Technical Note: Application of artificial neural networks in groundwater table forecasting – a case study in Singapore swamp forest

Yabin Sun [a, *], Dadiyorto Wendi [a], Dong Eon Kim [a], Shie-Yui Liong [a]

[a]: Tropical Marine Science Institute, National University of Singapore, 18 Kent Ridge Road, Singapore 119227, Singapore

[*]: Corresponding author

E-mail: tmssy@nus.edu.sg; Phone: +65-96379168; Fax: +65-67761455

Abstract

Accurate prediction of groundwater table is important for the efficient management of groundwater resources. Despite being the most widely used tools for depicting the hydrological regime, numerical models suffer from formidable constraints, such as extensive data demanding, high computational cost and inevitable parameter uncertainty. Artificial neural networks (ANNs), in contrast, can make predictions on the basis of more easily accessible variables, rather than requiring explicit characterization of the physical systems and prior knowledge of the physical parameters. This study applies ANN to predict the groundwater table in a freshwater swamp forest of Singapore. The inputs to the network are solely the surrounding reservoir levels and rainfall. The results reveal that ANN is able to produce accurate forecast with a leading time of 1 day, whereas the performance decreases when leading time increases to 3 days and 7 days.

Keywords**:** Artificial neural networks; Groundwater table forecasting; Freshwater swamp forest

1. Introduction

Physical-based numerical models are commonly used in groundwater table simulation. Different numerical models have been developed for different regions with different objectives, such as to describe regional groundwater flow patterns, and to understand local hydrological processes. (e.g. Matej et al., 2007; Pool et al., 2011; Yao et al. 2014). Numerical models solve the deterministic equations to simulate the groundwater systems based on the knowledge of the system characteristics, initial conditions, system forcings, etc. To develop a groundwater numerical model, essential data include: topography, geological coverage, soil properties, land use map, vegetation distribution, evapotranspiration information, hydrologic and climatic data, etc. Extensive data demanding makes numerical models highly data dependent and data sensitive. Fitting a physical model is not possible when data are not sufficient, and the accuracy of the numerical model to a great extent depends on how accurate the model inputs are. Numerical models are also less competent in forecast as most of the system forcings (e.g. evapotranspiration, rainfall) are less predictable. As a result of aforementioned constraints, numerical models tend to produce imperfect results in spite of the perfect knowledge of the governing laws (Sun et al., 2010).

To combat the deficiencies of the numerical models, artificial neural networks (ANNs) have emerged as an alternative modelling and forecasting approach with a variety of applications in hydrology research (e.g. French et al., 1992; Maier and Dandy, 2000). Unlike the traditional physical-based models, the ANN-based approach does not require explicit characterization of the physical properties, or accurate representation of the physical parameters, but rather simply determines the system patterns based on the relationships between inputs and outputs mapped in the training process. ANNs typically use input variables that are more accessible to make predictions, and therefore circumvent the data reliance inherent to the numerical models. As compared to classical regression techniques, e.g. linear regression model, ANNs are capable of taking into account of the nonlinear dynamics of the hydrological processes and hence result in superior modelling and forecasting performance.

ANNs in recent years have also been successfully applied in groundwater table modelling. Yang et al. (1997) utilized ANN to predict groundwater table variations in subsurface-drained farmland. Coulibaly et al. (2001) calibrated three different ANN models using groundwater recordings and other hydro-meteorological data to simulate groundwater table fluctuations. Lallahem et al. (2005) showed the feasibility of using ANN to estimate groundwater level in an unconfined chalky aquifer. Daliakopoulous et al. (2005) examined the performance of different ANN architectures and training algorithms in groundwater table forecasting. Taormina et al. (2012) developed a two-step ANN model to simulate the groundwater fluctuations in a coastal aquifer using past observed groundwater levels and external inputs, i.e., evapotranspiration and rainfall. Most of above studies, however, focus on applying ANN in large-scale semiarid or arid watersheds, where groundwater table is less variable and long-term groundwater table variation (e.g. monthly, annually) is of more concerns. In addition, these studies use historical groundwater tables as inputs to the network, requiring continuously long groundwater table recordings which can be a luxury for many regions.

