

# A comparative study of shallow groundwater level simulation with three time series models in a coastal aquifer of South China

Q. Yang<sup>1</sup> · Y. Wang<sup>1</sup> · J. Zhang<sup>1</sup> · J. Delgado<sup>2</sup>

Received: 5 November 2014 / Accepted: 23 March 2015 / Published online: 8 April 2015  
© The Author(s) 2015. This article is published with open access at Springerlink.com

**Abstract** Accurate and reliable groundwater level forecasting models can help ensure the sustainable use of a watershed's aquifers for urban and rural water supply. In this paper, three time series analysis methods, Holt–Winters (HW), integrated time series (ITS), and seasonal autoregressive integrated moving average (SARIMA), are explored to simulate the groundwater level in a coastal aquifer, China. The monthly groundwater table depth data collected in a long time series from 2000 to 2011 are simulated and compared with those three time series models. The error criteria are estimated using coefficient of determination ( $R^2$ ), Nash–Sutcliffe model efficiency coefficient ( $E$ ), and root-mean-squared error. The results indicate that three models are all accurate in reproducing the historical time series of groundwater levels. The comparisons of three models show that HW model is more accurate in predicting the groundwater levels than SARIMA and ITS models. It is recommended that additional studies explore this proposed method, which can be used in turn to facilitate the development and implementation of more effective and sustainable groundwater management strategies.

**Keywords** Groundwater table · Time series analysis · Holt–Winters · ARIMA · ITS

## Introduction

Groundwater is often one of the major sources of water supply for domestic, agricultural, and industrial users. In some areas, it is taken as the only dependable source of supply because of its ready availability. However, groundwater supplies for agricultural, industrial, and municipal purposes have been overexploited in many parts of the world. Various consequences of unsustainable groundwater utilization and management have been of great concerns globally, especially in developing countries (Konikow and Kendy 2005). The consequences of aquifer depletion can lead to local water rationing, excessive reductions in yields, wells going dry or producing erratic groundwater quality changes, changes in flow patterns of groundwater in the inflow of poorer quality water and sea water intrusion in coastal areas, and other harmful environmental side effects such as major water-level declines, reduction in water in streams and lakes, increased pumping costs, land subsidence, and decreased well yields have been a great concern to the water managers, engineers, and stakeholders (Adamowski and Chan 2011; Konikow and Kendy 2005; USGS 2010). As a result, many watersheds are experiencing severe environmental, social, and financial problems (Tsanis et al. 2008). As water demand will likely increase in the short and long term, there will be increasing pressures on groundwater resources (Sethi et al. 2010). Therefore, a constant monitoring of the groundwater levels is extremely important. Meanwhile, groundwater systems possess features such as complexity, nonlinearity, being multi-scale and random, all governed by natural and/or anthropogenic factors, which complicate the dynamic predictions. Therefore, many hydrological models have been developed to simulate this complex process. Models based on their involvement of physical characteristics

✉ Q. Yang  
qyang@udc.es

<sup>1</sup> Key Laboratory of Groundwater Resources and Environment Ministry of Education, Jilin University, Changchun 130021, People's Republic of China

<sup>2</sup> Escuela de Ingenieros de Caminos, Universidad de A Coruña, Campus de Elviña, 15192 A Coruña, Spain

generally fall into three main categories: black box models, conceptual models, and physical-based models (Nourani et al. 2010). The well-forecasted water levels in advance may help the administrators to plan better the groundwater utilization. Also, for an overall development of the basin, a continuous forecast of the groundwater level is required to effectively use any simulation model for water management (Nayak et al. 2006). The common used models for simulating groundwater level include BP neural network, wavelet random coupling model, the gray time series combination model, and time series model. Among which, time series model is the most common and suitable one, however, with different features, such as the integrated time series (ITS) model, the autoregressive moving average (ARMA) model, the autoregressive integrated moving average (ARIMA) model, the seasonal autoregressive moving average (SARMA) model, and the periodic autoregressive (PAR) model (Ahn 2000; Wong et al. 2007; Yang et al. 2009). One of the widely used time series models is the ARMA model, which provide a parsimonious description of a (weakly) stationary stochastic process autoregression and the second for the moving average (McNeil et al. 2005). Autoregressive integrated moving average model (ARIMA) and seasonal autoregressive integrated moving average (SARIMA) models are extensions of ARMA class in order to include more realistic dynamics, in particular, respectively, non-stationarity in mean and seasonal behaviors (Behnia and Rezaeian 2015).

This paper demonstrates a case study on how to utilize time series analysis to predict groundwater table in a coastal island, South China. We evaluate and compare the potential of the three time series models [Holt–Winters (HW), SARIMA and ITS] in the study area. The objectives of the present study are (1) to apply and compare the advantages and disadvantages of these three models on simulating groundwater levels and (2) to provide some useful insights and for the reasonable exploitation and sustainable utilization of groundwater.

## Methodology

### Holt–Winters model

Exponential smoothing methods are among the most widely used forecasting techniques in industry and business, in particular the HW methods that allow us to deal with univariate time series which contain both trend and seasonal factors. Their popularity is due to their simple model formulation and good forecasting results (Gardner 1985).

HW is the label we frequently give to a set of procedures that form the core of the exponential smoothing

family of forecasting methods. The basic structures were provided by Holt in 1957 and his student Winters in 1960 (Holt 1957). Its basic idea is to decompose a time series into a linear trend component, seasonal variation component, and random change component, incorporating the exponential smoothing algorithm; the long-term trend ( $S_t$ ), trends incremental ( $b_t$ ), and seasonal changes ( $I_t$ ) are estimated; and then, a predictive model is established to extrapolate the predicted value. Holt's method is widely used for forecasting as reported in the literatures (Bermúdez et al. 2010; Gelper et al. 2010). It is an extended single exponential smoothing, which allows forecasting data with non-constant trends and seasonal variations. Thus, it is also capable for detecting trend in different time periods. Kamruzzaman et al. (2011) applied Holt–Winters seasonal forecasting method to find the evidence of non-stationarity in rainfall and temperature. It was claimed that the Holt–Winters method is capable in tracking changes in the level, trend, and also seasonality; the influence of the random motion can be moderately filtered. Therefore, it is particularly suitable for time series prediction containing the trend and seasonal variation.

