The Forecasting of Groundwater Fluctuations using Time Series Analysis

Authors: Amirhossein Najafabadipour, Gholamreza Kamali , Hossein Nezamabadi-pour

Paper Link: https://link-springer-com-christuniversity.knimbus.com/article/10.1134/S0097807822020026

Journal: Springer

Groundwater fluctuation forecasting aids in the management of groundwater resources to avoid additional costs. The paper uses Auto-Regressive Integrated Moving Average (ARIMA) and Holt-Winters Exponential Smoothing (HWES) along with appropriate weights to develop a model for the forecasting of the groundwater fluctuations in six piezometer wells around the Gohar Zamin Iron Ore Mine. For this purpose, 250 non-seasonal data of daily groundwater level were used, of which 200 data are for modeling, and 50 data are for water level forecasting.

Groundwater is a natural resource that has costly adverse effects on mining operations. Too much water entering the mining environment may delay the project or impede production. Increased equipment failure, lack of access to part of the mining area and use of explosives are some of the undesirable impacts of groundwater entering an operation.

Data-driven techniques have been a useful tool for predicting groundwater fluctuations. Lack of need for information about aquifer hydrogeological parameters is the most crucial advantage. Data-driven models overcome uncertainty of parameters and data constraints to predict groundwater fluctuations, say researchers.

Linear models are one of the most widely used methods for predicting groundwater fluctuations. The need for less computational time and effort for training is an advantage of linear models. Struggling to handle non-linearity is the principal disadvantage of using a data-driven model. To develop linear models:

I) The input time series must be stationary.

II) Residuals must be white noise

There should be no cross-correlation and no autocorrelation between the input and the residuals; also, the mean of time series should be zero. Data driven models can be a costly and time consuming process so we use linear models using a small amount of time series data in a short time span of 250 days.

In time series analysis, the variable changes are explained based on the current and past changes of the other variables. The most critical aspect of a series analysis is when it helps eliminate or reduce the inherent distribution of measured valleys in any particular period over time. The purpose of time series analysis:

1. Create input models that express the random behaviour of a variable over time using a probabilistic model

2. Prediction to obtain values in forwarding time steps using the developed model

Due to the lack of the necessary theory for the time series, statistical concepts are used based on the distribution functions to extract the essential regression to model the time series. Besides, standard regression cannot be used due to the serial correlation of time series due to the inability to estimate the coefficients of the regression equation. The equation of the ARIMA model for a non-seasonal time series model as:

$$y_t = c + \varnothing_1 y_{d(t-1)} + \varnothing_p y_{d(t-p)} + \ldots + \theta_1 e_{t-1} + \theta_q e_{t-q} + e_t$$

Where y_t is groundwater fluctuations value at time t. y_d means y differenced d times. e_t is the error of the model as a mixture of the previous error. $\emptyset_k$. θ_k are the Auto-Regressive and moving average constant coefficients of the model at lag k.

Due to the uncertainty of the number of parameters identified for the Auto-Regressive Integrated Moving Average (ARIMA) model, information criterion methods such as Akaike (1987) Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan and Quinn Information Criterion (HQIC) can be used to estimate the optimal number of parameters. After determining the required number of parameters of the model, the model should be evaluated using these values. The validity of the developed model must be confirmed in the last phase of the model construction. Some unique statistical methods, such as Maximum Likelihood, can be used to determine the model. In the case of residual correlation, the number of model parameters must be increased.

The validity of the developed model must be confirmed in the last phase of the model construction. Residual diagnostics includes statistical calculations of the autocorrelation of residuals. Spikes in the correlogram signs that the residuals may be correlated, which the model created is not satisfactory.

The Exponential Smoothing approach is especially suitable for short-term forecasting. In this method, weight factors are used for the past values and reduced in the view with a distance from the previous values. This approach makes possible formula of the prediction algorithm that requires only a few recent data with fewer calculations.

The smoothed series y_t is given by:

$$y_{t+k} = a + bk$$

a is the permanent component, and b is the trend. The following recursions characterize these two coefficients:

$$a(t) = \alpha\, yt + (1-\alpha)\,(a(t-1) + b(t-1))$$

$$b(t) = \beta\,(a(t) - a(t-1)) + 1 - \beta b(t-1)$$

Where α and β are damping factors fluctuating between 0 and 1 depending on the specific time series characteristics. Time series patterns and smoothing objectives can affect the optimal value of these parameters. α values between 0.1 and 0.3 are more commonly used because they rely on many last observations to predict. Values close to 1 are seldom used because they make predictions that depend on recent data. For example, $\alpha=1$ shows the forecast is equal to recent observations.

Different weights are used to determine complex weight schemes during the forecast period. The following is a list of commonly used methods that use constant weights through time:

Simple mean: In this method, the arithmetic mean of forecasts at each observation within the forecast sample is calculated. During this method, every forecast is given an identical weight.

Simple median: The median of the prediction at each observation within the prediction sample will be estimated for the simple median method. The implicit (0, 1) weights are time-varying as every forecasting method is also the median for a few observations.

Least-squares weights: To use this method, the actual values of the forecasted variable for a few of the forecast periods must be identified. This method is calculated by regressing the forecasts against particular values and using regression coefficients as weights. In the least squares method, the underlying individual forecasts can be unbiased, and also the resulting average can be outside the scope of the underlying forecasts.

Mean squares error weights: Mean squares error (MSE) weighting during the forecast period

compares the actual values with the individual forecasts (Stock & Watson, 2002). After computing each forecast, individual forecast weights are formed using the following equation

$$w_i = \frac{\frac{1}{MSE_i^k}}{\sum_{j=1}^{N} \frac{1}{MSE_j^k}}$$

k is employed to change the MSE to different powers. The most comprehensive power for k is 1.

Groundwater fluctuations for mines can be forecasted by short-term analysis that shows the impact of groundwater recharging and harvesting from water storage in an aquifer. The records of piezometer No. 2 in the western part of Gohar Zamin Iron Ore Mine were reviewed from 22-Nov-2018 to 29-Jul-2019 to study groundwater fluctuations. In this research, the predictive ability of Holt-Winters Exponential Smoothing (HWES) and non-seasonal Auto-Regressive Integrated Moving Average (ARIMA) models at the time series of daily groundwater fluctuations around the Gohar Zamin Iron Ore Mine pit was compared.

Since each prediction has its characteristics, this research showed that combining predictions made with different methods has better results than individual predictions. Diebold and Marianos test with a p-value above 5% indicates a similar accuracy of the two predictive methods. An encompassing test with a p-value of about 0.66 for Auto-Regressive Integrated Moving Average (ARIMA) prediction indicates a higher efficiency of this method than Holt-Winters Exponential Smoothing (HWES) in short-term predictions. Also, the results of the prediction combination show a higher accuracy of the least-squares combination method than other prediction methods. Forecasts made by simulation of time series indicated that if the current situation is still around the pit, it is predicted that the groundwater fluctuations will decrease, indicating the activity of pumping wells around the pit. Due to the acceptable results of the prediction combination, which shows the high accuracy of this method, it is recommended to predict the short-term time series.

The research results show that among all the forecasting, the least-squares method in the innovation of combining models has the highest accuracy for forecasting the groundwater fluctuations in the short term. All the forecasts in this period time show a decrease in the groundwater level, indicating the effects of pumping wells in the Gohar Zamin Iron Ore Mine area.

-Ananya Kaushal

1940233

6CMS