



EPiC Series in Engineering

Volume 3, 2018, Pages 2042–2050

HIC 2018. 13th International
Conference on Hydroinformatics



Uncertainty Estimations in Different Components of a Hybrid ANN – Fuzzy - Kriging Model for Water Table Level Simulation

Evdokia Tapoglou¹, Emmanouil A. Varouchakis¹, George P. Karatzas¹

¹School of Environmental Engineering, Technical University of Crete, GR73100, Chania, Greece.

Corresponding author: karatzas@mred.tuc.gr

Abstract

The purpose of this study is to examine the uncertainty of a combined artificial neural network (ANN), kriging and fuzzy logic methodology, which can be used for spatial and temporal simulation of hydraulic head in an aquifer. This methodology was applied in the past, while the verification of the model was performed by implementing it in a new study area, in Miami – Dade County, FL, USA. The percentile methodology was applied as a first approach in order to define the ANN uncertainty. As a second approach, the uncertainty of the ANN training is tested through a Monte Carlo procedure. The model was executed 300 times using different training set and initial random values each time. The training results constituted a sensitivity analysis of the ANN training to the kriging part of the algorithm. The training and testing error intervals for the ANNs and the kriging prediction intervals calculated through this procedure can be considered narrow compared to the complexity of the study area. For the third and final approach used in this work, the uncertainty of kriging parameter was calculated through the Bayesian kriging methodology. The results derived can prove that the simulation algorithm can provide consistent and accurate results.

1 Introduction

Data driven models and statistical models have both been extensively used in groundwater modelling, especially in the last decades (Delbari, Amiri, & Motlagh, 2014; Maiti & Tiwari, 2014). A new concept allowed for the combination of two separate methodologies, Artificial Neural Networks (ANNs) and kriging for the spatial and temporal simulation of hydraulic head, in a large scale area. Moreover, the use of a fuzzy logic system, in order to connect these two methodologies, was proved effective, providing better simulation results (Tapoglou, Karatzas, Trichakis, & Varouchakis, 2014). In this study, the examined study area is small, but with a complex geological background. Simulation of the current and past states followed the same methodology as in the aforementioned work.

To provide also a measure of the uncertainty, an analysis comprising of three different methodologies were performed, corresponding to Kriging parameters for the first one and uncertainty derived by the ANNs for the two remaining.

1.1 Simulation algorithm

The simulation algorithm used in this work was based on the one presented in Tapoglou et al. (2014) and involved the use of ANNs, kriging and a fuzzy logic methodology in order to simulate the water table changes in a temporal and spatial scale. To summarize the conceptual framework, the ANN was used for the temporal simulation of the hydraulic head, while kriging was used for the spatial interpolation.

The temporal simulation took place in multiple locations where data were available, diversifying the inputs as necessary in order to acquire the optimal ANN results. For every location with available hydraulic head data one ANN was developed. Data prerequisites included information on meteorological and hydrological parameters with a daily time step as input data, and hydraulic head measurements in the same time step as the output parameter. The selection of input parameters focused on those linked directly or indirectly to the water budget. The ANNs used available data for training and simulation of the current state, while the trained ANNs are also able to predict future status of the aquifer, when evaluating different scenarios.

The results of all ANNs were subsequently interpolated using the kriging methodology. Initially, a simple ordinary kriging had been used and the linear, exponential and power-law variograms were tested. The interpolation was conducted in the convex polygon of the wells. This grid is a rectilinear grid, where all observation points fall exactly at a node. The nodes of the grid where observed data are not available are the prediction points. In order to exclude from the simulation grid nodes that fall far outside the study and far from any observation point, Delaunay triangles were created from the known data locations. Nodes that fell outside the area covered by at least one triangle were not considered as prediction points.

