Comparison of Parametric and Nonparametric Techniques for Water Consumption Forecasting

Bisher M. Iqelan

Abstract— Forecasting water consumption is an important foundation for the development of water demand management strategy. In this paper, a parametric time series exponential smoothing state space (ETS) method was employed to monthly water consumption for Gaza City/Palestine. It is compared with nonparametric Reweighted Nadaraya-Watson (RNW) method. The root mean square error (RMSE) and mean absolute scaled error (MASE) are used for comparing the forecasting accuracy. This paper finds statistically significant evidence showing that the ETS state space model outperforms RNW approach at forecasting water consumption.

Index Terms— Exponential smoothing, Nadaraya-Watson estimator, Nonparametric, Parametric, Reweighted, State space models, Water consumption forecast.

1 INTRODUCTION

With the rapid growth of world population, problems of water consumption are threatening and involving the attention of every human being. The implementation of long-term water consumption forecasting stands as one solution to the problem of water consumption. A good path to water management and design is through water prediction and water consumption forecasting. They can give crucial water data required in future development and predict other types of water problems. As a result, water consumption forecasting holds a significant theoretical and practical value that raises the predictive accuracy of water consumption. There are different conventional approaches to water consumption forecasting including parametric and nonparametric techniques.

Located along the Mediterranean Sea, the Gaza Strip is a very small Palestinian territory. It is 362-kilometer square and is home to a population of more than 2 million people. Therefore, it ranks as one of the most densely populated places around the world. As the siege continues to escalate, normal life would not be viable since the standards of living are declined. This paper focuses on the city of Gaza as it contains most of the population. The water crisis in Gaza is resulted from over-pumping of the aquifer, water loss, and salt accumulation. Water consumption forecasts can provide crucial insights into the future needs of Gaza’s life requirement. This paper attempts to provide a coherent water consumption forecasting model for Gaza policy makers as well. The objective of this study is to evaluate the application of two different forecasting techniques for Gaza City. These techniques include both parametric exponential smoothing state space (ETS) and nonparametric Reweighted Nadaraya-Watson (RNW).

The literature on forecasting water consumption is rich since it is a very important field of study. Maidment [20] applied short-term Box and Jenkins models on daily municipal water use study. These ARIMA time series models were used as a function of rainfall and air temperature.

The technique of artificial neural networks (ANN) was used in several civil engineering applications. In addition, [27] conducted a study to forecast daily municipal water demand using time series models. Jowitt and Xu [17] used exponential smoothing and autoregression to forecast water demand. Shvarster et al. [26] provided a model for hourly water use forecast based on time series analyses and pattern recognition. Moreover, [15] presented a technique to forecast daily water use. This technique includes two steps; the first step uses an exponential smoothing algorithm to forecast monthly average water use. The second step contains the monthly average forecast that is employed to obtain a daily forecast. Zhou [33] developed time series models for daily water consumption in Melbourne, Australia. Khan and Coulibaly [19] provided a comparative study between time series models in the forecasting of the water level of a lake. Horielova and Zadachyn [10] concluded that ARIMA models and ANN models are most adequate in forecasting water consumption of large cities.

Data Source: We use the data set of water consumption in Gaza City / Gaza Strip/ Palestinian Territories from Municipality of Gaza https://www.gaza-city.org. We considered monthly data sets for water consumption from January 1990 to July 2016.

The section above contains a summary of water consumption problem with an application on Gaza City. In addition, an overview of previous literature is also included. The remainder of the paper is organized as follows. Section 2 presents the parametric exponential smoothing state space models and the nonparametric Reweighted Nadaraya-Watson methods used in this paper. Section 3 discusses the methodology used in the selection of parametric and nonparametric techniques and forecasting evaluation. Data description and empirical results are provided in section 4. Section 5 presents the summary and conclusion of this paper.

2 MATERIALS AND METHODS

2.1 The Reweighted Nadaraya-Watson Estimator

Let \( (X_t, Y_t), t = 1, 2, \cdots, n \) be observations of a bivariate random variable \((X, Y)\) and assume that all are continuously...
distributed with joint probability density function \( f(x,y) \). This section presents a kernel regression estimator known as the reweighted Nadaraya-Watson estimator.

