

Medium-term Forecasting for City Water Demand and Revenue

Hosam H. A. Mukhairez, Alaa M. Al-Halees

Faculty of Information Technology Islamic University of Gaza
Palestine

Abstract

The combination of forecasting and planning has increased prominence within few decades and now receives considerable attention from both academics and practitioners. Palestine municipalities are facing complex circumstances because of the climate change (lack of rainfall) and due to the Israeli occupation. Water resources are decreasing by the passage of time. Therefore, it is necessary to focus on water service projects for best exploit of resources given the growth of population. In this research, we forecasted the consumption used by citizen in the medium-term future and forecasted revenues from water service. We conducted the forecasting using the following forecasting algorithms: Auto-Regressive Integrated Moving Average (ARIMA), Hybrid ARIMA, Singular Spectrum Analysis (SSA), and Linear Regression. We applied them on dataset collected from Khan Younis municipality (KHM) department of customer services. We found that the best algorithm is Hybrid ARIMA which gave the Mean Absolute Percentage Error (MAPE) of 17.38%. Finally, we forecasted the whole city of Khan Younis city. Generally, we found that the maximum water consumption for the overall city after five years will increase to about 8.4% compared to 2017, but the minimum water revenue will decrease to about 3.8% compared to 2017.

1. Introduction

Many countries are facing the real risk in lack of water resources because of the changing in the earth climate with the passage of time. Water resources in Palestine are limited according to low groundwater level due to lack of rainfall, increasing of population, also due to the Israeli occupation which prevents the Palestinian authorities to use many of the ground wells, and the unorganized drilled wells by citizens to bring up the underground water.

Many water service projects (for example water desalination and drilling of underground wells) were built by local and international organizations in many regions in Gaza strip and West Bank to support water service to citizens, but these projects aren't

established according to preconceived data belongs to the increasing of population, current per-capita water consumption and other factors may affect the process of distributing water to people.

Forecasting task is analyzing the past behavior of a variable to predict its future behavior or trend. It is one of the oldest known predictive analytics techniques. The idea is to use historical data to make forecasts and prediction about future data. Forecasting is one of the most common techniques of time series data analysis. It's used to predict future trends in water demands, retail sales, economic indicators, weather forecasting, stock markets, and many other application scenarios [1].

The combination of forecasting and planning has increased to prominence within a few decades and now receives considerable attention from both academics and practitioners. By operating forecasting tasks we'll be able to get more accurate results with less usage of resource which means save available resource for future by supporting effective decision making and for water researchers seeking to extend the current knowledge in the field. Moreover, to properly plan, it's important to be able to forecast your resource availability. In the past, resource forecasting relied on probability techniques. While metrics can help contribute to the process, prioritizing projects through simple Return of Investment (ROI) scoring can be a gamble in agency life [2].

Water demands represent the driver of water distribution systems, and reliable demand forecasting represents a valid aid in simulating and managing such systems. Depending on the forecasting time horizon term, water demand forecasts can be classified, as suggested by [3] into: (1) Long-Term which forecasts decades, (2) Medium-Term which forecasts years to decade, (3) Short-Term which forecasts up to year.

Khan Younis municipality as a case study of our research, is well prepared municipality technically and graphically and has a large dataset belongs to all its services on the city. Daily and weekly there are hundreds and thousands of records and data are collected from all distributed station (for example water distribution pumps) belongs to Khan Younis

municipality especially to water service. So, these data scientifically should be used in planning for water demand in the future.

In our research, we are going to collect some related data to water demands, consumptions and revenues used in Khan Younis municipality applications Medium-term Forecasting for City Water Demand and Revenue to prepare it for the forecasting algorithms to present useful results and show it to managers and water service specialists depending on the estimation of water consumption and new expected water revenues. In this work we are going to use some forecasting algorithms and choose the best one depending on the mean percentage error ‘MPE’.

The rest of this paper is organized as follows: Section 2 reviews some related works. Section 3 shows the methodology of our research. Section 4 shows experimental process and results. Section 5 presents discussion. Section 6 the conclusion. Finally, section 7 presents future work.

