Water Availability Forecasting Using Univariate and Multivariate Prophet Time Series Model for ACEA (European Automobile Manufacturers Association)

Prismahardi Aji Riyantoko [1]*, Tresna Maulana Fahrudin[2], Kartika Maulida Hindrayani[3], Amri Muhaimin[4], and Trimono[5]

[1,2,3,4,5] Department of Data Science, UPN "Veteran" Jawa Timur; prismahardi.aji.ds@upnjatim.ac.id, tresna.maulana.ds@upnjatim.ac.id, kartika.maulida.ds@upnjatim.ac.id, amri.muhaimin.stat@upnjatim.ac.id, and trimono.stat@upnjatim.ac.id
* Correspondence: prismahardi.aji.ds@upnjatim.ac.id;

Abstract: Time series is one of method to forecasting the data. The ACEA company has competition with opened the data in the Water Availability and uses the data to forecast. The dataset namely, Aquifers-Petrignano in Italy in water resources field has five parameters *e.g.* rainfall, temperature, depth to groundwater, drainage volume, and river hydrometry. In our research will be forecast the depth to groundwater data using univariate and multivariate approach of time series using Prophet Method. Prophet method is one of library which develop by Facebook team. We also use the other approach to making the data clean, or the data ready to forecast. We use handle missing data, transforming, differencing, decomposition time series, determine lag, stationary approach, and Augmented Dickey-Fuller (ADF). The all approach will be uses to make sure that the data not appearing the problem while we tried to forecast. In the other describe, we already get the results using univariate and multivariate Prophet method. The multivariate approach has presented the value of MAE 0.82 and RMSE 0.99, it's better than while we forecast using univariate Prophet.
Keywords: Time Series, Univariate, Multivariate, Prophet, Water Availability, Forecasting

1. Introduction

The oceans cover nearly three-quarters of the earth's surface, where weather and climate are affected by evaporation in seawater [1]. Water is the most valuable global commodity with its endless uses like drinking for humans, animals, and plants, for the others to industrial production, irrigation, and the production of fish, waterfowl, and shellfish [2]. The freshwater management that provides many benefits like flood control, transportation, waste processing, power plant, and habitat for aquatic life [3]. In the others, water also have some benefits like irrigation, and hydroelectric power plant, are achieved only by management and path of water flow from dams and water diversions [4]. Degradation of water resources has become a social concern by an environmental researcher. Therefore, researchers have investigated the wobbly distribution of water resources [5-7]. The well-being of humankind is closely linked with water resource management. Global climate and anthropogenic activities, especially over-exploitation of groundwater resources [8-9]. In the most places have increasing temperature, hence the groundwater recharge will decrease, however, excessive irrigation flow can make groundwater recharge increase in some areas [10]. Referring to the ecological value of groundwater, the groundwater-dependent ecosystems can be

commonly classified into three types: (a) aquifers and cave ecosystems; (b) springs, wetlands, rivers, estuaries; and (c) terrestrial vegetation ecosystems [11]. The water systems have many data resources to analysis and forecast for the future, like depth to groundwater, flow rate, hydrometry, etc.

In the Europe already exist ACEA company which focus on the water availability. They open the competition to forecast the opened data. In the other side, water availability in the recent years to be an highest issue to solve. In addition, the ACEA Group is one of the leading Italian multiutility operators to manages water resources, electricity networks, and environmental services. According to the competition they want to focus only on the water sector to help ACEA group preserve precious waterbodies. Each dataset represents a different kind of waterbodies, and they divided into four different type of waterbodies including water springs, lakes, rivers, and aquifers. In our research will be focus on the aquifers in depth to groundwater forecast using time series. Time series is one of method to forecast to get the best model with compare actual and predicted data. We will try to forecast using one of time series model, Prophet Model. The Prophet model already develop by Facebook team to forecast. In the other side, we also make sure the data is ready to proceed. To make sure this condition, we will use data evaluating using handle missing data, transforming, differencing, decomposition of time series approach, determine lag, and Augmented Dickey-Fuller (ADF).

2. Related Works

According to the related research about waterbodies prediction and forecasting method, it's will be breakdown in this subsection. In the research [12] already presented long and short-term prediction and analysis using techniques like detrended fluctuation analysis, auto-regressive models, and persistence algorithm. The modeled of prediction problem come up from analysis of dammed water level in the hydropower reservoir which processing using machine learning regression approach. They used many kinds of methods including Support Vector Regression or Gaussian techniques process. Based on the problem and method, they collecting and using the data from hydropower located in Galicia, Spain with variables from upstream measuring stations. For the long-term analysis showed that high accuracy and short-term analysis which solved by using Machine Learning Regression showed that the algorithms are able to obtain extremely accurate. Wetlands become research focuses by [13] which show up a vital role in the ecologic and hydrologic communities. But they have an issue in the collecting data of wetland water level because of a few studies conducted. The research already developed using various machine learning models including Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), and Artificial Neural Network (ANN). The data was collected using daily water level since 2009 until 2015 from UPO areas in South Korea, this data is used as dependent variables, but for the independent variables, they used meteorological and upstream water level gauge data. Furthermore, the predictive results show that Random Forest.