Let \( f(y|x) = \frac{f(x,y)}{g(x)} \) be the conditional density of \( Y_t \) given \( X_t \), where \( g(x) = \int f(x,y)dy \) is the marginal density of \( X_t \). The simple nonparametric regression function for \( Y_t \) on \( X_t \) is defined as

\[
m(x) = E(Y_t | X_t = x) = \int \frac{yf(x,y)}{g(x)}dy
\]

The basic idea is to estimate this non-parametrically, with minimal assumptions about \( m(x) \). Suppose the relationship between the dependent variable \( Y \) and the independent variable \( X \), the regression equation can be written as,

\[
Y_t = m(X_t) + \varepsilon_t, \quad t = 1, 2, \ldots, n
\]

Where \( \varepsilon_t, t = 1, 2, \ldots, n \) are the measurement errors such that \( E(\varepsilon_t | X_t) = 0 \) and \( E(\varepsilon_t^2 | X_t = x) = \sigma^2(x) \). Using the kernel estimation, the regression mean function \( m(x) \) is estimated by \( \hat{m}(x) \), where

\[
\hat{m}(x) = \int yf(y|x)dy
\]

A class of kernel-type estimator is called the Nadaraya-Watson (NW) estimator which is one of the most widely known and used for estimating \( f(y|x) \). The NW estimator has been suggested independently by [29] and [22].

Let \( \hat{f}_{NW}(y|x) \) stands for Nadaraya-Watson kernel estimation and defined as,

\[
\hat{f}_{NW}(y|x) = \frac{\sum_{i=1}^{n} K_h(x-X_i)Y_i}{\sum_{i=1}^{n} K_h(x-X_i)},
\]

Where \( k(.) \) is a kernel function, \( K_h(.) = \frac{1}{h}K\left(\frac{.}{h}\right) \) and \( h = h_n > 0 \) is the bandwidth. Consequently, the NW estimator, \( \hat{m}(x) \), of the regression mean function \( m(x) \) can be defined as

\[
\hat{m}_{NW}(x) = \frac{\sum_{i=1}^{n} K_h(x-X_i)Y_i}{\sum_{i=1}^{n} K_h(x-X_i)}.
\]

It is worth noting that there is some weakness in applying the NW estimator especially when the estimation is done near the boundary points or when the design density is highly clustered. For more details see [8], [4], [5], and [32].

To overcome the weakness of the NW estimator, a new class of the kernel estimation of the conditional density function and regression mean function has been proposed by [8]. The new estimator is called the Reweighted Nadaraya-Watson (RNW) estimator, see [4], [6]. The RNW estimator of \( f(y|x) \) is denoted by \( \hat{f}_{RNW}(y|x) \) and defined as

\[
\hat{f}_{RNW}(y|x) = \frac{\sum_{i=1}^{n} \tau_i(x) K_h(x-X_i)K_h(y-Y_i)}{\sum_{i=1}^{n} \tau_i(x) K_h(x-X_i)},
\]

where \( \tau_i(x) \) are probability weights functions satisfying the following properties

1. \( \tau_i(x) \geq 0 \),
2. \( \sum_{i=1}^{n} \tau_i(x) = 1 \),
3. \( \sum_{i=1}^{n} \tau_i(x)(X_i-x)K_h(x-X_i) = 1 \).

\[
\tau_i(x) = \frac{1}{n(\lambda(X_i-x)K_h(x-X_i)+1)}
\]

Where \( \lambda \) is the unique minimizer of \( L(\lambda) \), where \( L(\lambda) \) is defined by

\[
L(\lambda) = \sum_{i=1}^{n} \log(1 + \lambda(X_i-x)K_h(x-X_i)).
\]

Moreover, the Reweighted Nadaraya-Watson estimator, \( \hat{m}_{RNW}(x) \), of \( m(x) \) is given by

\[
\hat{m}_{RNW}(x) = \frac{\sum_{i=1}^{n} \tau_i(x) K_h(x-X_i)Y_i}{\sum_{i=1}^{n} \tau_i(x) K_h(x-X_i)}.
\]

Recently, [25] studied the RNW and proved the joint asymptotic normality of it estimated at a finite number of conditional points.

