Indian Journal of Geo Marine Sciences
Vol. 49 (01), January 2020, pp. 44-50

Ensemble based groundwater level prediction using neural network pattern fitting

Ajith Kumar S[*] & Vidhya R

Institute of Remote Sensing, Anna University, Tamil Nadu, Chennai-600025, India

[*][E-mail: ajith.ajith@gmail.com]

Received 23 July 2018; revised 10 September 2018

Prediction of groundwater level is implemented using Time-series prediction model and combined prediction model for learning the pattern and trend in groundwater level fluctuation, result show that the combined prediction model using, groundwater level time series and precipitation time series as input predictors is a better predictor. Study also shows that prediction is dependent on the pattern and trends at a particular location as every dataset depends on the dynamics of the location namely the geomorphology of the aquifer, the drainage inside the aquifer and pumping from the aquifer. Ensemble based forecasting is studied to fix the upper and lower limit of the prediction. Ensembles helped in fixing a range for the forecast instead of relying on a single unique value.

[**Keywords**: Combined model; Data mining; Ensemble; Forecasting; Groundwater level prediction; Time series]

Introduction

Water managers to help schedule the cropping in arid climatic zones where dependency of rainfall is not reliable can use groundwater level information. Groundwater being a depleting resource we need to monitor and plan the pumping activity in order to sustain the groundwater resource. For sustainability, the rate of abstraction from water table should be less than the recharge; sustainability ensures the water demand for present and future use. Groundwater modelling need hydro/geological characterization to understand the flow into and out of groundwater system, these mathematical models required number of equations to describe the phenomenon under consideration in order to forecast and predict the future availability of groundwater.

Madhumita Sahoo et al.[1,2], predicts groundwater level (GWL) outcome using the time series to changes in model input, using mathematical model require monitoring and characterizing number of parameters on a spatial scale and these observations are very costly to capture and implement. Hence, a data mining based simulation computational model that makes use of sparse input parameters to predict the outcome next time-step GWL will be a good alternative.

Ensemble Forecasting

Ensemble forecasting is the process of creating different models to predict an outcome either by using different algorithms or by using different dataset, each model will be an ensemble member. Ensemble forecasting reduces the generalization error of prediction using one model. Schaake et al.[3] discussed the procedures for developing ensemble systems making use of multidisciplinary collaboration, ensembles are needed due to uncertainty in initial conditions[4], and consequences of drought can be mitigated by predicting drought in advance using ensemble learners[5]. Ensemble model developed by changing the base model; bagging is the process of changing the training set of every base model that may contain duplicate records called boot strapping. Boosting is similar to bagging but they concentrate on records hard to classify, and over represent them in the training set for next iteration[6]. Krzysztofowicz[7] described the danger of providing deterministic, single forecasts to decision makers; such situation can lead to disaster. Rule-based strategies on ensemble prediction consist of a combination of heuristic rules and pre-defined anticipatory actions, e.g. if this then do that, else do something else[8].

With different choice of input for the same output, we can develop number of models that can predict the outcome with same accuracy; one forecast that contains a number of alternative predictions for the same forecast period, one prediction is an ensemble member.

In ensemble method, the model run several times with different initial conditions or parameters, each ensemble member has equally likely probability. The differences in forecast of ensemble members give the uncertainty of the particular forecast. The differences in the ensemble member provide information about the uncertainty of the particular forecast.

For irrigation proactive management to be effective the actions taken using all ensemble members. Based on all the information acquired, operational management decisions can be taking into account the uncertainty in the forecast.

Artificial Neural Network[9-11] (ANN) is a massively parallel-distributed information-processing system that can recognize patterns and learn from their interactions has been successfully used in predictive modeling[12,-14] by data mining[15,16]. Case studies of data-driven application using ANN in hydrology found in article[17,18].

Ensemble members

The current study makes use of all the prediction models with satisfactory prediction accuracy for the same outcome as ensemble members, the difference in forecast of prediction model is analysed to find the upper limit and lower limit of the forecast, and the difference between the upper limit and lower limit give control interval of the forecast to take decisions.

