Comparison of three forecasting models for groundwater levels: a case study in the semiarid area of west Jilin Province, China
Zhao Ying, Lu Wenxi, Chu Haibo and Luo Jiannan

ABSTRACT
As groundwater is a critical source of water for both drinking and agriculture in Jilin Province, China, it is important to investigate and understand groundwater level dynamics in this area. Time-series analysis and artificial neural networks (ANN) are commonly used for analysing and forecasting groundwater levels. The integrated time-series (ITS) and auto-regressive integrated moving average (ARIMA) are the most commonly used models for time-series analysis. Among ANN approaches, the radial basis function neural network (RBFNN) is a widely used model for making empirical forecasts of hydrological variables. There are no previous reports comparing the ITS, ARIMA and RBFNN models together in groundwater-level dynamics literature. An attempt has been made in this study to investigate the applicability of these three models for the prediction of groundwater levels based on root mean squared error, the Nash-Sutcliffe coefficient and mean absolute error. The results indicated that all three models reproduced groundwater levels accurately. In addition, the RBFNN model was more reliable than ITS and ARIMA. This provides a choice in the selection of models for analysis and prediction of groundwater levels. The predicted results also provide a basis for rational exploitation and sustainable utilization of groundwater resources.

Key words | forecasting, groundwater level, Jilin Province, radial basis function neural networks, time-series analysis

INTRODUCTION
Western Jilin Province, located in the northeast of China, is one of China’s most important agricultural and livestock areas. With a semiarid continental climate, it is among the largest grain-producing provinces and is one of the regions designated for future crop expansion. The normal annual precipitation is about 400 mm, and the mean annual potential evapotranspiration for the study area is generally taken as 1,900 mm (Bian et al. 2012). Groundwater in the region is crucial for crop growth and production, and knowledge of the dynamics is necessary for ensuring a sustainable water resource. Many unauthorized wells are used in the region and pumping volume is not regulated, resulting in overexploitation of the aquifer and lowering of the water table over many years (Ma et al. 2012).

Although conceptual and physically based models are the main tools for depicting hydrological variables and understanding the physical processes taking place in such a system, they have their practical limitations. When insufficient data is available, and making accurate predictions is more important than understanding the physics, empirical models remain a reasonable alternative, providing useful results without taking up costly calibration time (Daliakopoulos et al. 2005; Nikolos et al. 2008). Time-series models and artificial neural network (ANN) models are such ‘black box’ models whose properties are ideally suited to dynamic system modelling.

Time-series models based on stochastic theory include the integrated time-series (ITS) model, the auto-regressive
moving average (ARMA) model, the auto-regressive integrated moving average (ARIMA) model, the seasonal auto-regressive moving average model and the periodic auto-regressive model (Ahn 2000; Wong et al. 2007; Yang et al. 2009). A major advantage of time-series models is that they are easy to use without requiring a great deal of other data, and the usefulness of their results can be accurately assessed.

ANN models have been used for groundwater level prediction (e.g. Krishna et al. 2008; Tsanis et al. 2008; Banerjee et al. 2009; Sreepan et al. 2009; Sethi et al. 2010; Heesung et al. 2011; Yi et al. 2012; Mohanty et al. 2013), aquifer parameter determination (Samani et al. 2007; Karahan & Ayvaz 2008) and groundwater quality monitoring (Milot et al. 2002; Farmaki et al. 2013). An important feature of ANN models is their ability to detect patterns in complex systems (Adamowski & Chan 2011).

There have been a number of comparative studies including time-series models and ANN (Yang et al. 2009; Claveria & Torra 2014). They either compare ARIMA with ANN or contrast ITS with back propagation ANN. In the present study, the potential of two often-used time-series models (ITS and ARIMA) and the radial basis function neural network (RBFNN) model were evaluated for predicting groundwater levels. There are no previous reports in the groundwater-level dynamics literature that compare ITS, ARIMA and RBFNN models together. The aims of the present study were (1) to apply and compare the advantages and disadvantages of the three models for forecasting groundwater levels in areas where the water table has been steadily falling over the past decade due to overexploitation and (2) to provide reliable data on which to base reasonable exploitation that would ensure sustainable utilization of groundwater.

**METHODOLOGY**

**Time-series analysis**

**ITS models**

The principle of ITS models is to separate a time series into four major components (trend, seasonal, periodic and random), and then add these together to get the final forecasting model. The basic equation is

\[ H_t = X_t + S_t + P_t + R_t \]  \hspace{1cm} (1)

where \( H_t \) is the time series; \( X_t \) is the trend component; \( S_t \) is the seasonal component; \( P_t \) is the periodic component; and \( R_t \) is the random component.