This study, for the first time, applies ANN to forecast the groundwater table in a tropical wetland – the Nee Soon Swamp Forest (NSSF) in Singapore. Being nourished with water supply from reservoirs and precipitation, the groundwater table in the NSSF is close to the ground level and extremely sensitive to the changes in hydro-meteorological conditions. This study selects surrounding reservoir levels and rainfall as inputs to the network, avoiding the requirement on continuously long groundwater table recordings. The forecast is made with 3 leading times, i.e., 1 day, 3 days and 7 days, which provides sufficient reaction time for human intervention to maintain favorable hydrological conditions for conserving local ecosystem. The methodology, application, results and conclusions are elaborated in the following sections.

2. Methodology

2.1 Overview

As defined by Haykin (1999), artificial neural networks (ANNs) are massively parallel distributed processors made up of simple processing units, known as neurons, which have a natural propensity for storing experiential knowledge and making it available for use. ANNs are inspired by biological neural networks to emulate the way in which human brains function.

The fact that neurons can be interconnected in numerous ways results in numerous possible topologies that can be divided into two basic classes, i.e., feedforward neural networks (FNNs) and recurrent neural networks (RNNs; Graves et al., 2009). In FNNs information flows from inputs to outputs in only one direction, whereas in RNNs some of the information can flow not only in one direction from inputs to outputs but also in opposite direction.

There are many algorithms for training neural network models, most of which employ some form of gradient descent using backpropagation to compute the actual gradients (Werbos, 1974). The backpropagation algorithm is implemented by taking the derivatives of the cost function with respect to the synaptic weights and then changing the weights in a gradient-related direction (Sexton and Dorsey 2000; Mandischer, 2002).

This study opts for a standard FNN and a quasi-Newton training algorithm, more specifically a multilayer perceptron (MLP) trained with the Levenberg-Marquardt (LM) algorithm, attributing to its superior accuracy in groundwater table forecasting (Daliakopoulous et al., 2005).

2.2 Multilayer perceptron

Multilayer perceptron (MLP) was developed for pattern classification by Rosenblatt (1958). The architecture of a typical MLP consists of an input layer, one hidden layer and an output layer. In mathematical terms, a computational neuron in the hidden or output layers can be described by following pair of equations:

$$u = \sum_{i=1}^{n} w_i x_i \tag{1}$$

and

$$y = \varphi\left(u + b\right) \tag{2}$$

where x_1, x_2, …, x_n are the input signals to the neuron, w_1, w_2, …, w_n are the synaptic weights, u is the linear combiner of the input signals, b is the bias, y is the output signal of the neuron, whereas $\varphi(\cdot)$ is the activation function to limit the amplitude of the output signal and to create a mapping between the input and output signals.

The universal approximation theorem states that every continuous function defined on a closed and bounded set can be approximated arbitrarily closely by a MLP provided that the number of neurons in the hidden layers is sufficiently high and that their activation functions belong to a restricted class of functions with particular properties (Hornik et al., 1989).

2.3 Levenberg-Marquardt algorithm

The Levenberg-Marquardt (LM) algorithm, independently developed by Levenberg (1944) and Marquardt (1963), provides a numerical solution to the problem of minimizing a nonlinear function. The update rule of the LM algorithm can be presented as follows:

$$w_k = w_k - \left(J_k^T J_k + \mu_k I\right)^{-1} J_k e_k \tag{5}$$

where k is the iteration index, J is the Jacobian matrix, μ is the combination coefficient, I is the identity matrix and e is the error vector.

The LM algorithm essentially blends the steepest descent method and the Gauss–Newton algorithm. The optimization process is guided by the combination coefficient μ. Around the error surface with complex curvature, the LM algorithm switches to the steepest descent algorithm with a bigger μ, whereas if the local curvature is proper to make a quadratic approximation, μ can be decreased, giving the LM algorithm a step closer to the Gauss–Newton algorithm. The LM algorithm is faster, more stable and less easily trapped in local minima than other algorithms (Toth et al., 2000).

3. Application

3.1 Study case

Figure 1 shows the geographical location of the study area – the Nee Soon Swamp Forecast (NSSF) in Singapore. The NSSF is located in the northern part of the Singapore central catchment nature reserve bounded by the Upper Seletar, Upper Peirce and Lower Peirce reservoirs. As the only substantial freshwater swamp forest remaining in Singapore Island, the NSSF houses a diversity of flora and fauna some of which are found nowhere else in Singapore or the world (Karunasingha et al., 2013).