Materials and Methods

Description of study location

Prediction of groundwater level (GWL) from sparse dataset is the task. Prediction simulation for Veppaneri, K. V. Kuppam (Block), Vellore District located at latitude 12.955 and Longitude 78.994 north of Palar river and Anicut, Vellore district south of Palar River located at latitude 12.877, and longitude 78.988 at an altitude of 206 m above MSL, different models are simulated using ANN and tested for the dataset (Fig. 1).

Dataset

For the current study, groundwater level (GWL), is predicated for future period seasonal (quarterly) time horizon using the available historical groundwater level data recorded seasonally between 1998 to 2014 for study location Vellore provided by Central Ground Water Board (CGWB, India), accessed from the URL http://cgwb.gov.in/GW-data-access.html. Two wells located on Palar river basin near Vellore is studied. Different prediction models tested using the GWL and precipitation information for the region.

Data pre-processing and modelling

Data pre-processing of the historical dataset done to fit the different time series prediction model and combined prediction model to fit the input predictors used for learning the patterns and trends to predict next season groundwater level. The different models implemented for the current study summarized in Table 1 and Table 2 for time series and combined model prediction.

Evaluation Criteria

Pearson correlation coefficient (R) is used to evaluate the model. Correlation coefficient is a

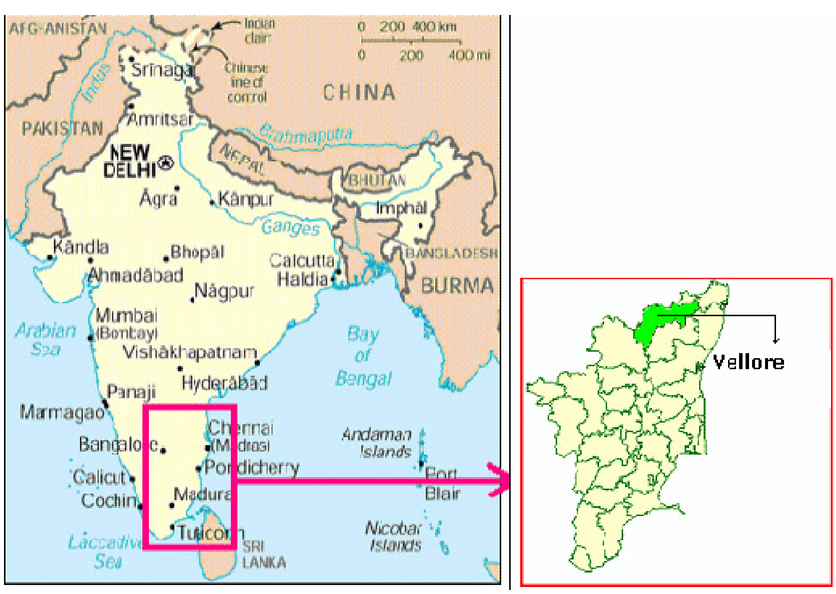

Fig. 1 — Study location map

Table 1 — Time Series GWL Prediction Model

Time Series Model (1 Season before prediction)	VEPPANERI Correlation Coefficient (R)	ANICUT Correlation Coefficient (R)
GWT_Level[y+1] = f{GWL[y]}	0.8	0.38
GWT_Level[y+1] = f{GWL[y], GWL[y-1]}	0.79	0.48
GWT_Level[y+1] = f{GWL[y], GWL[y-1], GWL[y-2]}	0.75	0.53
GWT_Level[y+1] = f{GWL[y], GWL[y-1], GWL[y-2], GWL[y-3]}	0.77	0.63
GWT_Level[y+1] = f{GWL[y], GWL[y-1], GWL[y-2], GWL[y-3],GWL[y-4]}	0.77	0.7