In this study, multiplicative form of HW model is used to make simulations, which consists of the following three components:

$$S_t = \alpha \frac{y_t}{I_{t-l}} + (1 - \alpha)(S_{t-1} + b_{t-1}) \quad (1)$$

$$b_t = \beta(S_t - S_{t-1}) + (1 - \beta)b_{t-1} \quad (2)$$

$$I_t = \gamma \frac{y_t}{S_t} + (1 - \gamma)I_{t-l} \quad (3)$$

where  $\alpha$ ,  $\beta$ ,  $\gamma$  is the smoothing coefficient, ranges between 0 and 1.  $S_t$  describes the smoothed series, which is the average of exponential smoothing eliminating the seasonal change impact on the time series;  $b_t$  is used to estimate the trend, which refers to the average of exponential smoothing of change trend time series;  $I_t$  is to estimate seasonality, which refers to the average of seasonal factor exponential smoothing;  $y_t$  represents actual value; and  $l$  represents the length of the season or time period. Thus, the prediction formula of HW seasonal exponential smoothing model can be written as:

$$f_{t+m} = (S_t + mb_t)I_{t-l+m} \quad (4)$$

where  $m$  is the time intervals from the moment to the predicted time.

Due to the influence of various factors, uncertainty and randomness exist in groundwater table depth. The different selection of smooth coefficients  $\alpha$ ,  $\beta$ ,  $\gamma$  will inevitably lead to the unreliability of prediction. Zhao and Wang (2007) suggested using “optimal smooth coefficient”, try any combination of three smoothing coefficient ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) using

exhaustive algorithm in EViews 6.0 software, the corresponding smoothing coefficient is selected as “optimal smoothing coefficient” when the sum of square relative error reaches the minimum, which is expressed as Eq. (5).

$$\sum_{i=1}^n (y_i - \bar{y}_i)^2 = \min_{\alpha, \beta, \gamma} \left\{ \sum_{i=1}^n (y_i - \bar{y}_i)^2 \right\} \quad (5)$$

### ARIMA model

ARIMA, also known as Box–Jenkins models (Box and Jenkins 1976), has been a very popular type of time series forecast models in hydrological field. It has the function of transforming a non-stationary time series into a stationary time series, by regressing the lag of independent variable, the present value, and lagged values of random error. ARIMA model, depending on the smoothness of the original sequence and the different part in regression model, includes the moving average (MA), autoregressive process (AR), autoregressive moving average process (ARMA), and ARIMA process.

In general, an ARIMA model is characterized by ARIMA ( $p, d, q$ ), where,  $p, q$ , and  $d$  denote the order of autoregression, integration (differencing), and moving average, respectively. The corresponding seasonal multiplicative ARIMA model is represented by ARIMA ( $p, d, q$ )  $\times$  ( $P, D, Q$ )<sup>s</sup> with  $P, D$ , and  $Q$  denoting the seasonal autoregression, integration (differencing), and moving average, respectively (see Box and Jenkins 1976; Tankersley and Graham 1993).

An AR( $p$ ) model can be expressed in the form (6)

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t \quad (6)$$

where  $\varphi_0$  is a constant;  $\varphi_1, \varphi_2, \dots, \varphi_p$  are autoregressive coefficients;  $p$  is the order of autoregressive model; and  $\varepsilon_t$  is the noise series that has a normal distribution with zero mean and constant variance  $\sigma^2(t)$  (Wong et al. 2007).

### Moving average model MA ( $q$ )

$Q$  order moving average model is denoted as MA ( $q$ ), satisfying the following equation:

$$y_t = \theta_0 + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_q y_{t-q} + \varepsilon_t \quad (7)$$

ARMA model, derived from combining AR and MA models, has the form (8):

$$y_t = c + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 y_{t-1} + \dots + \theta_q y_{t-q} \quad (8)$$

ARIMA models cannot really cope with seasonal behavior; we see that, compared with ARMA models, ARIMA ( $p, d, q$ ) only models time series with trends.

We will incorporate now seasonal behavior and present a general definition of the seasonal ARIMA models. The idea behind the seasonal ARIMA is to look at what are the best explanatory variables to model a seasonal pattern.

ARMA model is only applicable to the analysis of stationary sequences. However, in practice, many time series data contain a seasonal periodic component, which repeat every observation. To deal with seasonality, the ARIMA model is extended to a general multiplicative seasonal ARIMA ( $p, d, q$ )  $\times$  ( $P, D, Q$ )<sup>s</sup>, where the time sequence demonstrates both trend and seasonal trend; non-stationary sequence is transformed into a smooth one via differential transformation.

$$\varphi_p(B)\phi_p(B^s)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (9)$$

with

$$\begin{aligned} \phi_p(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \Phi_P(B^s) &= 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps} \\ \theta_q(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \\ \Theta_Q(B^s) &= 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs} \end{aligned}$$

where  $Z_t$  is appropriately transformed in period  $t$ , while  $(1-B)^d$  and  $(1-B^s)^D$  are the non-seasonal and seasonal different operators, respectively.  $B$  denotes the backward shift operator, and  $\varepsilon_t$  denotes the purely random process. If the integer  $D$  is not zero, then the seasonal differencing is involved. The above model is called a SARIMA model or seasonal ARIMA model ( $p, d, q$ )  $\times$  ( $P, D, Q$ ). If  $d$  is nonzero, then there is a simple differencing to remove trend. The seasonal differencing,  $(1-B^s)^D$  may be used to remove seasonality. Basically,  $d$  and  $D$  values are usually zero or one but rarely being two. The meaning of the parameters is explained below.