In the other research in the time series forecasting using Prophet approach by [15], they using the Prophet forecasting model along 730 days around two years. They predict in many categories in e-commers like sales of professional outdoor supplies, organizing outdoor educational courses, seminars, and competitions. They uses implementation of decomposition time series, forecast, and accuracy of measurement to be one model and get greatly simplifies. This project to be easier to

understand by analyst and manager. The measure has showed the accurate result, hence they get more good results in the accuracy coverage and RMSE. In the Ref [16], they using Prophet Algorithm to forecasting the sales data based on real-world data. The train and test 581 items in product of portfolio. They use monthly and quarterly sales to forecast, the get great potential to classifying the product into several categories. They get MAPE less than 15% in quarterly, it's mean they understanding the best sell items of the aforementioned retail company, and they can share more 80% annual financial condition. It's statement more than satisfactory in real-world scenario of sales forecasting.

3. Experiment and Analysis

3.1. Data Collection

In this subsection will be describes about the data resources. One of the big companies in the Italy, it is The ACEA Group creates competition analytics to processing data of ACEA Smart Water Analytics dataset. We know that ACEA Group is one of the leading Italian multiutility operators to manages water resources, electricity networks, and environmental services. According to the competition they want to focus only on the water sector to help ACEA group preserve precious waterbodies. Each dataset represents a different kind of waterbodies, and they divided into four different type of waterbodies including water springs, lakes, rivers, and aquifers.

Table 1 ACEA Dataset Resource

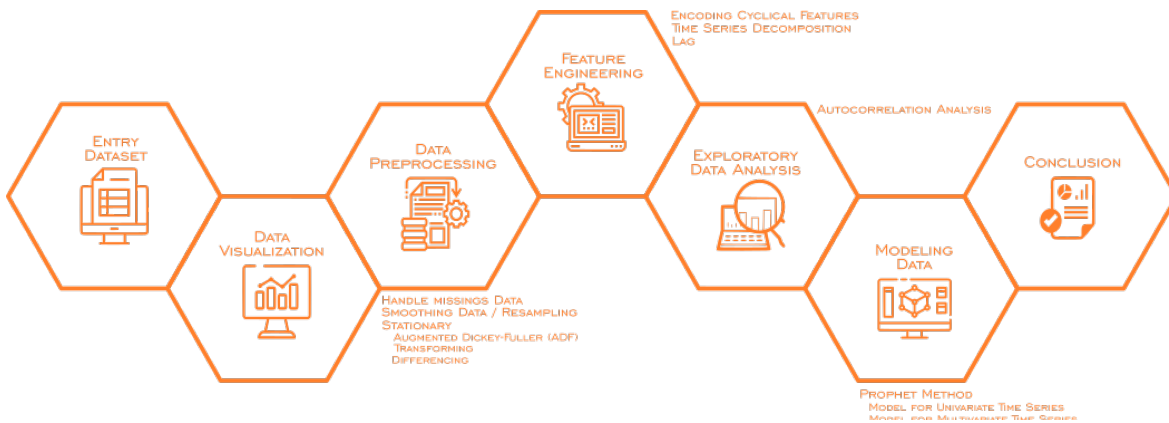| Waterbodies Types	Submission Evaluation Criteria	
	Input	*Feature for Forecast*
Aquifers	Auser Doganella Petrignano Luco	Depth to Groundwater
Water Spring	Amiata Madonna di Canneto Lupa	Flow Rate
Lake	Bilancino	Hydrometry (Lake Level), Flow Rate
River	Arno	Hydrometry (River Level)

Figure 1 Proposed System Design of Time Series using Prophet Method

3.2. Prophet Method

Prophet is one of kind for forecasting model which developed from Facebook's Core Data Science and conducted research by Ref [14]. Prophet based open sources library for decomposable models and provides good accuracy using simple parameters. Prophet method have three components including trend, seasonality, and holidays. The following is given the equation of the Prophet method

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{1}$$

Where the parameter will be provides below,

For $g(t)$, is notation for trend function and model non-periodic changes in the value of the time series, $s(t)$ is periodic changes (*e.g.,* daily, weekly, monthly, yearly, and seasonality), $h(t)$ is the effects of holidays. The error is represented by ϵ_t which are not accommodated by the model. The Prophet formulation provides a number of practical advantages for

- Flexibility: We can easily to make analysist with different assumptions about the trends and accommodate the seasonality with multiple periods.

- Compared with Autoregressive Moving Average (ARIMA) models, to find the measurements must not need to be regularly spaced, and interpolate missing parameters *e.g.* from removing outliers.

- The data fitting very fast for interactively explore many model specifications, for example in a Shiny Application [15].

- To get the model from forecasting has easily can interpretable parameters that can changed by the analyst to improving the assumptions while the forecast process. Therefore, the analyst have many typical to do have experience with regression and easier to extend the model to including the new components.