### 2.2 Exponential Smoothing State Space Models

A common class of forecasting models are exponential smoothing techniques. It’s simple but very helpful of adjusting time series forecasting. R. G. Brown [3], [9], and [31] have initially introduced these methods in their early works. The idea behind forecasting using exponential smoothing is to assign exponentially declining weights to observations as they go back in age, i.e., recent observations have a larger weight than the old ones.
Pegels [24] was the first to classify exponential smoothing techniques and propose a taxonomy of the trend component and seasonal component. Pegels’ taxonomy was later extended by [7], who added damped trend to the classification. This extension is then modified by [15], before the final extension is proposed by [28] who extended the classification to include damped multiplicative trends.

In the state space approach for the exponential smoothing methods, the error terms are smoothed. The corresponding models are called innovation state space models. Note that it is an innovation because all equations in this type use the same error term \( \varepsilon_t \). For more details, see [1], [2], and [11]. Hyndman [15], [23] and [28] showed that there are two possible underlying state space models, for each of the 15 methods in Table 1. There is one with additive errors and one with multiplication errors. A list of all the 30 possible specifications are described in [12]. An extra letter for error is added in the method notation. Each stated space model is labelled as ETS(\( \cdot, \cdot, \cdot \)) for (error, trend, seasonal). So ETS(M, A, A) refers to a model with multiplicative error, additive trend and additive seasonality. To illustrate the idea of the multiplicative error model ETS(M, A, A), let first \( \mu_t = \ell_t + s_{t-m} \) denote the one-step forecast of \( y_t \), supposing that all parameters are known. Also, let \( \varepsilon_t = (y_t - \mu_t) / \mu_t \), so that \( \varepsilon_t \) is a relative error. The general form of the resulting state equations is (See [12] pp. (19-22))

\[
\begin{align*}
\ell_t &= (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)
, \\
b_t &= b_{t-1} + \alpha (\ell_{t-1} + b_{t-1} + s_{t-m}) \varepsilon_t,
\end{align*}
\]

By defining the state space vector as \( x_t = (\ell_t, b_t, s_t, s_{t-1}, \ldots, s_{t-m+1})' \), the system (2.1) can be written in standard state space notation:

\[
\begin{align*}
y_t &= [1 \ 1 \ 0 \ \cdots \ 0 \ 1] x_{t-1} (1 + \varepsilon_t),
\end{align*}
\]

where \( m \) is the number of seasons in a year (e.g. for quarterly data \( m = 4 \), and for monthly data \( m = 12 \)). \( \ell_t \) is the level, \( b_t \) is growth, \( s_t \) is the seasonal component, \( \hat{y}_{t+h|t} \) denotes the \( h \) -step-ahead forecast and \( h_m^* = (h - 1) \text{ mod } m \) + 1 which establishes that the estimates of the seasonal indices for forecasting come from the last year of the sample. (The notation \( u \text{ mod } v \) means the largest integer not greater than \( u \).

The smoothing parameters \( \alpha, \beta \) and \( \gamma \), show how fast the level, trend and seasonality, respectively, fit to new information. For more details regarding the component for all fifteen exponential smoothing methods, see [12] pp.18.
3 MODEL SELECTION

3.1 Bandwidth Selection

The RNW estimator, \( \hat{m}_{RNW}(x) \) is a nonparametric approach which has no need for a distributional assumption. The goodness of the kernel regression estimate depends decisively on the selection of bandwidth, \( h \). When the bandwidth is very small, the estimate will be very close to the original data. If the bandwidth is very large, the estimate will be very smooth, lying close to the mean of the data. Many suggestions for bandwidth selection have been taken by [30]. Formally, the suitable bandwidth is the one minimizing the mean square error (MSE) of the RNW estimator, \( \text{MSE}(\hat{m}_{RNW}(x)) \).

