11	19.11	*17.84*
October	19.94	19.19	19.09	18.96	19.80	19.80	*18.59*
November	24.27	23.11	23.92	23.12	23.69	23.69	*21.57*
December	59.82	57.08	58.69	57.28	59.94	59.94	*53.49*
January	37.24	37.04	35.26	36.30	36.45	36.45	*34.28*
February	55.12	56.53	53.29	53.73	54.51	54.51	*52.07*
March	25.35	23.85	25.07	24.16	24.28	24.28	*23.55*

Table A7. *Accuracy of different methods at predicting nitrate nitrogen concentration*

	RMSE						
	IDW	**GPI**	**RBF**	**LPI**	**OK**	**UK**	**EBK**
April	3.449	3.572	3.563	3.526	3.349	3.349	*3.192*
May	3.035	2.890	3.188	2.884	2.794	2.794	*2.735*
June	2.519	2.349	2.527	2.449	2.256	2.256	*2.170*
July	2.684	2.685	2.829	2.670	2.529	2.529	*2.490*
August	0.801	0.840	0.773	0.805	0.778	0.778	*0.750*
September	2.334	2.395	2.416	2.382	2.247	2.247	*2.119*
October	4.982	5.147	5.200	5.209	4.904	4.904	*4.718*
November	4.537	4.770	4.691	4.694	4.344	4.344	*4.299*
December	2.162	2.195	2.265	2.157	2.039	2.039	*1.969*
January	2.181	2.246	2.260	2.217	2.067	2.067	*2.027*
February	2.166	2.190	2.267	2.148	2.049	2.049	*1.985*
March	3.316	3.147	3.507	3.164	3.052	3.052	*2.906*

Table A8. *Accuracy of different methods at predicting phosphate concentration*

	RMSE						
	IDW	**GPI**	**RBF**	**LPI**	**OK**	**UK**	**EBK**
April	*0.0383*	0.0411	0.0393	0.0400	0.0419	0.0419	0.0387
May	0.0550	0.0579	*0.0429*	0.0539	0.0522	0.0522	0.0544
June	0.0521	0.0554	*0.0510*	0.0561	0.0547	0.0547	0.0519
July	0.2372	0.2508	0.2351	0.2505	0.2299	0.2299	*0.2289*
August	0.2376	0.2847	*0.2137*	0.2559	0.2203	0.2203	0.2185
September	0.0791	0.0845	*0.0665*	0.0772	0.0761	0.0761	0.0784
October	0.1630	0.1643	0.1583	0.1641	0.1666	0.1666	*0.1543*
November	0.1601	0.1547	0.1578	0.1577	0.1561	0.1561	*0.1456*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

December	0.1233	0.1252	0.1184	0.1251	0.1214	0.1214	*0.1165*
January	0.0511	0.0513	0.0494	0.0514	0.0505	0.0505	*0.0478*
February	0.0644	0.0642	0.0660	0.0638	0.0678	0.0678	*0.0609*
March	0.8531	0.8177	0.8526	0.8143	0.8521	0.8521	*0.7667*

Table A9. *Accuracy of different methods at predicting bicarbonate concentration*

	RMSE						
	IDW	**GPI**	**RBF**	**LPI**	**OK**	**UK**	**EBK**
April	7.635	8.537	7.152	7.084	7.484	7.484	*7.791*
May	10.472	12.321	10.672	10.795	10.583	10.583	*10.177*
June	25.261	25.468	25.984	26.111	25.386	25.386	*23.049*
July	26.454	28.259	27.438	28.560	27.723	27.723	*25.567*
August	26.394	25.871	27.128	26.459	25.518	25.518	*23.879*
September	21.248	21.488	21.703	21.815	21.398	21.398	*20.118*
October	24.846	24.494	24.940	24.767	26.945	26.945	*22.679*
November	37.413	36.044	36.765	36.308	36.299	36.299	*33.808*
December	23.242	22.315	23.540	22.811	*21.880*	*21.880*	22.271
January	23.791	22.593	24.270	23.372	22.774	22.774	*21.953*
February	13.513	13.121	13.770	13.571	13.889	13.889	*13.124*
March	16.755	16.211	16.609	16.734	16.073	16.073	*15.043*

Table A10. *Accuracy of different methods at predicting chloride concentration*

	RMSE						
	IDW	**GPI**	**RBF**	**LPI**	**OK**	**UK**	**EBK**
April	523.00	472.00	514.00	479.00	*473.57*	*473.57*	474.63
May	437.00	399.00	432.00	405.00	397.85	397.85	*377.13*
June	299.00	260.00	303.00	266.00	256.10	256.10	*243.52*
July	339.00	295.00	340.00	303.00	292.21	292.21	*275.89*
August	340.00	310.00	340.00	316.00	301.81	301.81	*290.23*
September	380.00	352.00	380.00	358.00	349.04	349.04	*331.90*
October	349.00	320.00	349.00	324.00	317.13	317.13	*301.21*
November	348.00	323.00	350.00	327.00	319.33	319.33	*302.74*
December	467.00	426.00	473.00	433.00	423.95	423.95	*401.40*
January	347.00	321.00	351.00	327.00	322.52	322.52	*304.87*
February	366.00	341.00	368.00	347.00	341.23	341.23	*325.94*
March	332.00	301.00	336.00	308.00	301.53	301.53	*288.31*