3.3. Data Visualization

In this subsection, we will describe and discuss about the features of data to visualization. If we refer into Table 1 in Waterbodies type Aquifers-Petrignano, the dataset has five parameters

specifically rainfall, depth to groundwater, temperature, drainage volume, and river hydrometry. They have many different viz when we visualize into graphic and chart.
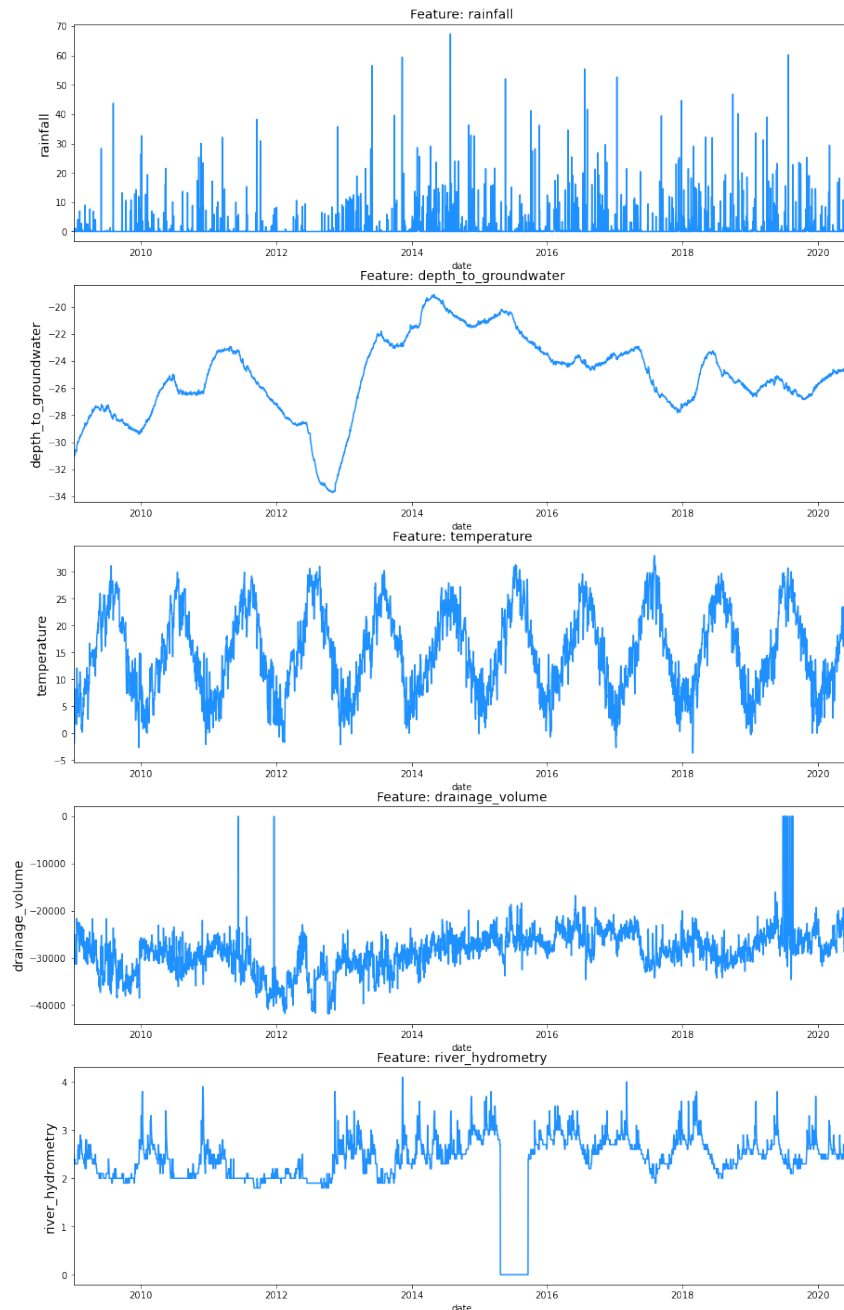

Figure 2 The Features of Dataset from Aquifers-Petrignano to forecast the depth to groundwater parameter

Refer into the Figure 2, we can describe many types of parameters, but we only focus on the depth to groundwater parameter to forecasting them. All parameters have indicated if the data collecting between around 2010 until 2020, it means the data has collected about ten years ago. The rainfall chart almost in every year has the same path, but we can show the highest value in around latest of 2014 that the value of data is 69. The depth to groundwater has negative value, why this condition happened? Because the keywords in the depth, begin from surface of water (up) into depth of water (down). The depth to groundwater has volatile or ups and downs condition every year, it

has occurred because of the sediment of soil always change in the river in Petrignano. The deepest value it happened in around of early 2013, almost into -34, but the shallowest data it occurred in around middle 2014 year, almost -19. The deep and swallow condition mostly caused flooding, therefore to prevent it, we tried to forecasting the data using our best method. The interested data it happened in the temperature chart, they present oscillation condition, like sinus graph. This condition be affected by climate change every year, because we know if in Europe have four weathers, especially in the Italy-Petrignano. The anomaly condition it happened in the drainage volume and river hydrometry, but it's normally. The un-normally pattern present in the river hydrometry, the graph show 0 value almost in every month at 2015. All the condition of parameters has interplay between one and each other, hence we only focus on depth of groundwater, in the other hand we still look in the other parameters.

