A Comparative Study of Data-driven Models for Groundwater Level Forecasting

R. Sarma[1] · S. K. Singh[1]

Abstract

Irregular rainfall patterns and limited freshwater availability have driven humans to increase their dependence on groundwater resources. An essential aspect of effective water resources management is forecasting groundwater levels to ensure that sufficient quantities are available for future generations. Prediction models have been widely used to forecast groundwater levels at the regional scale. This study compares the accuracy of five commonly used data-driven models–Holt–Winters' Exponential Smoothing, Seasonal Autoregressive Integrated Moving Average, Multi-Layer Perceptron, Extreme Learning Machine, and Neural Network Autoregression for simulating the declining groundwater levels of three monitoring wells in the National Capital Territory of Delhi in India. The performance of the selected models was compared using coefficient of determination (R^2), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Results indicate that Multi-Layer Perceptron had high R^2 while fitting the training data and least RMSE and MAE during testing, thus proving to be more accurate in forecasting than the other models. Multi-Layer Perceptron was used to forecast the groundwater level in the study wells for 2025. The results showed that the groundwater level will decline further if the current situation continues. Such studies help determine the appropriate model to be used for regions with limited available data. Additionally, predictions made for the future will help policymakers understand which areas need immediate attention in terms of groundwater management.

Keywords Groundwater level forecasting · Holt-winters' exponential smoothing · Seasonal ARIMA · Multi-layer perceptron · Extreme learning machine · Neural network autoregression

---

Highlights

- A number of different data-driven models exist for forecasting groundwater levels.
- In the present study, Multi-Layer Perceptron was the most accurate model for groundwater level forecasting.
- Accuracy measures - coefficient of determination, Root Mean Squared Error and Mean Absolute Error were compared to determine the most precise model.
- Multi-Layer Perceptron forecasts for 2025 showed a decline of 2-21 mbgl in the study wells.

✉ S. K. Singh
sksinghdce@gmail.com

[1] Department of Environmental Engineering, Delhi Technological University, Delhi, India

1 Introduction

With rising dependence on groundwater to meet domestic, irrigation and industrial demands, trend analysis and simulation of water table behaviour have become important fields of study (Noori and Singh 2021). Researchers worldwide have developed and applied a number of different models to forecast groundwater levels, particularly for areas where recharge is low and extraction is high. Although models available for forecasting are abundant, selecting a suitable model that accurately simulates groundwater behaviour is complex (Takafuji et al. 2019). Several factors have to be taken into account for model choice and performance–objective of the study and data availability being primary (Salvadore et al. 2015; Barzegar et al. 2017).

Numerical modeling techniques have gained immense popularity in recent years, mainly due to their ability to forecast the likely impacts of water management solutions (Singh 2014; Sarma and Singh 2021a). Although largely popular, numerical models may be limited by large data requirements and complex computations (Aguilera et al. 2019). Univariate forecasting can be beneficial for data-scarce study areas. Zhang et al. (2018) present a comprehensive review of commonly used data-driven models for hydrological processes. They classified data-driven models as conventional, AI-based and hybrid models for streamflow prediction.

Conventional time series models such as Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) are easy to use and widely popular (Mirzavand and Ghazavi 2015; Choubin and Malekian 2017; Rahaman et al. 2019;). The ARIMA model assumes that data at time t will directly correlate with previous data at t-1, t-2, …. and associated errors (Narayanan et al. 2013). Many studies around the world have compared the performance of ARIMA or SARIMA with other models such as Holt-Winters' Exponential Smoothing (HWES), Integrated Time Series model (ITS) and Artificial Neural Networks (ANN) (Shirmohammadi et al. 2013; Aguilera et al. 2019; Sakizadeh et al. 2019). The HWES model forecasting is based on Holt and Winters' basic structures (Holt 1957; Winters 1960) and can be applied on data with non-constant trends and seasonal variations (Yang et al. 2017). In the study by Yang et al. (2017) for a coastal aquifer in South China, HWES outperformed ITS and SARIMA.

In recent years, ANNs have seen widespread applications in forecasting hydrological parameters (Chen et al. 2020; Mozaffari et al. 2022). ANN replicates biological neuron processing and consists of an input layer, one or more hidden layers and an output layer. Lallahem et al. (2005) evaluated the feasibility of using ANNs for groundwater level predictions and concluded that ANNs, particularly MLPs with minimal lags and hidden nodes, gave the best simulation results. Yang et al. (2009) reported that Back-Propagation Artificial Neural Network outperformed ITS in simulating groundwater levels in China. Aguilera et al. (2019) compared the performance of the Prophet model with other forecasting techniques–seasonal naïve, linear model, exponential smoothing, ARIMA and neural network autoregression (NNAR) for groundwater level data in Spain. Recent years have seen the growing use of a novel method called Extreme Learning Machine (ELM) in hydrology (Kalteh 2019; Parisouj et al. 2020). In a study by Natarajan and Sudheer (2020), ELM had the best performance compared to ANN, GP and SVM for groundwater level prediction at six locations in Andhra Pradesh, India. Poursaeid et al. (2022) compared the performance of some mathematical and Artificial Intelligence (AI) models and concluded that the ELM method showed the best performance for groundwater level simulation.

In this study, two conventional models – Holt-Winters' Exponential Smoothing (HWES), Seasonal ARIMA and three AI-based models – Multi-Layer Perceptron (MLP), Extreme Learning Machine (ELM) and Neural Network Autoregression (NNAR) were applied on historical groundwater records of three monitoring wells in National Capital Territory of Delhi, India. The models were trained and tested as per standard procedure. The forecasting performance of each model was compared using accuracy measures–coefficient of determination (R^2), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Finally, the best model was used to predict groundwater depth in 2025. To the best of our knowledge, this comparative analysis has been done for the first time in the study area.