Table 2 — Combined GWL Prediction Model

Combined Model(1 Season before prediction)	VEPPANERI Correlation Coefficient (R)	ANICUT Correlation Coefficient (R)
GWT_Level[y+1] =f{ P[y],GWL[y]}	0.81	0.7
GWT_Level[y+1] =f{ P[y], GWL[y], GWL[y-1]}	0.79	0.68
GWT_Level[y+1] =f{ P[y], GWL[y], GWL[y-1], GWL[y-2]}	0.78	0.66
GWT_Level[y+1] =f{ P[y], GWL[y], GWL[y-1], GWL[y-2],GWL[y-3]}	0.776	0.66
GWT_Level[y+1] =f{ P[y], P[y-1], GWL[y]}	0.83	0.69
GWT_Level[y+1] =f{ P[y], P[y-1], GWL[y], GWL[y-1]}	0.82	0.689
GWT_Level[y+1] =f{ P[y], P[y-1], GWL[y], GWL[y-1], GWL[y-2]}	0.8	0.662
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2] , GWL[y]}	0.829	0.756
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2], GWL[y], GWL[y-1]}	0.81	0.81
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2], GWL[y], GWL[y-1], GWL[y-2]}	0.788	0.725
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2], GWL[y], GWL[y-1], GWL[y-2]}	0.789	0.732
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2], P[y-3], GWL[y]}	0.83	0.651
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2], P[y-3], GWL[y], GWL[y-1]}	0.815	0.551
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2], P[y-3], GWL[y], GWL[y-1], GWL[y-2]}	0.788	0.677
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2], P[y-3], GWL[y], GWL[y-1], GWL[y-2]}	0.8	0.699

measure of the strength and direction of the linear relationship between observed and forecast values, which defined as the covariance of the observed and forecast variables divided by the product of their standard deviations.

Pearson correlation coefficient (R) = Correlation of (Observed, Forecasted) Values

Correlation = covariance (observed, predicted)/ {standard deviation (observed) * standard deviation (predicted)}

The result of the correlation accuracy of observed and predicted GWL is summarized in Table 1 and Table 2.

Results

Time-series prediction model

Time series of GWL seasonal data considered as model input for predicting next season water level. Prediction model makes use of one or more time step-ahead GWL data for predicting the next season GWL. Different input model considered and validated using the Pearson correlation coefficient (R) accuracy achieved for observed GWL and predicted outcome GWL summarized in Table 1, for the two well locations Veppaneri and Anicut. The dataset simulated using the MLP neural network machine-learning algorithm.

Results of the simulation show that two well locations depicted different pattern to learning the sparse historical dataset as dynamics of water level variation being different. Veppaneri well location needed only the previous season water level data to predict the next season outcome with an R-value of 0.8, whereas Anicut well location needed five season before information to learn the well dynamics to predict the outcome with an R-value of 0.7.

Combined prediction model

Prediction model making use of previous GWL and precipitation seasonal data considered as model input for predicting next season water level. Prediction model made use of one or more time step-ahead precipitation and GWL data for predicting the next season GWL. Different input prediction model considered and validated using the Pearson correlation coefficient (R) accuracy achieved for observed GWL and predicted outcome GWL summarized in the Table 2. The two well location Veppaneri and Anicut dataset simulated using the MLP neural network machine-learning algorithm.

Results show that two wells depict different pattern to learn the sparse historical dataset, dynamics of water level variation being different; the learning and prediction accuracy was different for the two well location using different prediction models. The performance of the combined prediction model is better than the time series prediction model as inclusion of precipitation was able to better capture the cycle in the data series by pattern fitting. Prediction model that showed the best prediction accuracy shown in Figure 2 and Figure 3 for Veppaneri (Model 5) and Anicut (Model 9) outcomes.

Figure 2 for Model 5 shows the variation between the observed and predicted GWL for Veppaneri, the outcome shows peak water level drops is not captured by the simulation models.

Figure 3 for Model 9 shows the variation between the observed and predicted GWL for Anicut, here the outcome show the simulation models do not capture that peak water level drop.

Computational simulation prediction models for predicting GWL shows a fair understanding of well dynamics using the combined prediction model, decision makers can use this advisory of GWL with other weather information and experiences in planning for groundwater use policies.