3.2 Automatic Forecasting

The parametric ETS method is an automatic forecasting method which is applied to perform forecasting using ets() function via the forecast package in R software. The suitable exponential smoothing model is selected among 30 ETS models using maximum likelihood estimator (MLE) and selecting the best model based on information criteria AIC, AICc, and BIC. The model which minimizes the criteria is chosen as preferable for the data. For more details, see [15], [12], [11].

3.3 Forecast Evaluation

There are many different accuracy measures which can be used to select the forecasting model perform carefully. Denote the actual observation at time \( t \) by \( y_t \) and its forecasted value by \( \hat{y}_t \). The following accuracy measures are considered in Table 2.

<table>
<thead>
<tr>
<th>Acronyms</th>
<th>Definition</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>Mean square error</td>
<td>( \sum_{i=1}^{n} e_i^2 / n )</td>
</tr>
<tr>
<td>MAE</td>
<td>Mean absolute error</td>
<td>( \sum_{i=1}^{n}</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root mean square error</td>
<td>( \sqrt{\text{MSE}} )</td>
</tr>
<tr>
<td>MAPE</td>
<td>Mean absolute percentage error</td>
<td>( \sum_{i=1}^{n} 100</td>
</tr>
</tbody>
</table>

where \( n \) is the number of observations in the sample. It is worth noting that MSE, MAE and RMSE are scale dependent measures and because of that they should not be used in comparing models estimated from different data sets. Moreover, MAPE measure is based on percentage error which has the benefit of being scale independent but is not valid for series that are very sparse and scattered, as they are undefined or infinite for \( y_t = 0 \). This makes MAPE inappropriate in this application.

The accuracy measure, suggested by [13], proposes comparing the errors given by the forecasting models to the errors from naïve models, i.e., a method where the forecast for the observation time \( t \) is simply the observation at time \( t - 1 \). The scaled error would then be defined as,

\[
q_t = \frac{1}{n-1} \sum_{i=2}^{n} |y_t - y_{t-1}| \]

Measures based on scaled errors have benefit of being neither scale dependent nor implicated by series fluctuating around zero. The mean absolute scaled error (MASE) is given by

\[
\text{MASE} = \frac{\sum_{i=1}^{n} |q_i|}{n} .
\]

Moreover, the MASE is very flexible and has a simple interpretation. If the value is less than one, it means the forecast is more accurate than the naïve forecast.

4 APPLICATION TO REAL DATA

4.1 Data

We have analyzed the monthly water consumption series in Gaza city/Gaza strip/ Palestinian territories from January 1990 to July 2016. We have transformed the data by dividing it by 1000000. Table 3 presents some descriptive statistics for this data set.

| TABLE 3 |
| DESCRIPTIVE STATISTICS FOR MONTHLY WATER CONSUMPTION DATA 1990:1-2016:7 |
|----------|----------------|---------------|----------|----------|
| n        | Min. | 1st Qu | Mean | Std.dev | 3rd Qu | Max |
| 319      | 1.379 | 1.880 | 2.262 | 0.528 | 2.519 | 3.906 |

Figure 1 displays a graphical representation of the decomposed series of water consumption. Clearly by studying the plot of the water consumption in the first panel of Figure 1, it is easy to realize that the series does not have a constant mean. This reflects that the water consumption is a non-stationary series. It gives us an indication of the seasonality of water consumption. The series displays periodic attitude every 12 months. Moreover, the third panel in Figure 1 displays the trend of the series. It is the same series, but with seasonality extracted.
We use the first 307 observations from January 1990 to July 2015 as training sample for model prediction, and the remaining 12 observations from August 2015 to July 2016 as post-sample for forecast evaluation.