In most published articles on this subject (e.g. Zhao et al. 2007; Zhou et al. 2007; Yang et al. 2009; Lu et al. 2012), time series are divided into three parts – trend, periodic and random components. The seasonal component \( S_t \) was added in the present study, as it reflects the influence of rainfall variation, which is an important consideration in the analysis of groundwater levels, particularly in the semi-arid regions.

\( X_t \) stands for the general trend in a single series. It can be determined in many ways, including smoothing or polynomial fitting techniques. In this study, polynomial fitting was chosen because it is readily done in Excel software using the ‘add trendline’ tool.

\( S_t \) indicates how the series varies throughout the year and was found to be strongly correlated with rainfall. It was determined by a multi-year averaging method: after extracting the trend component, the series of \( H_t - X_t \) is obtained; then the average value of \( H_t - X_t \) is calculated for corresponding months in different years. From this, 12 weighting values – one for each month – were calculated. After determining the seasonal component, it was then removed, so that \( H_t - X_t - S_t \) becomes the series to be analysed when determining the periodic component.

\( P_t \), the periodic component, signals the interannual variability of the series. The harmonic wave analysis method of Olsson & Eklundh (1994) was used to determine its value. This method supposes that the periodic component should comprise many different cyclic waves, and therefore may be expressed as the Fourier series

\[ \hat{p}_t = \frac{a_0}{2} \sum_{k=1}^{L} \left[ a_k \cos \frac{2\pi k t}{n} + b_k \sin \frac{2\pi k t}{n} \right] \]  \hspace{1cm} (2)

where \( \hat{p}_t \) is the estimated value of \( p_t \); \( a_0 \) is the average of \( p_t \); \( L \) is the magnitude of the waves; \( k \) is the number of waves;
\( a_r, b_h \) are coefficients; and \( n \) is the number of samples. After it has been calculated, the periodic component is removed. \( H_t = X_t - S_t - P_t \) then becomes the series to be analysed when determining the random component.

\( R_t \), the random component, is the last to be calculated. It is influenced by many uncertain factors, such as noise. It was extracted by the auto-regression method (Gemiti & Stefanopoulos 2011)

\[
\hat{R}_t = \Phi_0 + \Phi_1 r_{t-1} + \Phi_2 r_{t-2} + \ldots + \Phi_p r_{t-p} + e_t, \quad (3)
\]

where \( p \) is the model order determined by the Akaike information criterion (AIC) and \( \Phi_1 \) is the autoregression coefficient. The \( p \) value is selected when the AIC(\( p \)) takes a minimum value from

\[
AIC(\rho) = n \ln \hat{\sigma}^2 + 2p, \quad (4)
\]

where \( n \) is the amount of data; \( \hat{\sigma}^2 \) is the variance of the residuals of AR(\( p \)); and \( e_t \) represents the residuals of evaluation (Akaike 1969).

**ARIMA models**

For more than half a century, ARIMA models have dominated many areas of time-series forecasting. In the ARIMA \((p,d,q)\) model, the future value of a variable is assumed to be a linear extrapolation of several past observations and random errors. That is, the underlying process that generates the time series with a mean value \( \mu \) has the form

\[
\Phi(B)\nabla^d(y_t - \mu) = \theta(B)a_t, \quad (5)
\]

where \( y_t \) and \( a_t \) are the actual value and random error for the time period \( t \), respectively; \( \Phi(B) = 1 - \sum_{i=1}^{p} \varphi_i B^i \), \( \theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j \), are polynomials in \( B \) of degree \( p \) and \( q \); \( \varphi_i \), \( i = 1, 2, \ldots, p \) and \( \theta_j \), \( j = 1, 2, \ldots, q \) are the model parameters; \( \nabla = 1 - B \); \( B \) is the backward shift operator; \( p, q \) are the orders of the model; and \( d \) is the order of differencing. Random errors \( a_t \) are assumed to be independently and identically distributed with a mean of zero and a constant variance of \( \sigma^2 \).

**Radial basis function neural networks**

The RBFNN was proposed by Moody and Darken in the late 1980s and has been widely used for classification or function approximation since then (e.g. Chu et al. 2013). It mainly consists of three layers: an input layer, a hidden layer and an output layer. A single output RBFNN with \( K \) hidden layer neurons is expressed as:

\[
y_k = \sum_{i=1}^{m} w_{ik} R_i(x) + \theta_k \quad (6)
\]

where \( y_k \) is the \( k \)th output node on the output layer; \( w_{ik} \) is the weight connection between \( i \)th hidden and \( k \)th output nodes; and \( \theta_k \) is the threshold value of the \( k \)th output node.