- $p$  the order of the local or regular AR term
- $d$  the number of local differences
- $q$  the order of local or regular MA term
- $P$  the order of periodic or stationary AR term
- $D$  the number of periodic differences
- $Q$  the order of periodic or stationary MA term
- $s$  the time period of the series.

The modeling steps of Box–Jenkins ARIMA Model involve an iterative five-stage process as follows:

- (i) Preparation of data including transformations and differencing.
- (ii) Identification of the potential models by looking at the sample autocorrelations and the partial autocorrelations.
- (iii) Estimation of the unknown parameters.

- (iv) Checking the adequacy of fitted model by performing normal probability plot, and creating a model.
- (v) Forecast future outcomes based on the known data.

### Integrated time series model

Application of ITS model to groundwater forecasting considers the groundwater system as a “black box” or “gray box” (Castellano-Méndez et al. 2004). The principle of ITS model is to decompose a time series into three parts, which includes trend component, periodic component, and random component. And then, adding these three could be the final forecasting model. The basic equation is written as:

$$H(t) = T(t) + P(t) + R(t) \quad (10)$$

where  $H(t)$  represents time series,  $T(t)$  represents trend component,  $P(t)$  represents periodic component, and  $R(t)$  represents random component.

The process of modeling involves extracting the components from the known sequence  $H(t)$  ( $t = 1, 2, 3, \dots, n$ ). The extraction order is the trend component with periodic component, followed by the random component. After the mathematical model has been developed and overlaid linearly, model (10) can be obtained (Yang et al. 2009).

For the trend component  $T(t)$  is available polynomial approximation, namely

$$T'(t) = c_0 + c_1t + c_2t^2 + c_3t^3 + \dots + c_kt^k = \sum_{k=0}^k c_kt^k \quad (11)$$

Multiple regression method can be used to determine the undetermined coefficient  $c_0, c_1, c_2, \dots, c_k$  and order  $k$ . The method is to use Excel software regression analysis templates to implement. To test the fitting result, the trend curve fitting correlation coefficient  $R$  at a certain level of significance is needed to be calculated.

$P(t)$  represents the inter-annual variability of series. Harmonic wave analysis method is adopted to extract the periodic component, which considers that periodic component is composed of many different cycle waves and can be expressed by Fourier series. Its estimate is expressed as:

$$P'(t) = a_0 + \sum_{j=1}^k a_j \cos\left(\frac{2\pi}{n}jt\right) + \sum_{j=1}^k b_j \sin\left(\frac{2\pi}{n}jt\right) \quad (12)$$

where  $k$  is the number of harmonic; the integer part is  $n/2$ ;  $J$  is usually referred to as the wave number, generally  $K$  wave with  $n/1, n/2, \dots, n/k$  cycle, i.e., the  $J$  wave frequency is  $j/n$ ;  $a_0, a_j, b_j$  are Fourier coefficient, the calculation formula:

$$\begin{cases} a_0 = \frac{1}{n} \sum_{t=1}^n X(t) \\ a_j = \frac{2}{n} \sum_{t=1}^n X(t) \cos \frac{2\pi j(t-1)}{n} \quad j = 1, 2, \dots, k \\ b_j = \frac{2}{n} \sum_{t=1}^n X(t) \sin \frac{2\pi j(t-1)}{n} \quad j = 1, 2, \dots, k \end{cases} \quad (13)$$

In practical applications, the estimation of the first six harmonics can meet the precision requirement.

If  $s_j^2 = a_j^2 + b_j^2 > 4s^2 \frac{\ln i}{n}$ , the wave was not significant; otherwise, where  $s^2$  is the variance, the calculation formula is:  $s^2 = \frac{1}{n-1} \sum_{t=1}^n (X_t - \bar{X}_t)^2$ ;  $\alpha$  is the significance level test (generally 5 %).

Random component is the last one to be extracted. It can be influenced by many uncertain factors, such as noise. It can be extracted with autoregression method.

Its autoregression model is:

$$R'(t) = \varphi_0 + \varphi_1 R(t-1) + \varphi_2 R(t-2) + \dots + \varphi_p R(t-p) \quad (14)$$

where  $p$  is the order of the model and  $\varphi_i$  is coefficient,  $i = 0, 1, 2, \dots, p$ . Autoregressive models of a similar order, multiple regression calculation can be obtained from the regression coefficient  $\varphi_i$ . To determine the order of the model, this paper uses AIC rules (Akaike 1969).

