International Journal of Global Research Innovations & Technology (IJGRIT)
e-ISSN : 2583-8717, Impact Factor: (Cosmos:6.972& I2OR: 5.002), Volume 04, No. 01, January-March, 2026, pp 159-164

Study of Year Wise Annual Average Value of Gound Water Level Data for Orissa Alongwith Determination of Trend of Same for Some Future Years by Neural Network(AI/ML)

Sumana Chatterjee*

Ph.D. Scholar (Computer Science), Nirwan University Jaipur, Rajasthan.

*Corresponding Author: sumana.spssaha.chatterjee@gmail.com

Citation: Chatterjee, S. (2026). Study of Year Wise Annual Average Value of Gound Water Level Data for Orissa Alongwith Determination of Trend of Same for Some Future Years by Neural Network(AI/ML). International Journal of Global Research Innovations & Technology, 04(01), 159–164. https://doi.org/10.62823/IJGRIT/4.1.8645

ABSTRACT

This paper is based on data analysis by python programming language code, data executed on google collaborator platform associated with the study of ground water level, data collected from the site of WRIS, "INDIA WATER RESOURCES INFORMATION SYSYEM". From the site searching for ground water level data, the historical data since 1990 to recent past as available there was collected for data analysis. The data collected was of different stations, different districts under the state Orissa as available there and also data with different data type, either manual observation or GPRS data. All these collected data files were then merged into one single file, data from 1990 to 2024. The rows with GPRS observational data were discarded, on account of considering only the manual observational data for analysis, otherwise analysis could generate error having problem of data divergence. Then created one new data column with year wise average value for ground water level data. All the locations of observation as collected were considered as same location for the purpose of having sufficient large data set without any missing year as well as the reason behind that all locations are in the same state Orissa and so could be considered as same data source. The main objective of data analysis was to monitor year wise change of average value of level of ground water for Orissa state. Apart from this study another aim was to determine the trend of yearly average value of ground water level of Orissa for future few years by neural network model. Visualization by bar plot was made to get insights of status of annual average of ground level water for the state Orissa along with future trend for few years.

***Keywords**: WRIS, Ground Water Level, Python, Google Collaborator, Neural Network, Bar Plot.*

Introduction

The necessity of analysis of ground water level is eternal. In world all countries depend on the ground water which is the necessity for daily life of human as public water supply as well as for all animals. Irrigation for agriculture, business and industrialization, rural as well as urban life all depends on the ground water source. The different sources of ground water are rivers, lakes, ponds, springs, snow and glacier melting. Rain water is the vital source of recharging the ground water. Filtering the ground water, drinking water can be provided which is nothing but the synonym of life. So to study the trend of the ground water level for future years is very essential and also a factor for sustainable environment. From the point of realization of the importance of the ground water, to study the trend of ground water level from the historical past to present years along with prediction of ground water level for some future years, was felt as necessary by researchers since past years. The study of analysis with the help of artificial intelligence and machine learning with the help of historical data is one of the most effective methods. Generally ARIMA time series, RNN, CNN, Random forest, Neural Network etc. all these are used for the purpose. In this research study basically LSTM Neural Network method were applied to understand the trend from historical year and trend of predicted ground water level for some future years till 2032. In some cases ARIMA time series was applied for comparison purpose.