The most commonly used function in the hidden layer is the Gaussian function, given by

\[
R_i(x) = e^{-\|x - c_i\|^2 / 2\sigma_i^2} \quad i = 1, 2, \ldots, m \quad (7)
\]

where \( x \) is the \( n \)-dimensional input vector; \( c_i \) is the centre of the \( i \)th radial basis function; \( \sigma_i \) is the spread of the radial basis function in the \( i \)th hidden node that indicates the radial distance from the RBF centre; \( m \) is the number of hidden nodes; and \( ||x - c_i|| \) is the radial distance from \( X \) to the RBF centre.

The training process for RBFNN may be divided into two stages. The first is forward propagation, where data is processed from the input layer to the hidden layer and finally to the output layer. Through the calculations in every layer, the actual outputs were obtained. The second stage is reverse propagation of the error. The error between the network output and the actual value is calculated. If the error does not fall within a permissible value, the network weight and spread of the RBF are adjusted. The error is propagated backwards from the output layer to the hidden layer and then to the input layer. Forward and reverse propagation are performed in a single iteration, which is repeated until the error reaches its permissible value, or until the learning cycle time has elapsed.

The number of hidden neurons is found to have a large influence on the output of the RBF. The optimal number of hidden neurons is able to be identified by the root mean squared error (RMSE). When the RMSE drops to the
lowest point, the corresponding number is the optimal number of hidden neurons.

Criteria of evaluation

Three standard statistical measures, that is, RMSE, the Nash–Sutcliffe coefficient (NS) and mean absolute error (MAE), were used to evaluate the performances of the forecasting models. The RMSE evaluates the residual between observed and forecast values and NS evaluates the capability of the model to forecast groundwater level away from the mean. MAE is a measure of the difference between observed and forecast values. The closer the NS value is to 100, the more accurate the forecasting. The closer the RMSE and MAE values are to 0, the more accurate the forecasting. These are calculated from

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (f_i - F_i)^2},
\]

\[
NS = 1 - \left[ \frac{\sum_{i=1}^{n} (f_i - F_i)^2}{\sum_{i=1}^{n} (\bar{f}_i - \bar{F})^2} \right] \times 100
\]

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} (f_i - F_i)^2
\]

where \( f \) is observed data; \( F \) is the predicted data; and \( \bar{f} \) is the mean observed value.

STUDY AREA AND DATA DESCRIPTION

The three models were tested with data taken from the western part of Jilin Province, China (Figure 1). Two typical wells located in Baicheng City and Songyuan City were chosen because groundwater levels at those locations had declined in the recent decades due to overexploitation. The level was measured monthly from 1986 to 2013. The time series was divided in two data subsets, one for establishing the time-series models and training the neural network (1986–2005), the other for model validation (2006–2013).

MODELLING

ITS model

All tests and results were obtained by using MATLAB 2011b and Excel 2007 software. The trend component (Figure 2) showed that the depth of the water table fell yearly during that period. The groundwater depth should not increase or decrease continuously under natural (unexploited) conditions. It should be steady or variable within a certain range. Thus the trend component reflects the degree of exploitation of the water resource by residents.
The seasonal component (Figure 3) was greater in January–June than July–December, consistent with the beginning of the rainy season in June. The rise in the water table was delayed because soil in the unsaturated zone absorbs water at the beginning of the rains. As a consequence, the water table began to rise in July. When the rainy season ended, the level dropped to its former state. Thus, the seasonal component implies the fluctuation of groundwater level influenced by rainfall.

From the periodic component (Figure 4), it can be seen that the groundwater levels at both locations had a 7- to 9-year periodicity. Almost three complete cycles were distinguished in the 19-year period from 1986 to 2005. The periodicity is believed to be driven by the cycle of solar activity and the Earth’s rotation and revolution (Zheng 1989). Sunspot activity is thought to influence the alternation of the rainy and dry seasons. The periodicity reflects the influence of natural climatic factors, and is also believed to be driven by the periodicity of rainfall (Hao et al. 2012). In the present study, the hydrological environment was karst, in which the interrelationship of surface and subsurface is strong, but for porous media, and the effect of unsaturated zones, the relationship may be not as obvious. In addition, the periodicity of precipitation in western Jilin Province is 2, 11 and 30 years (Wang et al. 2011). What is more, solar activity and the Earth’s rotation are also thought to influence rainfall, so the driving forces still need further study.

The random component (Figure 5) is influenced by many uncertain factors such as earthquakes and is not discussed further in this study. The fitted graphs are shown in Figure 6.

