Avrupa Bilim ve Teknoloji Dergisi
Sayı 41, S. 229-239, Kasım 2022

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
No. 41, pp. 229-239, November 2022

Research Article

Forecasting of Occupancy Rate of Dams in İstanbul

Mehmet Eren Nalici[1*], Ayhan Akbaş[2]

[1*] Abdullah Gül University, Faculty of Engineering, Industrial Engineering Department, Kayseri, Turkey, (ORCID: 0000-0002-7954-6916), mehmeteren.nalici@agu.edu.tr
[2] Abdullah Gül University, Faculty of Engineering, Computer Engineering Department, Kayseri, Turkey (ORCID: 0000-0002-6425-104X), ayhan.akbas@agu.edu.tr

(First Received 8 March 2022 and in Final Form 26 October 2022)

(**DOI:** 10.31590/ejosat.1084484)

ATIF/REFERENCE: Nalici, M., Akbaş, A., (2022). Forecasting Of Occupancy Rate of Dams in İstanbul. *Avrupa Bilim ve Teknoloji Dergisi*, (41), 229-239.

Abstract

Drinking water is becoming a crucial problem all over the world because of global warming. In crowded metropolises such as Istanbul, the problem of drinking water is a serious problem. In this study, it is aimed at developing a forecasting model by using the occupancy rates of Istanbul's dams between 2011 and 2020. The occupancy rate of each dam is then estimated using the best model for the years 2021 and 2022. According to the results of the estimation model, a decrease in the occupancy level of the dams in Istanbul is predicted. Therefore, it is thought that necessary measures should be taken to avoid water shortages.

Keywords: Forecasting, ARIMA, Fbprophet, Exponential Smoothing.

İstanbul Barajlarının Doluluk Oranı Tahmini

Öz

Küresel ısınmanın etkisiyle tüm dünyada içme suyu sorunu ciddi bir problem olmaya başlamaktadır. İstanbul gibi kalabalık metropollerde içme suyu sorunu ciddi bir sorundur. Bu çalışmada, İstanbul barajlarının 2011-2020 yılları arasındaki doluluk oranları kullanılarak bir tahmin modeli geliştirilmesi amaçlanmıştır. "Hareketli Ortalama", "ARIMA", "Fbprohet" ve "Üssel Düzgünleştirme" tahmin modelleri "Ortalama Karesel hatasının kareökü" ve "Ortalama karesel hata" değerlerine göre karşılaştırılmış ve her bir barajın 2021 ve 2022 yılları için doluluk oranı tahmin edilmeye çalışılmıştır. Tahmin modelinin sonuçlarına göre İstanbul'daki barajların doluluk oranında bir düşüş öngörülmektedir. Bu nedenle 2022 yılı için su kıtlığının yaşanmaması için gerekli önlemlerin alınması gerektiği düşünülmektedir.

Anahtar Kelimeler: Tahminleme, ARIMA, Fbprophet, Üssel Düzgünleştirme.

[*] Corresponding Author: mehmeteren.nalici@agu.edu.tr

1. Introduction

Climate change has become a major issue on a global scale as time goes on. According to Paterson (2013), global warming was first mentioned as a "serious problem" in 1988 in the US. The reason why global warming was mentioned as a problem in the 1990s was the abnormal meteorological events and the decrease in precipitation. Due to its geographical location, Turkey might be seriously affected by global warming. Climate change will affect different parts of Turkey in different ways and to varying degrees. The Aegean and Mediterranean areas, which lack adequate water, will be particularly hard hit (Öztürk, 2002). Due to the population density, the greatest need for drinking water is in Istanbul. Istanbul will suffer the most from a possible shortage of drinking water. Currently, Istanbul's drinking water needs are supplied from 10 different dams, such as "Ömerli, Elmalı, Terkos, Alibey, Büyükçekmece, Sazldere, Kazandere, Pabuçdere and Istrancalar" and "Yeşilçay" and "Melen" regulators. The total capacity of these dams is 868,683,000 m3 and the remaining water requirements are met by the regulators. In 2021, approximately 626 million m3 of water was used by these regulators. In the last 10 years, the yearly average water consumption of citizens in İstanbul was 994,030,000 m3 in total (İSKİ, 2022). The decrease in the occupancy rates of these dams may cause a serious water shortage in Istanbul.

Forecasting might be important for different areas because of predicting future outcomes. Necessary studies could be done, or measures can be taken according to the estimation results. In the literature, forecasting studies have been carried out in different areas. For instance, Yilmaz et al. (2021) claim that forecasting of COVID-19 cases are important to take precautions before spreading viruses. In this article COVID-19 case numbers is used in Brazil, America, Russia and India and Correlated Additive Model (CAM), Auto-Regressive Integrated Moving Average (ARIMA) and Back Propagation-Based Artificial Neural Network (BP-ANN) were used as prediction models. Moreover, Buyrukoğlu (2021) claims that estimated the market values of cryptocurrencies. Decision support models are created for investors by estimating the future market values of cryptocurrencies. In addition to this, deep learning methods were used as the prediction model. On the other hand, Ayyıldız et al (2022) is using LSTM networks for forecasting occupancy rate of Dams in İstanbul.

The purpose of this article is to forecast the occupancy rates of Istanbul's dams for the years 2021 and 2022. Thus, it is to ensure that necessary precautions are taken before detecting a possible decrease in water levels. 'Moving average', 'ARIMA', 'Exponential Smoothing' and 'Fbprophet' models were used as forecasting models.

2. Material and Method

2.1. Data Analysis

To estimate the occupancy rates of dams in Istanbul, daily occupancy rates between 2011 and 2020 were used. The dataset was taken from the Istanbul Metropolitan Municipality Open Data Portal (Data.ibb, 2022). In the data set, there are daily occupancy rates of 10 different dams in Istanbul between the years 2011-2020 as a percentage. Table 1. shows an example of a dataset for use in forecasting. In this table, dam number shows the name of dams, respectively, "Ömerli, Elmalı, Terkos, Alibey, Büyükçekmece". In this article, the "Python" programming language is used to visualize the data and create the forecasting model. Furthermore, "pandas", "numpy", "plotly", "seaborn", "matplotlib", "statsmodels", and "sklearn" libraries in the Python programming language are used in the visualization and analysis of the data.