3.4. Handle Missing Data, Stationary, Augmented Dickey Fuller (ADF), Transforming, and Differencing

In this subsection we will describe, present, and discuss many steps to explore, view, and correcting the missing data. Based on the data, as we can see all the data categories has some null (empty) values, hence we handle the missing data with clean them and replacing by NaN value and filling them afterwards.

Table 2 Handle Missing Data

Categories	Submission Evaluation Criteria	
	Missing Data	*Replacing Data*
Rainfall	0	0 (Nothing)
Depth to Groundwater	27	27 NaN
Temperature	0	0
Drainage Volume	1	1 NaN
River Hydrometry	0	0

Refer in to Table 2, we present 27 missing data in the depth to groundwater and 1 missing data in the drainage volume. All the missing data replacing by filling NaN value. How to handle the missing data? We will present and describe into four options, they are filling NaN, with zero or outlier value, mean value, last value or approach value, then linearly interpolated value. They have many different ways to filling NaN value, almost different for each option. In this case, we handle missing data using interpolate, because the interpolated values require knowledge with approached the neighboring value.

Stationarity uses for checking the constant mean, variance, and covariance the data with present visually, basic statistics, and statistical test. The mean, variance, and covariance be affected by not time-dependent. The stationary mean and covariance will be present straight graph value in *y-axis* and established oscillation pattern without out from the *y-axis* value. In the other condition of non-stationary mean and stationary variance will be present concave down graph. Then we present stationary mean and non-stationary variance form large oscillations into small ones towards a single point.
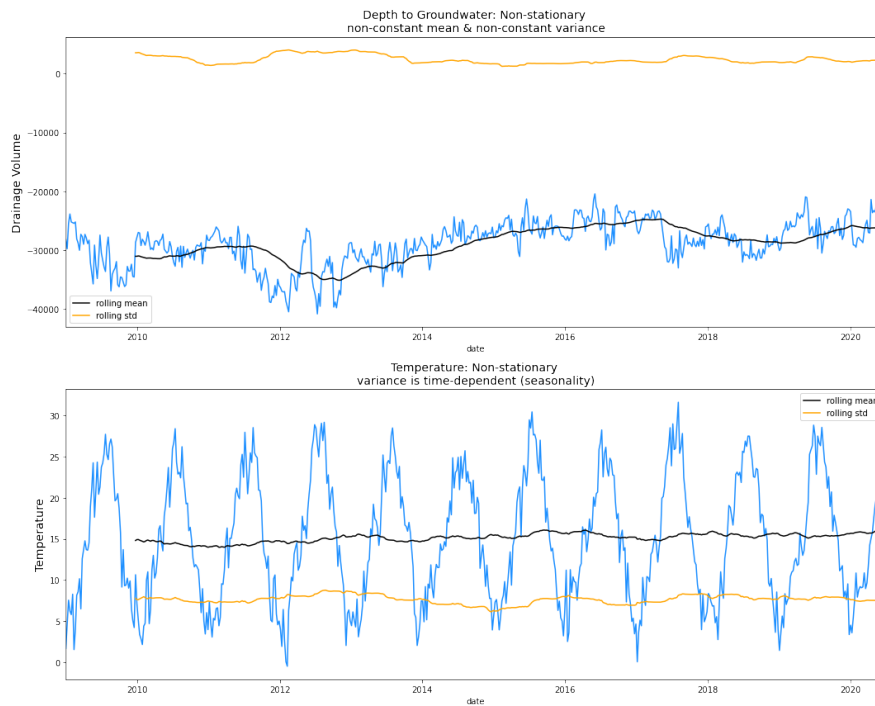
Figure 3 The Stationarity (1) Depth of Groundwater present Non-Stationary (2) Temperature present Non-Stationary

