Leveraging Hadoop Framework for Predicting Underground Water Levels

Raguvaran.S
Assistant Professor,
Department of Computer Science and Engineering
KPRIET, Coimbatore, India-641407

Joyce Beryl Princess.P
Department of Computer Science and Engineering
KPRIET
Coimbatore, India-641407

***Abstract*—Life, both animal and plant, is impossible without water. Water scarcity is a vital scenario to be considered in the current world. In India with the increasing population and growth of industries the water availability is becoming inadequate to meet up the requirements of people. So predicting the water level is significant to know about water scarce in the impending years. Here we use Big Data Analytics in the study of developing forecasting models for predicting underground water levels. In this prediction first the data of present underground water level is collected as an ingest data operation, which is next moved into the big data storage. Then dynamic linear modelling algorithm is used to observe the pattern of historical data and predict the future underground water levels by applying data-driven analytics and data mining concepts.**

Keywords: **Analytics, Big data, Prediction, Regression, Underground water.**

I. INTRODUCTION

At present India's population density records 446 per km^2 and it has the second largest population in the world [5]. Yearly per capita accessibility of renewable freshwater has contracted distressingly. Around 5,277 cubic meters in 1955 it plunged to less than 1,820 cubic meters in 2001 shows the drastic water depletion [6]. As India majorly bank on underground water for sustaining the needs of people. Presently, underground water is now mainstay of the Indian economy and subsequently, a life-threatening factor in India's water future.

Due to deteriorating water levels in quite a lot of parts of the country we are close to grueling of our ground water reserves, which are dwindling faster than they can be replenished. So the forecast of water level is an important issue to plan, strategize and build the standby structures and for the management of underground water for water supply purposes. Besides there are many environmental aspects distressing underground water level it is essential to cultivate models for simulation of the extreme or unusual level differences in order to bring about forthcoming underground water level fluctuations.

II. UNDERGROUND WATER LEVEL DATA

Ground water level information's are time-based and progressive in nature. It is mainly measured by rainwater patterns. A huge share of India's agronomic rests on ground water obtainability. The existing water level data for 30 states and union territories dated from 1994 to 2015 is received from Central Ground Water Board have been used for this analysis.

Additionally, water level data have been clustered conferring to four seasons: post-monsoon rabi (January to March), pre monsoon (April to June), Monsoon (July to September) and post-monsoon kharif (October to December) [6].

III. EMPLOYED MECHANISM

Concerning the consumption of population , underground water level prediction is done through manual process and the former altitudes of water obtainability for predicting the forth coming water level patterns is not considered .It aided in better processing and confronts the deficit of the water to a bigger extent.

Prediction methodologies can be categorized into kinds -

1. Perception in Qualitative (enduring prediction) - Here mathematical model is of no use, because the data accessible is not concerned to be subsidizing to the impending values.
2. Perception in Quantitative - Here past data is accessible. The analysis is centered on the past data having the precise variable of time series and other time series interconnection. Cause-and-effect associations are inspected of one type of variable with other related variables

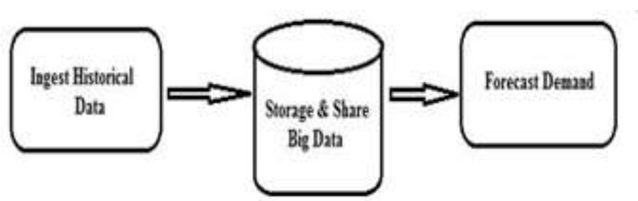
IV. PERCEPTION IN TIME SERIES

A time varying single variable whose impending values in some form to its historical values are certainly related [1].

To answer this problem a technical elucidation is proposed using dynamic linear modelling algorithm centered over time-series. This study can be certainly used with fewer charges and has the gain of further topping up by comprising abundant underground water data without distressing the performance. To conclude the claim of water prediction is done which can be used to adopt plan by the verdict makers in the underground water resource divisions.