With an estimated area of about 750 ha, the NSSF covers the lower area of shallow valleys with slow-flowing streams and a few higher grounds with dryland forests. The elevation of NSSF ranges between 1 m to 80 m above mean sea level (MSL). The aquifer depth in the NSSF is from 20 m to 40 m, and the major soil type features silty sand with a hydraulic conductivity of 4.05 x 10^{-5} m/s. Figure 1 also depicts the locations of the 4 piezometers installed for groundwater table monitoring. The piezometers are deployed near the streams, where the observed groundwater tables vary between 0 to 1 m below the ground level.

3.2 ANN setup

The surrounding reservoirs serve as important fresh water storage for Singapore, with reservoir levels being kept at relatively high levels ranging from 10 to 40 m above MSL. Singapore has a typical tropical rainforest climate with abundant rainfall; the annual rainfall at the NSSF region can be as high as 3,000 mm. Despite being another important influential factor for the groundwater, observed evapotranspiration is not available due to the constraints imposed from setting up monitoring stations in the protected forest, and hence it is excluded in the ANN setup. Reservoir levels and rainfall, as the major water source and driving force, are fed to the networks as inputs, while the output is the observed groundwater tables with a leading time of 1 day, 3 days and 7 days (i.e., future observed groundwater tables after 1 day, 3 days and 7 days).

A multiple-input multiple-output (MIMO) network is selected over 4 multiple-input single-output (MISO) ANNs for 2 reasons: (1) it is easier to implement; and (2) cross-correlation exists in the observed groundwater tables, e.g. the synchronous response to dry and wet conditions; targeting the groundwater table measurements at 4 locations simultaneously, the cross-correlation impact can be captured in the synaptic weights of the trained ANN and hence a better performance is expected. The MIMO network is composed of an input layer with 4 input neurons (including 3 reservoir levels and one rainfall), a hidden layer with 10 neurons (supported by the universal approximation theorem and determined by trial and error), and an output layer with 4 output neurons (future observed groundwater tables at the 4 piezometers). The logistic function and threshold function are respectively adopted as the activation functions for the hidden layer and the output layer.

Daily observed data, i.e., reservoir levels, rainfall and groundwater tables, are available in 2012 and 2013. The data set is divided into 3 subsets as follows:

- Training data (January 2012 to December 2012)

Training data are used for adjusting the synaptic weights in the network. An entire year's data are selected as the training data, so as to expose the network to a complete annual cycle for a robust training.

- Cross validation data (January 2013 to June 2013)

Cross validation data are used for avoiding overfitting. When the errors between the predicted values and desired values in the cross validation data begin to increase, the training stops and this is considered to be the point of best generalization. Half a year's data are selected as the cross validation data.

- Testing data (July 2013 to December 2013)

Testing data are used for evaluating the performance of the network. Once the network is trained, the weights are frozen; the testing set is fed into the network and the network output is then compared with the desired output. Remaining half a year's data are selected as the testing data.

4. Results and discussion

Figure 2 illustrates examples at P1 of the observed groundwater tables, the forecasted groundwater tables from a multiple linear regression (MLR) model and the ANN model. Due to the complicated geological characteristics and hydrological processes, the relationship between the input (reservoir level, rainfall) and the output (groundwater table) is highly nonlinear. Therefore, the MLR model is not suitable to serve our study purpose and produces inferior forecasting results, especially at the extreme values. In contrast, the ANN-forecast successfully resolves the rising and falling tendencies of the groundwater tables, resulting in rather reasonable groundwater table forecast. The scatter plots of the observed groundwater tables and the ANN-forecast are presented in Figure 3. The response of the groundwater tables to the system forcings, for such a confined and wet catchment, is rapid and sensitive. The correlation fades out between the inputs and outputs when the leading time progresses; this leads to the model performance deterioration at 3 days and 7 days. The groundwater tables experience a drastic drop in July and August 2013, caused by a continuous two-month drought. As such a drought condition does not exist in the training data, the ANN tends to over-predict the groundwater tables for that period.

Figure 4 and Figure 5 respectively present the groundwater table time series and scatter plots at P4. P4 is located near the Upper Seletar reservoir, and the groundwater table is affected by the spillway discharge released from the reservoir. Failing to include the spillway information makes the ANN less competent in capturing the groundwater table extreme values caused by the spillway discharge, and hence results in the lower forecast accuracy at P4.