In order to improve the initial results, a fuzzy logic system was used so as to combine the two methodologies. More specifically, the neighborhood used in kriging was defined through fuzzy logic, combining the distance between the prediction point and data points with the ANN training and testing error in every location. For every prediction point involved in the process and for every time step, a different variogram is calculated and the hydraulic head is simulated. A detailed description of the functionality and the initial methodology followed is found in Tapoglou et al. (2014).

1.2 Uncertainty Methodology

Model uncertainty can be attributed to many reasons. In this work, we have identified and grouped in two main categories the sources of uncertainty; uncertainty derived by the ANN training process and uncertainty due to the interpolation methodology.

Monte Carlo simulation and bootstrap methods are very common and can give a perspective of the uncertainty involved. The objective of an uncertainty analysis is to show the effects of model inputs or parameters on the model simulation results (Eslamian, 2014). In this study, different methodologies were used, for the determination of uncertainty derived from different parts of the developed algorithm (Figure 1). As far as the ANN part of the simulation algorithm is concerned, two methodologies were followed; the percentile methodology and a Monte Carlo simulation for the uncertainty of the ANN's training. The results of the second methodology became the input for the Kriging part of the algorithm, constituting a sensitivity analysis of the kriging interpolation method to the ANN training. Uncertainty due to the Kriging parameters is also studied using the Bayesian kriging methodology.

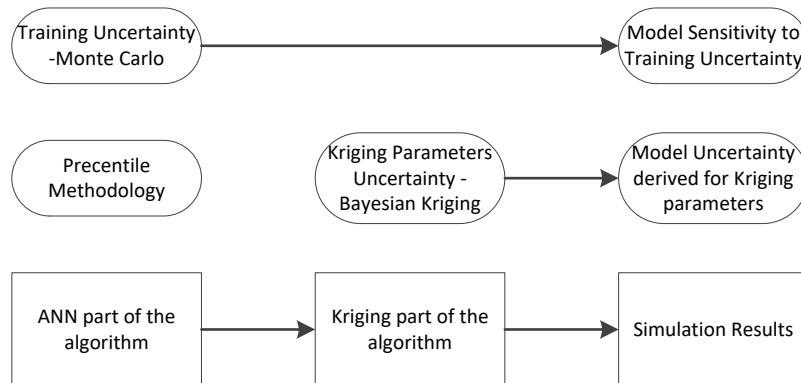


Figure 1: Methods tested in each simulation step

1.2.1 Uncertainty in ANN results

Uncertainty due to the ANN model has been extensively studied in the past, using both Bootstrap-based methods (Kasiviswanathan & Sudheer, 2013; Trichakis, Nikolos, & Karatzas, 2011) and Monte Carlo simulation methods (Dehghani, Saghafian, Saleh, Farokhnia, & Noori, 2014; Jiang, Nan, & Yang, 2013).

In ANNs, the source of uncertainty may be attributed to two reasons; model structure and training data (Dybowski & Roberts, 2001). In this study the model structure is fixed and only the uncertainty derived from the training dataset is considered. This uncertainty can be attributed to not only the imperfection in data collection tools or techniques, but also to the available amount of training data and how representative of the reality they are. Moreover, not all the possible realizations of the variables involved are available in the training set, as well as the training set itself is only one of a large number of possibilities. In this study two methodologies were used for the calculations of the prediction intervals, the first one based on a statistical analysis and the second one using Monte Carlo simulation results.

Percentile Prediction Intervals

In this approach the prediction intervals can be calculated through the statistical characteristics of the model errors that occurred when reproducing the observed in the field data (Shrestha & Solomatine, 2006).

This approach takes a large number of simulated parameters of interest and orders them from smallest to largest. A 90% Prediction Interval (PI) is then computed by identifying the lower PI as the 5th percentile and the upper PI as the 95th percentile. This leaves 90% of the simulated estimates within this range, while dividing the remaining 10% of the simulated values equally into the upper and lower tails. The advantage of this method is that it does not make any distributional assumptions and it does not require the distribution to be symmetric. However, evaluating the PI using this methodology yields larger confidence intervals compared to other methodologies.

