Istanbul Dam Water Levels Forecasting Using ARIMA Models

Judi Sekban
MIS Department
Kadir Has University
Istanbul, Turkey
20171102007@stu.edu.khas.tr

Mustafa Omar Mohamed Nabil
Computer Engineering Department
Kadir Has University
Istanbul, Turkey
mustafaomar.nabil@stu.edu.khas.tr

Huseyin Fuat Alsan
Computer Engineering Department
Kadir Has University
Istanbul, Turkey
huseyinfuat.alsan@stu.khas.edu.tr

Taner Arsan
Computer Engineering Department
Kadir Has University
Istanbul, Turkey
arsan@khas.edu.tr

2022 Innovations in Intelligent Systems and Applications Conference (ASYU) | 978-1-6654-8894-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ASYU56188.2022.9925418

***Abstract*— River, sea, reservoir, and dam water levels are constantly measured by organizations and governmental bodies because of their environmental effects as well as their influence on human behavior. In this study, the monthly dam levels in Istanbul, Turkey, were predicted. Different models and configurations were compared to each other, and the best-performing model was identified. The models were based on conventional auto-regressive models (AR), moving average models (MA), auto-regressive moving average (ARMA), and ARMA with Exogenous variables (ARIMAX).**

Keywords—Dams, water levels, forecasting, time series prediction, ARMA, ARIMA models

I. Introduction

Water has several characteristics that make it useful across different applications. Its utility as a source of life is obvious. Energy and power generation from the mechanical motion of water, both from natural and artificial sources, cannot be overlooked. River, sea, reservoir, and dam water levels are constantly measured by organizations and governmental bodies because of their environmental effects as well as their influence on human behavior. Depending on the body of water being measured, the water level metrics predict different effects. For example, tracking the water levels in rising rivers predicts the tendency of flooding in certain localities. Sea level fluctuations affect the rising and falling of the tide and consequently affect naval logistics. Reservoir water levels can assist engineers in predicting the frequency of maintaining its water retaining structures. Studying dam water levels provide information that is used to optimize power generation and efficient water distribution. In order to study and understand the changes in dam water levels, it is important to study the drivers that affect this increase or decrease. There are two factors that influence the rise and fall of dam water levels: (a) Environmental and (b) Human-related factors.

The environmental factors that affect the rise and fall of dam water levels comprise meteorological and hydro-geomorphological effects such as topography, altitude, precipitation, wind speed, humidity, and rainfall levels, among others. The human-related factor mainly stems from the human consumption of dammed water resources. Studying the relationship between environmental and human-related factors and the dam water levels allows us to build data-driven models that are able to predict changes in the water levels. Another approach to building these prediction models is to use historical measurements of dam water levels to develop time-series prediction models. Hence, dam water level prediction can be modeled through: (a) time-series analysis and (b) statistical machine-learning techniques or (c) hybrid approaches incorporating both. In addition, recent literature also emphasizes the value of pre-processing data methods such as wavelet transform in improving the quality of data used along with the models so as to improve prediction confidence, model accuracy, and predictive capacity.

This study aims at examining the dam water levels in Istanbul and the drivers that might contribute to the changes that consistently occur in the water levels. In this work, a case study is provided not only to describe but to facilitate the understanding of the properties and modeling processes of dam water level time series data. Therefore, the research questions (RQ) are presented as follows:

- **RQ1:** What are the specifications of the best-performing model?
- **RQ2:** Are the external exogenous variables significant in the forecasting process?

II. Literature Review

The rapid advancement of computing and distributed systems brought with it an increased performance in time-series predictive and forecasting models. The literature reveals two approaches to water level predictions and forecasting: (a) ARIMA-based models, (b) ML-based models, and (c) a hybrid of both. Hybrid models combine ARIMA (or its variants) with machine learning models such as artificial neural network (ANN), support vector regression (SVR), and random forest (RF). ARIMA was mostly used for univariate forecast and prediction. There were hardly any cases in which a multi-variate ARIMA was applied. Even then, the exogenous predictors were not used to define any parameters in the model. For instance, [1] precipitation, surface runoff, and evapotranspiration in order to predict groundwater levels. However, the training was applied to each variable individually, after which it was aggregated in order to generate a prediction. Their approach resulted in a sufficiently good model as a result of a high predictor correlation with the target variable. [2] approached the problem of groundwater level forecasting by applying several variants of the ARIMA model (AR, MA, ARMA, ARIMA, SARIMA). The authors relied on a systematically robust clustering approach to run their comparative study. In order to define the parameters of their models, they used common systematic statistical tests such as ACF and PACF plots. The study concluded that the AR (with lag 2) model showed the best performance.