Table 1. Example Dataset of Dams Occupancy Rate

Dams / Date	1	2	3	4	5	6	
2011-01-01	80.07	84.86	96.25	91.87	42.48	85.99	
2011-01-02	80.07	85.02	96.67	91.68	42.71	86.17	
2011-01-03	80.07	85.07	96.88	91.68	42.79	86.55	
2011-01-04	80.07	85.18	95.00	92.44	43.34	86.74	
2011-01-05	80.61	85.98	95.21	92.63	45.81	87.30	...
...	...	...	...	...	...	...	
2020-12-27	19.17	48.52	23.70	22.68	30.65	17.42	
2020-12-28	18.76	48.34	23.54	22.54	30.58	17.42	
2020-12-29	18.30	48.04	23.45	22.40	30.45	17.27	
2020-12-30	17.89	47.79	23.29	22.12	30.18	17.42	
2020-12-31	17.43	47.49	23.26	21.84	30.05	17.57	

2.1.1. Analysing Yearly Average

To make the data more meaningful, the annual average values of each dam were calculated.

Figure 1. Average Occupancy Rate of Each Dams by Years

Figure 1. shows the annual average occupancy rate for each dam. Each line on the chart represents a different threshold. When figure 1. is examined, an annual linear trend could not be determined in the annual dam occupancy rates. The average dam occupancy rate is around 50 percent. Moreover, dam occupancy rates decreased significantly in 2014. The reason for this decrease is thought to be the lack of precipitation due to global warming.

2.1.1. Analysing Seasonal Average by Years

The precipitation rates will be different in different seasons. This might affect the occupancy rate of the dams. As a result, it is necessary to investigate the seasonal variation in dam occupancy rates. Table 2. shows the average occupancy rate by season and year. Moreover, Figure 2 represents the average occupancy rate by season by year graphically.

Table 2. Average Occupancy Rate of Each Season by Years

Year	Spring	Summer	Autumn	Winter
2011	95.956228	85.523511	62.209978	78.656022
2012	92.119380	72.364359	49.000198	71.194209
2013	88.403750	65.615837	37.101253	62.721367
2014	26.541174	22.684326	33.062758	37.718133
2015	95.707359	79.798174	68.054275	79.211467
2016	81.285902	57.350696	36.955110	59.842956
2017	78.898989	62.739728	56.017725	66.724389
2018	87.218880	72.171250	53.453824	72.714022
2019	91.767043	67.938130	41.857571	69.768911
2020	55.564185	43.913315	27.057604	38.266505

Figure 2. Average Occupancy Rate of Each Seasons

According to Figure 2, spring is the season when the dams are most full. It is estimated that there is an increase in the occupancy rate with the melting of the snow falling in winter. Average occupancy rates are approximately the same in the winter and summer seasons. In autumn, it has the lowest occupancy rate of all the seasons. The reason for this is the low amount of precipitation in the summer season. Furthermore, average occupancy rates for all seasons significantly decreased in 2014. The reason for this situation is thought to be less precipitation in 2014.

2.2. Proposed Methodology

"Moving Average", "ARIMA", "fbprophet" and "Exponential Smoothing" models were used to forecast the occupancy rates of dams in İstanbul. For these models, "root mean squared error (RMSE)" and "mean squared error (MSE)" values were found for each dam, and the RMSE and MSE values of the models were compared for each dam. "fbprophet" and "statsmodels" libraries were used while creating the forecasting model and analysing the error rates.

Before modelling the data, it is necessary to check its stationarity. Van Greunen, Heymans, van Heerden, and van Vuuren (2014) claim that the capacity to render a time series to the correct form of stationarity can lead to false results, while the inability to render a time series to the correct form of stationarity can lead to erroneous results. The statistical features of a process that generates a time series do not vary over time, which is known as stationarity. This isn't to say that the series doesn't change over time; it just means that the method by which it changes doesn't change (towardsdatascience, 2022). The "dickey-fuller test" method was used to determine the stationarity of the data. Dolado, Gonzalo and Mayoral (2002) state that the Dickey and Fuller test statistic is one of the most used methods for determining whether a process is "I (1)" or "I (0)." Its broad application is owing to its computational simplicity as well as its flexibility to more generic setups such as serial correlation in residuals, seasonality, breaking trends, and so on. The Dickey-Fuller test is applied to each dam. If the P-value of the test is less than or equal to 0.05, the data will be stationary. Otherwise, the data is not stationary.

Table 3. P-values of each dam

Dams	P-value	Stationary or Not Stationary
Ömerli	0.000528	Stationary
Darlık	8.231593e-05	Stationary
Elmalı	0.211310	Not Stationary
Terkos	0.001289	Stationary
Alibey	0.000914	Stationary
Büyükçekmece	0.003648	Stationary
Sazlıdere	0.030055	Stationary
Kazandere	0.093578	Not Stationary
Pabuçdere	0.047259	Stationary
Istrancalar	0.000738	Stationary

According to Table 3, all data except "Elmalı" and "Kazandere" are stationary. For 30 days, the "shift" method will be applied to "Elmalı" and "Kazandere" data to make it stationary. Shifting or "lagging" numbers back and forth in time is a typical operation on time series data, such as computing the percent change from sample to sample (Weiming, 2015). Then, apply the Dickey-Fuller test for shifting data, and shifted data will become stationary (Table 4.).