2. Related Work

Many authors researched on water demand forecasting and its effects on urban water systems design, for example Donkor et. al in [4] reviewed most of water demand forecasting papers published between 2000 and 2010 to identify the useful methods and model for specific water utility decision making process.

Ajbar & Ali in [5] developed a neural network model for forecasting the monthly and annual water demand for Mecca city, Saudi Arabia. They used historical data of water production and estimated visitors' distribution to calibrate a neural network model for water demand forecast. They observed that NN prediction outperforms that of a regular econometric model. They developed two models for the forecast of monthly and annual water demand in the city. They designed the model to combine the impact of monthly-varying and yearly-varying explanatory variables. The neural network model was found to be superior to the econometric model in the sense of capturing the water consumption dynamics both in the short and long terms. Also, the NN model was useful in investigating the influence of the changes occur in the number of visitors specially in AL Haj season on the overall water demand of Mecca city.

RISH in [6] forecasted the electricity consumption in the Khan Younis province from 2000 to 2010 using exponential smoothing and Box-Jenkins ARIMA methods. She compared between these two models to detect which model is more accurate to forecast electricity consumption of Khan Younis province. She concluded that exponential smoothing has MAPE 5.92% while Box-Jenkins has MAPE 6.24%.

Rinaudo in [7] wrote a chapter that reviewed long term water demand forecasting methodologies. The

literature review is enriched by economists, engineers and system modelers. The author illustrated how different tools can be used depending on the regulatory context, the water scarcity level, the geographic scale at which they are deployed and the technical background of water utilities and their consultants. His chapter represented three main challenges, the integration of land use planning with demand forecasting, accounting for climate change, and dealing with forecast uncertainty.

Liu et. al in [8] also used data mining techniques to identifying determinants of urban water. Their approach is based on a genetic programming (GP) data mining technique, which automatically optimizes the structure of the function and parameters simultaneously. With historical urban water use as the target, the GP model identified the most relevant factors for 47 cities in northern China. They assume that housing area, water price and rainfall are affecting the forecasting model. They claim that the new model in their study could be helpful for urban water management, especially for cities that experience water scarcity.

3. Material and Methods

The four selected forecasting algorithms ARIMA, Hybrid ARIMA, SSA and Linear Regression are the most famous and recent algorithms in this field. We used four tools to build our model and test the selected algorithms over, these tools are RapidMiner Studio, R project, Anaconda and WEKA.

3.1. ARIMA

An ARIMA model typically consists of three parts (1) auto regression AR (order ‘p’ the number of terms used in the auto-regression), (2) moving average MA (order ‘q’ the number of terms used in the moving average) and (3) differencing to strip off the integration of the series (order ‘d’ the number of differencing terms) and then form ARIMA (p, d, q). Some of the limitation of ARIMA is, it handles time series data in linear form and must be stationary and the accuracy of ARIMA forecasting model is significantly affected by the noise of time-series dataset. We used RapidMiner “Arima Trainer” Operator with values (p=1, d=0, q=1).

3.2. Hybrid ARIMA

In this evaluation task we combined ARIMA with Neural Network (NN) to have Hybrid ARIMA with equal weight for each. While ARIMA may not be adequate with complex nonlinear problems and on the other hand NN yielded mixed results with linear problems. So, the combination handles effectively vary types of dataset nature.

The motivation of the hybrid ARIMA comes from the following perspectives [9] (1), it is often difficult in practice to determine whether water consumption series is linear or nonlinear nature or whether such method is more accurate than the other in forecasting. Also, many potential factors such as sampling variation, model uncertainty, and structure change may affect in choosing the best model for future use. (2), real-world time series contains both nonlinear and linear patterns are rarely found linear or nonlinear. If this is the case, then neither ARIMA nor ANNs can be generalized the best method in modeling and forecasting time series since ARIMA model cannot deal with nonlinear patterns while the NN model alone is not able to handle both linear and nonlinear patterns equally well. Hence, by combining ARIMA with ANN models, complex autocorrelation structures in the data can be modeled more accurately. (3), it is almost universally known that no such algorithm is best in every situation. For example, in the literature of time series forecasting with neural networks, most studies use the ARIMA models as the benchmark to test the effectiveness of the ANN model with mixed results.