2 Materials and Methods

2.1 Data Driven Models

2.1.1 Holt-winters' Exponential Smoothing

The simple smoothing methodology given by Brown (1959) was limited by its inability to support a time series with trend components (Sakizadeh et al. 2019). Holt (1957) proposed the exponential smoothing method which incorporated a trend component given by the following equations:

Level

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \tag{1}$$

Trend

$$b_t = \beta^* \left(l_t - l_{t-1} \right) + (1 - \beta^*) b_{t-1} \tag{2}$$

Forecast

$$y_{t+h|t} = l_t + b_t h \tag{3}$$

where, l_t: level of time series at t time step, b_t:slope of the series, $y_{t+h|t}$: forecast for next h time steps, α, β^*: smoothing parameters (between 0 and 1). For a time series with variable seasonal fluctuations, simulations are made using the multiplicative form:

Level

$$l_t = \alpha \frac{y_t}{S_{t-m}} + (1 - \alpha)(l_{t-1} + b_{t-1}) \tag{4}$$

Trend

$$b_t = \beta^* \left(l_t - l_{t-1} \right) + (1 - \beta^*) b_{t-1} \tag{5}$$

Seasonal

$$s_t = \frac{\gamma y_t}{\left(l_{t-1} + b_{t-1} \right)} + (1 - \gamma) s_{t-m} \tag{6}$$

Forecast

$$y_{t+h|t} = (l_t + b_t h)s_{t-m+h_m^+} \tag{7}$$

where, m: length of time series, h_m^+: (h-1)mod $m + 1$, S_t: seasonal component at time t, γ: a smoothing coefficient (between 0 and 1).

2.1.2 Seasonal ARIMA Model

An ARIMA model contains autoregressive (AR), integrated (I) and moving average (MA) parts which are expressed as ARIMA (p, d, q); where p is autoregressive part, d is integrated part and q is moving average part (Box and Jenkins 1976). AR describes the relationship between present and previous variables in the time series. If p = 1, then each variable is a function of only one last variable, i.e.

$$Y_t = c + \varnothing_1 Y_{t-1} + e_t \tag{8}$$

where, Y_t: observed value at time t, Y_{t-1}: previous observed value at time t-1, e_t: random error, c and φ_1: constants. For p > 1, other observed values of the series can be included as:.

$$Y_t = c + \varnothing_1 Y_{t-1} + \varnothing_2 Y_{t-2} + \cdots + \varnothing_p Y_{t-p} + e_t \tag{9}$$

The I part of the model denotes the stationarity of the series. For a non-stationary series, differencing has to be done. For linear trend in the time series, first-order differencing is done (d = 1), for a quadratic trend, d = 2 and so on. The MA part of the model identifies the relationship between the variable and previous q errors. If q = 1, each observation is a function of only one previous error i.e.

$$Y_t = c + \theta_1 e_{t-1} + e_t \tag{10}$$

where, c: constant, e_t: random error at time t, e_{t-1}: previous random error at time t–1. For q > 1, other errors can be included as:

$$Y_t = c + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} + e_t \tag{11}$$

The combined equation for non-seasonal ARIMA model of order (p, d, q) for a standard normal variable (Y_t) is:

$$\varnothing(B)(1 - B)^d Y_t = \theta(B)e_t \tag{12}$$

where, B: backshift operator. To account for seasonality, the ARIMA model is represented by ARIMA (p,d,q)×(P,D,Q)$_s$ with P, D and Q denoting the seasonal autoregression, integration (differencing), and moving average, respectively.

$$\varnothing_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D Y_t = \theta_q(B)\Theta_Q(B^s)e_t \tag{13}$$

where, Y_t: original time series, e_t: normal independently distributed white noise residual series with mean zero and variance σ^2, Φ and Θ: ARIMA structures between the seasonal observations, φ and θ: non-seasonal ARIMA structure, $\phi_p(B)$ and $\Phi_P(B^s)$: non-seasonal and seasonal autoregressive operators of order p and P, respectively, $\theta_q(B)$ and $\Theta_Q(B^s)$: non-seasonal and seasonal moving average operators of order q and Q, respectively, ∇^d

and ∇_s^D: non-seasonal and seasonal differencing operators of orders d and D. The Seasonal ARIMA procedure involves three major steps –

(a) Model identification: the time series is first analysed for stationarity and normality and accordingly differenced and/or log-transformed. Autocorrelation (ACF) and partial autocorrelation (PACF) functions of the original and differenced series are examined. The best model is identified based on least Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The AIC or BIC is written in the form [-2logL + kp], where L: likelihood function, p: number of parameters in the model, and k: 2 for AIC and log(n) for BIC (Akaike 1974; Schwarz 2007).
(b) Parameter estimation and diagnostic checking: autoregressive and moving average parameters of the identified model are estimated. Diagnostic checking is carried out from the residuals and accuracy measures to examine the suitability and assumptions of the model.
(c) Forecasting: the selected model is used to forecast the variables for future time periods.

2.1.3 Multi-layer Perceptron

Several AI-based deep learning frameworks are available that can simulate time series data. Of these, MLP is a simple feed-forward layered technique that works on the principle of the backpropagation algorithm (Hecht-Nielsen 1989; Aslam et al. 2020). The data are processed in the forward direction by feeding into input nodes. These are then multiplied by a given weight and passed onto one or more hidden layer nodes. The hidden layer nodes add the weighted inputs, calculate loss or bias and then pass it on through a transfer function to give the result. The output nodes perform the same operations. During the training phase, the weights and biases are adjusted to minimise the errors via the backpropagation algorithm and a comparison is made between target outputs at each output node and output network (Lallahem et al. 2005). The backpropagation algorithm keeps repeating this iteration till the maximum improvement is achieved. Fig. S1 demonstrates a three-layer feed-forward MLP structure. The equation for MLP is given by:

$$y_k = S_1(\sum_{j=1}^{J} w_j S_2(\sum_{i=1}^{I} w_i x_i + W_j) + W_k) \tag{14}$$