Ensemble forecasting of GWL

Ensemble members selected from among the best-correlated models from Table 1 and Table 2. The computational models considered as ensembles for Veppaneri well location is listed in Table 3 and the computational models considered as ensembles for

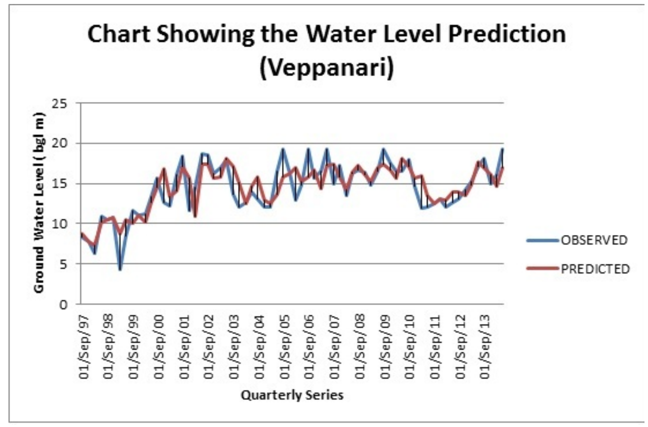

Fig. 2 — Graph showing the Observed and Predicted GWL variation (Model 5)

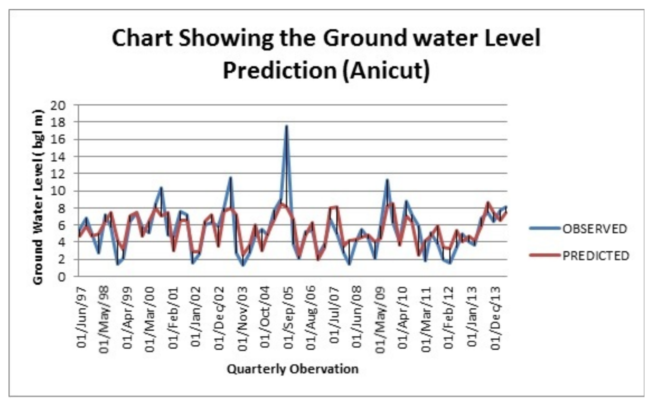

Fig. 3 — Graph showing the Observed and Predicted GWL variation (Model 9)

Table 3 — Prediction model chosen as ensemble's for VEPPANERI location

Prediction Model	Coefficient of Correlation(R)
GWT_Level[y+1] = f{GWL[y]}	0.8
GWT_Level[y+1] = f{GWL[y], GWL[y-1]}	0.79
GWT_Level[y+1] =f{ P[y],GWL[y]}	0.81
GWT_Level[y+1] =f{ P[y], GWL[y], GWL[y-1]}	0.79
GWT_Level[y+1] =f{ P[y], P[y-1], GWL[y]}	0.83
GWT_Level[y+1] =f{ P[y], P[y-1], GWL[y], GWL[y-1]}	0.82
GWT_Level[y+1] =f{ P[y], P[y-1], GWL[y], GWL[y-1], GWL[y-2]}	0.8
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2] , GWL[y]}	0.829
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2], GWL[y], GWL[y-1]}	0.81
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2], P[y-3], GWL[y]}	0.83
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2], P[y-3], GWL[y], GWL[y-1]}	0.815
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2], P[y-3], GWL[y], GWL[y-1], GWL[y-2]}	0.788
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2], P[y-3], GWL[y], GWL[y-1], GWL[y-2]}	0.8

Anicut well location is listed in Table 4, upper limit (UCL) and lower limit (LCL) predicted using the good performing model is given in Table 5 and Table 6 for Veppaneri and Anicut location GWL prediction.

Figure 4 and figure 5 shows the graphical variation of the upper and lower limit of the prediction along with the observed GWL value for Veppaneri groundwater well. Similarly figure 6 and figure 7 shows the graphical variation of the upper and lower limit of the prediction along with the observed GWL value for Anicut groundwater well.

Discussion

Analysis of the prediction results (Fig. 2 and Fig. 3) showed, a maximum deviation of about 4 m for Veppaneri water level prediction and a maximum deviation of 4.7 m for Anicut water level prediction indicating that the simulation models do not capture peak drop and rise in water levels. Accuracy was improved using a longer dataset to capture the trend and dynamics and involved some more input parameters that may influence the well dynamics like pumping information.

Prediction accuracy of the model is completely depend upon well dynamics. If we can include input parameters that are unique for a particular well then we may get better prediction accuracy.

The ensemble prediction accuracy Table 6 and Table 7 gives ground water level prediction sensitivity showed a maximum variance of 2.18 m and 3.85 m for the well location Veppaneri and Anicut, respectively. Well located at Veppaneri showed a maximum under and over prediction accuracy of 1.08 m and 3.25 m, respectively during June 2014.