3.3. SSA

Singular Spectrum Analysis (SSA) is one of the modern non-parametric method for the analysis of time series. It provides a set of fast and reliable implementations of various routines to perform forecasting, decomposition and reconstruction.

The birth of SSA is usually associated with the publication of papers by Broomhead [10] while the ideas of SSA were independently developed in Russia (St. Petersburg, Moscow) and in several groups in the UK and USA. At present, the papers dealing with the methodological aspects and the applications of SSA number several hundred.

SSA is a very useful tool which can be used for solving the following problems [11]: (1) finding trends of different resolution. (2) smoothing. (3) extraction of seasonality components. (4) simultaneous extraction of cycles with small and large periods. (5) extraction of periodicities with varying amplitudes. (6) simultaneous extraction of complex trends and periodicities. (7) finding structure in short time series. (8) change-point detection.

3.4. Linear Regression

Linear regression is a statistical method to model the relationship between two variables by fitting a linear equation to observed data. One variable is an explanatory variable, and the other is a dependent variable. Linear regression method can be used for forecasting under the assumption of continuing the correlation between the variables in the future. A linear regression model is as follows:

$$Y_i = \alpha + \beta x_i$$

where α and β are coefficients. The most common use of regression in business is to predict events that have yet to occur. Demand analysis, for example, predicts how many unit's consumers will purchase. Many other key parameters other than demand are dependent variables in regression models, however. Predicting the number of shoppers who will pass in front of a billboard or the number of viewers who will watch the Super Bowl may help management assess what to pay for an advertisement. Insurance companies heavily rely on regression analysis to estimate how many policy holders will be involved in accidents or be victims of burglaries [12].

3.5. Proposed Model

The proposed model is about forecasting water consumption and water revenue within customers and units' historical profiles registered in Khan Younis municipality KHM. The overall model methodology outlined in the thesis is shown in Figure 1.

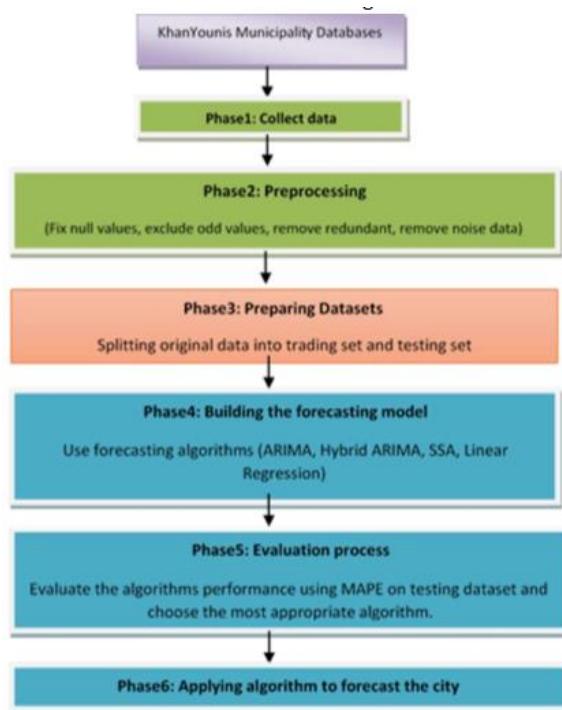


Figure 1. Proposed water consumption and revenue forecasting process

3.5.1. Data Collection At this stage related data was collected from available databases in Khan Younis municipality. At Khan Younis municipality there are many internal systems, and more than such system deals and effects the water service data. So, related tables and column that belong to our case study were selected and combined into one dataset to prepare it for the next phase.