The trend of ML-based models for prediction and forecasting definitely extended to this application. For

978-1-6654-8894-5/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: UNIVERSIDAD DE GUADALAJARA. Downloaded on February 21,2024 at 19:14:14 UTC from IEEE Xplore. Restrictions apply.

instance, [3] compared the use of ARIMA/ARMA models with a static and dynamic autoregressive ANNs. Results suggested that ARIMA was useful in forecasting short-term time frames while the dynamic autoregressive ANN proved superior for timeframes forecasting much larger timeframes (5 years). [4] forecasted daily reservoir water levels by decomposing signal measurements and feeding them into an ANN as well as a neuro-fuzzy model. Although, the study concluded the potential improvement of applying wavelet decomposition, the authors haven't addressed whether or not the ML models selected could have contributed to the results. The authors in [5] compare between the use of Genetic Programming (GP) and ANN in predicting the sea water levels along the coast of Australia. The study does not indicate the overall superior performance of any of the models that were compared. Rather, the results show that depending on the measurements, one model may perform better than the other, suggesting that a "one-size-fits-all" approach to water-level forecasting may not be useful. [6] studies the use of dynamic ANNs and its variants to predict inundation levels of water for flood-prevention. Rainfall, and floodwater storage pond levels were used as predictors. Results showed that the models are useful in predicting inundation levels. [7] fed over thirty years of water level data from sixty-nine lakes in Poland into a Feed-forward Neural Network (FFNN) and Deep Learning (DL) model to predict the next time-series of monthly lake water levels. Findings suggested that the (FFNN) and the DL models performed sufficiently well for predicting the lake water levels, with only insignificant differences Other researches such as [8], [9], and [10] also extensively study ML models in forecasting water levels, strongly suggesting its significance in this application.

As suggested by [5], the approach toward a combination of different models seems to be a step in the right direction; what may be defined as hybrid models in this study. More specifically, a combination of ARIMA and ML-based models in water level prediction has produced substantially performant models. [11] developed a model that predicts the river water levels based on historical time-series univariate measurements. The authors trained and tested a hybrid model incorporating ARIMA and other non-parametric ML methods. Residuals from the ARIMA model coupled with real data was used to train the ML model. Pairing ARIMA with multiple ML models, served to capture both linear and non-linear components present in the data. Findings suggested that the combinations of ARIMA-Random Forest surpassed other model combinations in terms of reliability. A recent study, [12], combines SARIMA (seasonal ARIMA) with ANN in an attempt to establish any marginal improvement in performance compared to using only SARIMA. Results from the study showed that SARIMA-ANN model displays better performance in forecasting the reservoir water levels. Similarly, [13] combines Recurrent Neural Network (RNN), and ARIMA to produce an ARIMA-RNN model that predicts water levels. Results from the study also indicate a performant model for forecasting.

It is observed from the literature that despite serious efforts in studying the performance of hybrid models, not much has been invested in improving ARIMA model predictions through better methodological approaches. More specifically, the use of hyper-parameter space search in ARIMA for water-level forecasting has not been explored. Furthermore, the use of external predictors in conjunction with ARIMA (ARIMAX) is relatively scarce in the literature.

III. Study Implementation Phases

In time series forecasting problems, future values of timestamped data (usually a continuous variable) are predicted based on historical data (past values of the same variable). Through conducting time series forecasting, it is not always the case that the output is an exact prediction for the target variable. Instead, it can simply provide the general trend (direction) of the data at a future time. The most popular example of such problems is forecasting stock prices for the current month based on prices of the previous months. In a similar manner, this study attempts to forecast the future dam water levels based on measured values of the past. A number of limitations surface when dealing with any unknown or unpredictable variables, and forecasting the unknown time series is no exception. This is the reason why forecasting is generally practiced in conjunction with an analysis process that involves statistics and modelling to gain a prior understanding of the data in order to inform strategic decision-making. Forecasting problems can be approached utilizing a common process pipeline used in pure machine learning-based problems. Fig. 1. below shows the pipeline which will be further expanded in the following sections within the context of attempting to answer the aforementioned research question. Each item of the pipeline is a phase or a stage comprising several activities, and is not simply a single step.