where, y_k are outputs from the network, x_i are inputs, w_i are weights connecting input and hidden layer nodes, w_j are weights connecting hidden and output layer nodes, I are input nodes, J are hidden nodes, K are output nodes, W_j is bias for jth hidden neuron and W_k is bias for kth output neuron. S_1 and S_2 are activation functions. Commonly, the logistic sigmoid function is used for activation:

$$S(x) = \frac{1}{1 + e^{-x}} S : R \rightarrow] -1, 1[\tag{15}$$

2.1.4 Extreme Learning Machine

ELM was developed to overcome traditional neural networks' high computational cost and time (Huang et al. 2006). In the three-layered structure (single hidden layer feed-forward network) of ELM, weights between inputs and hidden nodes and bias values in the hidden layer are generated at random which are frozen during model training. The weights

between hidden nodes and outputs are calculated using Moore–Penrose generalized inverse of the hidden-output matrix. This process makes ELM much faster than traditional ANNs because there is no iteration in learning. For a training set, ELM is expressed as:

$$\sum_{i=1}^{\widehat{N}} \beta_i g\left(w_i x_j + b_i\right) = o_j, j = 1, \dots, N \tag{16}$$

where, x_i: input, $\widehat{N}$: hidden nodes, w_i: weight vector from input to i^{th} hidden node, b_i: bias of ith hidden node, β_i: weight vector from i^{th} hidden node to output, o_j: output and g(x): non-linear activation function in hidden layer (Parisouj et al. 2020).

2.1.5 Neural Network Auto-regression

In an NNAR model, lagged values are used as inputs to a neural network. NNAR(p,k) indicates that there are p lagged inputs and k nodes in hidden layer. For seasonal data, last observed values from same season are used as inputs. It is denoted by NNAR(p, P, k)$_m$ which has (y_{t-1}, y_{t-2}, …, y_{t-p}, y_{t-m}, $y_{t-2\,m}$, …, y_{t-Pm}) as inputs and k neurons in hidden layer. Unlike SARIMA, NNAR does not require stationarity of time series. Mathematically NNAR is:

$$y_t = f\left(y_{t-1}\right) + \varepsilon_t \tag{17}$$

where, f: neural network, y_{t-1}: vector containing lagged values of series and ε_t: error series (assumed to be homoscedastic) (Hyndman and Athanasopoulos 2018).

2.2 Study Area and Data Used

The five forecasting models were applied on groundwater records of monitoring wells in the National Capital Territory (NCT) of Delhi. The NCT of Delhi in North India lies between 28° 24′ 15″ N to 28° 53′ 00″ N and 76° 50′24″ E to 77° 20′ 30″ E. It covers an area of 1483 km^2, 75% of which is urbanized (Central Ground Water Board 2016). It is divided into 11 districts. Much of Delhi's rain occurs during the monsoon months – July to September, bringing high humidity levels. Delhi's geological formations vary from Quartzite to Older and Younger Alluvium making the aquifer geology complex (CGWB 2021a).

The Central Ground Water Board (CGWB) has over a 100 monitoring stations spread over Delhi's alluvial and quartzitic area. Historical records of groundwater levels for monitoring stations in NCT of Delhi were obtained from CGWB for winter (January), pre-monsoon (May), monsoon (August) and post-monsoon (November) seasons. Mann–Kendall test on pre-monsoon and post-monsoon groundwater levels (in mbgl) for Delhi at the district level showed an increasing trend indicating that there has been a decline in the groundwater depth in the last two decades (Sarma and Singh 2021b).

2.3 Data Pre-processing and Methodology of Study

To demonstrate the comparison between the time series models, data from three wells were chosen because of their declining groundwater level – Haiderpur (GW1), Bhatti (GW2) and Kitchner Road (GW3). Locations of these wells are presented in Fig. 1. The time period for study was selected depending on data availability – GW1 (1999–2019), GW2 (1996–2019) and GW3 (1983–2017). The datasets were analysed for completeness
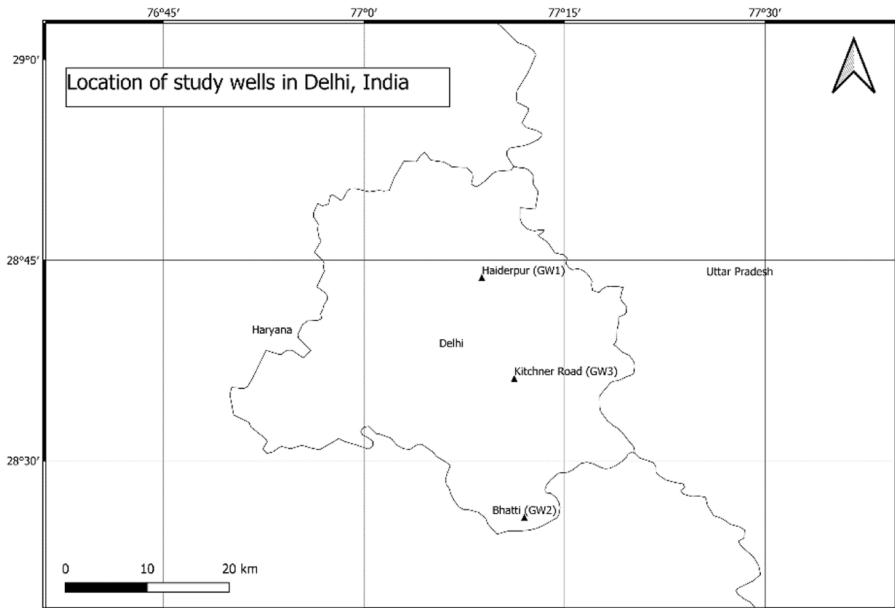
Fig. 1 Locations of selected monitoring wells for the study – Haiderpur (GW1), Bhatti (GW2) and Kitchner Road (GW3)

and a small number of missing values were imputed using the Multivariate Imputation by Chained Equations (MICE) package in R software (van Buuren and Groothuis-Oudshoorn 2011). Multiple imputation has an advantage over single imputation methods in that correlations error obtained from the imputations are not overestimated due to the inclusion of uncertainty because of missing data (Lee and Carlin 2010; Gao et al. 2018). The dataset was divided into two subsets for applying the time series models–80% for training and 20% for testing. To avoid larger values from overriding smaller ones and prevent saturation of hidden nodes, the dataset was normalized between 0 and 1 for MLP and ELM (Eq. (18)).