Table 4 — Prediction model chosen as ensemble's for ANICUT location

Prediction Model	Coefficient of Correlation (R)
GWT_Level[y+1] = f{GWL[y], GWL[y-1], GWL[y-2], GWL[y-3],GWL[y-4]}	0.7
GWT_Level[y+1] =f{ P[y],GWL[y]}	0.7
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2] , GWL[y]}	0.756
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2], GWL[y], GWL[y-1]}	0.81
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2], GWL[y], GWL[y-1], GWL[y-2]}	0.725
GWT_Level[y+1] =f{ P[y], P[y-1], P[y-2], GWL[y], GWL[y-1], GWL[y-2]}	0.732

Table 5 — Ensemble Predicted Upper control limit (UCL) and Lower control limit (LCL) and Observed GWL for VEPPANERI location

Predicted Dates	PREDICTED (UCL) (m)	OBSERVED GWL (m)	PREDICTED (LCL) (m)	Difference (UCL - LCL) (m)
31-Mar-09	16.34861	14.79	14.9478	1.400811
30-Jun-09	16.96921	16.49	15.25145	1.717759
30-Sep-09	17.70769	19.19	16.09069	1.617
31-Dec-09	17.37095	17.59	16.59584	0.775117
31-Mar-10	16.50845	16.45	15.41273	1.095721
30-Jun-10	18.12195	16.5	15.96568	2.156263
30-Sep-10	17.66115	17.89	16.36219	1.29896
31-Dec-10	16.53295	14.64	14.79089	1.742064
31-Mar-11	15.94583	11.85	14.07145	1.87438
30-Jun-11	14.53205	12	12.37718	2.154876
30-Sep-11	13.30886	12.5	11.99315	1.315703
31-Dec-11	13.90335	13	12.41199	1.491362
31-Mar-12	14.13513	12.04	12.14625	1.988873
30-Jun-12	14.59488	12.6	12.67181	1.923067
30-Sep-12	14.75335	13	13.57579	1.17756
31-Dec-12	14.19401	14.19	12.35104	1.842971
31-Mar-13	15.37317	15.24	13.4432	1.929964
30-Jun-13	17.69248	17.09	15.54391	2.148572
30-Sep-13	17.87155	18.04	16.4348	1.436747
31-Dec-13	16.9576	15	16.05643	0.90117
31-Mar-14	15.88173	16.09	14.27743	1.604302
30-Jun-14	18.10716	19.19	15.93866	2.168503

Table 6 — Ensemble Predicted Upper control limit (UCL) and Lower control limit (LCL) and Observed GWL for Anicut location

Predicted Dates	PREDICTED (UCL) (m)	OBSERVED GWL (m)	PREDICTED (LCL) (m)	Difference (UCL - LCL) (m)
31-Mar-09	4.00	2.07	2.54	1.46
30-Jun-09	6.77	5.9	4.42	2.34
30-Sep-09	8.22	11.3	6.63	1.59
31-Dec-09	8.53	6.25	5.17	3.36
31-Mar-10	6.33	4.03	3.64	2.69
30-Jun-10	7.24	8.75	6.48	0.76
30-Sep-10	8.24	7.24	6.41	1.83
31-Dec-10	5.21	5.86	2.47	2.74
31-Mar-11	4.25	1.8	2.97	1.28
30-Jun-11	6.54	5.15	4.58	1.95
30-Sep-11	6.85	3.85	5.82	1.02
31-Dec-11	3.36	2	2.35	1.00
31-Mar-12	3.30	1.65	2.01	1.29
30-Jun-12	6.21	3.35	4.62	1.58
30-Sep-12	6.07	5	4.11	1.95
31-Dec-12	4.69	4	3.35	1.34
31-Mar-13	4.17	3.65	2.02	2.14
30-Jun-13	6.55	6.85	5.20	1.34
30-Sep-13	8.62	7.45	6.72	1.89
31-Dec-13	7.50	6.35	3.64	3.85
31-Mar-14	6.93	7.75	5.14	1.79
30-Jun-14	8.47	8.05	7.37	1.09