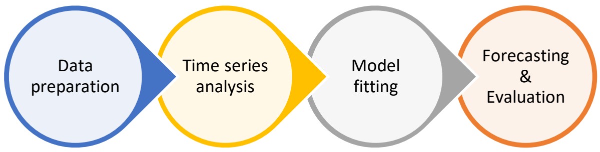

Figure 1 Flow chart that represents the Study Implementation Phases

IV. Data Sources

Five raw datasets are loaded from the Istanbul municipality open data repository: General Dam Occupancy Rate (DS-1), Total Rainfall on Istanbul Dams (DS-2), Monthly Water Consumption (DS-3), Istanbul Dams Daily Occupancy Rate (DS-4), and Meteorological Stations Data (DS-5). These datasets are initially loaded to be inspected and assessed in terms of quality and usefulness. Consequently, appropriate datasets that are suitable for use in this study are selected. The quality and usefulness of the datasets are measured in terms of the time availability of the data, and their theoretical relevance in predicting dam water levels. Brief summaries of each of the datasets are further presented:

A. *General Dam Occupancy Rate (DS-1)*

This dataset includes the daily changes in the occupancy rates of Istanbul's dams. There are just under 6,000 records collected across 16 years, from January 2005 up until April 2021. Each data record has an identification key, a date timestamp, general dam occupancy rate (GDOR), and the corresponding dam reserved water rate (DRWR). The measurements for GDOR are taken as a percentage of Istanbul's dam capacities. It is assumed that the values of

Authorized licensed use limited to: UNIVERSIDAD DE GUADALAJARA. Downloaded on February 21,2024 at 19:14:14 UTC from IEEE Xplore. Restrictions apply.

GDOR and DRWR are the average aggregate for all the dams present in Istanbul.

B. *Total Rainfall on Istanbul Dams (DS-2)*

The dataset includes the daily total precipitation amount measured in kilogram per square meter for ten dams in Istanbul. It is assumed that the data for GDOR and DRWR in DS-1 are for the same Istanbul dams present in this dataset. Unlike DS-1, however, the timespan for the records starts from January 2011 up to December 2020. The features for each observation in this dataset also include a date timestamp and an identification key column.

C. *Water Consumption (DS-3)*

The dataset includes the total monthly water consumption data of the Istanbul Water and Sewerage Administration throughout Istanbul. The water consumed is measured in the volumetric measurement of a cubed meter. It is assumed that the measure for water consumption is based on water consumed from the dams in Istanbul. The records in this dataset are collected from January 2010 to December 2020. The timestamp for the records, unlike the previous two (DS-1 and DS-2), is taken on a monthly basis. Hence, the dataset contains drastically fewer data points. In addition, there are only two features in this dataset: the timestamp and the water consumption measurement.

D. *Istanbul Dams Daily Occupancy Rates (DS-4)*

This dataset shows the daily dam occupancy rates of 10 dams in Istanbul for the last ten years, starting from January 2011 to December 2021. The data in this dataset is very similar to that in DS-1. It also contains the daily occupancy rate for dams in Istanbul, however, disaggregated. Hence, each record contains the occupancy in percentage for each of the ten dams in Istanbul. Similar to DS-2, there are 12 features per record, whereby 10 are the occupancy rate for the dams, and the other two comprise the identification key and the timestamp.

E. *Meteorological Station Data (DS-5)*

The data set includes meteorology data such as temperature, humidity, and wind speed. In addition, the data for each meteorological observation is recorded based on the observatory and is taken on an hourly basis. Both the observations and the features in this dataset are many and detailed. Hence, the dimensions of this dataset are extremely large (the largest) compared to the others. However, the time period in which the data was collected is extremely small compared to the other datasets under consideration: meteorological data for under two years was recorded, from January 2020 to April 2021.

V. Data Preparation

Raw datasets that are used for this study undergo several processes (data wrangling) to transform them into a format that is useful and usable for training and testing the model. Fig. 2. below describes the pipeline followed in the data preparation phase.

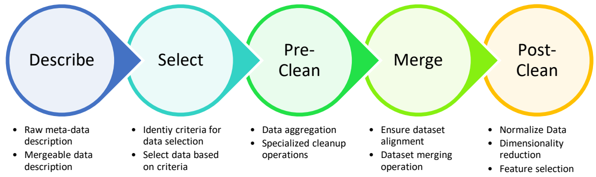

Figure 2 Flow chart of Data preparation pipeline

A. *Data selection*

The criteria for selecting the right datasets were based on relevance and availability. All datasets were picked to be relevant based on the literature review. The more difficult aspect of this selection was the availability of the data. The individual time-series datasets did not span the entire time period. More importantly, the datasets did not all come from the same time period (Fig. 3 below demonstrates the availability of each dataset). The main aim was to select the time period which contained the most data points. Ultimately, the datasets that satisfied this condition were DS-1, DS-2, and DS-3; other datasets were not selected for developing our models.