$$x_n = \frac{(x_i - x_{min})}{(x_{max} - x_{min})} \tag{18}$$

where, x_n: normalized data, x_i: actual value, x_{min}: minimum value and x_{max}: maximum value in each dataset (Shirmohammadi et al. 2013). The HWES, NNAR, SARIMA, MLP and ELM algorithms were run on the training data using their respective packages in R software – *forecast* (Hyndman et al. 2008, 2019; Hyndman and Athanasopoulos 2018), *astsa* (Shumway and Stoffer 2016) and *nnfor* (Crone and Kourentzes 2010; Kourentzes et al. 2014). The complete methodology of the study is depicted in Fig. 2.

2.4 Evaluation of Model Performance

After fitting the training data, each model was used to forecast the groundwater level for the testing period. To assess model performance, accuracy measures–root mean square
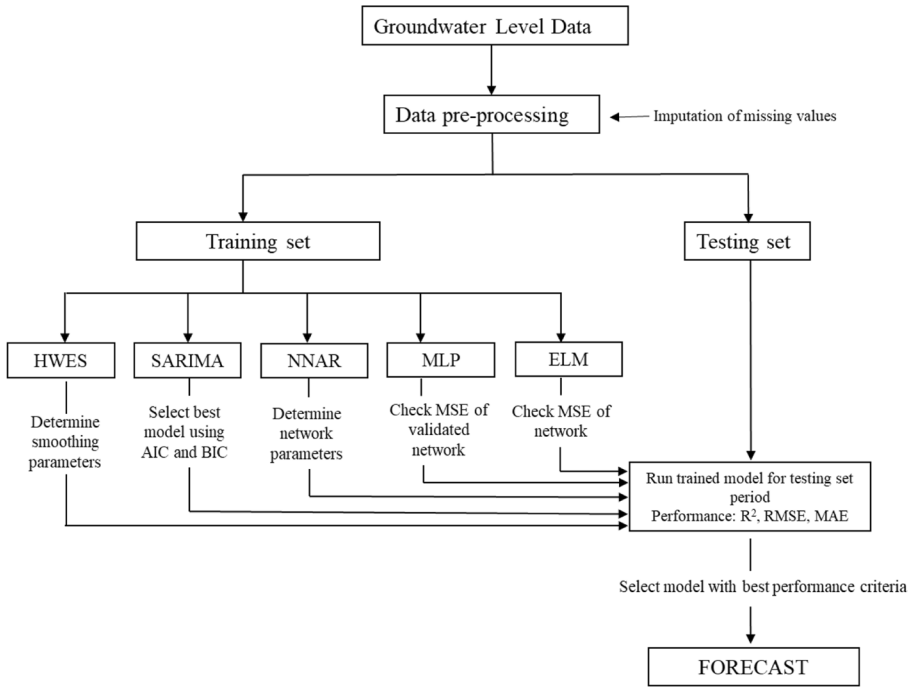
Fig. 2 Methodology of study

error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) were compared. The model with the least RMSE and MAE and maximum R^2 during the training and testing phases was determined as the best model for forecasting groundwater levels in the study area.
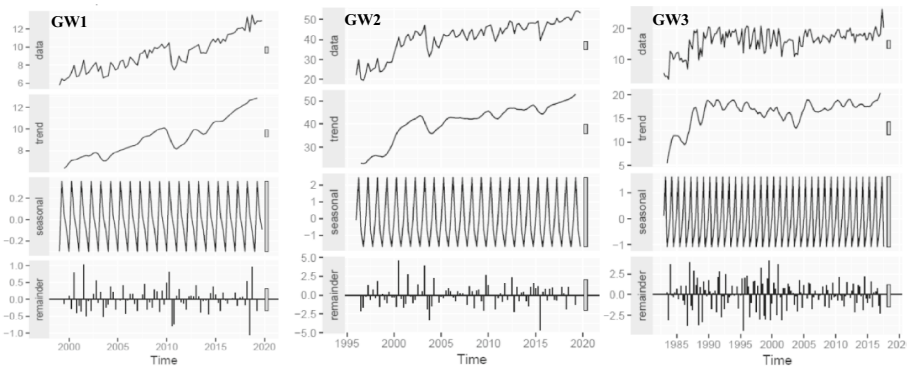
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2} \tag{19}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(O_i - P_i)| \tag{20}$$

$$R^2 = [\frac{\sum_{i=1}^{n}(O_i - \underline{O})(P_i - \underline{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \underline{O})^2 \sum_{i=1}^{n}(P_i - \underline{P})^2}}]^2 \tag{21}$$

where, n: number of data points, O_i: observed variables with mean $\underline{O}$ and P_i: predicted variables with mean $\underline{P}$ (Choubin and Malekian 2017; Yan and Ma 2016).