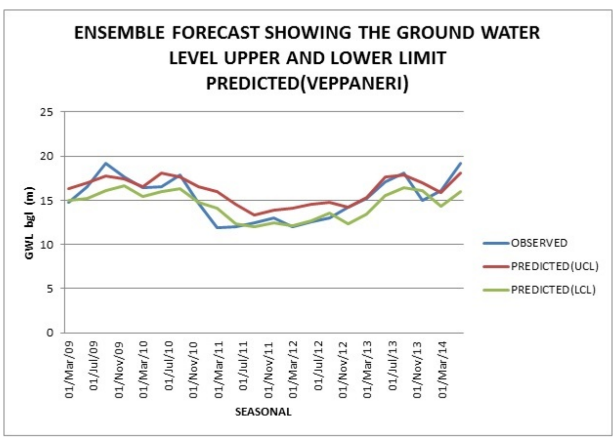

Fig. 4 — Line Graph Showing Observed, Upper and Lower limit Predicted (GWL)

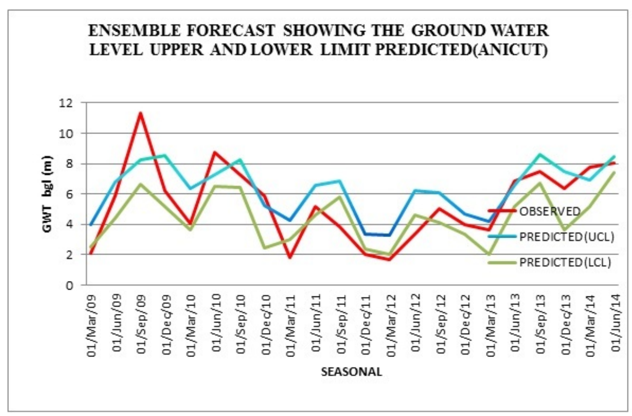

Fig. 6 — Line Graph Showing Observed, Upper and Lower limit Predicted (GWL)

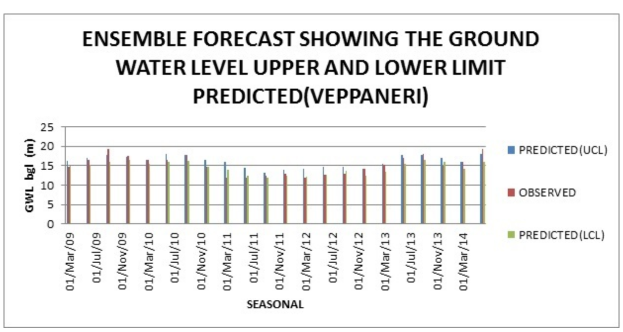

Fig. 5 — Bar Chart Showing Observed, Upper and Lower limit Predicted (GWL)

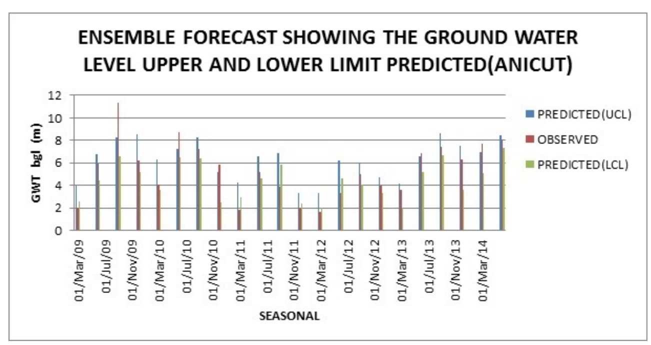

Fig. 7 — Bar Chart Showing Observed, Upper and Lower limit Predicted (GWL)

Well located at Anicut showed a maximum under and over prediction of 3.07 m and 4.67 m during September 2009.

The control limits of water level prediction can give an idea of the water stress forecasted for a location.

Conclusion

Prediction statistical accuracy shows that prediction is independent of ground water well location, and the prediction is dependent on the pattern and trends in a particular well location dataset. Every well dataset will be unique as the data depends on the dynamics of the well namely the geomorphology of the aquifer, the drainage inside the aquifer and pumping from the aquifer.

Prediction accuracy comparison of time-series prediction model and combined prediction model shows that the combined prediction model with groundwater level in time series and precipitation time series as input predictors is a better predictor model to predict next time horizon ground water level, as the combined information is required for better capturing the pattern and trend in the well dynamics.