Monte Carlo method

Conventionally, a single ANN is developed through training and once completed the ANN parameters are considered fixed. Such an ANN becomes deterministic model where a given input data vector will always produce a specific output value. In order to calculate the uncertainty in this prediction, Monte Carlo methods can be used, which are typically computer based techniques for brute force numerical simulation of probabilistic processes.

The uncertainty of the ANNs results in this case can be assessed mainly by the random initialization of values (i.e. the ANN synaptic weights), as well as in the random selection of training and testing datasets. In order to investigate the uncertainty in the predictions of the ANN, multiple different executions of the algorithm, with random initial values for the neural weights and different training and testing sets were performed. The examination of the uncertainty attributed to the architecture and structure of the ANNs would require large computational cost, hence it is not examined in this study.

1.2.2 Uncertainty in Kriging results

The uncertainty of the kriging algorithm was evaluated following two separate methodologies. The first one included the use of the Monte Carlo results of the ANN uncertainty analysis and can be characterized also as a sensitivity analysis of the kriging to the ANN training. The second methodology followed is the one of Bayesian kriging, which can be used to estimate the uncertainty due to the kriging parameters.

Monte Carlo method

In this case the results of the Monte Carlo simulation in training the ANNs were used. More specifically the multiple different training results (different neural weights) were used for the evaluation of the hydraulic head, resulting in multiple different values for every timestep and for every prediction point. Using the 5th and 95th percentile it is possible to determine the 90% prediction interval. These results were then passed forward to the kriging part of the algorithm, generating a range within which the simulation results are, for every prediction point and time step separately.

Bayesian kriging

The process of interpolating spatial data using kriging can be realized in two steps. First, the covariance function/semivariogram is calculated and then the kriging interpolation takes place. When using the classical approach the assumption that the model parameters are known is made, and the uncertainty in the calculation of range and sill is considered as minimal. The purpose of the Bayesian kriging is to quantify the uncertainty of these parameters and hence the uncertainty of the covariance estimation. In Bayesian kriging (Pilz & Spöck, 2008) the posterior distribution of the parameters involved (semivariogram and/or transformation parameters) is specified by means of simulations, which corresponds to the uncertainty of covariance parameters.

In order to implement Bayesian kriging methods, first the appropriate variogram model must be fitted to the experimental data, using the standard procedure followed when using the variogram estimation (1st step in kriging interpolation methodology). Using the parameters derived and the covariance function of the model, the covariance matrix is constructed. A common method for testing a statistical model is the use of artificial data. In order to do so, a set of properties have to be embedded in the dataset and then the model must be examined for the presence of these effects and how they behave under different experimental conditions.

Uniform random numbers generation is the first step in this procedure. A generally approved method of correlating random numbers, with a known covariance matrix (C), is by finding a matrix U , by using decomposition method. Matrix decomposition is commonly used in Monte Carlo method for simulating systems with multiple correlated variables. Using this matrix, correlated random numbers R_c can be generated from uncorrelated numbers R , by multiplying them with this matrix U . These correlated random values have the same statistical characteristics with the initial, observed in the field data. These random values are then fitted into a variogram and the value in an unknown location is estimated using kriging. Cholesky decomposition is one way of providing these random correlated values. The covariance matrix of a vector Y can be given as $C = E(Y Y^T)$. If X is a random vector, consisting of uncorrelated random values uniform in $[0,1]$, then $E(X X^T) = I$. The Cholesky decomposition of the

correlation matrix is given as $C = L \cdot L^T$. Note that it is possible to obtain a Cholesky decomposition of C since by definition the covariance matrix C is symmetric and positive definite.

The distribution of the simulated values using the kriging of different correlated random values can be used in order to derive the confidence intervals of the process. In this way the distribution of the parameters involved (sill, range etc.) in the process is used as a prior knowledge, while the posterior distribution reflects the uncertainty of the covariance estimation (Pilz & Spöck, 2008).