4.2 Empirical Study

This section establishes the empirical results on the training sample for fitting models of water consumption in Gaza City by using the two different methods, the nonparametric RNW kernel estimation method and the parametric ETS method. R statistical software was used for fitting the models and to forecast water consumption in Gaza City for 12 months from August 2015 to July 2016. The procedure for the nonparametric RNW is based on selecting the suitable bandwidth, \( h \), which minimizes \( \text{MSE}(\hat{m}_{\text{RNW}}(x)) \). Table 4 shows different values of bandwidth and the corresponding \( \text{MSE}(\hat{m}_{\text{RNW}}(x)) \). Comparing the values to obtain that the minimum value of \( \text{MSE}(\hat{m}_{\text{RNW}}(x)) \) is equal to 0.1898 corresponding to the bandwidth \( h = 0.4939 \).

<table>
<thead>
<tr>
<th>MSE(( \hat{m}_{\text{RNW}}(x) ))</th>
<th>The bandwidth ( h )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10974</td>
<td>0.21949</td>
</tr>
<tr>
<td>0.36144</td>
<td>0.26491</td>
</tr>
<tr>
<td>0.49385</td>
<td>0.18975</td>
</tr>
<tr>
<td>0.54872</td>
<td>0.99956</td>
</tr>
</tbody>
</table>

When fitting an exponential smoothing model with a state space approach, the ets() function in forecast package was used. It is used to choose the suitable model automatically on the basis of maximum likelihood method (MLE) and then to calculate the point forecasts for water consumption in Gaza City for the 12 months from August 2015 to July 2016. When applying the function, the results showed that the best performing model was \( \text{ETS}(M,A,A) \), that is a model with multiplicative error, additive trend and additive seasonality. The general equation of this model was explained in section 2.2 by the system 1.2.

4.3 Forecasting Models

Table 5 compares the actual with the point forecast for the 12 months in the period from August 2015 to July 2016 for water consumption in Gaza City in the Palestinian Territories based on ETS and RNW methods.

<table>
<thead>
<tr>
<th>Month</th>
<th>Actual Data</th>
<th>ETS(M,A,A)</th>
<th>RNW method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aug 2015</td>
<td>3798939</td>
<td>3973588</td>
<td>3298502</td>
</tr>
<tr>
<td>Sep 2015</td>
<td>3994249</td>
<td>3821226</td>
<td>3700635</td>
</tr>
<tr>
<td>Oct 2015</td>
<td>3842813</td>
<td>3922643</td>
<td>3085275</td>
</tr>
<tr>
<td>Nov 2015</td>
<td>3361297</td>
<td>3678210</td>
<td>3243449</td>
</tr>
<tr>
<td>Dec 2015</td>
<td>3956254</td>
<td>3583795</td>
<td>3850790</td>
</tr>
<tr>
<td>Jan 2016</td>
<td>3589814</td>
<td>3489380</td>
<td>3505749</td>
</tr>
<tr>
<td>Feb 2016</td>
<td>3657156</td>
<td>3516415</td>
<td>3492082</td>
</tr>
<tr>
<td>Mar 2016</td>
<td>4229867</td>
<td>3607954</td>
<td>3699882</td>
</tr>
<tr>
<td>Apr 2016</td>
<td>3795809</td>
<td>3694175</td>
<td>3266582</td>
</tr>
<tr>
<td>May 2016</td>
<td>4151937</td>
<td>3865150</td>
<td>4255222</td>
</tr>
<tr>
<td>Jun 2016</td>
<td>3972153</td>
<td>4019894</td>
<td>3544637</td>
</tr>
<tr>
<td>Jul 2016</td>
<td>3837802</td>
<td>3917415</td>
<td>3094285</td>
</tr>
</tbody>
</table>

It is quite clear that the forecast error of \( \text{ETS}(M,A,A) \) technique is much less than that of the RNW technique. This outcome points out that \( \text{ETS}(M,A,A) \) method fits water consumption data much better than the RNW kernel estimator method. It is interesting to note that the parametric exponential smoothing model outperforms the nonparametric Re-weighted Nadaraya-Watson kernel estimator method. Comparing the results of the measures of forecast accuracy in Table 6, the \( \text{ETS}(M,A,A) \) model still performs best based on all the measures values.