Acknowledgement

Groundwater level data was provided by Central Ground Water Board (CGWB, India), accessed from the URL http://cgwb.gov.in/GW-data-access.html. Precipitation data for this study is provided by Indian Meteorological Department (IMD).

References

1 Madhumita Sahoo, Tanmoy Das, Komal Kumari & Anirban Dhar, Space–time forecasting of groundwater level using a hybrid soft computing model, *Hydrol. Sci. J.* 62 (4) (2017), 561-574

2 Madhumita Sahoo, Anirban Dhar, Aman Kasot, & Amlanjyoti Kar, Space-Time Co kriging Approach for Groundwater-Level Prediction with Multi attribute Multi resolution Satellite Data, *J. Hydro. Eng.*, April 2018 DOI: 10.1061/ (ASCE) HE.1943-5584.0001670

3 Schaake, JC, Franz, K, Bradley, A & Buizza,R, The Hydrologic Ensemble Prediction Experiment (HEPEX), *Hydrol. Earth Sys. Sci. Discuss.* , 3(2006), 3321-3332

4 Berliner, L.M., Monte Carlo based ensemble forecasting, Statistics and Computing (2001) 11: 269

5 Patel, P, Predicting the future of drought prediction, IEEE (Institute of Electrical and Electronics Engineers) Spectrum (2012)

6 Vijay kotu & Bala Deshpande, *Predictive Analytics and Data Mining*, (Morgan Kaufmann imprint of Elsevier 225 Wyman street, Waltham, M 02451, USA), 2015, pp. 1-16, pp. 305-327

7 Krzysztofowicz, R., The case of probabilistic forecasting in hydrology, *J. hydro.*, 249(2001), 2-9

8 Andel, S.J. van, Price, R.K., Lobbrecht, A.H., Kruiningen, F. van, & Mureau, R., Ensemble precipitation and water-level forecasts for Anticipatory water-system control, *J. Hydrometeor.*, 9(2008), 776-788

9 Haykin, S., *Neural Networks: A Comprehensive Foundation*, (PEARSON Prentice Hall imprint of Pearson Education), 1999, pp. 23-277

10 Sandhya Samarasinghe, *Neural Networks for applied sciences and engineering*, (Auer Bach Publications Taylor & Francis Group Boca Raton New York), 2007, pp. 11-67

11 Priddy L. Kevin & Paul E. Keller, *Artificial Neural Network: An introduction*, (SPIE Press Book, Washington), 2005, pp. 107-124

12 Elshorbagy, A., Corzo, G., Srinivasulsu, S., & Solomatine, D.P., Experimental investigations of the predictive capabilities of data driven modelling techniques in hydrology –Concepts and methodology. Hydrology Earth System Sciences 14(2010), 1931-1941 and 1943-1961

13 Rani R. & Elangovan K., An emerging intuitionistic fuzzy based groundwater level prediction, *Indian J. Mar. Sci.*, Vol. 46(06), June 2017, pp. 1213-1219

14 Nabil Bessaih, Mohsin Qureshi, Fatima Salem Al-Jabri, Iman Rashid Al-Harmali & Zahra Ali Al Naamani, Groundwater Water Level Prediction in Wadi El Jezzy Catchment Using ANN, Proceedings of the World Congress on Engineering 2014 Vol I, WCE 2014, July 2 - 4, 2014, London, U.K.

15 Han Jiawei, Micheline Kamber & Jian Pei, *Data Mining Concepts and Techniques*, (Morgan Kaufmann imprint of Elsevier), 2012, pp. 1-35, pp. 243-276

16 Mohammed J. Zaki & Wagner Meira Jr., *Data Mining and Analysis: Fundamental concepts and algorithms*, (Cambridge University Press.), 2014, pp. 1-36

17 Solomatine DP & Ostfeld A, Data-Driven modelling: some past experiences and new approaches, *J. Hydroinform.*, 10 (2008), 3-22 DOI: 10.2166/hydro.2008.015

18 Remesan R., & Mathew J., *Hydrological Data Driven Modelling A case study Approach (*Earth Systems Data and Models, vol. 1 Springer, Cham), 2015, pp. 1-17, pp. 71-110