2 Study Area

The area in which the proposed methodology is applied is located in Miami – Dade County, Florida, USA (USGS, 2017; Fish & Stewart, US Department of the Interior, US Geological Survey). A total number of 30 wells within Biscayne aquifer inside or around the urban area of Miami were used for the water table simulation. The data available covered the time period from 25/4/2010 to 22/2/2014, with daily timestep and included the hydraulic head, the discharge and water level in four nearby rivers and meteorological data from two stations. The data used in this study regarding water levels were acquired through United States Geological Survey (USGS), while the meteorological data from the National Climatic Data center of the National Oceanic and Atmospheric Administration (NOAA). All the data used were directly or indirectly linked to the aquatic equilibrium.

3 Results

3.1 Preprocessing of available data

An important factor that can easily be neglected when choosing ANN parameters is that of the time lag; i.e. the time between an event happening (for example a rainfall occurrence) and the appearance of its effect in the dataset. The correlation coefficient between time series A for a parameter and time series B for the hydraulic head change can be used to determine the optimal lag between the parameter and the hydraulic head change.

The highest correlation lagged parameters will be used as input to the ANN. The input parameters to the ANNs were a combination of the parameters with the best correlation coefficient. For each one of the ANNs, the input parameters used are the Precipitation for 2 meteorological station (3 days), the soil and air temperature, the humidity, the surface water levels from 2 stations (2 days) and the discharge from 1 station (best correlated day).

There isn't a universal rule about the optimal architecture of an ANN. However, according to the rule of Fine (1999), using three times as many training patterns as network parameters (weights) is adequate to achieve good generalization. Taking into account that 9 input parameters were used, 1 output and 1100 training sets, two architectures were examined; the first one with 1 hidden layer with 32 hidden nodes (Architecture 1) and the second one with 2 hidden layers with 17 and 12 hidden nodes respectively (Architecture 2).

Three variogram models were tested for the best architecture as well, the linear, the exponential and the power-law, in order to choose the most appropriate for the data at hand. These variograms were chosen in order to have one variogram model per type; the most simple models, the linear, one differentiable, the exponential, and one non-differentiable, the power-law.

3.2 Simulation results

The ANNs performances for each one of the two possible architectures (Architecture 1 - 1 hidden layer with 32 hidden nodes and Architecture 2- 2 hidden layers with 17 and 12 hidden nodes) are presented in terms of Root Mean Squared Error (RMSE) in Table 1.

RMSE (m)	Architecture 1	Architecture 2
Average Training Error	$1.16 \cdot 10^{-3}$	$6.63 \cdot 10^{-4}$
Average Testing Error	$1.20 \cdot 10^{-3}$	$7.92 \cdot 10^{-4}$
Maximum Training/Testing Error	$2.19 \cdot 10^{-3}/3.42 \cdot 10^{-3}$	$1.78 \cdot 10^{-3}/2.98 \cdot 10^{-3}$
Minimum Training/Testing Error	$2.46 \cdot 10^{-4}/2.63 \cdot 10^{-4}$	$2.23 \cdot 10^{-4}/1.26 \cdot 10^{-4}$

Table 1: Training and testing RMSE for 2 architectures

In terms of average minimum and maximum values the second architecture has better results, both for the training and testing error. In order to evaluate the performance of the each one of the 30 ANNs developed they were divided into three categories, depending on their training and testing RMSE performance. The first category includes the ANNs with the best performance (training RMSE= $2.77 \cdot 10^{-4}$ m, testing RMSE= $3.13 \cdot 10^{-4}$ m, NSE=0.69), represented by well 11. The second category has average results, represented by well 3 (training RMSE= $5.68 \cdot 10^{-4}$ m, testing RMSE= $4.19 \cdot 10^{-4}$ m, NSE=0.65), while the third category has the worst RMSE results amongst the ANNs studied (training RMSE= $1.11 \cdot 10^{-3}$ m, testing RMSE= $6.23 \cdot 10^{-4}$ m, NSE=0.58). It should be noted that the NSE values are not very close the optimal value, which equals one ($NSE_{optimal} = 1$). This can be attributed to the low hydraulic head values that should be simulated by the model. As indicated by Schaepli and Gupta (2007), it is not possible to achieve very high NSE values when simulating small absolute values, however it is an indication of the performance of the model.