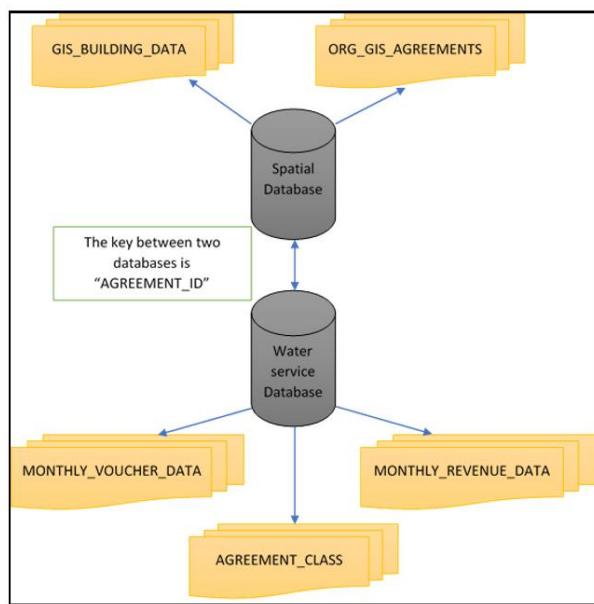


Figure 2. Overview of main databases used in Khan Younis Municipality belongs to water service system

Figure 2 presents a main major database belongs to water service system and how data collection process is done from different datastores. First, ‘Spatial Database’ related to geolocation data for every unit benefited from water service this data includes xcoordinated, y-coordinated, counting zone, building number, neighborhood label and agreement-id which is the main key with the water service database. This database contains one table which registers one record for every existing building (regardless of number of floors) and another table registers all water agreements exists at every house unit at this building.

Second, ‘Water Service Database’ related to monthly transactions of water service operations like reading water meter devices, number of cubes consumed, date of reading, the status of meter device and the class nature of this water consumption unit. This database contains one table which registers one record for every existing water subscription (At Khan Younis city, there are 20949 water subscription) including some related managerial data like owner name, beneficiary name, registration date, class label etc. Many of these data were discarded because it is not important in our forecasting task and does not effects on the water consumption and water revenue numbers, another table which registers all water monthly transactions including the water consumption and water revenue values.

Every dataset consists of four columns (water consumption ‘double’, water revenue ‘double’, voucher date ‘date’, class ‘string’). And to create this dataset for any level, we must read the water consumption and voucher date from the table which registers the monthly amount for every customer or

unit record, and we read the paid water revenue from customers or units from table which registers the monthly payment transaction, and we read where this customer or unit in which sub-area and what is the class labeled for it from tables belongs to GIS system.

3.5.2. Data Preprocessing At this stage preprocessing tasks were applied to prepare collected dataset to the forecasting algorithms, because data preparation is an important step for treating the data to be ready for the analysis and modeling steps. In some cases, some duplicated records for the same month with different values were found, this is due to the reading process from customers and units is done twice at the same month, and because our time series dataset assumes one single record for each month. So, we summed these duplicated records in one record only. For example, AlAmal region has two redundant records for the same month 2/2007, so we summed them into one single record which contains water consumption = 40597 and water revenue = 59442.5

3.5.3. Building Model At this stage we built the model by applying four forecasting algorithms over the monthly water consumption and water revenue for the whole city to choose the appropriate algorithm depending on the lowest mean percentage error MPE. The four selected forecasting algorithms are Auto Regressive Integrated Moving Average (ARIMA), ARIMA combined with Neural Networks (Hybrid ARIMA), Singular Spectrum Analysis (SSA) and Linear Regression forecast. These algorithms are the most famous and recent algorithms in this field. We used four tools to build our model and test the selected algorithms over, these tools are RapidMiner Studio, R project, Anaconda and WEKA.

First step, we divided our original data into two parts, water consumption data and water revenue data as shown in Figure 3, each part is separated into two datasets (training set and testing set). The training set for each starts from 01/2007 to 12/2015 which presents 83% from the original data and the testing set for each starts from 01/2016 to 10/2017 which presents 17% from the original data.