Table 4. P-values of Shifted Data

Dams	P-value	New P-value	Stationary or Not Stationary
Elmalı	0.211310	0.000153	Stationary
Kazandere	0.093578	0.000390	Stationary

Moreover, since seasonal estimation will also be made, the "dickey-fuller test" method has been applied to the seasonal average occupancy rate data for years. The same confidence interval as for previous data is used for seasonal average occupancy rate data by year. According to Table 5, all seasonal data might be stationary with a 0.95 confidence interval.

Table 5. P-values of Each Season

Season	P-value	Stationary or Not Stationary
Spring Occupancy	0.008896	Stationary
Summer Occupancy	0.007490	Stationary
Autumn Occupancy	0.021246	Stationary
Winter Occupancy	0.032537	Stationary

2.2.1. Moving Average

A "Moving Average" is a time series that is created by averaging many consecutive data points from some other time-series data. Because each average is determined by discarding the oldest observation and including the next, the technique is referred to as "moving average." The averaging "moves" across the time series until all observations are computed at each observation that has all the average's constituents accessible (Hyndman, 2011). The moving average is calculated by taking the arithmetic average of a given set of values over a given period. The number is added as the number of days determined and divided by the determined number. Thus, the estimated value on the specified day is found. This process is continued until the last day comes.

The number of "Moving Average" steps is taken as "4 days", "5 days" and "6 days" to forecast the occupancy of the dams. "Mean squared error" and "Root mean squared error" values were calculated separately for each model.

Table 6. RMSE and MSE Rate for Moving Average (4 days)

Dams	Root Mean Squared Error	Mean Squared Error
Ömerli	0.9962	0.9925
Darlık	0.9885	0.9771
Terkos	1.1265	1.2691
Alibey	1.1272	1.2707
Büyükçekmece	0.8921	0.7958
Sazlıdere	0.6334	0.4012
Pabuçdere	1.6302	2.6576
Istrancalar	5.1126	26.1387
Elmalı	2.0182	4.0731
Kazandere	2.8417	8.0753

Table 7. RMSE and MSE Rate for Moving Average (5 days)

Dams	Root Mean Squared Error	Mean Squared Error
Ömerli	1.2023	1.4457
Darlık	1.1929	1.4231
Terkos	1.3345	1.7809
Alibey	1.3521	1.8283
Büyükçekmece	1.0714	1.1479
Sazlıdere	0.7639	0.5836
Pabuçdere	1.9242	3.7025
Istrancalar	5.8641	34.3888
Elmalı	2.3226	5.3949
Kazandere	3.3041	10.9173

Table 8. RMSE and MSE Rate for Moving Average (6 days)

Dams	Root Mean Squared Error	Mean Squared Error
Ömerli	1.4023	1.9664
Darlık	1.3918	1.9371
Terkos	1.5311	2.3445
Alibey	1.5672	2.4562
Büyükçekmece	1.2457	1.5518
Sazlıdere	0.8906	0.7933
Pabuçdere	2.2042	4.8587
Istrancalar	6.5208	42.5212
Elmalı	2.6087	6.8058
Kazandere	3.7259	13.8824

The number of "Moving Average" steps is taken as "2 years", "3 years" to forecast the seasonal occupancy of the dams. "Mean squared error" and "Root mean squared error" values were calculated separately for each model.

Table 9. RMSE And MSE Rate for Moving Average (2 Years)

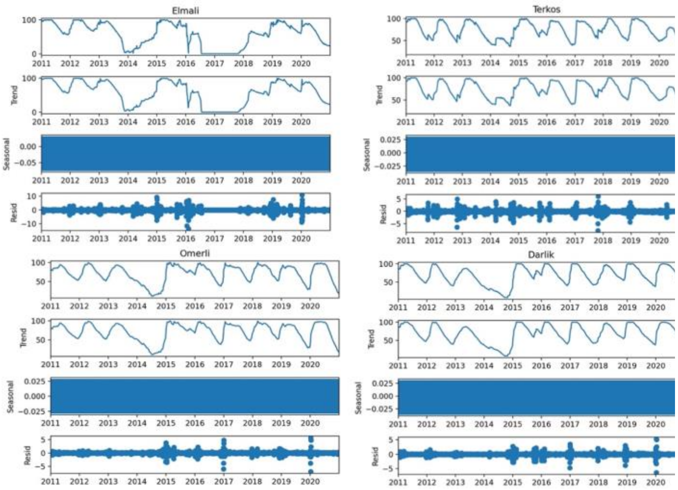
Season	Root Mean Squared Error	Mean Squared Error
Spring Occupancy	19.1098	463.4019
Summer Occupancy	15.0262	274.0968
Autumn Occupancy	10.2318	128.4443
Winter Occupancy	11.6593	182.9532

Table 10. RMSE and MSE Rate for Moving Average (3 years)

Season	Root Mean Squared Error	Mean Squared Error
Spring Occupancy	15.4878	239.8735
Summer Occupancy	12.6865	160.9482
Autumn Occupancy	11.9357	142.4621
Winter Occupancy	12.2521	150.1153

2.2.2. ARIMA

ARIMA techniques are used to evaluate time series and were formerly primarily employed for load forecasting because of their accuracy and mathematical soundness (Contreras et al., 2003). ARIMA models are those that are used on non-stationary series and then changed by differencing to a stationary state. Non-stationary linear stochastic models are those that are applied to non-stationary series and then made stationary via the use of differences. These models, which apply AR to series with d-degree differences, represent the value of the variable in the t-period as a linear function of a certain number of back-period values and the error term in the same period. The value of the variable in the t-period is also a linear function of a specific number of back-period error terms. They combine the MA models and are represented as the models are typically represented by ARIMA (p, d, q). In this case, the degrees of the moving average (MA) and autoregressive (AR) models, respectively, are p and q, and the degree of difference is denoted by d (Kaynar et al., 2009). To use "ARIMA", there has to be no seasonality in the data. If there is seasonality in the data, "SARIMA" should be used. According to this result, checking seasonality is essential for using ARIMA. After trend and seasonal elements have been removed, model fitting entails identifying and estimating parameters. Regression analysis that measures the strength of one dependent variable in relation to multiple fluctuating variables is known as an autoregressive integrated moving average model. Instead of using actual values, the model seeks to forecast future values by analysing discrepancies between values in the series. A preliminary autoregressive moving average (ARIMA) process is built based on the estimated autocorrelation function (ACF) and the estimated partial autocorrelation function (PACF) during the identification step (Nobre et al., 2001). For checking seasonality, the "seasonal decompose" method is used in "statsmodels" libraries. According to Figure 3., there is no seasonality each dam. Therefore, ARIMA will be used for forecasting.