The second part of the algorithm includes the fuzzy logic use for the neighborhood definition and the kriging interpolation. The number of neighbors for each prediction point is set to 20. This number should be chosen carefully so as to allow for the 30 pairs per distance class, which according to Journel and Huijbregts (1978) are necessary in order to acquire a good fitting of the experimental variogram to the theoretical model.

To evaluate the results, cross-validation was performed for all three variograms studied. For every data point, for 10% of the time steps the observed in that point data were ignored, and the algorithm follows as if the data point does not exist. In this way three values are available; the observed in the field, the ANN simulated and the kriging simulated. In the end of the process, the kriging simulated values are compared to the real and ANN simulated, in order to derive an independent measure of performance. The kriging interpolation tends to correct any misjudgements of the hydraulic head performed by the ANNs, improvement that is also reflected on the error indicators. More specifically, The RMSE, NSE and MAE values of the data used for the cross validation when using ANNs are 0.77, 0.58 and 0.59 while when using ANNs together with kriging are 0.49, 0.8 and 0.41 respectively

3.3 Uncertainty results

ANN prediction confidence intervals

Using the initial simulation results it is possible to calculate the maximum difference between observed and simulated values. By applying the 95% and 5% percentile error to all simulated values the 95% and 5% percentile of the simulation results can be calculated. The results for the three representative wells are presented in Figure 2. While the 90% prediction intervals were constructed using the training dataset, this doesn't necessarily mean that the testing data will always fall within

these intervals as well. In most cases, especially in wells 3 and 18, the observed values fall within the 90% prediction interval of the ANN prediction for the testing dataset. In well 11 in few exceptions, and more specifically at some of the lowest observed values, the real values do not fall within the prediction interval. The coverage of these intervals with respect to the observed values was 82% for well 11, 75% for well 3 and 79% for well 18.

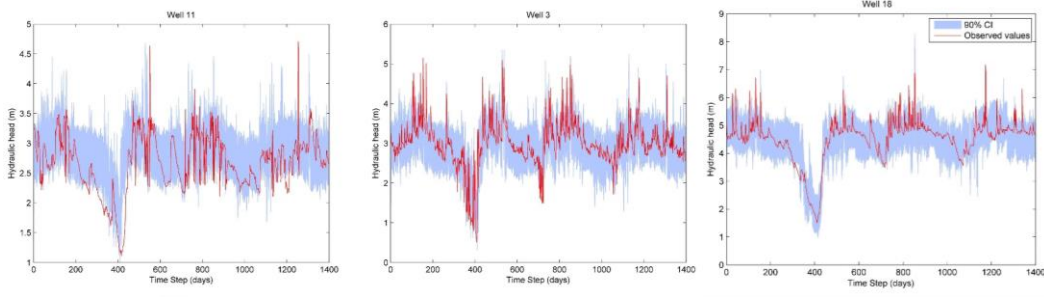


Figure 2: 90% confidence intervals for ANN simulation

ANN training model uncertainty

In order to evaluate the uncertainty of the ANNs training, the process described in section 2.2.1 was performed 300 times, with each case having different training and testing data set, as well as different initial random neural weights. For all wells used, the results of this process can be summarized in Figure 6, for the training and testing error respectively. The training error range, as expected, is smaller than the testing in most of the cases (26/30 wells). The results in terms of the representative wells are calculated and presented, together with the observed in the field values, in Figure 3.