### Criteria of performance evaluation

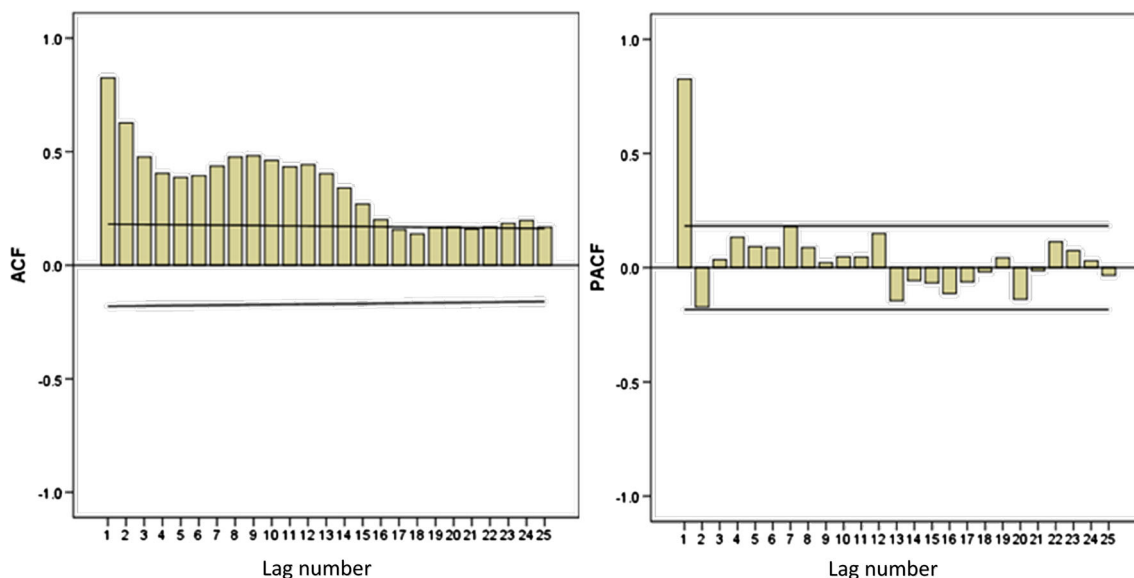
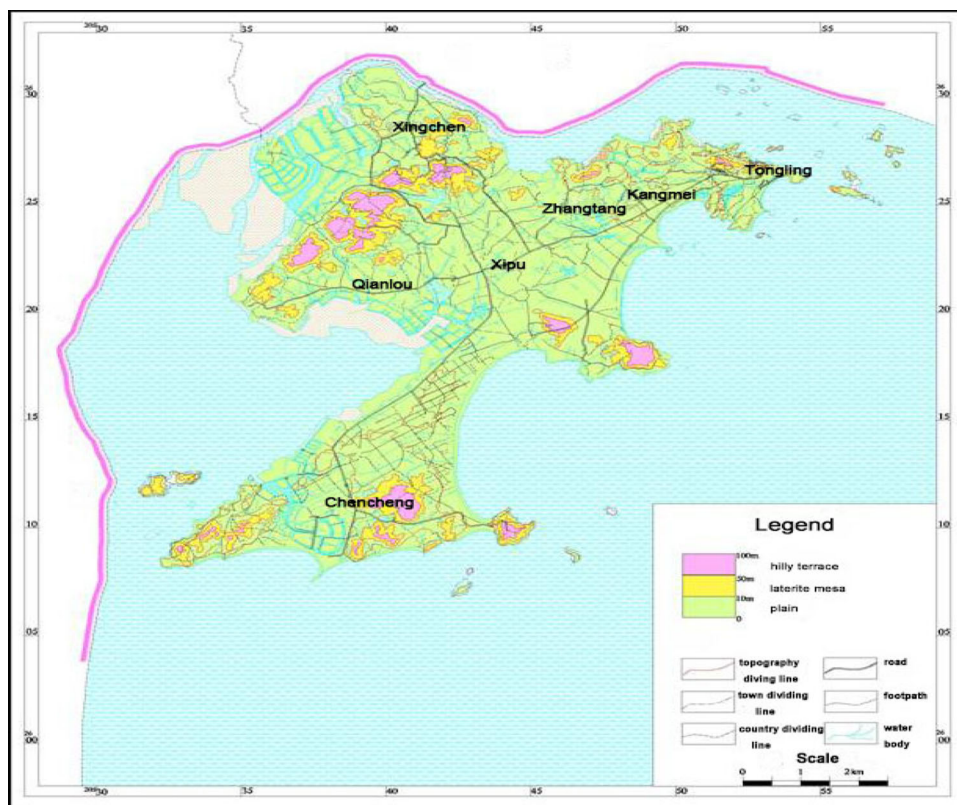
The performance of different forecasting models can be assessed in terms of goodness of fit once each of the model structures is calibrated using the training/validation data set and testing data set (Adamowski and Chan 2011). The coefficient of determination ( $R^2$ ), Nash–Sutcliffe model efficiency coefficient ( $E$ ), and root-mean-squared error (RMSE) were used in this study.  $R^2$  measures the degree of correlation among the observed and predicted values.  $R^2$  values range from 0 to 1. The coefficient of determination describes the proportion of the total variance in the observed data that can be explained by the model.  $R^2$  is given by Sreekanth et al. (2009):

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N y_i^2 - \frac{\sum_{i=1}^N \hat{y}_i^2}{N}} \quad (15)$$

The Nash–Sutcliffe model efficiency coefficient is used to assess the predictive power of hydrological models. It is defined as:

$$E = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (16)$$

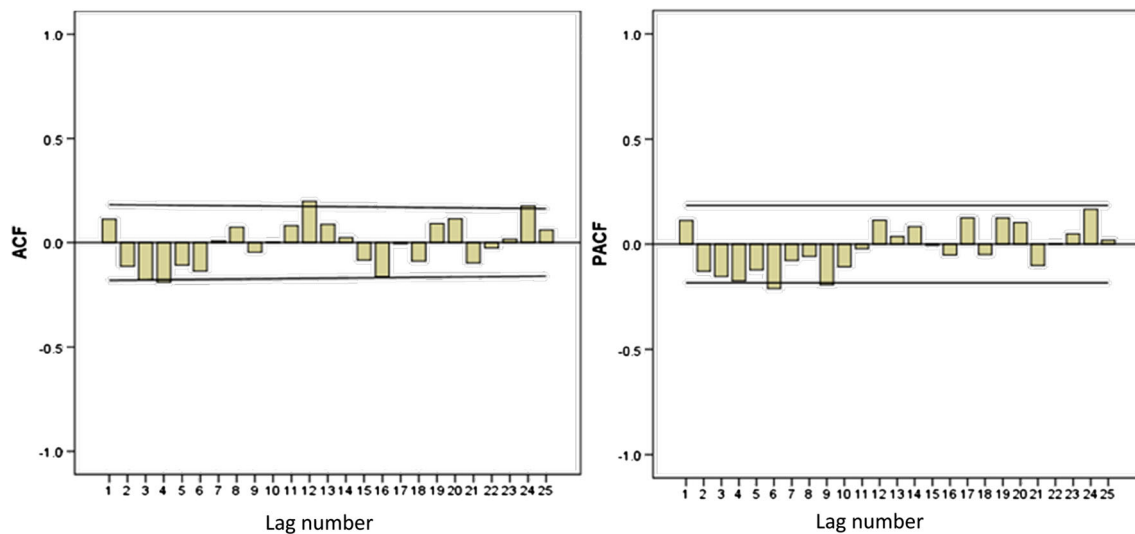
**Fig. 1** Outlined location map of the study area