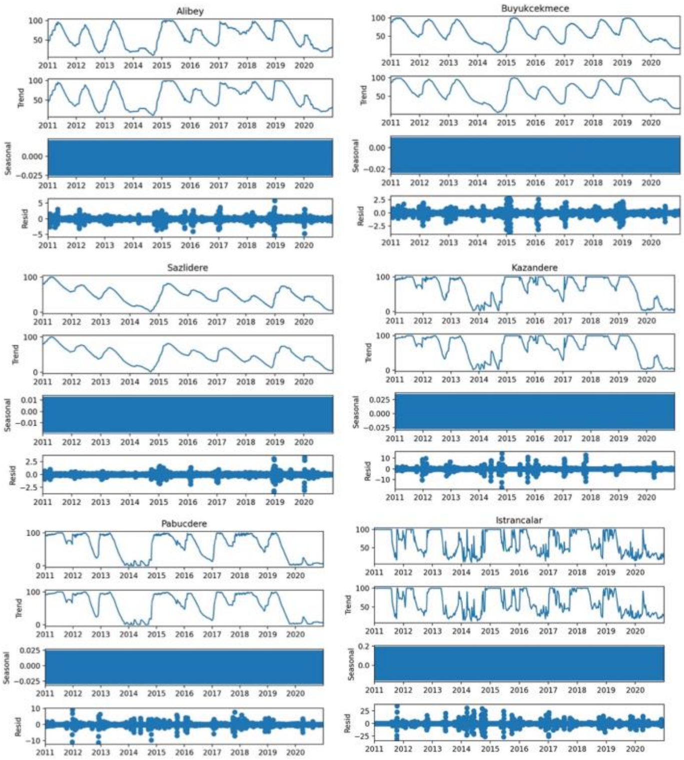


Figure 3. Seasonal_Decompose Result of Each Dams

ARIMA has three variables, such as "p", "d" and "q." "p" is the autoregressive model's order (number of time lags), "d" is the degree of differencing (the number of times the data has had past values subtracted), and "q" is the moving-average model's order. In this paper, different (p,d,q) combinations will try to fit the ARIMA model. For example, "p" and "q" have a range of "0-7" and "d" has a range of "0-4" for forecasting each dam. All possible (p,d,q) combinations will be tried to get the "Root Mean Squared Error" for each pair. For training the model, 0.8 of data for each dam is used. The remaining one (0.2) is used for verification. The ideal combination was chosen for the combination training model that gave the smallest RMSE value for each dam.

Table 11. Example of RMSE rate of each dam by each pdq pairs

(p,d,q)	Ömerli	Darlık	Terkos	Alibey	
(0, 0, 0)	24.3471	22.1796	23.4816	28.9917	
(0, 0, 1)	24.3446	22.1602	23.4793	28.9686	
(0, 0, 2)	24.3404	22.1499	23.4774	28.9465	
(0, 0, 3)	24.3414	22.1396	23.4743	28.9536	
(0, 0, 4)	24.3329	22.0857	23.4732	28.8911	
...	...	...	...	...	…
(6, 3, 2)	178.2897	298.5143	214.5498	1152.5791	
(6, 3, 3)	118.5196	652.6726	224.9215	300.7981	
(6, 3, 4)	86.7820	602.1260	608.4480	553.8868	
(6, 3, 5)	407.6319	329.7974	357.2996	615.1873	
(6, 3, 6)	427.4416	400.7443	370.0125	1235.7716	

The smallest (p,d,q) values among the error values, such as "root mean squared error" and "mean squared error", were found for each dam data set. These values will then be used to compare the error values of other models.

Table 12. The smallest RMSE value of (p,d,q) pairs of each dam

Dams	(p,d,q)	RMSE
Ömerli	(5, 0, 4)	22.3810
Darlık	(6, 0, 4)	17.0530
Terkos	(6, 0, 6)	20.1219
Alibey	(5, 0, 4)	16.1057
Büyükçekmece	(4, 0, 5)	24.0485
Sazlıdere	(0, 3, 1)	18.7186
Pabuçdere	(4, 2, 0)	23.7179
Istrancalar	(3, 0, 0)	34.7594
Elmalı	(6, 0, 4)	5.7194
Kazandere	(1, 0, 2)	6.6172

Table 13. The smallest MSE value of (p,d,q) pairs of each dam

	(p,d,q)	MSE
Ömerli	(5, 0, 4)	500,9092
Darlık	(6, 0, 4)	290,8048
Terkos	(6, 0, 6)	404,8909
Alibey	(5, 0, 4)	259,3936
Büyükçekmece	(4, 0, 5)	578,3304
Sazlıdere	(0, 3, 1)	350,3860
Pabuçdere	(4, 2, 0)	562,5388
Istrancalar	(3, 0, 0)	1208,2159
Elmalı	(6, 0, 0)	32.7115
Kazandere	(6, 0, 0)	43.7873

For seasonal data, "p" and "q" have the range of "0-2" for forecasting each season. The model is trained using 0.8 seasons of data. The remaining one (0.2) is utilized for verification purposes. For the combination training model, the best combination was found that yielded the least RMSE value for each season.