Literature Review

The research study is based on analysis with historical data of ground water level data for Orissa district. In one part, the bivariate analysis had been done with the historical data of ground water level to understand year wise change. In other part, for multivariate analysis, two other variables were taken under consideration, those were rainfall data and evapotranspiration data. The data for ground water level was collected from WRIS portal .Data collection process also was done in two methods. In one method the time series data for some random districts(mostly data of Mayurbhanj district) and in another method for all districts under the state of Orissa was downloaded. After collection for each method, sorted the data since historical years, ultimately merged into a single csv file for analysis .For the case of multivariate analysis, the rainfall data was collected from the data collection platform IMD Pune. The historical rainfall data was collected for the state Orissa with all districts for all months of years since 1990 to 2025, as available there to download, then these csv files merged into a single csv file to get yearly average rainfall data for Orissa district. The other related variable taken into consideration was evapotranspiration data. This data set was obtained by Nasa power API link by python code giving coordinates of all districts of Orissa and ultimately by feature engineering merged into a single csv file, ready to be downloaded from content folder with yearly average evapotranspiration data for Orissa district. These three final csv files obtained from three different sources as described earlier, ultimately joined together on the basis of year starting from 1990 to 2024, the available latest year common for each variable, for the purpose of data analysis to get insights of trend of yearly average ground water level for Orissa along with prediction for near future. The research papers associated with study and prediction of ground water level by artificial intelligence and machine learning as available on web are based on analysis with Random forest, for large data set, ARIMA time series for prediction of future data, RNN (Recurrent Neural Network),CNN (Convolution Neural Network), SVM (Support vector machine),GAN (Generative adversarial Network), PCA (Principal Component Analysis), LSTM (Long short term memory) etc. depending upon the way as well as data set for analysis. RNN is used to handle sequential data, CNN is used basically to handle image dataset, GAN is used for generation of new data similar to real data, PCA is used when dimensionality reduction is applicable to reduce the number of features. SVM is a supervised machine learning technique, works for classification of different groups separated by hyperplane. Between RNN and LSTM, LSTM is more efficient model as though both model can be applied for prediction of future data from application of historical data but RNN has some limitations like the factor of limitation with vanishing gradients and problem with long term dependencies. So this research study was done by LSTM mainly. LSTM (Long short term memory) is a recurrent neural network method, uses gate, understands the important data to remember and useless data to forget, remembers long term pattern, reduces vanishing gradient problem, tackles time series data well to give accurate output by repeated learning of machine. For this study with csv data file, to study historical data for ground water level as well as prediction for future years of the same, LSTM was one of the best fit models.

Research Gap

The source of data for ground water level was India WRIS portal (India Water Resources Information System). From this portal, data was downloaded from the menu ‘Time series data’. After clicking on the ‘Time series data’ menu, a search panel was opened. In this search panel several fields were to select and enter. Firstly random collection of files for Orissa ground water level data had been collected for some districts (mostly location with ‘Mayurbhanj’ districts) and then data for all districts were collected for analysis, ‘Tehsil’ and ‘Block’ remaining same as the name of district chosen each time. In both case, manual data was collected for analysis. Variation in sample collection as well as variation with inclusion of depending variables were applied to compare results and obtaining the ensured valuable insights. Other than ARIMA time series, mainly LSTM neural network method was applied for better accuracy. Different types of data set as well as different set of variables were used to be ensure with verified output in diverse environment.

Research Questions / Hypothesis

Due to continuous usage of ground water, the ground water source may face scarcity in future. So to understand the trend of change of ground water level for all states under India is very necessary for the future India. As for example in this research study Orissa state had been taken under consideration. By similar research study with same type of data sets, any district may be subjected to analysis to understand the trend of change of ground water from historical year to present years along with prediction of ground water level for near future years. The objective was to find out whether any

significant change will occur or not. Whether significant decrease of level of ground water level will be there for Orissa or not.