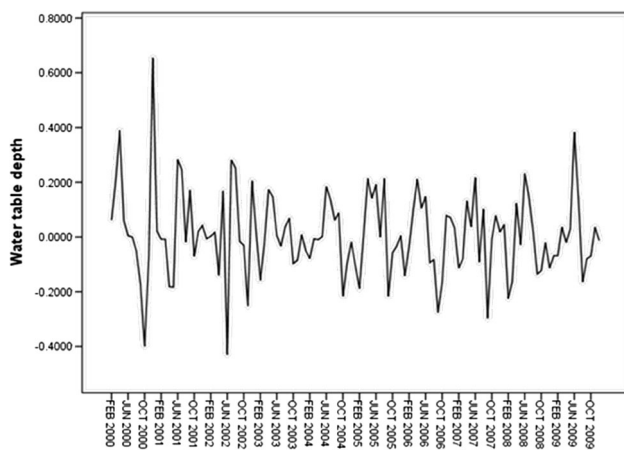
**Fig. 2** Residual ACF (left) and PACF (right) of groundwater table depth

The Nash–Sutcliffe model efficiency coefficient is used to assess the predictive power of hydrological models (Pulido-Calvo and Gutierrez-Estrada 2009). An efficiency of 0 ( $E = 0$ ) indicates that the model predictions are as

accurate as the mean of the observed data, whereas an efficiency less than zero ( $E < 0$ ) occurs when the observed mean is a better predictor than the model or, in other words, when the residual variance (described by



**Fig. 3** ACF (left) and PACF (right) plots for the first-order differential



**Fig. 4** Time series after the first-order differential

the numerator in the expression above) is larger than the data variance (described by the denominator). Essentially, the closer the model efficiency is to 1, the more accurate the model is.

RMSE is a frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed. The lowest the RMSE, the more accurate the prediction is. It is calculated by

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (17)$$

In Eqs. (15), (16), and (17):  $N$  is the number of data points used,  $\bar{y}_i$  is the average value of observed values over  $N$ ,  $y_i$  is the observed monthly groundwater level, and  $\hat{y}_i$  is the forecasted groundwater level from the model.

## Study area and data description

Groundwater level data in the study area was obtained by monitoring the groundwater level each 5 days. The monitoring period continued from January 2000 to December 2011.

The three models are tested with the data take from Dongshan hydrological station, Dongshan Town is a coastal island located at the most south of “golden delta” of Fujian Province, which lies between  $117^{\circ}17'E-117^{\circ}35'E$  longitude and  $23^{\circ}33'N-23^{\circ}47'N$  latitude, consisting Dongshan island and the rest 44 small islands, and covers an area of about  $248.34 \text{ km}^2$  (Fig. 1). The total length of coastline is around 200 m. It is influenced under the sub-tropical marine monsoon climate. The annual average temperature is about  $20.9^{\circ}\text{C}$  and varies between 13.1 in January and 27.3 in July. Annual average rainfall is about 1224.9 mm, most of which occur during May and September. A typical feature in the study area is frequent typhoon during July and September. Rural population accounts for approximately 80 % of the total population. Due to the topography, water body is not well developed within Dongshan Town, surface water is scarce, and groundwater has become a dependent source of water supply and servers in many aspects.

The main geological coverage of the study area is coastal plain, plateau, and hilly region. Of which alluvial plain consisting sand and gravel with thin clay accounts for more than 80 % where all the observations wells are located. The study area can be considered as an independent hydrogeological unit due to the sea surrounding on the four sides. Water yield property differs greatly because of lithology and thickness of the aquifer. Groundwater type is

coarse porous water, recharged predominantly by rainfall infiltration.

## Modeling

To analyze and forecast groundwater table in Dongshan County with three time series methods mentioned above, taking long-term observation well with number 3506260025 as an example, in which monthly average groundwater level was monitored during 2000–2011, the data set from 2000–2009 is used for model establishment, and those of 2010–2011 is used for predicting the dynamic change.

### Holt–Winters model

Multiplicative method is selected in EViews 6.0 software to establish HW model, the software automatically select the minimum of sum squared error (SSE) and mean square error (MSE) as the best prediction model. After calculation, when  $\alpha = 0.98$ ,  $\beta = 0$ ,  $\gamma = 0$ , the SSE and MSE were found to get the minimum; therefore, this combination of smoothing coefficient was considered as the optimal prediction model, fitting curve between observed values and the calculated values was shown in Fig. 10.