The predicted values using \( \text{ETS}(M,A,A) \) is highly accurate as \( \text{RMSE} = 0.2620172 \) and \( \text{MASE} = 0.613292 \); it has the lowest RMSE and MASE estimate.

<table>
<thead>
<tr>
<th></th>
<th>MAE</th>
<th>RMSE</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNW estimator</td>
<td>0.363108</td>
<td>0.4356086</td>
<td>1.070811</td>
</tr>
<tr>
<td>ETS(M,A,A)</td>
<td>0.2079781</td>
<td>0.2620172</td>
<td>0.613292</td>
</tr>
</tbody>
</table>

The smoothing parameters of the exponential smoothing state space model \( \text{ETS}(M,A,A) \) that adequately fit the water consumption data are listed as follows:

\[ \alpha = 0.1502, \quad \beta = 0.0026, \quad \lambda = 0.0001 \]

with initial states (the components of the state space vector \( x_{t - 1} \)) are listed as follows:

\[ \ell_{t - 1} = 1.7337, \quad b_{t - 1} = -0.0031 \]
\[ s = -0.1573, -0.0519, 0.2035, 0.1129, 0.2763, 0.0998, \]
\[ 0.2131, 0.0693, -0.0907, -0.166, -0.2465, -0.2626 \]
where $s$ above refers to the values of the seasonal initials $s_{t-1}, s_{t-2}, \ldots, s_{t-m}$. Putting all these values in the system (2.2) above to obtain the standard state space equations of ETS($M,A,A$) model.

Figure 2 shows the forecasts from ETS($M,A,A$) model. The thin blue indicates the fitted values, while the forecasts are shown as the dashed thicker red line. The prediction intervals are clarified by the light 95% and dark 80% grey confidence bands.

![ETS(M,A,A)](image)

Fig. 2. Forecast for $h=12$ months ahead with ETS($M,A,A$) model.

However, to make prediction intervals using exponential smoothing techniques, the prediction intervals require that the forecast errors should be uncorrelated and normally distributed with mean zero and constant variance.

The sample correlation in Figure 3 displays that the most sample autocorrelation coefficients of the residuals are within the confidence limits, consequently the residuals are white noise reflecting that ETS($M,A,A$) is adequate.

To confirm the evidence of autocorrelations, the Box-Ljung test p-value result in Table 7 displays that there is evidence of no autocorrelations in the forecasts errors. Furthermore, the Jarque Bera test p-value shows that there is a strong evidence for normality of forecasts errors. Moreover, the test of homoscedasticity p-value proves the evidence that the variance is constant.

![Residual Diagnostics Tests](image)

**TABLE 7**

<table>
<thead>
<tr>
<th>Residual Test</th>
<th>p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual Autocorrelation test</td>
<td></td>
</tr>
<tr>
<td>Box-Ljung test</td>
<td>7.695*10^{-65}</td>
</tr>
<tr>
<td>Residual Normality test</td>
<td></td>
</tr>
<tr>
<td>Jarque Bera Test</td>
<td>&lt; 2.2*10^{-16}</td>
</tr>
<tr>
<td>Homoscedasticity test</td>
<td></td>
</tr>
<tr>
<td>Box-Ljung test (Squared Residuals)</td>
<td>0.01018</td>
</tr>
</tbody>
</table>

5 CONCLUSION REMARKS

The main objective of this paper is to forecast water consumption in Gaza City/Palestinian Territories from January 1990 to July 2016. Two approaches are used attempting to obtain a suitable fit. They are a parametric exponential smoothing state space (ETS) model and a nonparametric Reweighted Nadaraya-Watson (RNW) kernel estimator method.

The analysis recognizes that the ETS state space model has good prediction results than the RNW approach based on both mean absolute error (MAE), root mean square error (RMSE) and mean absolute scaled error (MASE). However, we can note that although ETS state space models are not widely applied in forecasting of water consumption, the empirical results assert the importance of their application.

ACKNOWLEDGMENT

I would like to thank Raed B. Salha for the very useful discussion and suggestions he provided in this study. I would also like to thank Municipality of Gaza for providing the water consumption data.

REFERENCES