Second step, we choose the most appropriate algorithms by running these four algorithms over the first part (water consumption data) on training set with horizon value 22 which means until 10/2017 to compare forecasting results with the testing set, and then apply the equation of MPE between the actual water consumption values in testing set and predicted values from forecasting results.

3.5.4. Testing and Evaluating the Model At this stage we evaluated the model accuracy using mean absolute percentage errors MAPE.

$$(((\text{Actual} - \text{Predicted}) / \text{Actual}) * 100)$$

Also, the accuracy of forecasts is evaluated by comparing them with observed water demand and water revenue.

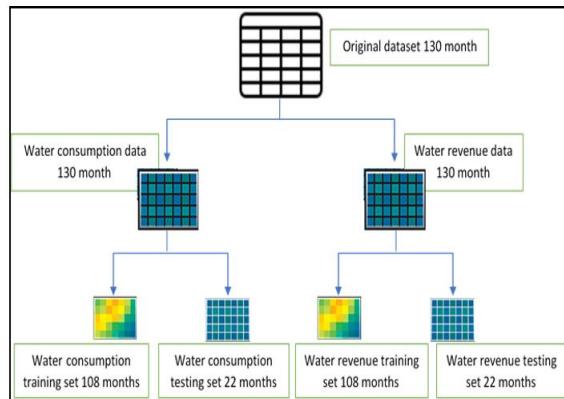


Figure 3. Original data splitting steps

This evaluation provides insights in recommending changes to existing models to reduce deviations in future forecasts. After computing and comparing MPEs from different algorithms, we found that Hybrid ARIMA was the lowest MPE for both parts (water consumption part and water revenue part), which means the most appropriate algorithm.

4. Experimental Results

The chosen forecasting algorithms can be labeled as the most recent and famous algorithms used in time series analysis field. These algorithms are Auto Regressive Integrated Moving Average (ARIMA), ARIMA combined with Neural Networks (Hybrid ARIMA), Singular Spectrum Analysis (SSA) and Linear Regression forecast. We are going to run these four algorithms over our dataset to calculate the Mean Percentage Error (MPE) to detect which algorithms gives the lowest MPE, which mean the more accuracy algorithm to be used in further.

Our data set is real data set collected from Khan Younis Municipality (KHM) – department of customer services, Khan Younis City, Gaza Strip, Palestine. The data set belongs to the monthly vouchers produced from Khan Younis municipality presenting the consumption amount of water in letters for each registered unit (double values), and the revenue fees of available water distributing service (double values) and voucher date in format month-year (mm-yyyy).

Our original data consists of four columns (Voucher Date, Area, Water Consumption, Water Revenue). The data sets containing almost eleven years of indices data (~11 years, about 130 months) was divided into training set, containing the first nine years of data 01-2007 to 12-2015(9 years, about 108 months), and testing set holding the remaining data which almost 2 years from 01-2016 to 10-2017 (~2

years, about 22 month). Models were trained on the training sets before being applied to the unseen data in the test sets. Table 1 gave a sample of training set for both attributes (consumption and revenue), and Table 2 gave sample of testing set for both attributes (consumption and revenue)

Table 1. Sample of the training set from 1-2007 to 12-2015

VOUCHER DATE	WATER CONSUMPTION (M ³)	WATER REVENUE (NIS)
01-2007	286273	450597.5
02-2007	293195	539147.2
03-2007	218840	429298.5
04-2007	312961	467045.8
05-2007	329750	510998.9
06-2007	357077	516393.4
07-2007	374079	538364.1

Table 2. Sample of the testing set from 01-2016 to 10-2017

VOUCHER DATE	WATER CONSUMPTION (M ³)	WATER REVENUE (NIS)
01-2016	459852	773740.5
02-2016	455574	777848.5
03-2016	264750	472531
04-2016	301128	488146.9
05-2016	388855	552890.7
06-2016	351510	510060.9

The testing set presents 17% of the overall dataset, and this is the period from 01/2016 to 10/2017 (22 months). To choose the appropriate forecasting algorithms we first evaluated and compared the results from each algorithm with the original data we have. We run four forecasting algorithms on the training set and with horizon value 22 to reach 10-2017 to compare it with the testing set, at this stage we will only forecast the attribute ‘Water Consumption’ as sum of all areas over every month. These four forecasting algorithms are (ARIMA, Hybrid ARIMA, SSA and Linear Regression).