Table 1 summarizes the ANN forecast efficiency through evaluating based on the testing data the root mean square error (RMSE) and the correlation coefficient (r). The forecast accuracy decreases slightly when the leading time increases due to the rapid and sensitive response of the groundwater tables to the system forcings. The RMSE is in general within 10 cm with the exception at P4 caused by the absence of the spillway information. Averaged over the 3 leading times, at P1 to P3 the RMSE is less than 8.0 cm with correlation coefficient r higher than 0.7, whereas at P4 the averaged RMSE and correlation coefficient r are respectively 13.8 cm and 0.67.

5. Conclusions

This study, for the first time, applies artificial neural networks (ANNs) to predict the groundwater table variations in a tropical wetland – the Nee Soon Swamp Forest (NSSF) in Singapore. The ANN model solely utilizes the easily accessible surrounding reservoir levels and rainfall as inputs to forecast the groundwater tables, without requiring any other prior knowledge of the system's physical properties. The ANN forecast shows in general promising accuracy, while its performance decreases when the leading time progresses due to the fading correlation between the network inputs and outputs.

In this study, surrounding reservoir levels and rainfall are selected as ANN inputs. The limited number of inputs eliminates the data demanding restrictions inherent in the numerical models. However, improvements are expected if more variables can be involved in the training/cross-validation/testing process; such variables, for example, are spillway discharge, evapotranspiration, soil properties, water level measurements. Less data demanding, lower computational cost and higher site-specific forecast accuracy are the advantages of the ANN-based approach over the physical-based numerical models. Numerical models, however, can be applied to describe the spatio-temporal variations of the system process over the entire model domain provided with sufficient information of the model inputs. Therefore, the ANN and numerical model can act as natural complements in such a way that ANN is more suitable for site-specific forecast while the numerical model provides a better spatial coverage.

Acknowledgements

This study forms part of the research project "Nee Soon Swamp Forest Biodiversity and Hydrology Baseline Studies—Phase 2" funded by National Parks Board (NParks), Singapore. The authors are also grateful of the data support from Public Utilities Board (PUB), Singapore for making this study possible.

References

Coulibaly, P., Anctil, F., Aravena, R., Bobee, B., 2001. Artificial neural network modeling of water table depth fluctuations. Water Resources Research 37 (4), 885–896.

Daliakopoulosa, I.N., Coulibaly, P., Tsanis, I.K., 2005. Groundwater level forecasting using artificial neural networks. Journal of Hydrology 309, 229–240.

French, M.N., Krajewski, W.F., Cuykendall, R.R., 1992. Rainfall forecasting in space and time using a neural network. Journal of Hydrology 137, 1–31.

Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., Schmidhuber, J., 2009. A novel connectionist system for unconstrained handwriting recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (5), 855–868.

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation. Prentice Hall, New Jersey.

Hornik, K., Stinchcombe, M., White, M., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.

Karunasingha, D.S.K., Chui, T.F.M., Liong, S.Y., 2013. An approach for modelling the effects of changes in hydrological environmental variables on tropical primary forest vegetation. Journal of Hydrology 505 (2013), 102–112.

Lallahem, S., Mania, J., Hani, A., Najjar, Y., 2005. On the use of neural networks to evaluate groundwater levels in fractured media. Journal of Hydrology 307, 92–111.

Levenberg, K., 1944. A method for the solution of certain problems in least squares. Quarterly of Applied Mathematics 5, 164–168.

Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. Environmental Modelling & Software 15, 101–124.

Mandischer, M., 2002. A comparison of evolution strategies and backpropagation for neural network training. Neurocomputing 42 (1-4), 87–117.

Marquardt, D., 1963. An algorithm for least-squares estimation of nonlinear parameters. SIAM Journal on Applied Mathematics 11 (2), 431–441.

Matej, G., Isabelle, W., Jan, M., 2007. Regional groundwater model of north-east Belgium. Journal of Hydrology 335, 133–139.

Pool, D.R., Blasch, K.W., Callegary, J.B., Leake, S.A., Graser, L.F., 2011. Regional Groundwater-Flow Model of the Redwall-Muav, Coconino, and Alluvial Basin Aquifer Systems of Northern and Central Arizona: USGS Scientific Investigation Report 2010-5180, v. 1.1, 101.

Rosenblatt, F., 1958. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review 65 (6), 386–408.

Sexton, R.S., Dorsey, E.D., 2000. Reliable classification using neural networks: a Genetic Algorithm and backpropagation comparison. Decision Support Systems 30, 11–22.

Sun, Y., Babovic, V., Chan, E.S., 2010. Multi-step-ahead model error prediction using time-delay neural networks combined with chaos theory. Journal of Hydrology 395 (2010), 109–116.