### SARIMA model

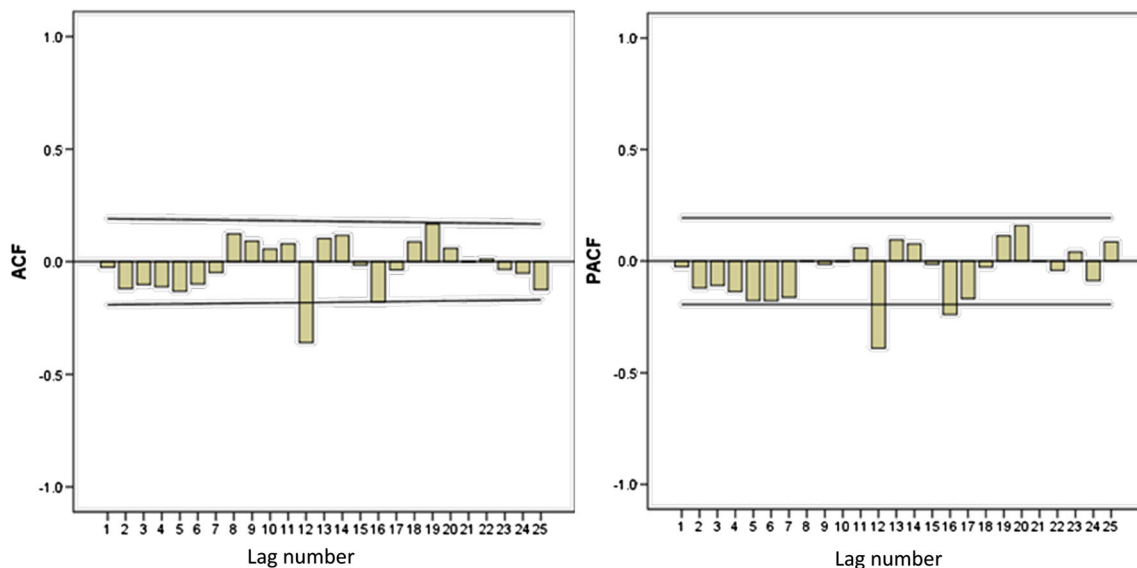
According to the description in “ARIMA model,” SARIMA  $(p, d, q) \times (P, D, Q)^s$  is created using SPSS 18.0 software with the monthly average groundwater table depth

monitored during 2000–2009. In identifying a seasonal model, the first step is to determine whether or not a seasonal difference is needed, in addition to or perhaps instead of a non-seasonal difference. Time series plots and auto-correlation function (ACF) and partial autocorrelation function (PACF) plots for all possible combinations of 0 or 1 non-seasonal difference and 0 or 1 seasonal difference should be considered. The detailed procedure of establishing ARIMA  $(p, d, q) \times (P, D, Q)^s$  is described in the following section.

It can be seen in Fig. 2 that the 25th order ACF and PACF exceeds the random interval and demonstrates a decreasing trend, which indicates that the time series contains strong trend character with slight seasonal fluctuations; PACF plot shows that the other orders are all zero except for the first-order function. In order to eliminate the trend component in time series and make other factors more obvious, the series is treated with the first-order differential, and  $d$  is set as 1. The corresponding ACF and PACF plot are shown in Fig. 3.

It can be seen from Fig. 3 that the trend component is eliminated totally and that the periodic fluctuation in ACF plot appears. In addition, time series in Fig. 4 demonstrates that the seasonality is visualized more clearly and also contains non-stationarity. Therefore, it is necessary to perform the first-order seasonal differential. Figures 5 and 6 demonstrate the corresponding plots of ACF and PACF, time series.

It can be observed from Figs. 5 and 6 that the seasonality is eliminated completely after the first-order seasonal differential and demonstrates a steady convergence. Thus,  $d$  and  $D$  values are set as 1 in the model, according to AFC



**Fig. 5** Residual ACF (*left*) and PACF (*right*) after the first-order seasonal differential

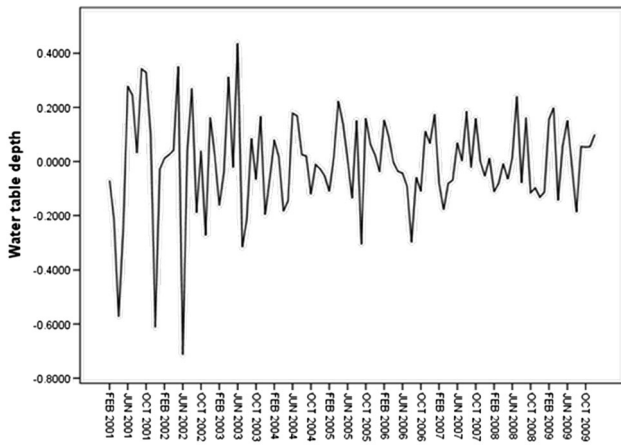


Fig. 6 Time series after the first-order seasonal differential

plot,  $q$  is equal to 0, while the 12th function is not 0, so  $Q = 1$ , and PACF plot gives the information that  $p = 0$  or 1. After analyzing the model fitness, the best model set is determined as ARIMA  $(0, 1, 0) \times (0, 1, 1)^{12}$ . The fitness between the observed the simulated values are shown in Fig. 7.

**Integrated time series model**

ITS model was implemented with Excel 2007. Through decomposition procedures of groundwater level series in “Integrated time series model,” the predicted groundwater table can be obtained by adding the three components; Fig. 7 demonstrates that the groundwater table shows an increasing trend year by year. In natural condition, the groundwater depth should be steady or vary to a certain extent and cannot increase or decrease continuously. So the trend component could reflect the degree of exploitation by human.