Fig. 3 Seasonal decomposition of groundwater level time series of GW1, GW2 and GW3 into trend, seasonal and random components

3 Results and Discussion

3.1 Time Series Decomposition

The time series of the three wells were first decomposed into its trend, seasonal and residual components in R software (Fig. 3). Since the data are in metres below ground level (mbgl), all 3 wells showed increasing trend, implying continuous groundwater depletion during the time period of study. The trend ranged over 7 mbgl for GW1, 30 mbgl for GW2 and 15 mbgl for GW3. The seasonal component has an oscillatory amplitude of 0.3 mbgl for GW1, 3 mbgl for GW2 and 2 mbgl for GW3. The seasonal component has a peak in November and trough in May, according to the rainfall patterns in Delhi. Much of the rainfall is received during the monsoon season (July – September) which recharges the groundwater during the post-monsoon months (October – November). During January to June, scanty rainfall is unable to compensate for the groundwater abstraction, thus leading to a lower recharge in May. The residual component fluctuates between ± 1 mbgl for GW1, ± 5 mbgl for GW2 and ± 3 mbgl for GW3. This decomposition showed that for all 3 study wells, the trend has the largest component, followed by the residual and seasonal parts.

3.2 Holt-winters' Exponential Smoothing

The Holt-Winters' model predicts a future value by combining the influences from the level, trend and seasonality. Each component has an associated smoothing parameter that provides information about its influence on the model's predictions (Table 1). The value

Table 1 Smoothing parameters from HWES with trend and additive seasonal component

	Smoothing parameters		
	α	$\beta*$	γ
GW1	0.8191	0.0321	0.2680
GW2	0.7383	0.0000	0.6628
GW3	0.4267	0.0897	0.2982

of α is close to 1 for GW1 and GW2, indicating that prediction of a current observation is mostly based on the immediate past observations. For GW3, lower value of α implies that older observations are weighted more than the most recent ones. All three study wells had β* close to or equal to 0 indicating that the trend does not change much over time and remains fairly constant during each prediction. GW1 and GW3 have γ values close to 0 indicating that like trend, seasonality also does not change. However, for GW2, a larger γ value implies a strong seasonal component. Thus, for GW2, seasonal component of the current observation is based on the seasonality of the most recent observations.

3.3 Seasonal ARIMA

The autocorrelation (ACF) and partial autocorrelation (PACF) plots for the time series of GW1, GW2 and GW3 were inspected to check the stationarity. The plots did not tail off to zero indicating that the series were not stationary and required appropriate differencing. First order non-seasonal differencing was applied on GW1 and GW3 while GW2 required first order seasonal. This corroborated the results from HWES as GW2 showed a strong seasonal component while GW1 and GW3 did not have much influence from seasonality. The Augmented Dickey-Fuller (ADF) test was applied before and after differencing and the p-value was compared. The p-value before differencing was computed as 0.4451, 0.1118 and 0.0419 for GW1, GW2 and GW3 respectively. After the appropriate differencing, the p-value as determined from the ADF test was 0.01, rejecting the null hypothesis that the series has a unit root and is thus stationary.

With the appropriate differencing, the value of d or D of the SARIMA model was determined. Different combinations of p, P, q and Q were tested on trial-and-error method. The ACF and PACF plots of each of these models were inspected and model selection criteria AIC and BIC were compared (Table S1). The best model was selected based on the minimum AIC and BIC values. The final models selected were $(1,1,1)(1,0,0)_4$, $(1,0,0)(0,1,1)_4$ and $(1,1,1)(1,0,1)_4$ for GW1, GW2 and GW3 respectively.

3.4 MLP, ELM and NNAR

The MLP networks were trained in R using a validation argument and specified number of lags to prevent overfitting the training data (Fig. 4). The grey inputs represent
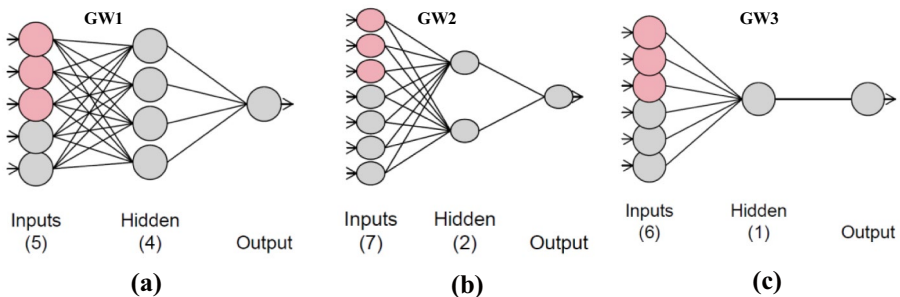
Fig. 4 MLP network for **a** GW1 (5 inputs, 1 hidden layer, 4 nodes) **b** GW2 (7 inputs, 1 hidden layer, 2 nodes) and **c** GW3 (6 inputs, 1 hidden layer, 1 node)

Table 2 ELM and NNAR model descriptions for the study wells

	ELM network	**Univariate lags**
GW1	Fit with 62 hidden nodes	1,3
GW2	Fit with 69 hidden nodes	1,2,3,4
GW3	Fit with 100 nodes	1,2,3
	NNAR model	**Description**
GW1	NNAR(1,1,2)$_4$	2–2–1 network with 9 weights
GW2	NNAR(1,1,2)$_4$	2–2–1 network with 9 weights
GW3	NNAR(9,1,5)$_4$	9–5–1 network with 56 weights

autoregressive lags and pink inputs represent seasonality. The mean square error (MSE) for the trained MLP networks were 0.0022, 0.0036 and 0.0127 for GW1, GW2 and GW3 respectively. The ELM networks were trained in R using the lasso regression method. The time taken for training was significantly lesser than the MLP networks. The mean square error (MSE) for the trained ELM networks were 0.0072, 0.0062 and 0.0156 for GW1, GW2 and GW3 respectively. The NNAR models were applied on the training data with Box-Cox transformation (lambda=0). This transformation ensured that the residuals were homoscedastic. The descriptions of the resultant ELM and NNAR networks are presented in Table 2.