Methods

This study was related with analysis by machine learning technique and specifically neural network model. The whole process of analysis was done in multiple manners so that the result and accuracy for each case can be compared with other patterns. Results obtained from analysis with diverse data and technique could give more dependable insights. In four cases this analysis had been done. The variation of pattern considered was related to the number of variables taken for data analysis, either bivariate or multivariate and also related to the volume of data set considered for analysis. Volume of data was dependent on either some arbitrarily chosen districts or all districts of Orissa. First and second case of analysis were done with some districts while third and fourth case were done with all districts of Orissa. Analysis for the first case, bivariate analysis had been done to get insights of year wise ground water level data, as per years 1990 to 2024, as available in the site WRIS. Along with visualization of year wise change of ground water level, prediction for ground water level till 2032 was visualized also by LSTM neural network. Ground water level data for various locations with different districts, chosen randomly for the state Orissa were downloaded .Selecting mandatory choice option in the search panel of data downloading, the files of ground water level data, for some districts of Orissa, with districts chosen randomly, along with selection of same tehsil and same block having same name as per chosen district, if those available there or if not, selecting any other available tehsil and block, data files were downloaded, under the fixed agency CGWB. In this first case of analysis, basically maximum data files of 'Mayurbhanj' district were collected. In this connection, this is also to mention hereby that two types of data were available there to download, depending on nature of acquisition, one form was manual and the other form was telemetry. In this study the collection of manual acquisition data had been used. In this first case, data files were downloaded randomly from the website India WRIS portal. Random download technique was applied to download data files with data types of both of manual and telemetry from various locations, mostly from 'Mayurbhanj' district. Later the rows of files with manual acquisition type data had been considered only for analysis .Each excel file downloaded had two sheets from which only sheet with ground water level data was considered for analysis. As mentioned earlier, the data obtained from randomly downloaded collection was basically from the district 'Mayurbhanj'. Few were from other districts of Orissa. LSTM Neural network model was built to understand trend of yearly average ground water level data since historical years 1990 with the predicted value of the same till 2032. Bar plot visualization was used for understanding. In the second case of analysis, with the same data set as done for case 1, analysis was done similarly by LSTM but the difference between the first case and the second case was regarding the number of variables. In the second case, the number of variables was more than one, analysis was multivariate, the other two influencing variables whose year wise change depending on the year wise change of ground water level were 'year wise average rainfall data' and 'year wise average evapotranspiration data'. Monthly rainfall data and evapotranspiration data for all districts of Orissa was collected from DSP IMD PUNE and by python API link to fetch from NASA POWER respectively. For third case of analysis, change of data set was done by the way of collection of new set of data, collected from the same portal WRIS, here the data for ground water level was collected for all districts of Orissa instead of random collection with some few districts as done for case 1 and case 2. Next as similar as variable taken for first case, year wise average ground water value, bivariate analysis was done to get the insights of trend of year wise average value of ground water level data for Orissa from historical year along with trend of the same for predicted years till 2032. In fourth case also, data was collected from WRIS portal for all districts of Orissa. Similar as all other cases of analysis only the sheet with ground level water data value was considered for multivariate analysis. Here other than the variable 'ground water level data' from WRIS portal, like case 2, other two variables were also taken under consideration. These were yearly average rainfall data and yearly average evapotranspiration data for all districts of Orissa. For all cases 1 to 4, separately all the excel files downloaded were subjected to feature engineering to form new combined csv files consisting of individual variable and ultimately for each case, formation of one single merged csv file was made, compatible for python analysis, to get insights of year wise trend of ground water level till available period as well as predicted value of the same till few future years, till 2032 by LSTM neural network model. The whole process of data sorting, filtering, data extracting, merging, discarding of useless sheets, all were done by python languages, necessary for the subject of study. Python language code, executed on google collaborator platform was used there as adequate to perform all the necessary work related with finding all the insights. For neural network method, generally used LSTM method with 'adam' optimizer, model compilation with 100 epochs,

activation function used was 'Relu'. For case 2, ARIMA time series was applied also for counter verification and almost same output as with LSTM model was obtained. Comparison between accuracy of all cases by LSTM neural network model, was done also for understanding of RMSE,MAE value. The figures of findings for all cases have been pasted herewith at the particular section with screenshots.

Significance of the Study

It is really beyond expression to express in language the importance of water in our every day life. Not only for us, water is nothing but life for all animals. There are several sources of water among which ground water is one of the vital sources of the Earth. Due to continuous usage, for the effect of global temperature rising continuously, on account of the adverse effect of climate change issue, there may be a factor of scarcity of the ground water. So from this point of view, to study the status of ground water level is very essential to restore sustainable environment for 'Future India'.