Taormina, R., Chau, K-W., Sethi, R., 2012. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. Engineering Applications of Artificial Intelligence, 25 (2012), 1670–1676.

Toth, E., Brath, A., Montanari, A., 2000. Comparison of short-term rainfall prediction models for real-time flood forecasting. Journal of Hydrology 239, 132–147.

Werbos, P.J., 1974. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. Ph.D. Thesis, Harvard University, Cambridge, MA.

Yang, C.C., Prasher, S.O., Lacroix, R., Sreekanth, S., Patni, N.K., Masse, L., 1997. Artificial neural network model for subsurface-drained farmlands. Journal of Irrigation and Drainage Engineering 123 (4), 285–292.

Yao, Y., Zheng, C., Liu, J., Cao, G., Xiao, H., Li, H., Li, W., 2015. Conceptual and numerical models for groundwater flow in an arid inland river basin. Hydrological Processes 29 (6), 1480–1492.

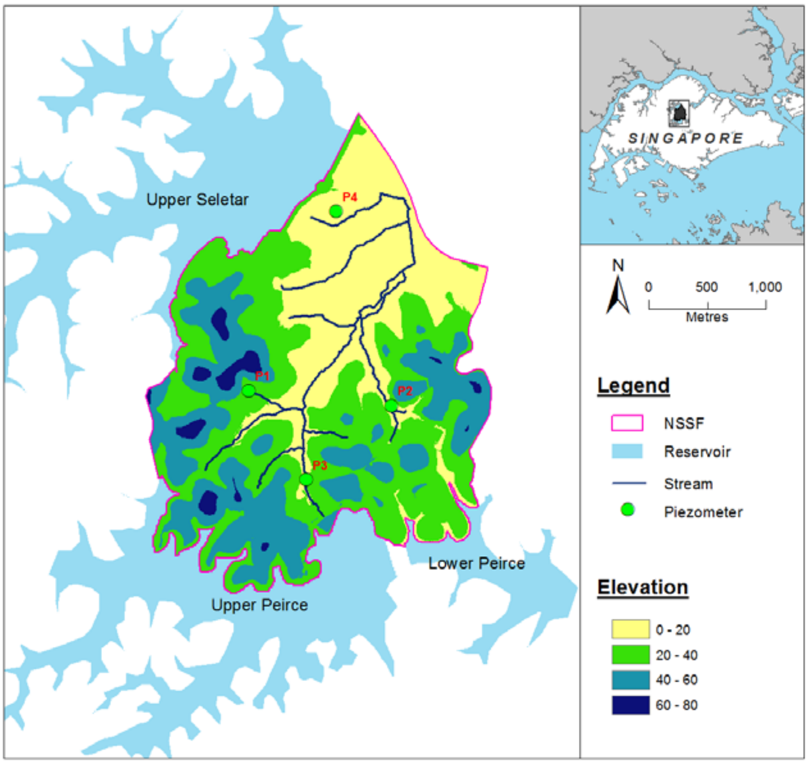

Figure 1 Geographical location of the Nee Soon Swamp Forest in Singapore

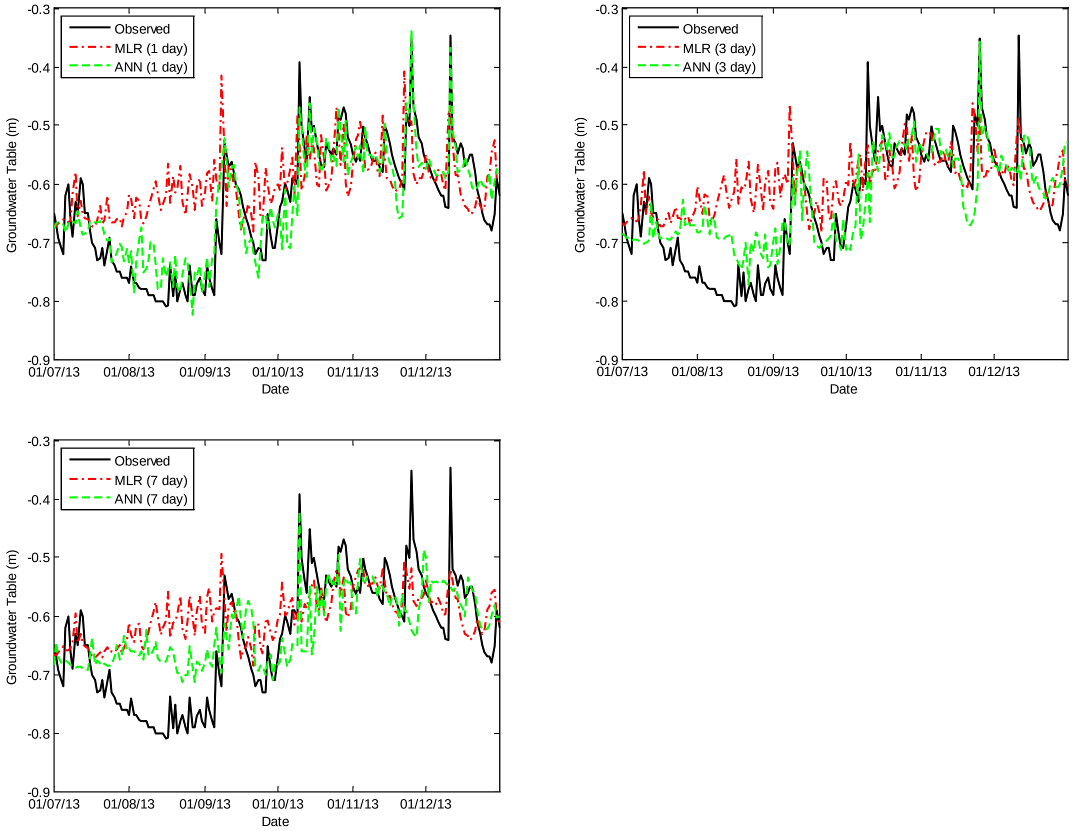

Figure 2 Observed vs. MLR- and ANN-forecasted groundwater tables (P1)