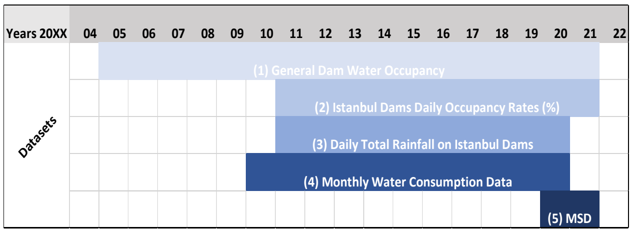

Figure 3: Gantt Chart of Data availability based on the years from 2004 until 2022 timeline.

B. *Data Pre-Cleaning*

The followed approach consisted of applying low-level operations on the datasets, followed by high-level. CSVed is used to perform low-level operations such as the deletion of non-relevant columns/features, character replacement for number formats standardizing, data-labels relabeling, and records sorting. For high-level operations of data wrangling, the Pandas and NumPy libraries were employed. These high-level operations encompass data transformation methods such as data aggregation, re-sampling, normalization, and dimensionality reduction. The objective is to prepare all datasets such that they are mergeable.

C. *Data Merging*

All datasets were merged into one for ease of representation and modeling. This stage unifies all data into a single structure and prepares the dataset for an exploratory analysis before the train/test split. An index series of timestamps was created from the dataset with the widest time period. Afterward, datasets were merged, convoluting each dataset upon the timestamp index.

D. *Data Post-Cleaning*

In order to improve the overall quality of the models, another round of data cleaning was implemented. Duplicated records and null values were all checked for. Visualizing the data was essential to inspect for any anomalies and to also understand the trend or observe any distinct features.

Authorized licensed use limited to: UNIVERSIDAD DE GUADALAJARA. Downloaded on February 21,2024 at 19:14:14 UTC from IEEE Xplore. Restrictions apply.

VI. ANALYZING TIME-SERIES DATA

Before doing any modeling, data understanding is required. Visually inspecting time series is essential and is the first step into a more detailed process of series analysis. Through visualization, components of the time series can be extracted to make useful insights. The most important topic for studying is the ***stationarity*** of data, which is a property of time series data. A time series is said to be stationary when values of the series are not functions of time, independent from time. In this case, the statistical properties, such as the autocorrelation, mean, and variance, are constant over time. The first step into time series forecasting is to transform the series into stationary series since most statistical forecasting methods assume the series is stationary. The stationarity of the data in this study is investigated through a) plots of the series; b) ACF and PACF plots behavior; and c) statistical tests.

A. Series plot

Fig. 4. shows the plot of the time series, where the following points are observed about the data:

-No trend component is present since no increasing or decreasing slope is observed in the series.

-Visually inspecting the series, a repeated pattern (seasonality) can be observed, where GDOR levels tend to increase in the early months of the year and decrease in the latest months. This is a repeated pattern over the years, except for the year 2014, where instead of a peak-bottom pattern, an opposite pattern of the bottom-peak pattern is observed.

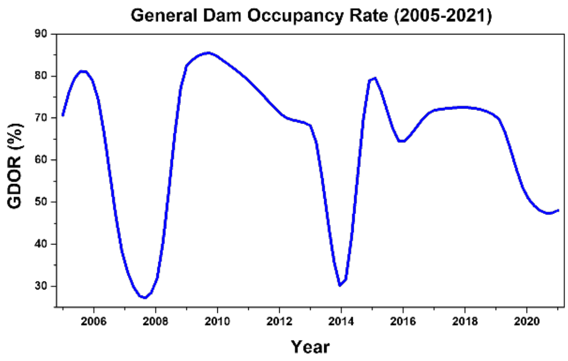

Figure 4 Plot of GDOR against YEAR

However, this pattern can be recognized rather as a ***cyclical behavior*** considering the irregularity of the peaks. While seasonality has strict regular changes over a given period, meaning that there is a precise amount of time between peaks in the series, the cyclical behavior is rather less strict, which can be observed in Fig. 5. Additionally, the rise and fall pattern might not be happening in fixed calendar-based intervals but instead is affected by other factors. The highest level recorded during the year, for example, may move (because of factors other than the time that influence the data). Cycles are not predictable, and thus they do not violate the assumption that for time-series data to be stationary, it should not contain any predictable patterns/components.

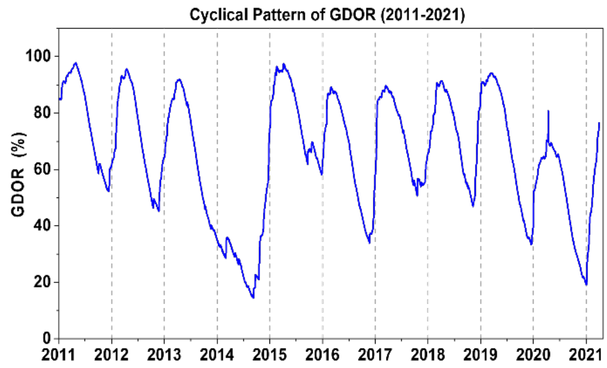

Figure 5 Graph of GDOR cyclical pattern over 10 years.