Table A11. *Accuracy of different methods at predicting electrical conductivity*

	RMSE						
	IDW	**GPI**	**RBF**	**LPI**	**OK**	**UK**	**EBK**
April	556.00	522.00	562.00	533.00	511.68	511.68	*485.51*
May	1014.00	972.00	1020.00	991.00	954.44	954.44	*900.32*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

June	901.00	789.00	915.00	809.00	781.29	781.29	*733.61*
July	889.00	775.00	894.00	796.00	760.98	760.98	*717.11*
August	966.00	861.00	974.00	883.00	846.27	846.27	*800.76*
September	20.00	24.00	19.00	19.00	19.22	19.22	*18.55*
October	1048.00	980.00	1049.00	999.00	961.50	961.50	*907.20*
November	1052.00	985.00	1057.00	1002.00	961.89	961.89	*914.46*
December	771.00	719.00	775.00	732.00	703.45	703.45	*669.21*
January	843.00	797.00	849.00	812.00	779.35	779.35	*739.80*
February	598.00	560.00	602.00	571.00	549.77	549.77	*522.34*
March	824.00	764.00	817.00	780.00	762.11	762.11	*719.89*

Table A12. *Accuracy of different methods at predicting sodium adsorption ratio*

	RMSE						
	IDW	**GPI**	**RBF**	**LPI**	**OK**	**UK**	**EBK**
April	1.571	1.667	*1.446*	1.546	1.459	1.459	1.639
May	2.298	2.442	2.197	2.306	*2.110*	2.110	2.231
June	1.394	1.367	1.435	1.421	1.449	1.449	*1.311*
July	2.046	1.950	2.118	2.040	2.057	2.057	*1.886*
August	2.458	2.420	2.526	2.475	2.779	2.779	*2.269*
September	2.568	2.591	2.601	2.673	2.564	2.564	*2.372*
October	2.177	2.187	2.217	2.261	2.253	2.253	*2.012*
November	2.027	2.122	2.047	2.086	2.225	2.225	*1.949*
December	2.150	2.186	2.160	2.196	2.142	2.142	*2.100*
January	1.788	*1.749*	1.843	1.817	1.829	1.829	2.148
February	1.921	2.008	*1.622*	1.872	1.781	1.781	1.694
March	2.168	2.319	*2.097*	2.157	2.104	2.104	2.159

Table A13. *Accuracy of different methods at predicting total hardness*

	RMSE						
	IDW	**GPI**	**RBF**	**LPI**	**OK**	**UK**	**EBK**
April	364.00	338.00	363.00	339.00	357.00	357.00	*349.00*
May	228.00	207.00	225.00	210.00	208.57	208.57	*195.55*
June	202.00	185.00	204.00	186.00	183.93	183.93	*170.94*
July	290.00	295.00	288.00	283.00	275.66	275.66	*267.19*
August	134.00	120.00	135.00	122.00	117.87	117.87	*113.19*
September	143.00	131.00	141.00	132.00	132.02	132.02	*123.10*
October	117.00	117.00	111.00	*103.00*	114.81	114.81	110.32
November	139.00	130.00	138.00	131.00	137.79	137.79	*122.21*
December	344.00	325.00	340.00	325.00	336.25	336.25	*307.04*
January	227.00	223.00	220.00	220.00	228.00	228.00	*222.00*
February	338.00	339.00	331.00	324.00	332.00	332.00	*318.69*
March	158.00	146.00	156.00	148.00	146.53	146.53	*144.43*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380

Table A14. *Summary of Whisker boxplot diagram*

Parameter	Measured					Predicted				
	Median	**Mean**	**Q1**	**Q3**	**Asym.**	**Median**	**Mean**	**Q1**	**Q3**	**Asym.**
Na	118.4	133.0	71.94	183.8	RS	126.5	132.9	114.8	145.7	RS
K	1.845	2.48	1.04	2.88	RS	2.30	2.44	1.81	2.96	RS
Ca	48.8	67.15	27.77	88.15	RS	55.85	67.77	36	87.4	RS
Mg	34.88	47.29	19.67	63.07	RS	41.88	48.26	29.35	66.51	RS
Cl	190.0	325.5	118.8	420.0	RS	320.4	356.5	285.4	339.6	RS
HCO3	85.5	86.7	75	96.7	RS	86.0	87.0	86.0	89.0	RS
NO3-N	0.72	1.54	0.41	1.44	RS	1.14	1.36	0.85	1.64	RS
PO4	0.085	0.234	0.03	0.35	RS	0.106	0.23	0.05	0.346	RS
pH	7.86	7.83	7.58	8.1	LS	7.89	7.83	7.65	7.8	RS
EC	1016	1259	1719	1350	RS	1239	1261	1165	1403	RS
SAR	3.12	3.42	1.96	4.31	RS	3.49	3.41	2.97	3.83	LS
TH	261.4	361.0	168.1	485.7	RS	294.3	364.9	222.7	497.8	RS

RS - right skewed; LS - left skewed; Q1 - 1st quartile; Q3 - 3rd quartile; Asym. - asymmetry

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1605_53595380