Refer into the Figure 3, we present visual checking of the depth to groundwater and temperature has non-stationary variance is time-dependent in seasonality. In addition, the two features have not constant mean and standard deviation. Therefore, we will describe the other method to change non-stationary to stationary condition using unit root test. The unit root test presence the time series for non-stationary. Once of method to solve this case using statistical test of unit root test called Augmented Dickey-Fuller (ADF). We can approach using Null (h0) and Alternate Hypothesis (h1) to present stationary condition.
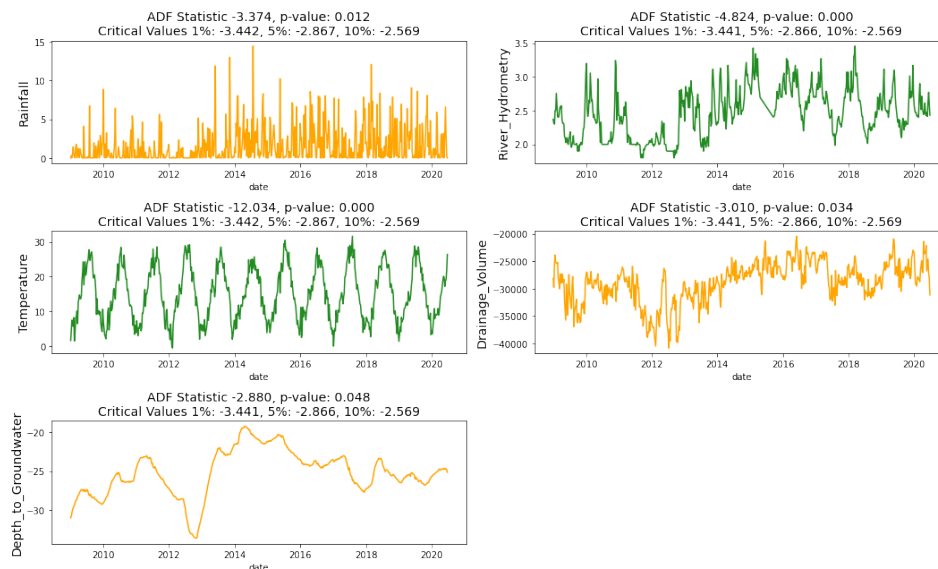

Figure 4 The analysis stationary of parameter using *Augmented Dickey-Fuller (ADF)* with *p-value* less than 5% and compare range the ADF with critical values.

We present each parameter to stationary in the Figure 3 using many options checking, there are checking ADF Statistics, *p-value,* critical values (for 1%, 5%, and 10%). To easier described the Figure 3 we will present into table with row and column below.

Table 3 The analysis stationary of parameter using *Augmented Dickey-Fuller (ADF)* with *p-value* less than 5% and compare range the ADF with critical values (1%,5%, and **10%)**

Categories	Evaluation Criteria to Stationary				
	ADF Statistics	*p-value*	Critical value		
			1%	5%	10%
Rainfall	-3.347	0.012	-3.442	-2.867	-2.569
Depth to Groundwater	-2.880	0.048	-3.441	-2.866	-2.569
Temperature	-12.034	0.000	-3.442	-2.867	-2.569
Drainage Volume	-3.010	0.034	-3.441	-2.867	-2.569
River Hydrometry	-4.824	0.000	-3.441	-2.867	-2.569

Refer into Table 3, we will describe and discuss one by one of the parameter, what the condition is stationary and non-stationary. Stationary or not, we can ensure if the ADF value less than critical value (ADF < critical value). We divide into two categories, the stationary condition is Temperature and River Hydrometry, and the non-stationary condition is rainfall, depth to groundwater, and drainage volume. To next ways, we will change non-stationary to stationary condition only in depth to groundwater, because this parameter will forecast in the future. The two common to change into stationary condition, we uses transformation *e.g.* stabilize the non-constant variance and differencing to subtract the current value from the previous value.

Table 4 The Stationary the Depth to Groundwater using Transforming and Differencing

Depth to Groundwater	Evaluation Criteria to Stationary				
	ADF Statistics	*p-value*	Critical value		
			1%	5%	10%
Depth to Groundwater (Non-Stationary)	-2.880	0.048	-3.441	-2.866	-2.569
Transforming	-2.947	0.040	-3.441	-2.866	-2.569
Differencing	-5.150	0.000	-3.441	-2.866	-2.569

In the Table 4, we can writes if the transforming and differencing methods can reduces ADF Statistics into less than the critical value, it mean the condition of the depth to groundwater to be stationary condition. This case reinforced with ADF Statistic always decreasing from -2.880 in the previous condition (initial non-stationary), -2.947 while transforming process, until less then critical value in the differencing condition in -5.150, the stationary condition is done.

3.5. Encoding Cyclical Features, Time Series Decomposition, and Lag

In this subsection, we will describe the data of depth to groundwater using analysis in encoding cyclical features, time series decomposition, and lag. We know that the time have cycling features, like Monday-Sunday, day 1st until day 30th, January-December in year, they will loop again. In the other side, month have difference day each year, while the difference between each month increments by 1 during the year, hence this condition can confuse a lots model. The cycle condition its equal with *y=cos(x),* they will begin from zero or *pi* until *2pi, or* in the other example like clock time.