B. ACF and PACF plots

The recommended approach to investigate the time series is not to rely only on visual inspection of time-series data since it can be misleading or insufficient to make decisions about the data. Hence next step is to utilize the autocorrelation function (ACF) and partial autocorrelation function (PACF) graphs/plots to further investigate the stationarity issue. Another reason behind the need to interpret the graphs produced by the ACF and PACF is that based on those, the order of autoregressive and moving average terms, AR and MA, can be determined, which is the first step in the modeling phase (discussed in the following section). Here, the ACF and PACF are calculated not only on the levels of the series but also on the 1st and 2nd difference from the series to study the effect.

Non-random data have at least one significant lag that can be observed in all of the autocorrelation plots. When the data is not random, it's a good indication to incorporate lags into a regression analysis to model the data appropriately. Fig. 5. shows the ACF and PACF of the data, and they both show a gradual decreasing pattern, which is an indication that an ARMA model should be considered for modeling, which will be discussed in the following section. From the ACF, it is observed that the autocorrelation coefficients fade out more quickly in the 1st differenced series. Another piece of evidence from the PACF that might support the theory that the repeated patterns observed in the series are rather of a cyclical nature is that no spikes are observed at the 12th lag (at the multiplications of 12 as well), which should have been the case as the data is monthly. It is observed that the PACF plots for the original series and for the 1st differencing of the series are similar to each other. According to both of the plots, the first lag is most significant, a fact that will be used in the modeling process.

C. Stationarity statistical test

In some cases, the stationarity of the data in not determined using the previous two methods where in such a case a statistical test can be run to decide on the data. The Augmented Dickey Fuller (ADF) method, which is one of the most popular methods for stationarity detection, tests the null hypothesis that a unit root is present in a time series sample. Fig. 6. below shows the output of running the test on the series where the p-value is close to 0 and the ADF statistic is below the 1% critical value. Hence, the null hypothesis that the series has a unit root and is not stationary is rejected. The series is therefore stationary.

Authorized licensed use limited to: UNIVERSIDAD DE GUADALAJARA. Downloaded on February 21,2024 at 19:14:14 UTC from IEEE Xplore. Restrictions apply.

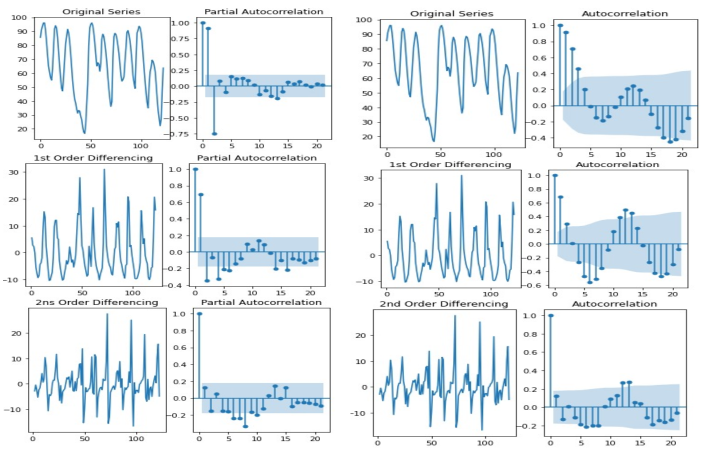

Figure 6 ACF & PACF plots

Table 1: ADF statistics highlight that the series is stationary.

ADF Statistics		-6.449542
P-Value		0.000000
Critical Values:	1%	-3.486
	5%	-2.886
	10%	-2.580

VII. Model fitting

ARIMA family of models was utilized in this study for the forecasting job. There are two types of these models, ARMA and ARIMA. The ARMA model stands for Autoregressive Moving Average, and it explains a given time series based on its previous values. While ARMA models work for stationary data, ARIMA is suitable for a non-stationary that can be transformed to be stationary by differencing the series. ARIMA models are used to further study the effect of differencing the series. ARIMA models are specified by three parameters: p; the order of the AR term, which is basically the number of lags of the predicted feature to be used as predictors; the d term; to determine how many differences are needed; and q; the order of the MA term, which refers to the number of lagged forecast errors that should go into the ARIMA Model. The ACF and PACF plots highlighted in the previous section were used in the modeling phase to determine the order of AR and MA terms.