4.1. Evaluating algorithms over water consumption

First step we evaluated the selected four algorithms over the attribute ‘Water Consumption’ and compute the mean percentage error MPE for each algorithm.

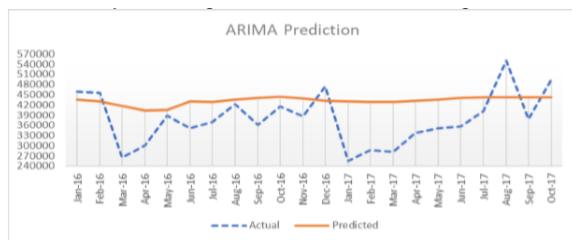


Figure 4. ARIMA Evaluation for water consumption

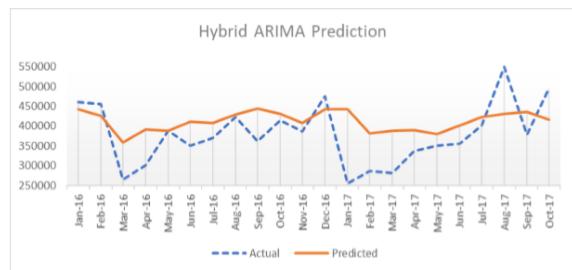


Figure 5. Hybrid ARIMA Evaluation for water consumption



Figure 6. SSA Evaluation for water consumption



Figure 7. Linear Regression Evaluation for water consumption

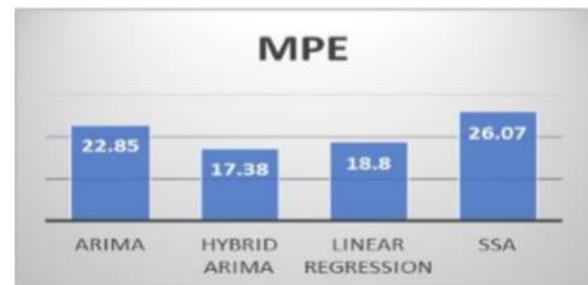


Figure 8. MPA percentages for four forecasting algorithms over water consumption attribute

From Figure 8 which summarize the MPE results of the four algorithms, we can notice that Hybrid ARIMA is the most appropriate algorithms with least MPE 17.38% which means the forecasting values of predicting set are close to data of the testing set, while SSA is out of consideration with very high MPE 26.07% which means the forecasting values of predicting set are far away from data of the testing set.

4.2. Evaluating algorithms over water revenue

After running the four algorithms on the attribute ‘Water Consumption’ and we noticed that hybrid ARIMA was the most accurate algorithm to be used in real forecasting, now repeated the past work on the attribute ‘Water Revenue’.

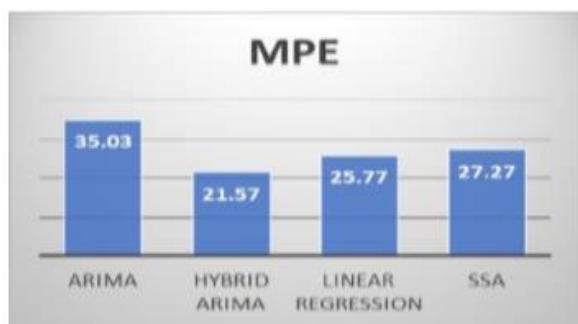


Figure 9. MPA percentages for four forecasting algorithms over water revenue attribute

From Figure 9 which summarize the MPE results of the four methods, we again noticed that Hybrid ARIMA is the most appropriate algorithms with least MPE 21.57% which means the forecasting values of training set are close to data of the testing set, while SSA is out of consideration with very high MPE 27.27% which means the forecasting values of training set are far away from data of the testing set.