Timeseries is cycle to analyze components condition *e.g.* levelling, trend, seasonality, and noise. The component condition has relation with decomposition which serve abstracted model will be use to solve timeseries with generally and easier to understanding time series analysis and forecasting. The Depth to Groundwater data has seasonality trend, hence in this case we will use seasonal decompose.
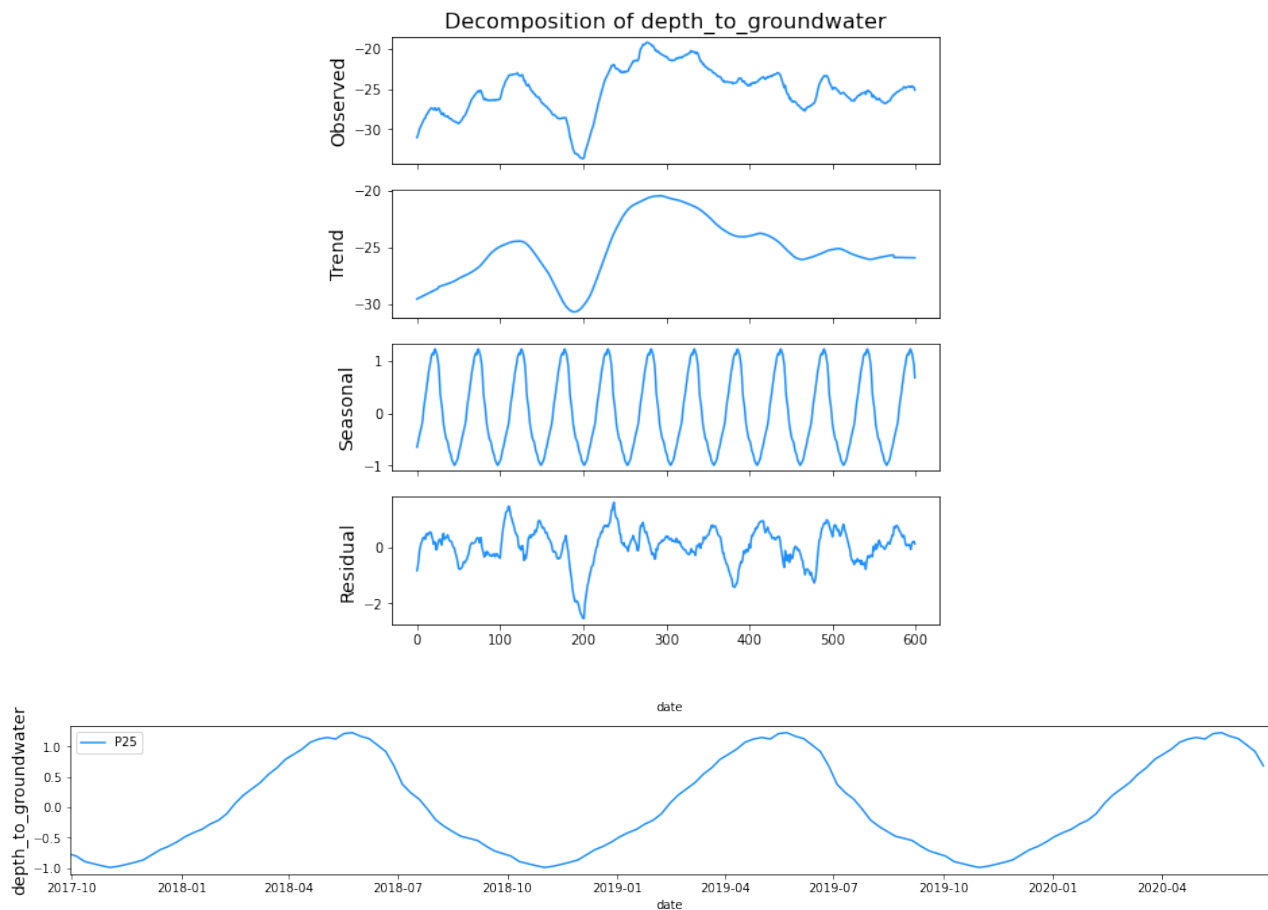

Figure 5 The Decomposition Timeseries of Depth to Groundwater with Seasonality, Observed, Trend and Residual

Refer into Figure 5, we present decomposition of the depth to groundwater four path, they are observed, trend, seasonal, and residual. The seasonal has best oscillation condition, and it will best path and good for forecasting in the future. To easier calculate and build a model, we make sure to compare the correlation with other variables. Hence, we will use lag four weeks in one month or monthly. The depth to groundwater is refer to Figure 5 (2) the maximum value is around May/June and the minimum value is around November.