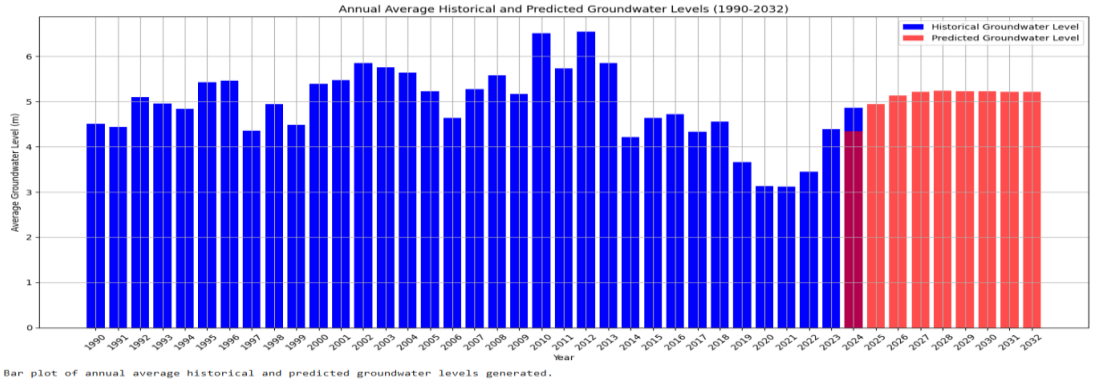
Timeline

The time taken for collection of ground water level data, for districts under Orissa state, one by one from WRIS portal, firstly files of data were collected for some districts, option of downloading chosen randomly, then for all districts under Orissa, data files were downloaded one by one for each set of selection associated with each district and the process was repeated till the end of selection of every district under Orissa. After completion of collection of all data files, downloaded from WRIS portal, merged all these excel data files into one single csv file consisting of year wise ground water level data. The whole process of merging, sorting, discarding and necessary feature engineering took sufficiently enough time associated with creation of suitable python code necessary for compatibility to perform the whole process of aggregation to get year wise average ground water level data for Orissa. Similar job was done for rainfall data of Orissa district collected from the site of the data supply platform, IMD Pune to collect all files of monthly rainfall data for all districts under Orissa state to merge into one single csv file with year wise average rainfall data .For collection of the evapotranspiration data for all districts under Orissa state, code with python API was applied, to collect data from NASA POWER, providing the adequate latitude and longitude of all districts for fetching the data one by one from that interface, followed by necessary algorithm to aggregate all data files into one single csv file with year wise evapotranspiration data. After preparation of each merged individual three data files, obtained from three different interfaces, consisting of year wise average value of ground water level, year wise average value of rainfall, year wise average value of evapotranspiration each for Orissa, ultimately these three csv files joined to form a new file to build model for multivariate analysis. For the case where bivariate analysis had been performed, data sorting, filtering, feature engineering, merging of files needed for only the particular variable there which was year wise ground water level data. However for the case of bivariate or multivariate, for random collection or for collection of data for all districts, for collection from only WRIS or also from DSP IMD PUNE and NASA POWER, sufficiently enough time was necessary to follow steps of feature engineering, data sorting, filtering, aggregation of data files for analysis to visualize the historical as well as predicted trend to gain insights of the trend of ground water level. The model building, model fitting, validation all needed for compatibility and successful execution, for model building with neural network, by machine learning, followed by processing with several epochs under suitable learning, sometimes supported by suitable activation function like ' Relu', optimizer by 'Adam', for all types of fine tuning to obtain better accuracy, sufficient time and effort was to be given.

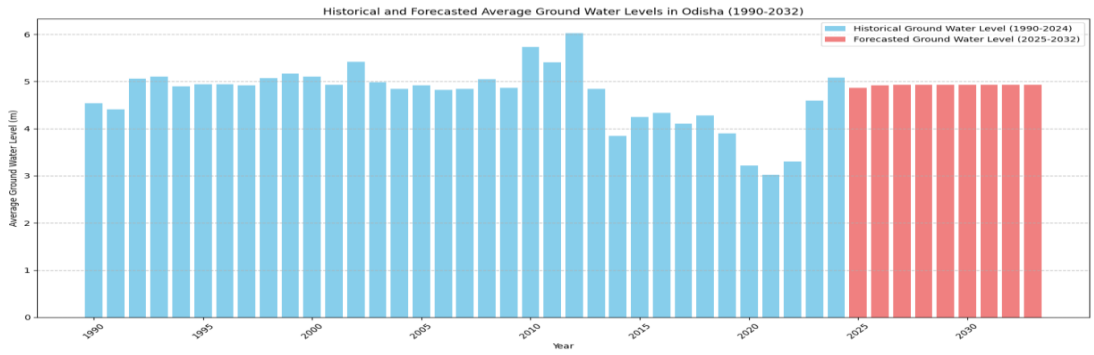
Conclusion and Future Work

From the research study as obtained from analysis, the insights can be noted down as mentioned herewith .The predicted value for future years till 2032 was obtained with comparatively higher value for analysis with case 1 and case 2, value around 5 meter approximately for ground water level, while the same for analysis with case 3 and case 4, value around 3.5 meter to 4.0 meter approximately. The RMSE and MAE value for case 4 is the lowest, so the result and analysis of this case is most acceptable. The case 4 was analysis with all districts of Orissa and with other depending variables rainfall and evapotranspiration. In future, this type of analysis will be done with other region also with same or different research study.