3.5 Comparison of Model Performance

The accuracy measures were compared for the models for each study well (Table 3). For GW1 and GW2, MLP performed the best for both the training and testing phases with highest R^2 and lowest RMSE and MAE. MLP best fitted the observed values with the predicted values during training and testing (Fig. 5). For GW3, HWES had the highest RMSE and MAE among all models. NNAR performed very well during training but gave unrealistic forecasts during testing. There may have been overfitting of data, indicated by the large parameters in the network, i.e. 9–5–1 network with 56 weights. The MLP network yielded low R^2 values for both training and testing. MLP network for GW3 had only 1 hidden node (Fig. 4) and compared to GW1 and GW2, MSE for GW3 was higher. ELM presented very high RMSE and MAE for both training and testing. SARIMA had slightly better R^2 but MLP had the lowest errors. Thus, MLP was considered the best performing model for GW3 as well.

The best model MLP was used to forecast the groundwater level for the year 2025. MLP forecasts for May 2025 showed that groundwater level will fall by 2 mbgl and 21 mbgl below the 2019 level for GW1 and GW2 respectively and 3 mbgl below the 2017 level for GW3.

3.6 Discussion

The results from applying the five models suggest that with validation in the training set, MLP can train the model with R^2 as high as 0.914. For GW1 and GW2, SARIMA and

	Training			Testing		
	R^2	**RMSE**	**MAE**	R^2	**RMSE**	**MAE**
GW1						
HWES	0.746	0.638	0.455	0.729	0.597	0.469
SARIMA	0.800	0.617	0.479	0.379	1.813	1.616
MLP	0.914	0.047	0.035	0.733	0.069	0.052
ELM	0.731	0.085	0.062	0.721	0.112	0.093
NNAR	0.793	0.561	0.431	0.516	2.266	2.111
GW2						
HWES	0.882	2.557	1.822	0.708	5.377	5.041
SARIMA	0.912	2.448	1.676	0.671	5.033	4.612
MLP	0.910	0.060	0.041	0.710	0.139	0.128
ELM	0.852	0.079	0.058	0.705	0.112	0.095
NNAR	0.873	2.541	1.924	0.633	5.166	4.595
GW3						
HWES	0.271	3.060	2.323	0.284	3.071	2.803
SARIMA	0.558	2.645	1.989	0.246	1.839	1.351
MLP	0.436	0.113	0.088	0.195	0.104	0.086
ELM	0.354	2.793	2.021	0.266	2.282	1.902
NNAR	0.931	0.904	0.659	0.017	2.263	1.475

Table 3 Comparison of accuracy measures for selected models

HWES also gave high R^2 values during training. For the testing period, the R^2 tends to decrease slightly for MLP and HWES and dramatically for SARIMA. Thus, fitting precision of a model may not necessarily imply accurate forecasting which makes evaluating the model forecasts against a testing subset a crucial step in the modeling process. ELM gave the lowest R^2 for GW1 and GW2 during training. NNAR had overall low efficiency which was also observed in the study by Aguilera et al. (2019). These models may perform better by selecting the appropriate activation functions, lagged variables as inputs or numerical procedure (Faraway and Chatfield 1998). MLP, ELM, HWES and SARIMA did not perform well on the GW3 dataset. Even though the dataset for GW3 was large with 139 observations over 1983–2017, there was overall low performance of the models. Groundwater levels are influenced by factors like rainfall, soil properties, surface water and abstraction (Lee et al. 2019) and during interpretation of model results, it may be pertinent to analyse these influencing factors as well.

MLP forecast for 2025 presents an alarming result for GW2, a monitoring well located in the southern part of the study area (Fig. 1). Declining trends in the South district of Delhi have been reported by the CGWB, particularly due to over-exploitation (CGWB 2016, 2021b). Results of aquifer response modeling for Delhi using MODFLOW also showed a similar decline for GW2 (Bhatti) (CGWB 2016), necessitating an urgent and effective groundwater management plan in that region.
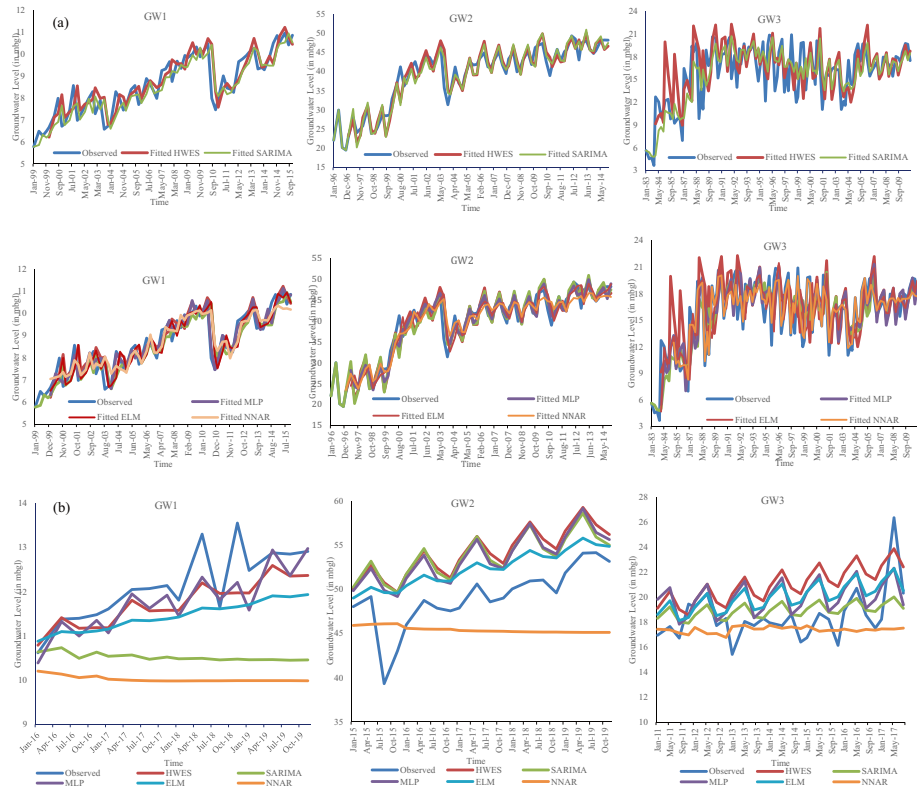
Fig. 5 Fitting the observed and simulated values for **a** training and **b** testing