Table 14. RMSE Rate of Each Season by Each pdq Pairs

(p,d,q)	Spring	Summer	Autumn	Winter
(0, 0, 0)	19.4443	14.9235	16.7478	19.8501
(0, 0, 1)	21.4657	15.2257	14.7616	20.1604
(0, 1, 0)	22.6131	20.2043	20.3866	24.4469
(0, 1, 1)	19.4424	14.9216	16.7469	19.8492
(1, 0, 0)	20.1171	15.2298	16.3097	20.1360
(1, 0, 1)	20.3859	14.3219	14.6632	19.4277
(1, 1, 0)	21.8028	18.3283	21.2040	23.3002
(1, 1, 1)	19.7451	14.9290	16.3080	19.8874

For each seasonal data, the least (p,d,q) values were found among the error values such as "root mean squared error" and "mean squared error." The error values of other models will then be compared to these values.

Table 15. The smallest RMSE value of (p,d,q) pairs of each season

	(p,d,q)	RMSE
Spring Occupancy	(0, 1, 1)	19.4424
Summer Occupancy	(1, 0, 1)	14.3219
Autumn Occupancy	(1, 0, 1)	14.663
Winter Occupancy	(1, 0, 1)	19.4277

Table 16. The smallest MSE value of (p,d,q) pairs of each Season

	(p,d,q)	MSE
Spring Occupancy	(0, 1, 1)	378,0069
Summer Occupancy	(1, 0, 1)	205,1168
Autumn Occupancy	(1, 0, 1)	215,0036
Winter Occupancy	(1, 0, 1)	377,4355

2.2.3. Fbprophet

"Fbprophet" is an additive model for forecasting time series data that fits non-linear patterns with yearly, weekly, and daily seasonality, as well as seasonal impacts. It works best with time series with substantial seasonal influences and historical data from several seasons. "Fbprophet" can withstand missing information and trend alterations, and it usually handles extremes well (Chikkakrishna, 2019). For the "Fbprophet" library, 0.8 of each dam's data is reserved for training, the rest for testing. It was run daily over 365-day period. Moreover, RMSE and MSE values were calculated for each dam.

Table 17 RMSE and MSE Rate for Fbprophet of Each Dam

Dams	Root Mean Squared Error	Mean Squared Error
Ömerli	18.2305	332,3511
Darlık	17.5149	306,7717
Terkos	17.2301	296,8763
Alibey	35.9049	1289,1618
Büyükçekmece	46.5420	2166,1577
Sazlıdere	38.3423	1470,1319
Pabuçdere	62.1590	3863,7412
Istrancalar	53.4463	2856,5069
Elmalı	10.1054	102,1191
Kazandere	7.0872	50,2284

For seasonal data, the model run yearly and in 1 year period. Furthermore, RMSE and MSE values were calculated for season.

Table 18. RMSE and MSE Rate for Fbprophet of Each Season

	Root Mean Squared Error	Mean Squared Error
Spring	22.7623	518,1223
Summer	15.3115	234,4437
Autumn	19.4409	377,9502
Winter	22.4733	505,0495

2.2.4. Exponential Smoothing

Exponential smoothing is a straightforward and practical method of forecasting that involves constructing a forecast using an exponentially weighted average of previous data. Most of the weight is given to the current observation, with less weight being given to the observation immediately before it, and even less weight being given to the observation before that, and so on. The objective behind exponential smoothing is to smooth the original series in the same manner as the moving average smooths it, and then use the smoothed series to anticipate future independent variables of importance. Typically, the exponential smoothing approach is founded on the idea that time series levels should vary around a fixed level or change gradually over time (Ostertagová, 2011).

For the model, Figure 3 is used to model dam data for seasonality. According to Figure 3., exponential smoothing with an additivity seasonality model is used for "Elmalı" "Kazandere" and "Pabuçdere" data forecasting. Exponential smoothing with a multiplicative seasonality model is used for other dams' data. 0.8 of the datasets were used to train the models, and the rest were used for testing the models. In both models, the seasonality period was taken as 365 days and the daily estimation of the 2-year data is made. RMSE and MSE are calculated for each dam's data.

Table 19 RMSE and MSE Rate for Exponential Smoothing of Each Dam

Dams	RMSE	MSE
Ömerli	22.4963	506,0835
Darlık	48.1110	2314,6683
Terkos	36.2185	1311,7797
Alibey	94.5095	8932,0456
Büyükçekmece	63.7549	4064,6873
Sazlıdere	18.2419	332,7669
Istrancalar	274.4037	752,9701
Pabuçdere	47.1646	2224,4995
Elmalı	23.9124	571,8029
Kazandere	18.7226	350,5358

For seasonal data, the model is fitted with exponential smoothing with multiplicative seasonality. Seasonal periods are selected every 2 years. The models are trained using 0.8 of the datasets and tested with the remaining datasets, and RMSE and MSE values are calculated.

Table 20. RMSE and MSE Rate for Exponential Smoothing of Each Season

	RMSE	MSE
Spring Occupancy	11.5633	133,7099
Summer Occupancy	9.4762	89,7983
Autumn Occupancy	15.0599	226,8005
Winter Occupancy	15.6953	246,3424

3. Results and Discussion

Root mean squared error and mean squared error values were calculated for each time series forecasting model. The "moving average" model was compared within itself. If the models are compared the "moving average" models according to the RMSE and MSE values, the model that gives the smallest values for all dams is the "moving average 4 days" model. This model is comparable to other models for the "moving average" model.