Finally, clearly illustrated that Hybrid ARIMA is the appropriate algorithms which will be used further in forecasting future 5 years over the attributes ‘Water Consumption’ and ‘Water Revenue’. We propose to take a combining approach to time series forecasting. The linear ARIMA model and the nonlinear ANN model are used jointly, aiming to capture different forms of relationship in the time series data. The hybrid model takes advantage of the unique strength of ARIMA and ANN in linear and nonlinear modeling. For complex problems that have both linear and nonlinear correlation structures, the combination method can be an effective way to improve forecasting performance. After choosing Hybrid ARIMA to continue our forecasting tasks, we forecasted the whole Khan Younis city.

4.3. Forecasting the overall city

Now we forecast the whole city in advance 5 years regardless of sub-areas or classes within Khan Younis city. Then we computed the deviation percentage for every forecasted year and compare it with year 2017.

From Figure 10 we can notice that the two curves almost having the same trend, which means when forecasted water consumption increases the forecasted revenue also increases and that is logical and vice versa. But there are some exceptions such as the interval from 01-2019 to 03-2019, the curve of water consumption is decreasing but the curve of revenue in increasing.

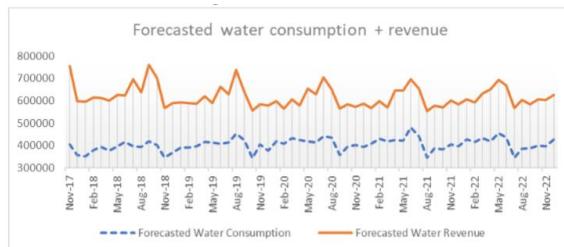


Figure 10. Five years forecasting water consumption and revenue for Khan Younis city.

We compared the last year of the original dataset with the forecasted data for both ‘Water Consumption’ and ‘Water Revenue’, to help decision makers in planning water distributing and production systems and give them future vision of how much Khan Younis city will be using water.

Table 3. Consumption deviation in comparison to 2017

Year	Water Consumption (M³)	Water Consumption Deviation %
2017	4555059	0
2018	4693144	3.03
2019	4764013	4.59
2020	4928643	8.20
2021	4938253	8.41
2022	4901387	7.60

Table 4. Revenue deviation in comparison to 2017

Year	Water Revenue (NIS)	Water Revenue Deviation %
2017	7527434	0
2018	7828069	3.99
2019	7367335	-2.13
2020	7306228	-2.94
2021	7242724	-3.78
2022	7396991	-1.73

From Table 3 we can notice that water consumption amounts will be increasing in all next 5 years in comparison to 2017, but on the other hand, from Table 4 water revenues are increasing only in 2018 but decreasing from 2019 to 2022. Generally, we can say that after 5 years the maximum water consumption will increase with 8.4% (1.7% per year) in comparison with 2017 but the minimum water revenue will be decreased with 3.8% (0.76% per year).

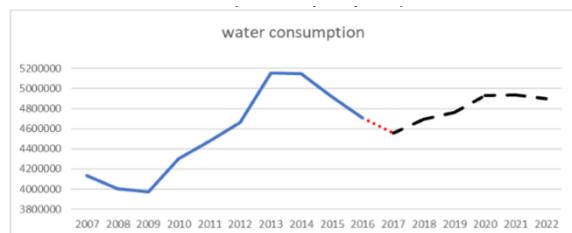


Figure 11 Water consumption data (Actual + forecasted) from 2007 to 2022

Figure 11 shows the whole yearly consumption data (actual + forecasted), the dotted line shows 2017 consumption data, while the dashed line shows the 5 years forecasted data from 2018 to 2022. By visual it's clear that the dashed line trend is going up with comparison to dotted line, which means the average of water consumptions for the next five years will increase.