4 Conclusion

This study used univariate time series forecasting methods HWES, Seasonal ARIMA, MLP, ELM and NNAR for prediction of groundwater levels in Delhi, India. Decomposition of the time series of three monitoring wells showed increasing trend, indicating that the groundwater level has continuously declined over the study period. The parameters of HWES and SARIMA were calculated and MLP, ELM and NNAR models were trained and validated. The accuracy of the models was assessed based on the R^2, RMSE and MAE. MLP had the least values of RMSE and MAE and highest R^2 for two wells during training and testing, indicating that the MLP approach was most accurate in forecasting the groundwater levels. For the third well, MLP had a slightly lower R^2 but least RMSE and MAE and was also concluded to be better than HWES, SARIMA, ELM and NNAR for making predictions.

Such univariate forecasting studies are particularly suitable for regions where large hydro-climatological data are unavailable. MLP was used to make further predictions for May 2025 for all three wells. Results indicate that groundwater level in the three wells will decline by 2–21 mbgl. This study has helped identify the areas that require urgent groundwater management decisions. Further research may be done on adding more arguments in the model algorithm to give the lowest errors. It is recommended that more comparison

studies should be made for other regions in India so that the most accurate models help in identifying the areas where groundwater is declining rapidly.

Declarations

Ethical approval Not applicable.

Consent to Participate Not applicable.

Consent to Publish The authors hereby grant all rights of publication of the manuscript to the publisher.

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

References

Aguilera H, Guardiola-Albert C, Naranjo-Fernández N, Kohfahl C (2019) Towards flexible groundwater-level prediction for adaptive water management: using Facebook's Prophet forecasting approach. Hydrol Sci J 64(12):1504–1518. https://doi.org/10.1080/02626667.2019.1651933

Akaike H (1974) A New Look at the Statistical Model Identification. IEEE Trans Autom Control 19(6):716–723. https://doi.org/10.1109/TAC.1974.1100705

Aslam M, Lee JM, Hong S (2020) A multi-layer perceptron based deep learning model to quantify the energy potentials of a thin film a-Si PV system. Energy Rep 6:1331–1336. https://doi.org/10.1016/j.egyr.2020.11.025

Barzegar R, Asghari Moghaddam A, Adamowski J, Fijani E (2017) Comparison of machine learning models for predicting fluoride contamination in groundwater. Stochastic Environ Res Risk Assess 31(10):2705–2718. https://doi.org/10.1007/s00477-016-1338-z

Box GEP, Jenkins GM (1976) Series analysis forecasting and control. Prentice-Hall Inc., London

Brown RG (1959) Statistical forecasting for inventory control. McGraw-Hill, New York

Central Ground Water Board (2016) Aquifer mapping and ground water management plan of NCT Delhi. New Delhi. http://cgwb.gov.in/AQM/AQM-Reports.html. Accessed 20 Oct 2021

Central Ground Water Board (2021) Groundwater Yearbook 2019-20. New Delhi. http://cgwb.gov.in/GW-Year-Book-State.html. Accessed 15 Nov 2021

Central Ground Water Board (2021b) Report: Dynamic ground water resources of NCT Delhi as on March 2020. New Delhi. http://cgwb.gov.in/Dynamic-GWResources.html. Accessed 20 Nov 2021

Chen C, He W, Zhou H, Xue Y, Zhu M (2020) A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin. Northwestern China Sci Rep 10:3904. https://doi.org/10.1038/s41598-020-60698-9

Choubin B, Malekian A (2017) Combined gamma and M-test-based ANN and ARIMA models for groundwater fluctuation forecasting in semiarid regions. Environ Earth Sci 76(15). https://doi.org/10.1007/s12665-017-6870-8

Crone SF, Kourentzes N (2010) Feature selection for time series prediction - A combined filter and wrapper approach for neural networks. Neurocomputing 73(10–12):1923–1936. https://doi.org/10.1016/j.neucom.2010.01.017

Faraway J, Chatfield C (1998) Time series forecasting with neural networks: A comparative study using the airline data. J R Stat Soc, C: Appl Stat 47(2):231–250. https://doi.org/10.1111/1467-9876.00109

Gao Y, Merz C, Lischeid G, Schneider M (2018) A review on missing hydrological data processing. Environ Earth Sci 77:47. https://doi.org/10.1007/s12665-018-7228-6

Hecht-Nielsen R (1989) Theory of the backpropagation neural network. In International 1989 Joint Conference on Neural Networks. Washington, DC, USA

Holt CC (1957) Forecasting trends and seasonals by exponentially weighted averages. In Carnegie Institute of Technology, Pittsburgh ONR memorandum no. 52

Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1):489–501. https://doi.org/10.1016/j.neucom.2005.12.126

Hyndman RJ, Koehler AB, Ord JK, Snyder RD (2008) Forecasting with Exponential Smoothing: The State Space Approach. Springer

Hyndman R et al (2019) Forecast: forecasting functions for time series and linear models. R package version 8.7. http://pkg.robjhyndman.com/forecast. Accessed 15 Oct 2021

Hyndman RJ, Athanasopoulos G (2018) Forecasting: principles and practice. Melbourne, Australia: OTexts. https://otexts.org/fpp2/. Accessed 20 Oct 2021

Kalteh AM (2019) Modular wavelet-extreme learning machine: A new approach for forecasting daily rainfall. Water Resour Manag 33:3831–3849. https://doi.org/10.1007/s11269-019-02333-5