FIGURE I

STAGES IN PROPOSED SYSTEM

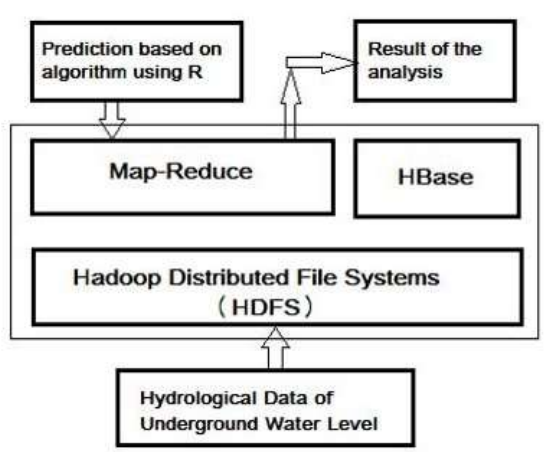


In the proposed solution data that is moved to various phases -

1. Data Ingestion Phase - The past data of underground water level over the years is collected from the Central Ground Water Board and before being shifted to the big data storage data is moved to the transitory storage.
2. Big Data Retention Phase - From the initial phase the data is warehoused into the database. It can be utilized in different phases and forms the main backup of the system,
3. Demand Prediction Phase - In this phase the data is sent to R for statistical future water level analysis.

The proposed system architectural representation is shown as below

FIGURE II ARCHITECTURE DIAGRAM



V. DYNAMIC LINEAR MODELLING

Dynamic Linear Models are a very general class of non-stationary time series models [7]. The approach is more influential to analyze and forecast the time-series modelling. In order to control certain progressive control systems the unknown quantities need to be estimated from noisy measurements. The position of the system transforms from one state to another according to a known transition equation, possibly including random commotion and interruption effects. The observations are derived from the state values by an observation equation, also possibly including random disturbances and interruption effects.

The dynamic linear models perform two successive operations filtering and smoothing. Assessing the state values at time t, using only observations up to (and considering) t-1 is the filtering operation. Contradictory to filtering, assessing the state values using the whole set of observations is smoothing [8]. Special cases of the DLM are the traditional time series models such as ARIMA. Prediction performance analysis for dynamic linear modelling varies: for long-term forecast DLM works very well, Holt winters works well for for short-term forecast. Besides non-stationary, time-varying parameters, multivariate time series, data from multiple sources, irregular temporal observations and missing data among other things can be integrated by dynamic linear model [9].

VI. DYNAMIC LINEAR MODEL PREDICTING UNDERGROUND WATER LEVELS FOR FUTURE DEMAND

Dynamic linear model generates data that progress over time. The state of this process at time t is expressed as a vector of parameters, θ_t.The time-series corresponding value is - the observation equation R_t - parameters of the linear function is specified by the regression vector F_t and a random noise term ν_t .The latter is assumed to be simply normally distributed noise with known variance.

The evolution process given by the progression matrix J_t, the previous period's state vector is pre multiplied. ω_t is summed up to the result to produce the state vector in the current period which is an evolution noise vector.

A normally distributed random variable with the given mean and variance indicated as $x \sim N[\mu,\sigma^2]$,and G'_t is the transpose of vector G_t :

The primary equation is given by $\theta_0 \sim M[n_0 , k_0]$,

The system equation is $\theta_t = J_t \theta_t + \omega_t$,

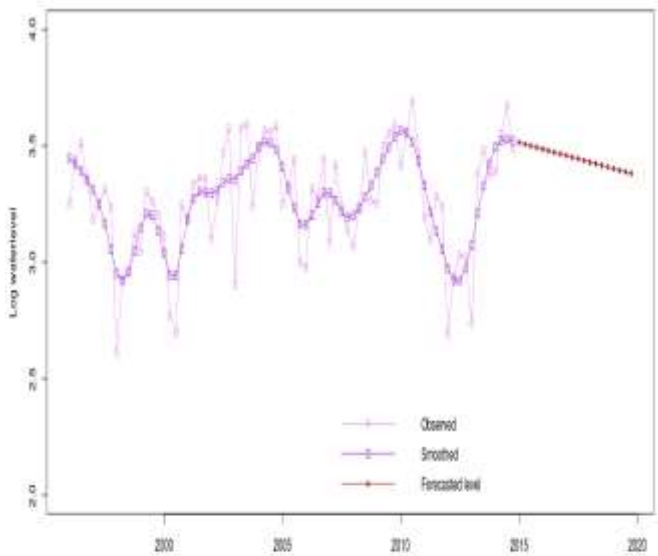
The observation equation $R_t = G'_t \theta_t + v_t$.