Table 21. RMSE values of Moving Average Models for Each Dam

Dams	Moving Average (4 days)	Moving Average (5 days)	Moving Average (6 days)
Ömerli	0.9962	1.2023	1.4023
Darlık	0.9885	1.1929	1.3918
Terkos	1.1265	1.3345	1.5311
Alibey	1.1272	1.3521	1.5672
Büyükçekmece	0.8921	1.0714	1.2457
Sazlıdere	0.6334	0.7639	0.8906
Pabuçdere	1.6302	1.9242	2.2042
Istrancalar	5.1126	5.8641	6.5208
Elmalı	2.0182	2.3226	2.6087
Kazandere	2.8417	3.3041	3.7259

When the RMSE values of all models are compared, the smallest value is in the "moving average 4 day" model. However, due to the nature of the model, trend and seasonality are not used in the "Moving Average" forecasting model. When the dam data is examined, it is seen that the data includes trend (Figure 3). Therefore, it is not considered appropriate to choose the "moving average of 4 days" model as the estimation model. A choice will be made between "ARIMA", "Fbprophet", and "Exponential Smoothing" as they include seasonality and trend data. When the RMSE values of the "ARIMA", "Fbprophet" and "Exponential smoothing" models are examined, the "ARIMA" model gives the lowest value for "Darlık", "Alibey", "Büyükçekmece", "Pabuçdere", "Istrancalar" and "Elmalı" dams. The "Fbprophet" model gives the lowest "rmse" value for "Ömerli" and "Terkos" dams. Finally, the "Exponential Smoothing" model gives the lowest RMSE value for the "Sazlıdere" dam. Since the comparison was made for the RMSE values, the MSE values were not compared again. Forecasts for 2021 and 2022 were created with the model that gave the lowest RMSE value for each dam.

Table 22. RMSE values of Models-1 (Each Dam)

Dams	Moving Average (4 days)	ARIMA with (p,d,q)	
Ömerli	0.9962	22.3810	(5, 0, 4)
Darlık	0.9885	17.0530	(6, 0, 4)
Terkos	1.1265	20.1219	(6, 0, 6)
Alibey	1.1272	16.1057	(5, 0, 4)
Büyükçekmece	0.8921	24.0485	(4, 0, 5)
Sazlıdere	0.6334	18.7186	(0, 3, 1)
Pabuçdere	1.6302	23.7179	(4, 2, 0)
Istrancalar	5.1126	34.7594	(3, 0, 0)
Elmalı	2.0182	5.7194	(6, 0, 4)
Kazandere	2.8417	6.6172	(1, 0, 2)

Table 23. RMSE values of Models-2 (Each Dam)

Dams	Fbprophet	Exponential Smoothing
Ömerli	18.2305	22.4963
Darlık	17.5149	48.1110
Terkos	17.2301	36.2185
Alibey	35.9049	94.5095
Büyükçekmece	46.5420	63.7549
Sazlıdere	38.3423	18.2419
Pabuçdere	62.1590	274.4037
Istrancalar	53.4463	47.1646
Elmalı	10.1054	23.9124
Kazandere	7.0872	18.7226

Since the comparison was made for the RMSE values, the MSE values were not compared again. Forecasts for 2021 and 2022 were created with the model that gave the lowest RMSE value for each dam.

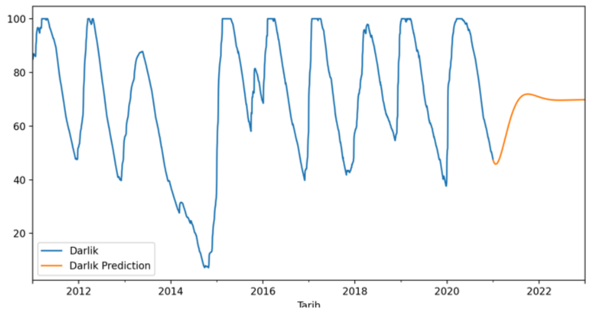


Figure 4. Prediction of Darlık between 2021-2022 (ARIMA (6,0,4))

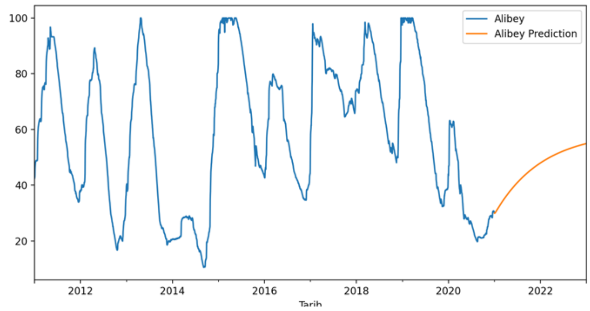


Figure 5. Prediction of Alibey between 2021-2022 (ARIMA (5,0,4))

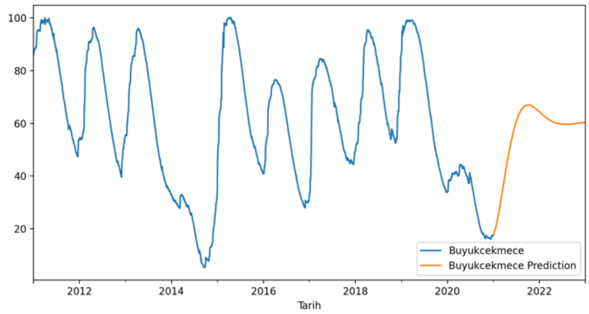


Figure 6. Prediction of Büyükçekmece between 2021-2022 (ARIMA (4,0,5))

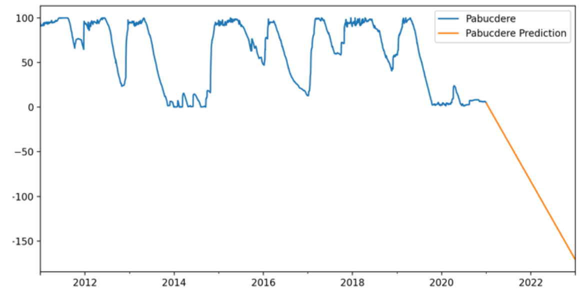


Figure 7. Prediction of Pabuçdere between 2021-2022 (ARIMA (4,2,0))