The steps outlined below show the pipeline that starts from investigating the series until the evaluation of 5 models:

1-Investigate the stationary of the data using the three methods mentioned in a previous section. In cases where it is not straightforward to decide on the stationarity of the data, one approach is to build models both on the levels and the first difference of the series.
2-Transform the data to be stationary, if necessary.
3-Investigate the ACF and PACF plots; according to the spikes in these plots, a decision can be taken on whether to build the model with AR terms only or MA terms only or to combine the two terms together.
4-Build different variations of the models.
5-Compare the models' performance on the training set.
6-Select the best two models based on step 5.
7-Make a prediction on the testing set and evaluate the performance of the two final models.

Any autocorrelation in a stationary series can be rectified by adding enough AR terms. Thus, if the value of the PACF of any particular month is more than a significant value, only those values will be considered for the model analysis. Hence, initially, the order of the AR term is set to be equal to as many lags that crosses the significance limit in the PACF plot, where it is observed that the 1st lag clearly crosses the limit and the 2nd lag is very close to the limit. So tentatively, Model #1, the AR term (p) = 2. This is a pure autoregressive model (no moving average term). For Model #2, the same approach was followed to determine the MA term based on those lags that cross the significance limit, MA term (q) = 2. This is a pure moving average model. Model #3, the two terms are combined at lower orders (p = 1, q=1). Model #4, The automated ARIMA function (auto_arima), which is created to find the optimal order and the optimal seasonal order (if applicable) based on determining criteria such as AIC, BIC. The optimal parameters turned out to be AR = 4 and MA = 5. Model #5 will include the two external predictors (exogenous variables) in addition to two terms; ar = 1 and ma = 1.

VIII. Results

This section highlights the result of five models that were measured in terms of performance metrics, residuals, and fit of models to the training data. Before running the forecasting on the testing set, an initial analysis of the performance of the models on the training set was conducted. The criteria for the performance of the models on the training set were based on the value of R-square, AIC, and RMSE, in addition to a discussion about the significance of the terms/exiguous variables included in each model.

When the models are compared to each other based on the performance of training data, the best two performers were models #4 and #1. R-square in the context of time series describes how well the model fits past data. Based on that, model #4, then model #1, fitted past values best compared to the other models. Moreover, the same two models recorded the lowest RMSE values. AIC, on the other hand, is usually used to estimate models' prediction power for a given dataset. For AIC, the lowers the score, the better, which is an indication of the best balance of model fit with generalizability. Again, and according to the AIC, models #4 and #1 were the best performers. From Table 1, where all metrics of the models are recorded, it is observed that model #5, which included the exogenous variables, was the worst performer on all metrics recorded, both on training and testing sets. There could be more than one reason, such as the lags of the same series being the most important predictor for future forecasting. Another reason that these exogenous variables could be just not the correct external predictors, particularly the water consumption dataset.

Finally, residual analysis shows the residual distribution for models 1, 2, and 3 to be very similar. Model 5 slightly deviates, but model 4 deviates the most compared to the other models. At the same time, as observed, model 4 is the best performing model. This pattern highlights that the GDOR data for the time period in which the model was

Authorized licensed use limited to: UNIVERSIDAD DE GUADALAJARA. Downloaded on February 21,2024 at 19:14:14 UTC from IEEE Xplore. Restrictions apply.

trained comprises random components that cannot be interpreted in the model. Both Table 1 and Fig. 8 summarize the details presented in this section.

Based on the filtering process for the best performer models on the training set, the forecasting performance on the testing set can be compared for the two models. Model #4 outperformed the other candidate, model 1, with a higher r-square score, lower RMSE, and MAPE

IX. Conclusion

In this study, the drivers that affect the water level in Istanbul dams were identified. Datasets of the general dam water levels, water consumption, and rainfall on the dams were used in the development of five variants of the ARIMA model. The five models were designed to forecast the water levels of the dams in Istanbul. According to the results discussed in the previous section, the model which was able to best predict and captures the pattern in the series is the model with an autoregressive term of order = 4 and moving average term = 5. This observation answers the research question about the parameter's specification of the best-performing model. On the other hand, it also can be concluded that the exogenous variables included in this study did not have a positive impact on the forecasting and predictive capabilities as observed with model 5; they were not able to enhance the ARMA model. Should the effect of external predictors be studied on the same dataset, another cycle of finding the candidate datasets to be integrated with the time series need to be conducted. Moreover, the approach followed for the integration of the datasets can be reviewed for further enhancement. The forecasting capacity of ARIMA and its variations have been studied, paving new approaches to the development of time-series models for water level prediction and forecasting.