Kourentzes N, Barrow DK, Crone SF (2014) Neural network ensemble operators for time series forecasting. Expert Syst Appl 41(9):4235–4244. https://doi.org/10.1016/j.eswa.2013.12.011

Lallahem S, Mania J, Hani A, Najjar Y (2005) On the use of neural networks to evaluate groundwater levels in fractured media. J Hydrol 307(1–4):92–111. https://doi.org/10.1016/j.jhydrol.2004.10.005

Lee KJ, Carlin JB (2010) Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. Amer J Epidemiol 171:624–632

Lee S, Lee KK, Yoon H (2019) Using artificial neural network models for groundwater level forecasting and assessment of the relative impacts of influencing factors. Hydrogeol J 27(2):567–579. https://doi.org/10.1007/s10040-018-1866-3

Mirzavand M, Ghazavi R (2015) A stochastic modelling technique for groundwater level forecasting in an arid environment using time series methods. Water Resour Manag 29(4):1315–1328. https://doi.org/10.1007/s11269-014-0875-9

Mozaffari S, Javadi S, Moghaddam HK, Randhir TO (2022) Forecasting groundwater levels using a hybrid of support vector regression and particle swarm optimization. Water Resour Manag. https://doi.org/10.1007/s11269-022-03118-z

Narayanan P, Basistha A, Sarkar S, Kamna S (2013) Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India. C R Geosci 345(1):22–27. https://doi.org/10.1016/j.crte.2012.12.001

Natarajan N, Sudheer C (2020) Groundwater level forecasting using soft computing techniques. Neural Comput Appl 32:7691–7708. https://doi.org/10.1007/s00521-019-04234-5

Noori AR, Singh SK (2021) Spatial and temporal trend analysis of groundwater levels and regional groundwater drought assessment of Kabul. Afghanistan Environ Earth Sci 80:698. https://doi.org/10.1007/s12665-021-10005-0

Parisouj P, Mohebzadeh H, Lee T (2020) Employing machine learning algorithms for streamflow prediction: A case study of four river basins with different climatic zones in the United States. Water Resour Manag 34:4113–4131. https://doi.org/10.1007/s11269-020-02659-5

Poursaeid M, Poursaeid AH, Shabanlou SA (2022) A comparative study of artificial intelligence models and a statistical method for groundwater level prediction. Water Resour Manag 36:1499–1519. https://doi.org/10.1007/s11269-022-03070-y

Rahaman M, Thakur B, Kalra A, Ahmad S (2019) Modeling of GRACE-derived groundwater information in the colorado river basin. Hydrol 6(1):19. https://doi.org/10.3390/hydrology6010019

Sakizadeh M, Mohamed MMA, Klammler H (2019) Trend analysis and spatial prediction of groundwater levels using time series forecasting and a novel spatio-temporal method. Water Resour Manag 33(4):1425–1437. https://doi.org/10.1007/s11269-019-02208-9

Salvadore E, Bronders J, Batelaan O (2015) Hydrological modelling of urbanized catchments: A review and future directions. J Hydrol 529(1):62–81. https://doi.org/10.1016/j.jhydrol.2015.06.028

Sarma R, Singh SK (2021a) Temporal variation of groundwater levels by time series analysis for NCT of Delhi, India. In: Mehta YA, Carnacina I, Kumar DN, Rao KR, Kumari M (eds) Advances in Water Resources and Transportation Engineering. Lecture Notes in Civil Engineering (149). Springer, Singapore. https://doi.org/10.1007/978-981-16-1303-6_15

Sarma R, Singh SK (2021b) Simulating contaminant transport in unsaturated and saturated groundwater zones. Water Environ Res 93(9):1496–1509. https://doi.org/10.1002/wer.1555

Schwarz G (2007) Estimating the dimension of a model. Ann Stat 6(2):461–464. https://doi.org/10.1214/aos/1176344136

Shirmohammadi B, Vafakhah M, Moosavi V, Moghaddamnia A (2013) Application of several data-driven techniques for predicting groundwater level. Water Resour Manag 27(2):419–432. https://doi.org/10.1007/s11269-012-0194-y

Shumway RH, Stoffer DS (2016) Time series analysis and its applications With R Examples EZ Edition. Springer Texts in Statistics. Springer Series in Statistics, EZ Edition, New York

Singh A (2014) Groundwater resources management through the applications of simulation modeling: A review. Sci Total Environ 499:414–423. https://doi.org/10.1016/j.scitotenv.2014.05.048

Takafuji EHDM, Rocha MMD, Manzione RL (2019) Groundwater level prediction/forecasting and assessment of uncertainty using SGS and ARIMA models: A case study in the Bauru Aquifer System (Brazil). Nat Resour Res 28(2):487–503. https://doi.org/10.1007/s11053-018-9403-6

van Buuren S, Groothuis-Oudshoorn K (2011) MICE: Multivariate imputation by chained equations in R. J Stat Softw 45(3):1–67. https://www.jstatsoft.org/v45/i03/

Winters PR (1960) Forecasting sales by exponentially weighted moving averages. Manage Sci 6(3):324–342. https://doi.org/10.1287/mnsc.6.3.324

Yan Q, Ma C (2016) Application of integrated ARIMA and RBF network for groundwater level forecasting. Environ Earth Sci 75(5):1–13. https://doi.org/10.1007/s12665-015-5198-5

Yang Q, Wang Y, Zhang J, Delgado J (2017) A comparative study of shallow groundwater level simulation with three time series models in a coastal aquifer of South China. Appl Water Sci 7(2):689–698. https://doi.org/10.1007/s13201-015-0282-2

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(4–5):487–492. https://doi.org/10.1016/j.jaridenv.2008.11.008

Zhang Z, Zhang Q, Singh VP (2018) Univariate streamflow forecasting using commonly used data-driven models: literature review and case study. Hydrol Sci J 63(7):1091–1111. https://doi.org/10.1080/02626667.2018.1469756