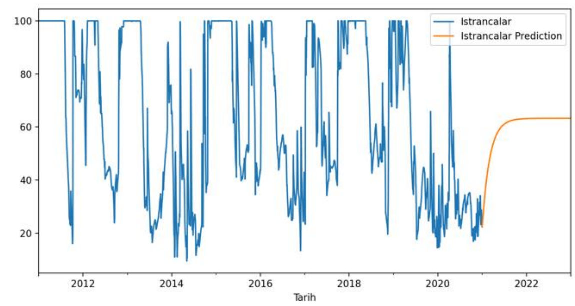


Figure 8. Prediction of Istrancalar between 2021-2022 (ARIMA (3,0,0))

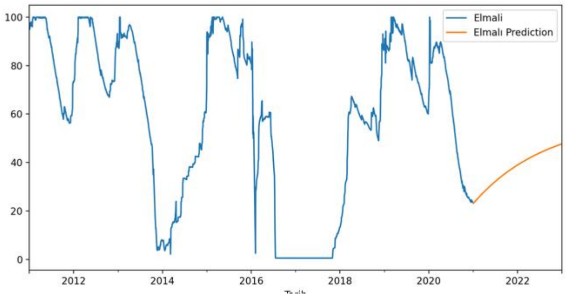


Figure 9. Prediction of Elmalı between 2021-2022 (ARIMA (6,0,4))

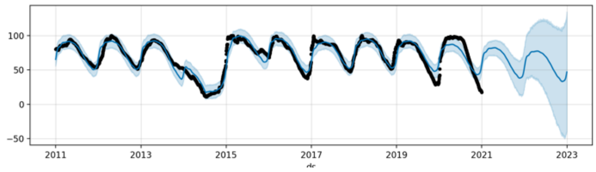


Figure 10. Prediction of Ömerli between 2021-2022 (Fbprophet)

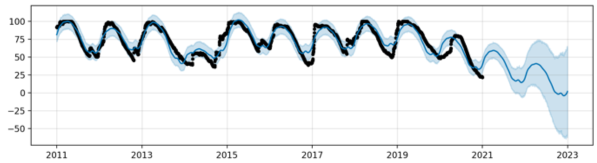


Figure 11. Prediction of Terkos between 2021-2022 (Fbprophet)

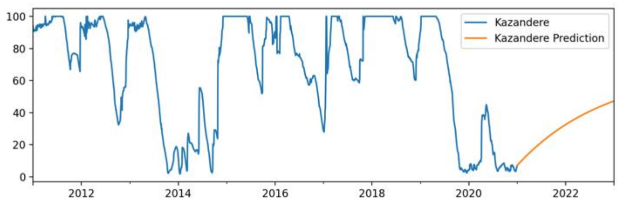


Figure 12. Prediction of Kazandere between 2021-2022 (ARIMA (1,0,2))

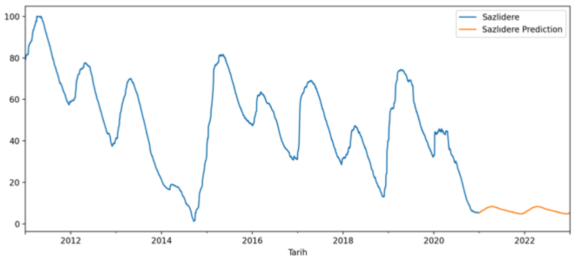


Figure 13. Prediction of Sazlıdere Between (2021-2022) (Exponential Smoothing)

For seasonal data, forecasting models are compared. The "Moving Average" model has given the smallest RMSE value, but it is thought that estimating with this model will not give accurate results, since this model does not contain trend and seasonality values. When the other 3 models were compared, it was found appropriate to estimate occupancy in spring, summer, and winter with "Exponential Smoothing" and to estimate occupancy in autumn with ARIMA (1,0,1) parameters.

Table 24. RMSE Values of Models-1 (Each Season)

	Moving Average (2 Years)	**Moving Average (3 Years)**
Spring Occupancy	19.1098	15.4878
Summer Occupancy	15.0262	12.6865
Autumn Occupancy	10.2318	11.9357
Winter Occupancy	11.6593	12.2521

Table 25. RMSE Values Of Models-2 (Each Season)

	Fbprophet	**Exponential Smoothing**
Spring Occupancy	22.7623	11.5633
Summer Occupancy	15.3115	9.4762
Autumn Occupancy	19.4409	15.0599
Winter Occupancy	22.4733	15.6953

Table 26. RMSE Values Of Models-3 (Each Season)

	ARIMA with (p,d,q)	
Spring Occupancy	19.4424	(0, 1, 1)
Summer Occupancy	14.3219	(1, 0, 1)
Autumn Occupancy	14.663	(1, 0, 1)
Winter Occupancy	19.4277	(1, 0, 1)

Since the comparison was made for the RMSE values, the MSE values were not compared again. Forecasts for 2021 and 2022 were created with the model giving the lowest RMSE value for each season.

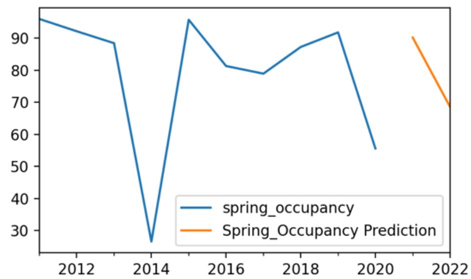


Figure 14. Prediction of Spring Occupancy for 2021-2022 (Exponential Smoothing)

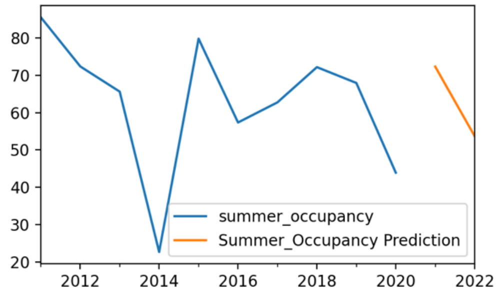


Figure 15. Prediction of Summer Occupancy for 2021-2022 (Exponential Smoothing)