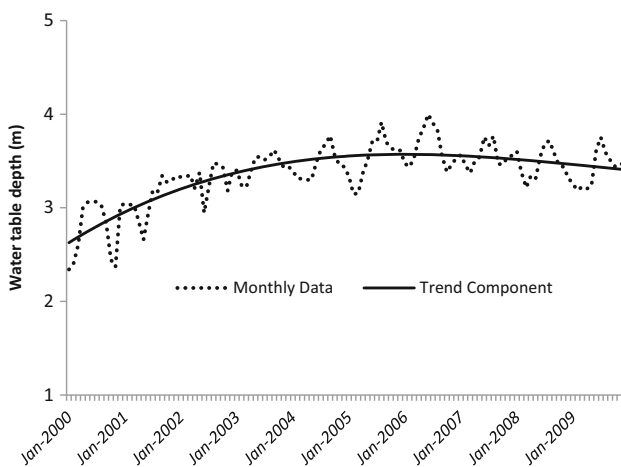


Fig. 7 Trend component

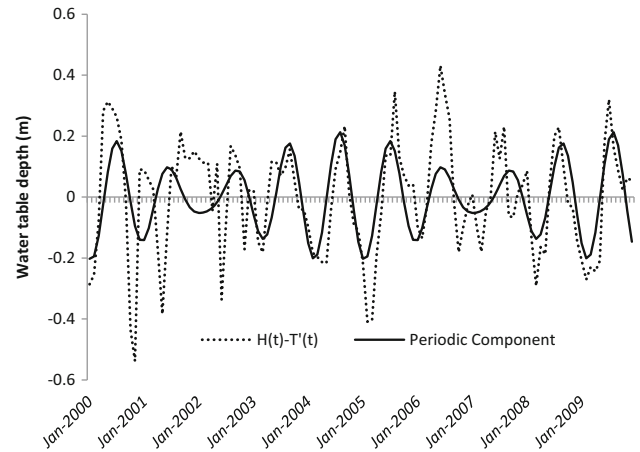


Fig. 8 Periodic component

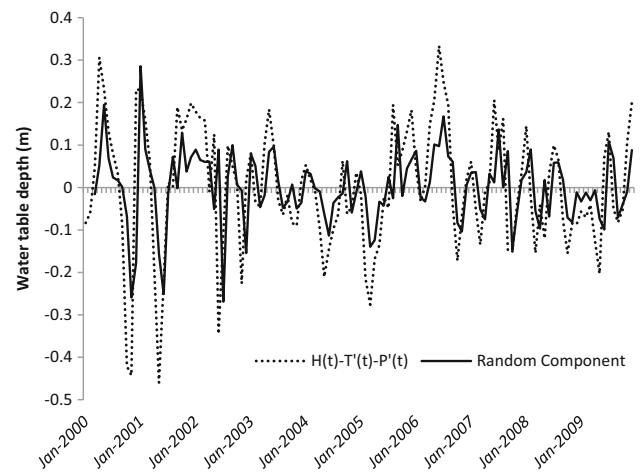


Fig. 9 Random component

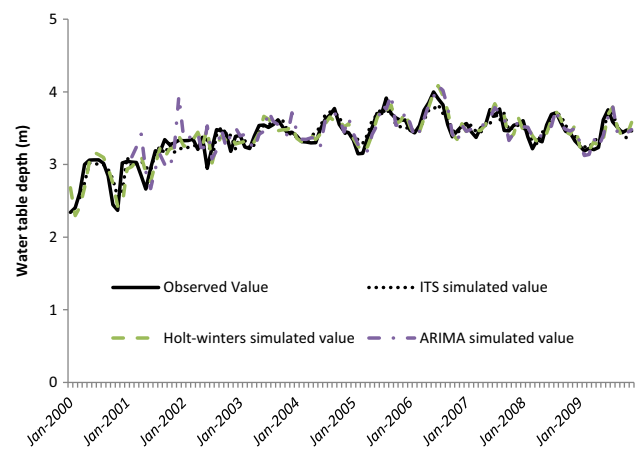


Fig. 10 Fitting graphs between observed and simulated values for three models



From periodic component (Fig. 8), it can be seen that periodicity exist in the groundwater level series. This periodicity is believed to be driven by the cycle of the solar activity and earth’s rotation and revolution (Zheng 1989). Sunspot activity can influence the alternation of the dry season and the rainy season. Thus, the periodicity reflects the influence of natural factor of climate.

The random component is affected by many uncertain factors. Therefore, in this study, no further discussion is presented. The random component is shown in Fig. 9. The fitting curve is shown in Fig. 10.

Then, Fig. 10 shows a comparison of observed and calculated groundwater levels for training period by three models. It can be seen that all three models reproduced the observed time series well enough.

### Results and discussions

Figure 11 and Table 1 summarize the results of validation for each method. Three common indexes, coefficient of determination ( $R^2$ ), Nash–Sutcliffe model efficiency coefficient ( $E$ ), and RMSE, are calculated to compare the performance of validation.

It can be seen in Fig. 11 that these three models are all suitable to predict groundwater levels. HW model was found to outperform SARIMA and ITS model, which can be observed both from the fitness between the observed and predicted values in Fig. 11 and statistical results of

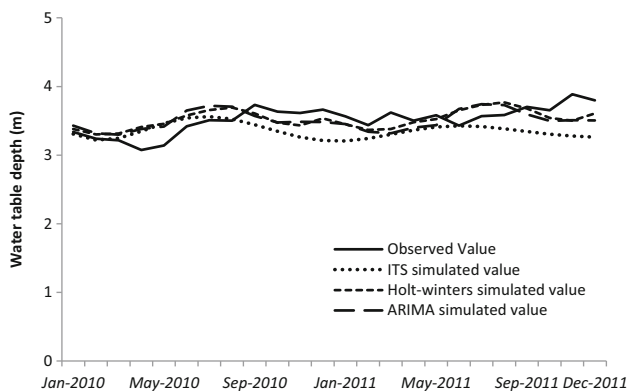


Fig. 11 Comparison of validation results for different models

Table 1 Statistical results of accuracy evaluation for groundwater level simulation during 2010–2011

Parameter	Holt–Winters	SARIMA	ITS
$R^2$	0.997249	0.996683	0.993105
$E$	0.167941	−0.003479	−1.09388
RMSE	0.180867	0.198626	0.286918

performance evaluation in Table 1. HW model has smaller RMSE than SARIMA and ITS models. The negative values of  $E$  for ARIMA and ITS models indicate that the observed is better than model results; however, ARIMA model performs better than ITS model since a smaller absolute  $E$  value is obtained.