3.6. Forecasting Data using Univariate and Multivariate Prophet Methods

In this subsection will be describe and discuss about our method using univariate and multivariate in Prophet-FB algorithm. We know if univariate is time series method using single time dependent variable and multivariate is time series method using multiple dependent variables. The first, we will present univariate time series using Prophet method. In the previous section (3.4), we already described the Prophet or Facebook-Prophet is open-source library. We will analyze three part, there are Mean Absolut Error (MAE), Root Mean Square Error (RMSE), and Graph.
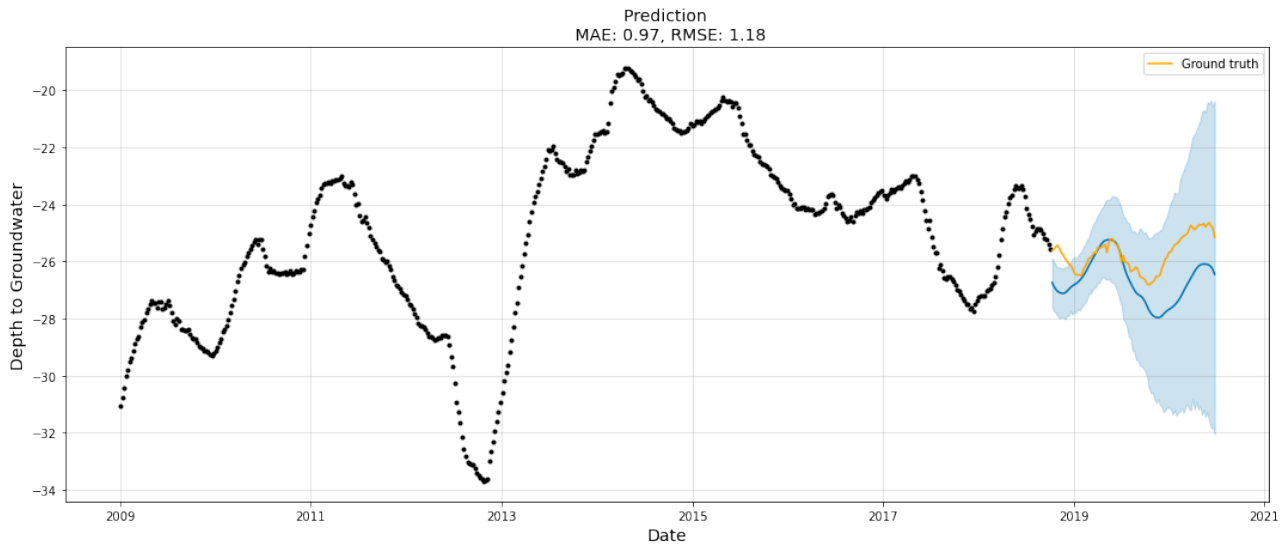

Figure 6 The Forecasting Data using *Univariate* Prophet Method with MAE and RMSE approach

Refer to Figure 6, the forecasting with univariate data using Depth to Groundwater, we present the MAE value around 0.97, and RMSE around 1.18. If we refer again in the right figure with blue highlight using yellow line, and blue line, we can describe if the yellow line is ground truth or actual value, and the blue line is predicted value. The predicted value almost far-way from actual value, but in around 2019 (middle, May-October) the predicted and actual value to be in one line, or it mean the actual value equal with predicted value.
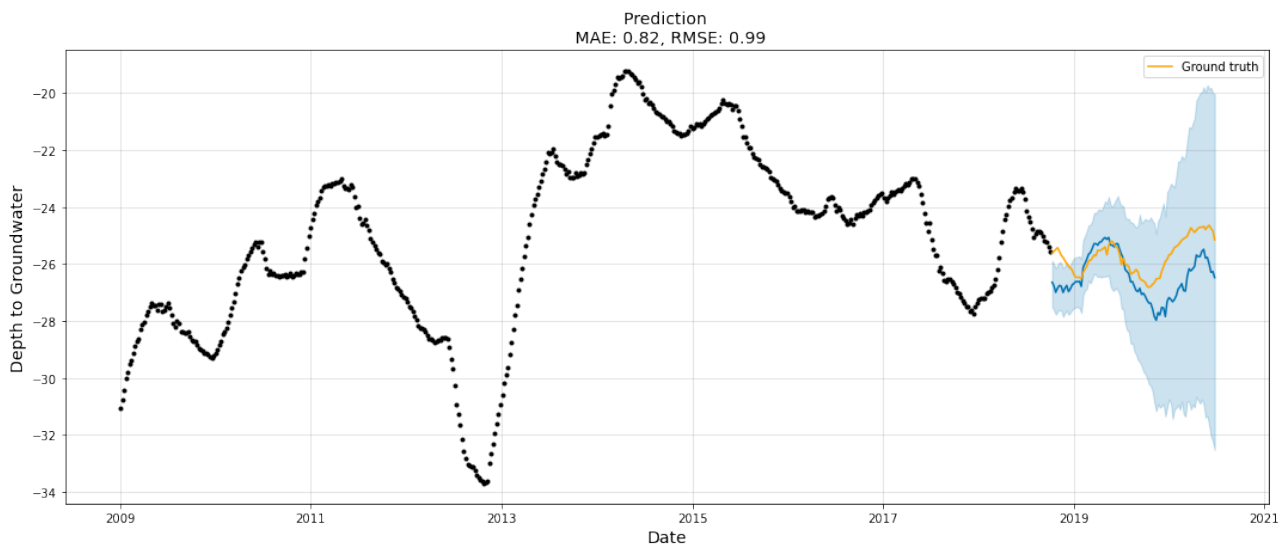

Figure 7 The Forecasting Data using *Multivariate* Prophet Method with MAE and RMSE approach

Refer to Figure 7, the forecasting with univariate data using Depth to Groundwater, we present the MAE value around 0.82, and RMSE around 0.99. If we refer again in the right figure with blue highlight using yellow line, and blue line, we can describe if the yellow line is ground truth or actual value, and the blue line is predicted value. The predicted value almost far-way from actual value, but in around 2019 (middle, May-October) the predicted and actual value to be in one line, or it mean the actual value equal with predicted value.