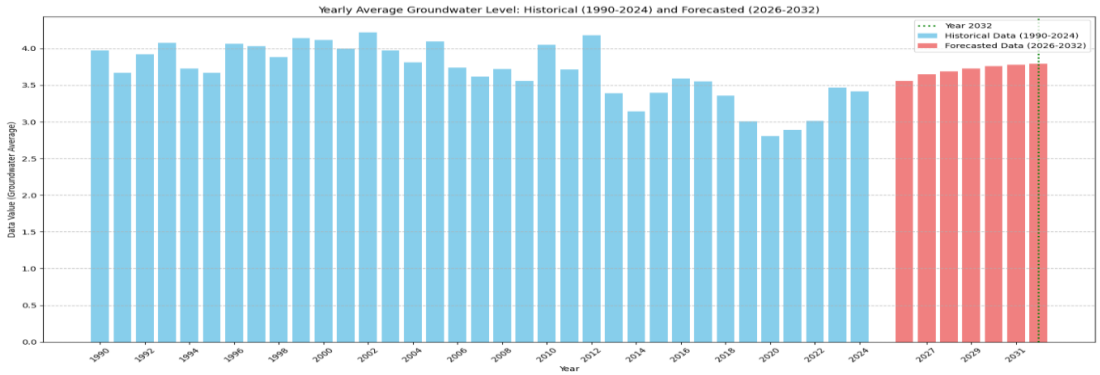
Some Screenshots of Analysis (Label for Each Figure Mentioned Below)



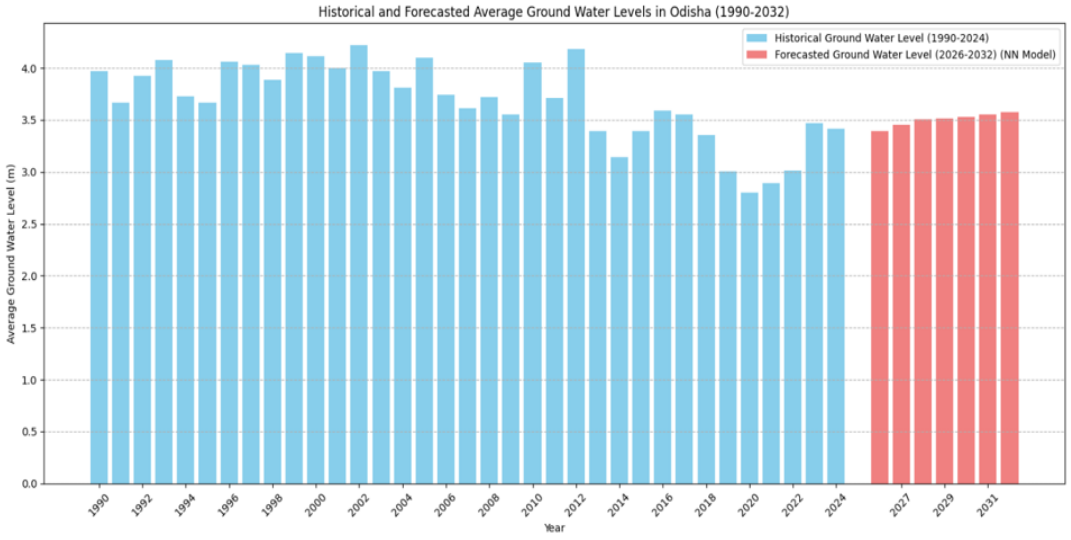
Case 1: Year Wise Ground Water Level Lstm Bivariate Analysis With Some Districts, Mostly Mayurbhanj (Rmse 0.6770, Mae 0.5738)



Case 2: Year Wise Ground Water Level, Multivariate Analysis, Lstm, Some Districts, Mostly Mayurbhanj (Rmse 0.87, Mae 0.78)



Case 3: Year Wise Average Ground Water Level, Lstm Neural Network Model, Bivariate Analysis, Data With All Districts (Rmse 0.4484, Mae 0.3830)



Case 4: Year, Ground Water,Rainfall,Evapotranspiration, Lstm Neural Network Model, Multivariate, Data With All Districts, Rmse 0.2687, Mae 0.2207

References

1. H.-Y. C., 2023. Groundwater Level Prediction with Deep Learning Methods. *MDPI,* 15(3118), pp. 1-21.
2. Eslam A. Hussein 1, *., . C. T. 2., . M. G. 1., . A. B. 1. M. V., 2020. Groundwater Prediction Using Machine-Learning Tools. *ResearchGate,* Issue November 2020, pp. 1-16.
3. Gabriela May-Lagunes a, b., 2023. Forecasting groundwater levels using machine learning methods: The case of California's Central Valley. *Journal of Hydrology X,* 21(DECEMBER 2023), pp. 1-14.
4. Hai Tao a, b., 2022. Groundwater level prediction using machine learning models:. *ELSEIVER,* 489(489(2022)), pp. 271-308.
5. S. Nagavli, S. D. S. P. C. P. S., 2024. Prediction of Ground Water Level using MachineLearning. *International Journal for Modern Trends in Science and Technology,* 10(9)(OCTOBOER), pp. 1-7.