**ARIMA model**

The historical data indicated that groundwater depth was a non-stationary sequence. For a non-stationary time series,
Figure 4 | Periodic component for Baicheng (a) and Songyuan (b).

Figure 5 | Random component for Baicheng (a) and Songyuan (b).
the differential makes it stationary, in effect. Using the augmented Dickey–Fuller unit root test, the order of differencing exponent, \( d \), was fixed at 1; that is, \( d = 1 \) in the methodology of the ARIMA \((p,d,q)\) model (see Equation (5)). Then the values of \( p \) and \( q \) were calibrated. From autocorrelation and partial correlation, the correlation diagram in Figure 6 was drawn. The default model was chosen as ARIMA \((1,1,1)\). The calibrated parameters are shown in Table 1 and the corresponding fitted graphs in Figure 7.

**RBFNN model**

All test results were obtained using MATLAB 2011b software. The optimum network and parameter configuration were then derived by trial and error. The number of hidden neurons was found to have a large influence on the output of the radial basis function. The optimal number of hidden neurons was able to be identified by the RMSE. Figure 8 shows the effect on the network accuracy of changing the number of hidden neurons from 2 to 12. The training errors (RMSE) were minimal when four hidden neurons were modelled for Baicheng, and three for Songyuan.

Accordingly, for Songyuan, the three-layer RBFNN model contained eight neurons in the input layer, three in the hidden layer and one in the output layer; the spread of the Gaussian function was 20. For the Baicheng model, the model contained eight neurons in the input layer, four in the hidden layer and one in the output layer; the spread of the Gaussian function was 50.

Figure 8 shows the performance of the ITS and AR models in the process of establishing models and the fitting of the RBFNN model in the process of training the network. It is clear that the three models simulated the observations adequately, and consequently the performance of all three was evaluated.
RESULTS AND DISCUSSION

Figure 9 and Table 2 summarize the results of validation of the three methods by the RMSE, NS and MAE. ‘Short-term’ and ‘long-term’ validations mainly refer to 1-year and 8-year validations. The comparisons show whether or not long-term prediction was acceptable in this study (Yang et al. 2009; Shiri et al. 2013).

It can be seen in Figure 10 that all three models were capable of predicting groundwater levels in the short term, but the ARIMA model was not effective in making long-term predictions. The RBFNN method demonstrated the closest agreement with observation, followed by the ITS and ARIMA models.

The same conclusion may be drawn from the information in Table 2. The RBFNN model has smaller RMSE and MAE and larger NS than the ITS and ARIMA models.

For both the 1-year and 8-year validations, a different result is seen for ITS and ARIMA. Validation accuracies of two models were almost identical for 1-year validation; however, significant differences exist for 8-year validation. For example, for Baicheng (Table 2) the RMSE, which evaluates the residual between observed data and model predictions, was respectively 0.20 and 0.19 for the 1-year validation of ITS and ARIMA. The NS, which indicates the deviation of the prediction from the mean, was 97 and 99 respectively.

### Table 1 | Calibrated parameters for Baicheng and Songyuan

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>T-statistic</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baicheng</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR(1)</td>
<td>1.005025</td>
<td>0.003533</td>
<td>284.4735</td>
<td>0.0000</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0.342756</td>
<td>0.061077</td>
<td>5.611857</td>
<td>0.0000</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.989533</td>
<td>Mean dependent variable</td>
<td>4.659205</td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.989489</td>
<td>SD dependent variable</td>
<td>2.001513</td>
<td></td>
</tr>
<tr>
<td>Standard error of squares</td>
<td>0.205206</td>
<td>AIC</td>
<td>$-0.321270$</td>
<td></td>
</tr>
<tr>
<td>Residual sum of squares</td>
<td>9.979972</td>
<td>Schwarz criterion</td>
<td>$-0.292178$</td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>40.39174</td>
<td>Hannah–Quinn criterion</td>
<td>$-0.309547$</td>
<td></td>
</tr>
<tr>
<td>Durbin–Watson statistic</td>
<td>1.971301</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Songyuan</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR(1)</td>
<td>1.000484</td>
<td>0.002914</td>
<td>343.2963</td>
<td>0.0000</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0.317003</td>
<td>0.061588</td>
<td>5.147190</td>
<td>0.0000</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.950668</td>
<td>Mean dependent variable</td>
<td>4.829791</td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.950460</td>
<td>SD dependent variable</td>
<td>0.750493</td>
<td></td>
</tr>
<tr>
<td>Std error of regression</td>
<td>0.167030</td>
<td>AIC</td>
<td>$-0.732953$</td>
<td></td>
</tr>
<tr>
<td>Residual sum of squares</td>
<td>6.612070</td>
<td>Schwarz criterion</td>
<td>$-0.703862$</td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>89.58792</td>
<td>Hannah–Quinn criterion</td>
<td>$-0.721230$</td>
<td></td>
</tr>
<tr>
<td>Durbin–Watson statistic</td>
<td>2.016551</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 7 | RMSE sensitivity to the number of hidden neurons.
for ITS and ARIMA. The MAE, which evaluates the difference between observed and predicted values, was respectively 0.15 and 0.12 for the ITS and ARIMA models. For 8-year validation, the ITS model was much more accurate than the ARIMA model. For example, for Baicheng (Table 2) the RMSE was 0.58 for the ITS model and 310 for the ARIMA model. The NS was 93 and 0.14, and the MAE was 0.47 and 89, respectively.