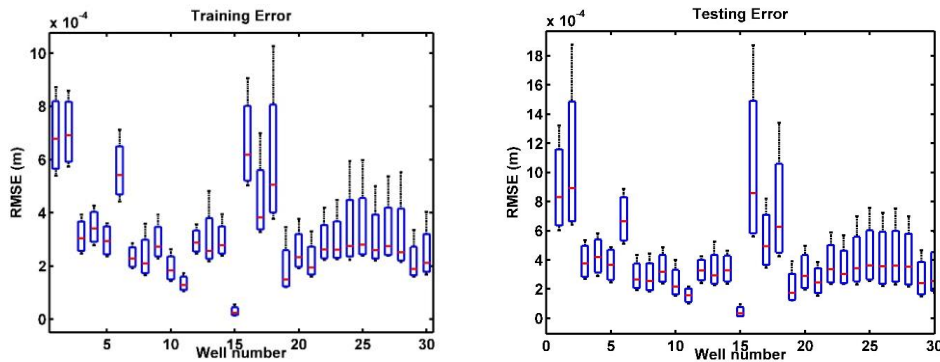


Figure 3: Training and testing error range for training uncertainty for 30 wells

For well number 11, the 5.5% of the observed values are outside the 90% prediction interval. As the ANN performance deteriorates this percentage grows, to 9.1% for well 3 and 11.2% for the worst performance ANN for well number 18. The average width of the prediction intervals for each one of the representative wells are 0.153 for well 11, 0.243 for well 3 and 0.452 for well 18. By comparing the prediction intervals widths for the two different methodologies it can be noted that the uncertainty derived for the training of the ANNs is only a small fraction of the model uncertainty calculated through the percentile methodology. This may also be attributed to the coarse nature of calculations in the percentile methodology.

Uncertainty in hydraulic head change due to kriging parameters

For every prediction point and time step, 500 different data samples and hence 500 different variograms were constructed. The uncertainty was calculated for the algorithm with the use of 20 neighbors defined by the fuzzy logic system and for the exponential variogram. The 90% prediction intervals (blue range) together with the prediction derived from the initially simulated values (red lines) for 3 prediction points are presented in Figure 4.

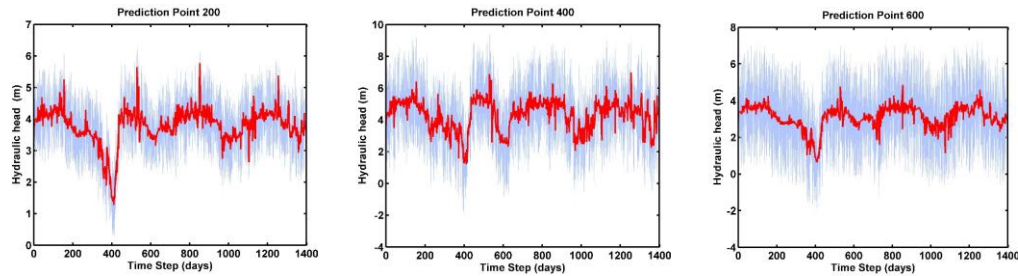


Figure 4: Kriging parameter uncertainty for three prediction points

In all cases the uncertainty of the results, derived from using the described methodology, can be characterized as small, hence the model is consistent and reliable and can be used successfully for the hydraulic head simulation in an aquifer.

4 Conclusions

In the present study, an uncertainty analysis is performed on a combined ANN-Fuzzy logic-kriging methodology for the hydraulic head simulation of an aquifer. The initial simulation of the hydraulic head in a complex study area, in Miami, Dade County, FL, USA had average RMSE training error m, average RMSE testing error m and cross validation RMSE of 0.962m. The uncertainty analysis proved that the methodology used is both consistent and accurate, especially considering the complexity of the case study and the methodology involved.