Since groundwater level dynamic is a complex response to many factors, time series analysis can reflect the influence of human behavior, rainfall and solar activity. With its advantage of being easily implemented, its potential in analyzing and forecasting groundwater level dynamic is overwhelming.

### Conclusions

Time series techniques have been widely used in environmental contexts. In particular, its application in hydrological forecasting has been a common practice. This paper explores the utilization of three time series models, HW, SARIMA model, and ITS model, and their potential for forecasting groundwater levels is investigated with a case study in a shallow aquifer of a coastal island, China. The monitored long-term observation monthly groundwater table depth data series from 2000 to 2011 are used in model setup and prediction. The capability to make precise predictions for each model was evaluated with statistical error criteria, coefficient of determination ( $R^2$ ), Nash–Sutcliffe model efficiency coefficient ( $E$ ), and RMSE. The results indicate that three models are all accurate in reproducing the historical time series of groundwater levels. The comparison of three models shows that HW model is more accurate in predicting the groundwater levels than SARIMA and ITS models.

HW model is a more sophisticated method of forecasting than methods of moving average and exponential smoothing. The method can include seasonality, which is important, and the overall trend. It is an extension of Holt method of a three-parameter exponential smoothing. It means the method is characterized by three parameters that must be selected to get the forecast.  $\alpha$ ,  $\beta$ , and  $\gamma$  are parameters that must be selected before the forecasting. The choice of parameters was carried out using “smart” enumeration and minimization of errors on the known data.

Time series analysis methods, HW, ITS, and SARIMA, are explored to simulate the groundwater level in a coastal aquifer, China. The error criteria are estimated using coefficient of determination ( $R^2$ ), Nash–Sutcliffe model efficiency coefficient ( $E$ ), and RMSE. The results showed that three models can accurately predict the water table. However, SARIMA model demonstrates more reliable capability compared with ITS and HW model.

**Acknowledgments** This research was financially supported by National Natural Science Foundation of China (Grant number 41402202) and Specialized Research Fund for the Doctoral Program of Higher Education (20130061120084).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Adamowski J, Chan HF (2011) A wavelet neural network conjunction model for groundwater level forecasting. *J Hydrol* 407:28–40
- Ahn H (2000) Modeling of groundwater heads based on second-order difference time series models. *J Hydrol* 234(1–2):82–94
- Akaike H (1969) Fitting autoregressive models for prediction. *Ann Inst Stat Math* 21(1):243–247
- Behnia N, Rezaeian F (2015) Coupling wavelet transform with time series models to estimate groundwater level. *Arab J Geosci*. doi:10.1007/s12517-015-1829-0
- Bermúdez JD, Segura JV, Vercheri E (2010) Bayesian forecasting with the Holt–Winters model. *J Oper Res Soc* 61:164–171
- Box GEP, Jenkins GM (1976) *Series analysis forecasting and control*. Prentice-Hall Inc., London
- Castellano-Méndez M, González-Manteiga W, Febrero-Bande M, Parada-Sánchez JM, Lozano-Calderón R (2004) Modelling of the monthly and daily behavior of the runoff of the Xallas river using Box–Jenkins and neural networks methods. *J Hydrol* 296(1–4):38–58
- Gardner ES Jr (1985) Exponential smoothing, the state of the art. *J Forecast* 4:1–28
- Gelper S, Fried R, Croux C (2010) Robust forecasting with exponential and Holt–Winters smoothing. *J Forecast* 29:285–300
- Holt CC (1957) Forecasting trends and seasonals by exponentially weighted averages. In: Carnegie Institute of Technology. Pittsburgh ONR memorandum no. 52
- Konikow LF, Kendy E (2005) Groundwater depletion, a global problem. *Hydrogeol J* 13(1):317–320
- McNeil AJ, Frey R, Embrechts P (2005) *Quantitative risk management: concepts, techniques, and tools*. Princeton University Press, Princeton
- Nayak P, Rao YRS, Sudheer KP (2006) Groundwater level forecasting in a shallow aquifer using artificial neural network approach. *Water Resour Manag* 20(1):77–90
- Nourani V, Ejlali RG, Alami MT (2010) Spatiotemporal groundwater level forecasting in coastal aquifers by hybrid Artificial Neural Network-Geostatistics model: a case study. *Environ Eng Sci* 28:217
- Pulido-Calvo I, Gutierrez-Estrada JC (2009) Improved irrigation water demand forecasting using a soft-computing hybrid model. *Biosyst Eng* 102(2):202–218
- Sethi RR, Kumar A, Sharma SP, Verma HC (2010) Prediction of water table depth in a hard rock basin by using artificial neural network. *Int J Water Resour Environ Eng* 2(4):95–102
- Sreekanth P, Geethanjali DN, Sreedevi PD, Ahmed S, Kumar NR, Jayanthi PDK (2009) Forecasting groundwater level using artificial neural networks. *Curr Sci* 96(7):933–939
- Tsanis IK, Coulibay P, Daliakopoulos N (2008) Improving groundwater level forecasting with a feedforward neural network and linearly regressed projected precipitation. *J Hydroinformatics* 10(4):317–330
- USGS (2010) <http://ga.water.usgs.gov/edu/gwdepletion.html>. Accessed 02 July 2010
- Wong H, Wai CI, Zhang RQ, Xia J (2007) Non-parametric time series models for hydrological forecasting. *J Hydrol* 332:337–347
- Yang ZP, Lu WX, Long YQ, Li P (2009) Application and comparison of two prediction models for groundwater levels. A case study in Western Jilin Province, China. *J Arid Environ* 73:487–492
- Zhao Y, Wang H (2007) Forecasting total circular volume of air railway transport with Holt–Winters model. *J Civ Aviat Univ China* 25(2):1–11
- Zheng DW (1989) Time series analysis and studies of earth rotation. *Prog Astron* 7(2):118–124