X. References

[1] M. Birylo, Z. Rzepecka, J. Kuczynska-Siehien, and J. Nastula, "Analysis of water budget prediction accuracy using ARIMA models," *Water Sci. Technol. Water Supply*, vol. 18, no. 3, pp. 819–830, 2018, doi: 10.2166/ws.2017.156.

[2] M. Mirzavand and R. Ghazavi, "A Stochastic Modelling Technique for Groundwater Level Forecasting in an Arid Environment Using Time Series Methods," *Water Resour. Manag.*, vol. 29, no. 4, pp. 1315–1328, 2015, doi: 10.1007/s11269-014-0875-9.

[3] M. Valipour, M. E. Banihabib, and S. M. R. Behbahani, "Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir," *J. Hydrol.*, vol. 476, pp. 433–441, 2013, doi: 10.1016/j.jhydrol.2012.11.017.

[4] Y. Seo, S. Kim, O. Kisi, and V. P. Singh, "Daily water level forecasting using wavelet decomposition and artificial intelligence techniques," *J. Hydrol.*, vol. 520, pp. 224–243, 2015, doi: 10.1016/j.jhydrol.2014.11.050.

[5] M. Ali Ghorbani, R. Khatibi, A. Aytek, O. Makarynskyy, and J. Shiri, "Sea water level forecasting using genetic programming and comparing the performance with Artificial Neural Networks," *Comput. Geosci.*, vol. 36, no. 5, pp. 620–627, 2010, doi: 10.1016/j.cageo.2009.09.014.

[6] F. J. Chang, P. A. Chen, Y. R. Lu, E. Huang, and K. Y. Chang, "Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control," *J. Hydrol.*, vol. 517, pp. 836–846, 2014, doi: 10.1016/j.jhydrol.2014.06.013.

[7] S. Zhu, B. Hrnjica, M. Ptak, A. Choiński, and B. Sivakumar, "Forecasting of water level in multiple temperate lakes using machine learning models," *J. Hydrol.*, vol. 585, no. February, p. 124819, 2020, doi: 10.1016/j.jhydrol.2020.124819.

[8] P. P. Hadiyan, R. Moeini, and E. Ehsanzadeh, "Application of static and dynamic artificial neural networks for forecasting inflow discharges, case study: Sefidroud Dam reservoir," *Sustain. Comput. Informatics Syst.*, vol. 27, p. 100401, 2020, doi: 10.1016/j.suscom.2020.100401.

[9] D. Zhang *et al.*, "Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm," *J. Hydrol.*, vol. 565, no. August, pp. 720–736, 2018, doi: 10.1016/j.jhydrol.2018.08.050.

[10] J. H. Wang, G. F. Lin, M. J. Chang, I. H. Huang, and Y. R. Chen, "Real-Time Water-Level Forecasting Using Dilated Causal Convolutional Neural Networks," *Water Resour. Manag.*, vol. 33, no. 11, pp. 3759–3780, 2019, doi: 10.1007/s11269-019-02342-4.

[11] T. T. H. Phan and X. H. Nguyen, "Combining statistical machine learning models with ARIMA for water level forecasting: The case of the Red river," *Advances in Water Resources*, vol. 142, p. 103656, Aug. 2020, doi: 10.1016/j.advwatres.2020.103656.

[12] A. S. Azad *et al.*, "Water Level Prediction through Hybrid SARIMA and ANN Models Based on Time Series Analysis: Red Hills Reservoir Case Study," *Sustain.*, vol. 14, no. 3, 2022, doi: 10.3390/su14031843.

[13] G. Xu, Y. Cheng, F. Liu, P. Ping, and J. Sun, "A water level prediction model based on ARIMA-RNN," *Proc. - 5th IEEE Int. Conf. Big Data Serv. Appl. BigDataService 2019, Work. Big Data Water Resour. Environ. Hydraul. Eng. Work. Medical, Heal. Using Big Data Technol.*, pp. 221–226, 2019, doi: 10.1109/BigDataService.2019.000 38.

Authorized licensed use limited to: UNIVERSIDAD DE GUADALAJARA. Downloaded on February 21,2024 at 19:14:14 UTC from IEEE Xplore. Restrictions apply.