These results are consistent with those of Yang et al. (2009) in that the RBFNN model is better suited than ITS for predicting groundwater levels although the comparative accuracies of the three-component ITS model in Yang et al. (2009) and the four-component ITS model in the present study cannot be definitely stated because different data was used in the two cases. The performance of each criterion in the Baicheng region was poorer than in Songyuan because the groundwater depth data for Baicheng indicated much stronger variation than for Songyuan. Added to that, the variation of groundwater depth quoted in Yang et al. (2009) lay somewhere between that of Baicheng and Songyuan. As a result of these considerations, the same data for three- and four-component models will be assessed in a further study; nevertheless, incorporating the fourth (seasonal) component in the present work has undoubtedly aided the analysis of groundwater level variation.

Figure 8 | Fitting graphs for Baicheng (a) and Songyuan (b).
In general, the RBFNN model is suited to non-linear dynamic systems, whereas the other two models are more suited to linear dynamic systems. However, groundwater level variation is quite complex, and a non-linear analysis may be the better of the two. Although it is recognized that time-series analysis has its limitations, it also has its advantages; for instance, it reflects the influence of human activity, rainfall and solar fluctuation.

For forecasting the dynamics of the groundwater level, the RBFNN method is preferable, but for analysing...
groundwater level variation, time-series analysis may be more appropriate. However, as expected, the efficiency of all three methods decayed as the prediction period increased. The results in Table 1 show that the 1-year validation was much superior to the 8-year validation; for the latter, the accuracy of the time-series analysis was relatively low. For long-term prediction, time-series analysis is not recommended compared with RBFNN.

**PREDICTION**

Three models were used to predict the groundwater level for the four years following the original experiment (i.e. 2014–2017). The results are shown in Figure 10. For Baicheng, because of the rapid groundwater variation, the ARIMA model was not capable of such a long-term forecast, so Figure 10(a) shows only the ITS and RBFNN forecasts. The results predict that the water table in Baicheng will continue to fall at the rate of 1 m/year if the present degree of exploitation continues.

Figure 10(b) shows small predicted changes in the depth of the water table in Songyuan, remaining at an average depth of 6.5 m.

From the results of this study, it is clear that the ARIMA model was unsuitable for long-term groundwater level prediction and should not be used for this purpose.

**CONCLUSIONS**

All three models accurately simulated the observed groundwater data. The RBFNN model produced the most accurate simulations, especially in long-term predictions, and it is concluded that the RBFNN model was much more reliable than ITS and ARIMA for this purpose.

Despite its greater reliability, the RBFNN model has limitations that should not be overlooked. Conversely, the ITS and ARIMA models have certain advantages; for example, the ITS model is more applicable to the analysis of the factors that regulate groundwater dynamics: the trend, season and periodicity components reflect the degree
of water exploitation, rainfall, solar activity and the Earth’s rotation and revolution around the Sun. The main advantages of the ARIMA model are the ease of calibrating its model parameters, and it is a sound choice for short-term prediction.

From the above results, it is recommended that groundwater exploitation in Baicheng be monitored in view of the continuous lowering of the water table over many years. As the predictions were all based on historical groundwater level measurement data, they all have certain limitations. If more relevant data were obtained, including rainfall data, knowledge of the relationship between the various factors would be improved.

ACKNOWLEDGEMENTS

This study is based on the work supported by the National Natural Science Foundation of China (41372237), National Water Pollution Control and Management of Science and Technology Major Projects (2012ZX07201), Doctoral Program of the Ministry of Education (20120061110058) and Project 2014025 supported by the Graduate Innovation Fund of Jilin University. The authors thank the editor and anonymous reviewers for their constructive comments and suggested revisions.

REFERENCES

Shiri, J., Kisi, O., Yoon, H., Lee, K.-K. & Nazemi, A. H. 2013 Predicting groundwater level fluctuations with meteorological...