Table 5 The Results to Compare Univariate and Multivariate Prophet Method

Prophet Methods	Evaluation Criteria Approach	
	MAE	*RMSE*
Univariate	0.97	1.18
Multivariate	0.82	0.99

Refer to Table 5, we can easier to understand the results between univariate and multivariate of Prophet Method. The accurate data can be measure with more increasing the MAE and RMSE value, the results will be un-accurately, and vice versa. The smallest MAE and RMSE have presented by multivariate method. To solving our problem using Multivariate Prophet method more give best results to forecasting the Depth of Groundwater data.

4. Conclusions

Our research is far from perfectly, we present time series to forecasting the data using Prophet Method give the good results. We divide into two parts with univariate and multivariate approach. We also present how to handle missing data, transforming, differencing, determine the time series decomposition and lag, they give formulate to increase the knowledge how to prepare the data to solve the time series data. Actually, the ACEA company has more data to forecasting but we only trial in one dataset, maybe in the future we will proceed the other data, and compare with other method. In this case, we have presented if the Multivariate Prophet Method give a good result with MAE 0.82 and RMSE 0.99. To develop our research in the future we will discuss with other ACEA data in Aquifers (Auser, Doganella, Petrignano, Luco) to forecast using time series with ARIMA, SARIMA, and LSTM approach.

References

[1] Rahmstorf, S, "Ocean Circulation and Climate during The Past 120,000 years," Nature, 419, pp. 207-214, 2002.

[2] Vidyasagar, D, "Global Minute: Water and Health – Walking for Water and Water Wars", J. Peritanol., 27, pp. 56-58., 2007.

[3] Postel, S., and Carpenter, S., "Freshwater Ecosystem Service"., In. G. C. Daily, editor. Nature's services: Societal Dependence on Natural Ecosystem. Island Press, Washington DC., USA. Pp. 195-214. 1997.

[4] Rosenberg, D.R., McCully, P., and Pringle., C.M. "Global-Scale Enviromental Effects of Hidrologycal Alterations: Introdution". BioScience, 55, pp 746-751. 2000.

[5] Alperovits, E., and Shamir, U. "Design of Optimal Water Distribution-Systems". Water Resour. Res 13., pp 885-900. 1977.

[6] Morgan, D.R., and Goulter, I.C., "Optimal Urban Water Distribution Design'. Water Resour. Res 21., pp 642-652. 1985.

[7] Cunha, M.D, and Sousa, J. "Water Distribution Network Design Optimization: Simulated Annealing Approach". J. Water Res. PIASCE, 125. Pp 215-221. 1999.

[8] Bogert, M.T., "Chemistry and the conversation of our water resources". Journal of Franklin Institute 169. Pp 385-388. 1910.

[9] Resh, V.H., Norris, R.H., and Barbour, M.T., "Design and implementation of rapid assesment approach for water-resource monitoring using benthic macroinvertebrates". Aus. J. Ecol., 20 pp. 108-121. 1995.

[10] Sawaya, K.E., Olmanson, L.G., Heinert, N.J., Brezonik, P.L., and Bauer, M.E., "Extending satellite sensing to local scales: lan and water resource monitoring using high-resolution imagery". Remote Sens. Environ., 88 pp 144-156. 2003

[11] Gahr, F., Hermanutz, F., and Oppermann, W. "Ozanation: An important technique to comply with new German laws for textile waste-water treatment". Water Sci. Technol. 30 pp 255-263. 1994.

[12] Dossantos, L.M.F., and Livingston, A.G. "Membrane-attached biofilms for VOC waste-water treatment". 1-Novel in-situ biofilm thickness measurement technique, Biotechnol. Bioeng-47. Pp 82-89. 1995.

[13] Oturan, M.A. "An ecologically effective water treatment technique using electrochemically generated hydroxyl radicals for in situ destruction of organic pollutans: Application to herbicide". D.J. Electrochem. 30-pp 475-482. 2000.

[14] Andreu, J., Capilla., and Sanchis, E. "Aquatool: a generalized decision-support system for water-resources planning and operational management". J. Hydrol. 177-pp 269-291. 1996.

[15] Navratil, M., and Kolkova, A. "Decomposition and Forecasting Time Series in Business Economy using Prophet Forecasting Model". Central European Business Reviwe. Volume 8, Number 4. 2019.

[16] Zunic, E., Korjenic, K., Hodzic, K., and Donko, D. "Application of Facebook's Prophet Algorithm for Successful Sales Forecasting Bases on Real-World Data". International Journal of Computer Science & Technology (IJCSIT). Vol 2, No 2, April 2020.