Table 2: Comparative description table for all models. [“AVG-MN”: Average Mean, “STD”: Standard Deviation, “R-SQ”: R-Sqaure, “AIC”: Akaike information criterion, , “RMSE”: Root mean square error, “MAPE”: Mean Absolute Percentage Error]

	Residuals		Model Fit (Training)			Forecast Performance		
Models	**AVG-MN**	**STD**	**R-SQ**	**RMSE**	**AIC**	**R-SQ**	**RMSE**	**MAPE**
1	0.151108	5.772358	92.7	5.745	628.605	92.5	5.157	0.093
2	0.098001	8.155090	85.4	8.114	701.538	91.2	5.600	0.112
3	0.004173	6.414579	90.9	6.382	653.512	90.0	4.964	0.113
4	1.143798	5.123453	94.2	5.100	618.910	94.0	4.634	0.084
5	0.187143	9.843211	78.7	9.795	745.114	81.2	8.183	0.157

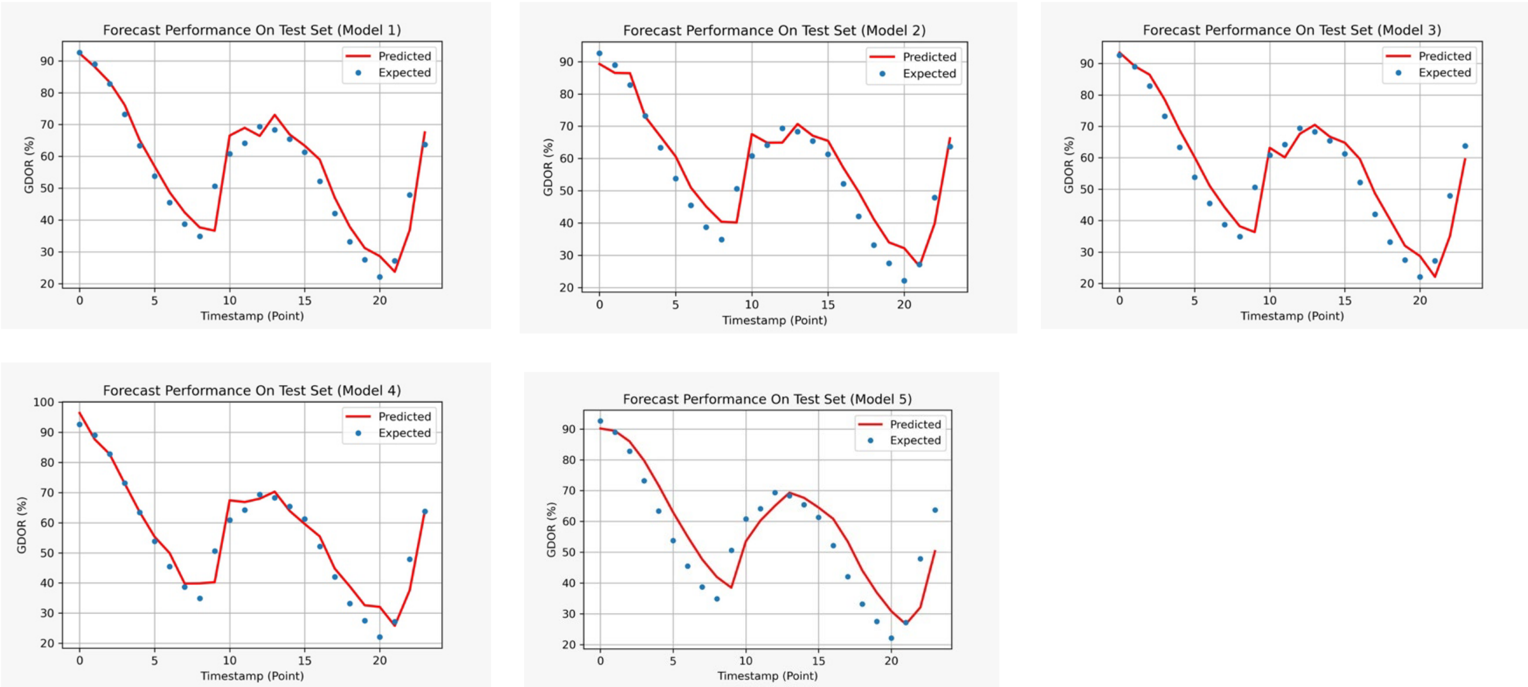

Figure 7: The plots above highlight the forecast performance tests executed for each model. Model 1 (top-left), a purely auto-regressive model with order=1. Model 2 (top-middle) is a purely moving -average model with order=1. Model 3 (top-right) is based on both model 1and 2 orders ARIMA (1,0,1). Model 4 (bottom-left) is the ARIMA (4,0 ,5) selected based on the hyper-parameter tuning. Model 5 (bottom-right) is similar to model 3 with the same order ARIMA (1,0,1) except that it is also trained and forecasted using exogenous predictors of RID and WC.

Authorized licensed use limited to: UNIVERSIDAD DE GUADALAJARA. Downloaded on February 21,2024 at 19:14:14 UTC from IEEE Xplore. Restrictions apply.