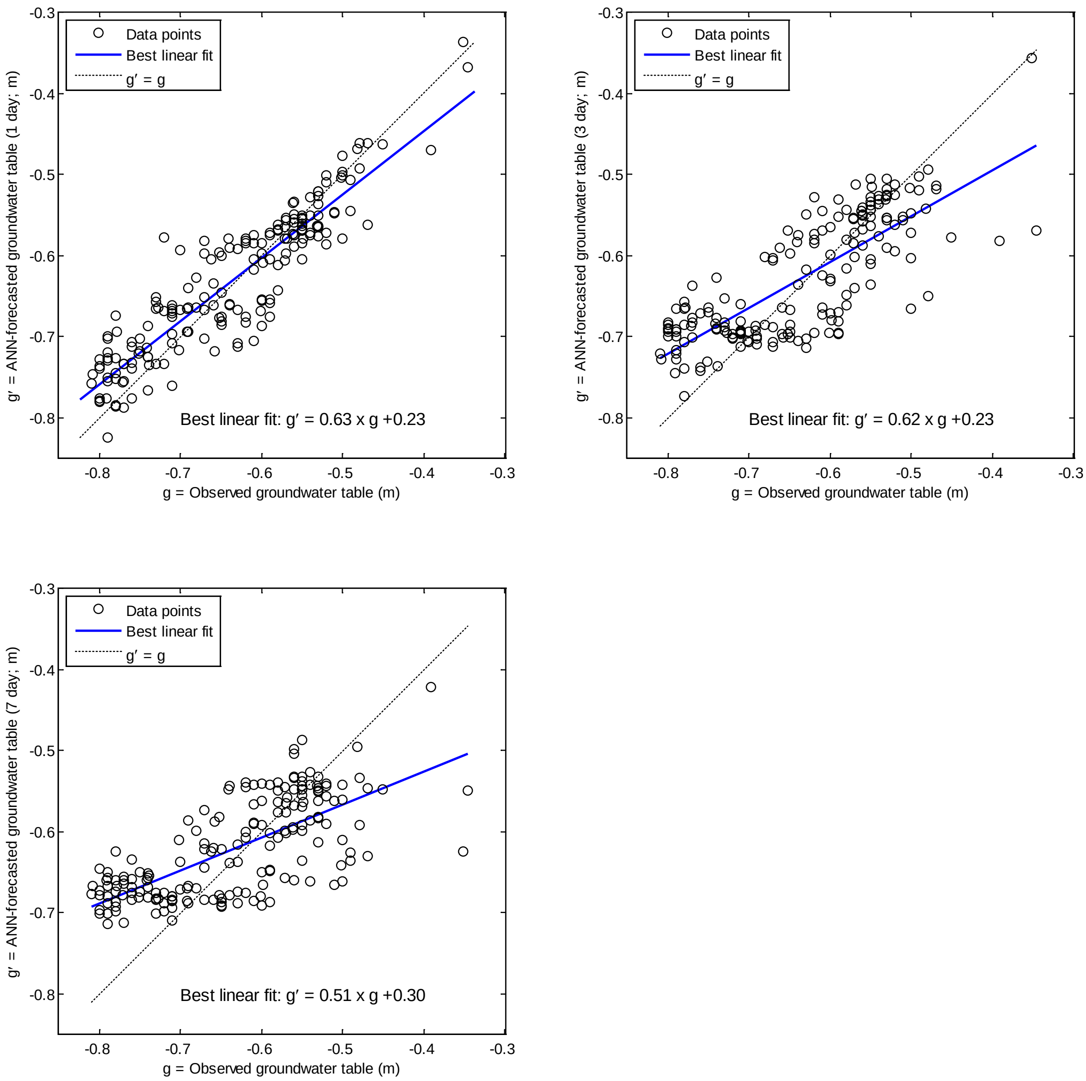

Figure 3 Scatter plots of observed and ANN-forecasted groundwater tables (P1)