Figure 12 Water revenue data (Actual + forecasted) from 2007 to 2022

Figure 12 shows the whole yearly consumption data (actual + forecasted), the dotted line shows 2017 consumption data, while the dashed line shows the 5 years forecasted data from 2018 to 2022. By visual, the dashed line trend is going down with comparison to dotted line, which means the average of water revenues for the next five years will decrease.

5. Discussion

From Figure 11 and Figure 12, we observed that water consumptions for the next five years will increase, and this is logical result, because the population is increasing by the time which means more people will demand water for basics needs and

day life use. But on the other hand, water revenues will decrease for the next five years, and this contrary with the previous observation about water consumption, while the nature relationship between water consumption and water revenue is positive relationship, which means when water demand increases, water revenues also should increase and vice versa.

But water revenues are decreasing maybe due to several reasons, (1) The suffocating economic crisis, which affecting directly on people financial status, which lets them not pay for the water voucher produced by Khan Younis municipality, (2) blockage from sea, land and air which effect on the society projects and income, (3) Political differences and their impact on employees' salaries, so much so that a class of employees received 40% of the salary for a long time, (4) Khan Younis municipality do not count the quantities of water it's buildings consume and which definitely affect their revenues in return. All these reasons and other effects on the municipality revenues from water service.

6. Conclusion

In this paper, the water consumption used by citizen and revenues were forecasted in the medium term future.

Four main forecasting algorithms were used ARIMA, Hybrid ARIMA, SSA and Linear Regression. We applied them on dataset collected from Khan Younis municipality (KHM)-department of customer services. We found Hybrid ARIMA is the best algorithm which gave the Mean Absolute Percentage Error (MAPE) of 17.38%. In general, for the next 5 years water consumption will increase with 8.4% (1.7% per year) but the water revenue will be decreased with 3.8% (0.76% per year).

7. Future Work

As future work, we can use long term forecasting approaches to predict the water consumption and water revenue for more than 10 years. Also, we can forecast data using more variables and attributes such as population.

More deep learning models can be used specially to solve our problem or achieve our goal. Also, other resources can be forecasted in Khan Younis municipality KHM.

This model could be applied for other municipalities datasets to forecast water demand and water revenue or any other resources available. Finally, we can use forecasting methods to predict individual water bills.

8. References

- [1] Aggarwal, C.C., Data mining: the textbook. 2015: Springer.
- [2] Ripley, R., Why Resource Forecasting is Important: Preparing for Projects. workamajig, 2015.
- [3] Billings, R.B. and C.V. Jones, Forecasting urban water demand. 2011: American Water Works Association.
- [4] Donkor, E.A., et al., Urban water demand forecasting: review of methods and models. Journal of Water Resources Planning and Management, 2012. 140(2): p. 146-159.
- [5] Ajbar, A. and E.M. Ali, Prediction of municipal water production in touristic Mecca City in Saudi Arabia using neural networks. Journal of King Saud University-Engineering Sciences, 2015. 27(1): p. 8391.
- [6] RISH, R.M.A.A., Electricity Consumption Forecasting in the Khan Younis Province Using Exponential Smoothing and Box - Jenkins Methods: A Modeling Viewpoint. 2015: p. 95.
- [7] Rinaudo, J.-D., Long-term water demand forecasting, in Understanding and Managing Urban Water in Transition. 2015, Springer. p. 239-268.
- [8] Liu, Y., J. Zhao, and Z. Wang, identifying determinants of urban water use using data mining approach. Urban Water Journal, 2015. 12(8): p. 618630.
- [9] Khashei, M. and M. Bijari, A new hybrid methodology for nonlinear time series forecasting. Modelling and Simulation in Engineering, 2011. 2011: p. 15.
- [10] Broomhead, D.S. and G.P. King, Extracting qualitative dynamics from experimental data. Physica D: Nonlinear Phenomena, 1986. 20(2-3): p. 217-236.
- [11] Elsner, J.B. and A.A. Tsonis, Singular spectrum analysis: a new tool in time series analysis. 2013: Springer Science & Business Media.
- [12] Ozyasar, H., Application of Regression Analysis in Business. 2011.