R_t is the observation equation obtained from the regression vector, for which the data is retrieved from the underground data. Θ_t provides the vector of parameters at time t for the seasonal variations. J_t is progression vector, which evolves over the time-series.

VII RESULT AND DISSCUSION

The collected underground water level data is processed. The data's over the years for post monsoon, premonsoon, monsoon, such period's ground level data and utilization is considered and underground water demand is predicted for a period of time.The model's capability confronts that it is an appropriate process for sudden fluctuation in the water level patterns.

FIGURE III GRAPH FOR LEVEL PREDICTION



The figure III shows the future logged underground water level graph from 2016 to 2019 for Coimbatore.

TABLE I TABLE FOR PREDICTED UNDER GROUND WATER LEVEL

YEAR	**PREDICTED UNDER GROUND WATER LEVEL**			
	Post-monsoon rabi	*Pre monsoon*	*Monsoon*	*Post-monsoon kharif*
2016	32.65314	32.42591	32.26662	31.97619
2017	31.75367	31.53278	31.31326	31.09536
2018	30.87897	30.66408	30.45070	30.23879
2019	30.02836	29.81940	29.61189	29.40582

TABLE I provides the predicted underground water level data for the years 2016 to 2019, it can also be predicted for further years, by means of predicted values the water demand can be managed well.

The prediction performance of dynamic linear modelling(DLM) is justified as follows, chiefly DLM delivers noble forecast entirely over long term predictions. Dynamic linear modelling does not provide noble forecasting for short-term prediction. On second note dynamic linear modelling in long-term prediction beats seasonal naïve model in all instances, in which it becomes only possible selection.Besides, NAIVE produces prediction based on recurrent dependences alone.

For handling huge amount of data Hadoop-based environment is utilized and it handles effortlessly. The data is stored in hadoop distributed file system and map reduced using map reducer, reduced data is stored in HBase and processed using R-tool and Dynamic Linear Modelling forecasting the underground water levels for water demand.

VIII. CONCLUSION

In this paper, for predicting underground water levels is a controllable and scalable scheme is discussed. Open source tools are exerted for the prediction of underground water levels and are employed in application that the execution charge certainly diminished and simultaneously answerprovided and applied in the current technologies.

Proposed DLM gives more accurate interpretation on the changeable underground water level data. Also the forecast intervals from our proposed DLM were added realistic than any other models. Out of sample inaccuracies specify that no model accomplished preeminent in terms of predicting overthe short and long prospect concurrently.

REFERENCES

[1]. Prashant Shrivastava, S. Pandiaraj and Dr. J. Jagadeesan "Big Data Analytics In Forecasting Lakes Levels", CSE Department, SRM University, Chennai, India, IJAIEM, March 2014.
[2]. Leixiao Li,Zhiqiang Ma, Limin Liu and Yuhong Fan College of Information Engineering, Inner Mongolia University of Technology Huhhot, China."Hadoop-based ARIMA Algorithm and its Application in Weather Forecast". International Journal of Database Theory and Application Vol.6, No.5 (2013).
[3]. Giovanni Petris, Sonia Petrone, Patrizia Campagnoli, "Dynamic Linear Models with R", 2014.
[4]. David Makowski, Lucie Michel, AgroParisTech, "Use of dynamic linear model for predicting crop yield trends in foresight studies on food Security", France, 2013.
[5]. http://www.worldometers.info/world-population/india-population/
[6]. http://www.india-wris.nrsc.gov.in/
[7]. http://www.math.unm.edu/~ghuerta/tseries/dlmch2.pdf
[8]. Cyrille Szymanski , "Package modeler", February 19, 2015.
[9]. Roberto Rivera, "A Dynamic Linear Model to Forecast Hotel Registrations in Puerto Rico Using Google Trends Data",December 2015.