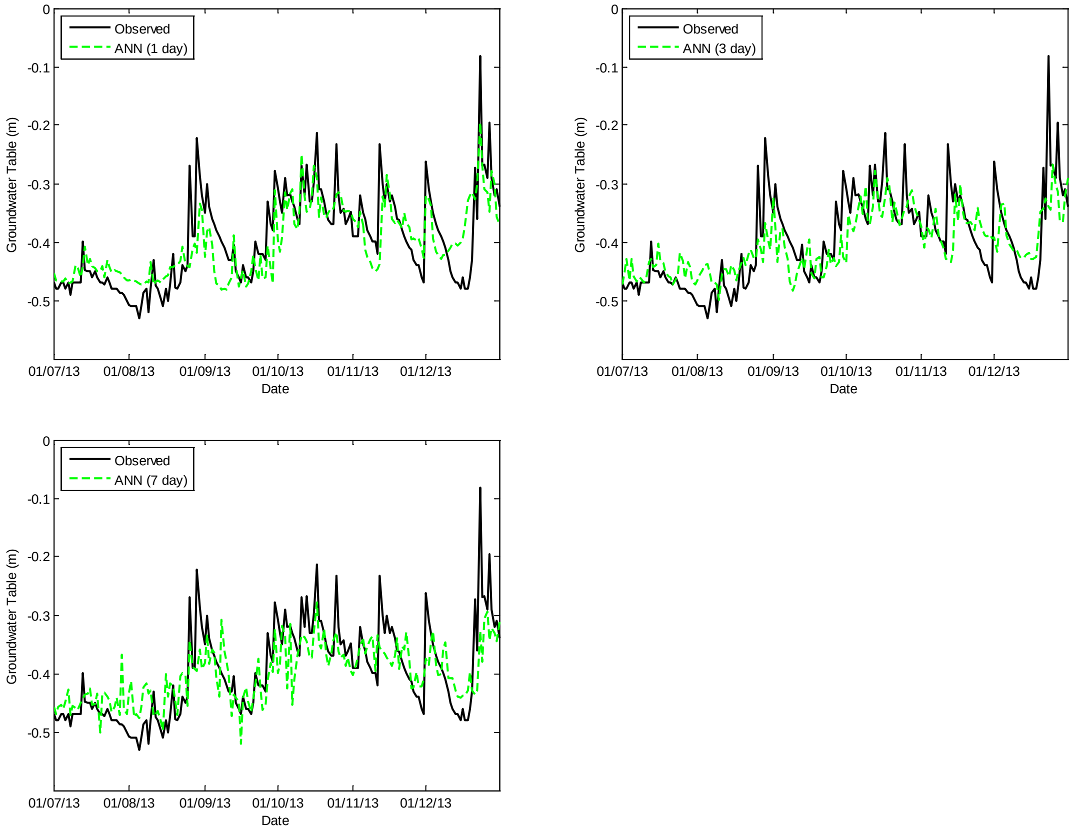

Figure 4 Observed vs. ANN-forecasted groundwater tables (P4)