Using the percentile methodology, the ANN uncertainty can be calculated. In this case the 90% prediction intervals are wider than those produced by all other methodologies involved in this study. This can be attributed to the coarse nature of the calculations in this methodology.

Monte Carlo method for assessing the uncertainty attributed to the ANN training and consecutively for sensitivity analysis of the kriging results to the ANN training is also performed. The range of training and testing error varied, without, however, having a large effect on the 90% prediction interval of the hydraulic head, in locations where data were available. The 90% prediction interval of the hydraulic head is also depicted in 4 prediction points, having narrow intervals in all cases.

Kriging parameter uncertainty reflects the uncertainty attributed to the observed data used. Using artificial data and the Bayesian kriging methodology, this aspect is examined. The results in terms of predicted hydraulic head intervals are close to the simulated results using the real data.

All the above mentioned uncertainty calculations indicate that the methodology used can provide consistent and reliable results under various conditions, hence it can be used for groundwater level simulation, especially in complex study areas where geological information are obscure, making conventional modelling unsatisfactory.

References

- Dehghani, M., Saghafian, B., Saleh, F. N., Farokhnia, A., & Noori, R. (2014). Uncertainty analysis of streamflow drought forecast using artificial neural networks and Monte-Carlo simulation. *Int J Climatol*, 34(4): 1169-1180.
- Delbari, M., Amiri, M., & Motlagh, M. (2014). Assessing groundwater quality for irrigation using indicator kriging method. *Appl Water Sci*, 1-11.
- Dybowski, R., & Roberts, S. (2001). *Confidence intervals and prediction intervals for feed-forward neural networks. Clinical applications of artificial neural networks*. Cambridge University Press.
- Eslamian, S. (2014). *Handbook of Engineering Hydrology: Environmental Hydrology and Water Management*. CRC Press.
- Fine, T. (1999). *Feedforward Neural Network Methodology*. Springer Science & Business Media, Springer-Verlag New York.
- Fish, J., & Stewart, M. (US Department of the Interior, US Geological Survey). *Hydrogeology of the surficial aquifer system, Dade County, Florida*. 1991.
- Jiang, Y., Nan, Z., & Yang, S. (2013). Risk assessment of water quality using Monte Carlo simulation and artificial neural network method. *J Environ Manage*, 122(0): 130-136.
- Journel, A., & Huijbregts, C. (1978). *Mining geostatistics*. Academic Press.
- Kasiviswanathan, K., & Sudheer, K. (2013). Quantification of the predictive uncertainty of artificial neural network based river flow forecast models. *Stoch Environ Res Risk Assess*, 27: 137-146.
- Maiti, S., & Tiwari, R. (2014). A comparative study of artificial neural networks, Bayesian neural networks and adaptive neuro-fuzzy inference system in groundwater level prediction. *Environ Earth Sci*, 71(7): 3147-3160.
- Pilz, J., & Spöck, G. (2008). Why do we need and how should we implement Bayesian kriging methods. *Stoch Environ Res Risk Assess*, 22(5): 621-632.
- Schaefli, B., & Gupta, H. (2007). Do Nash values have value? *Hydrol Processes*, 21(15): 2075-2080.
- Shrestha, D., & Solomatine, D. (2006). Machine learning approaches for estimation of prediction interval for the model output. *Neural Networks*, 19(2): 225-235.
- Tapoglou, E., Karatzas, G., Trichakis, I., & Varouchakis, E. (2014). A spatio-temporal hybrid neural network-Kriging model for groundwater level simulation. *J Hydrol*, 519, Part D(0): 3193-3203.
- Trichakis, I., Nikolos, I., & Karatzas, G. (2011). Comparison of bootstrap confidence intervals for an ANN model of a karstic aquifer response. *Hydrol Processes*, 25(18): 2827-2836.
- USGS. (2017). *Geological units in Miami-Dade county, Florida*. Retrieved from mrdata.usgs.gov/geology/state/fips-unit.php?code=f12086