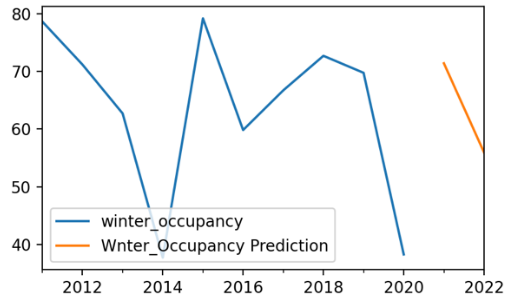


Figure 16. Prediction of Winter Occupancy for 2021-2022 (Exponential Smoothing)

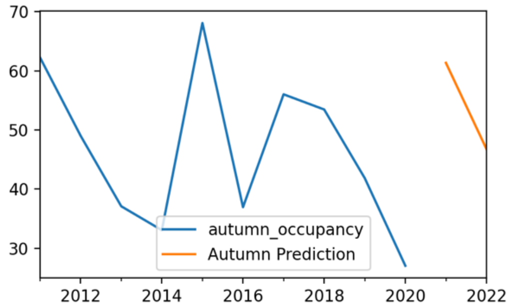


Figure 17 Prediction of Autumn Occupancy for 2021-2022 (ARIMA (1,0,1)

4. Conclusion and Recommendations

Global warming poses a serious threat to the entire world. Increasing global warming poses the danger of water scarcity. In Turkey, in crowded cities such as Istanbul that meet their drinking water needs from dams, the lack of precipitation due to global warming causes water scarcity. In this study, the daily occupancy rates of 10 different dams in Istanbul between 2011 and 2020 were examined. In line with this dataset, occupancy forecasts of dams in Istanbul for the years 2021 and 2022 were made using the "Moving Average," "ARIMA," "Fbprophet" and "Exponential Smoothing" models. A comparison of RMSE and MSE values of forecasting models was made. For each dam, the estimation model giving the lowest RMSE value was chosen. As a result of these models, it is estimated that the occupancy rate for the "Darlık", "Alibey", "Büyükçekmece" and "Istrancalar" dams for 2021 and 2022 will be approximately 60 percent. The occupancy rate of the "Elmalı" and "Kazandere" dams is expected to be around 40%, that of the "merli" dam is expected to be around 25%, and that of the "Terkos" dam is expected to be around 30%. On the other hand, it is estimated that the "Sazlıdere" and "Pabuçdere" dams will have an occupancy rate of almost 0 percent. If the forecasts of seasonal data are analyzed, it is estimated that the average dam occupancy rate in summer, winter, and spring will be around 50 percent in 2022. This rate, on the other hand, is around 70% during the spring season.

These forecasting models show that in 2022, the average dam occupancy rate will be 33.5 percent. In other words, it is estimated that Istanbul will have approximately 291,008,805 m3 of drinking water in 2022 for dams. There might be water shortages in Istanbul in 2022. It is estimated that there will not be enough water in the dams. Therefore, necessary precautions should be taken to ensure that the people living in Istanbul do not suffer from water shortages.

References

2022. [online] Available at: <https://towardsdatascience.com/stationarity-in-time-series-analysis-90c94f27322> [Accessed 11 January 2022].

2022. [online] Available at: https://www.iski.istanbul/web/

Buyrukoglu, S. (2021). Promising Cryptocurrency Analysis using Deep Learning. 2021 5Th International Symposium On Multidisciplinary Studies And Innovative Technologies (ISMSIT).

Chikkakrishna, Naveen Kumar, et al. "Short-term traffic prediction using sarima and FbPROPHET." 2019 IEEE 16th India Council International Conference (INDICON). IEEE, 2019.

Contreras, Javier, et al. "ARIMA models to predict next-day electricity prices." IEEE transactions on power systems 18.3 (2003): 1014-1020.

Data.ibb.gov.tr. 2022. Hoş Geldiniz - İBB. [online] Available at: <https://data.ibb.gov.tr>

Dolado, J., Gonzalo, J. and Mayoral, L., 2002. A Fractional Dickey-Fuller Test for Unit Roots. Econometrica, 70(5), pp.1963-2006.

E. Ayyıldız & M. Erdoğan (2022). Forecasting of daily dam occupancy rates using LSTM networks. World Journal of Environmental Research. 12(1), 33-42.

Hyndman, Rob J. "Moving Averages." (2011): 866-869.

K. Öztürk, "Küresel İklim Değişikliği ve Türkiye'ye Olası Etkileri," G.Ü. Gazi Eğitim Fakültesi Dergisi Cilt 22, Sayı 1 (2002) 47-65.

Kaynar O., Taştan S., ZAMAN SERİLERİ TAHMİNİNDE ARIMA-MLP MELEZ MODELİ, Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, Cilt: 23, Sayı: 3, 2009.

Nobre, Flávio Fonseca, et al. "Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology." Statistics in medicine 20.20 (2001): 3051-306

Ostertagová, Eva, and Oskar Ostertag. "The simple exponential smoothing model." The 4th International Conference on Modelling of Mechanical and Mechatronic Systems, Technical University of Košice, Slovak Republic, Proceedings of conference. 2011.

Paterson, M., 2013. Global Warming and Global Politics. Hoboken: Taylor and Francis.

van Greunen, J., Heymans, A., van Heerden, C. and van Vuuren, G., 2014. The Prominence of Stationarity in Time Series Forecasting. Studies in Economics and Econometrics, 38(1), pp.1-16.

Weiming, J., 2015. Mastering Python for Finance. Birmingham, UK: Packt Publishing.

Yilmaz, Y., & Buyrukoglu, S. (2021). Hybrid Machine Learning Model Coupled with School Closure For Forecasting COVID-19 Cases in the Most Affected Countries. Hittite Journal Of Science &Amp; Engineering, 8(2), 123-131.