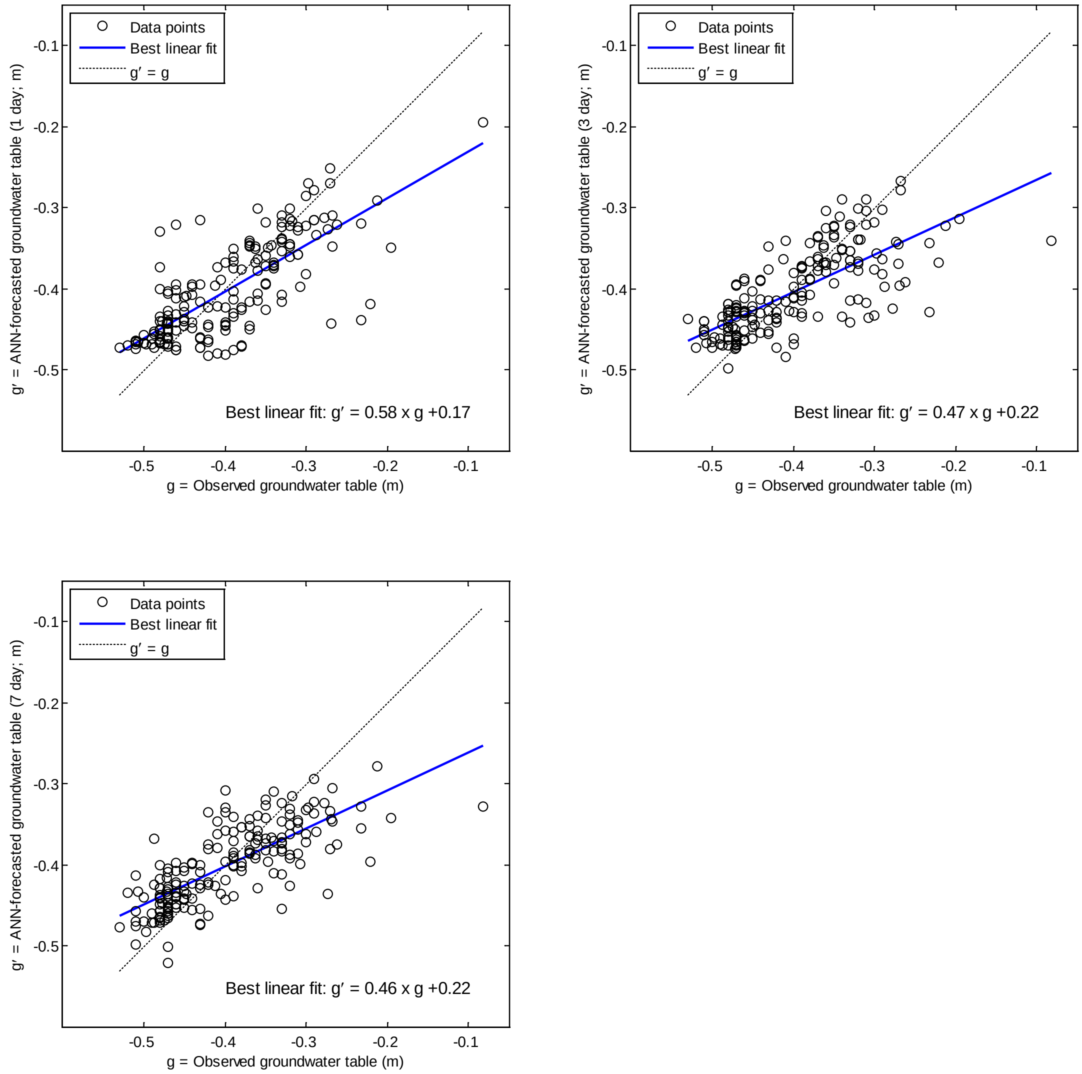

Figure 5 Scatter plots of observed and ANN-forecasted groundwater tables (P4)

Table 1 Evaluation statistics of the ANN forecast

	P1		P2		P3		P4	
	RMSE (cm)	r	RMSE (cm)	r	RMSE (cm)	r	RMSE (cm)	r
1day	5.4	0.88	6.4	0.78	5.2	0.77	12.2	0.69
3 day	8.2	0.76	7.1	0.76	6.6	0.71	13.3	0.68
7 day	9.9	0.64	9.2	0.72	8.6	0.67	15.8	0.65
Average	7.8	0.76	7.6	0.75	6.8	0.72	13.8	0.67