Accurate Forecasting of the Satellite-Derived Seasonal Caspian Sea Level Anomaly Using Polynomial Interpolation and Holt-Winters Exponential Smoothing

Moslem Imani, Rey-Jer You, and Chung-Yen Kuo*

Department of Geomatics, National Cheng Kung University, Tainan, Taiwan

Received 3 April 2012, accepted 12 October 2012

ABSTRACT

Polynomial interpolation and Holt-Winters exponential smoothing (HWES) are used to analyze and forecast Caspian Sea level anomalies derived from 15-year Topex/Poseidon (T/P) and Jason-1 (J-1) altimetry covering 1993 to 2008. Because along-track altimetric products may contain temporal and spatial data gaps, a least squares polynomial interpolation is performed to fill the gaps of along-track sea surface heights used. The modeling results of a 3-year forecasting time span (2005 - 2008) derived using HWES agree well with the observed time series with a correlation coefficient of 0.86. Finally, the 3-year forecasted Caspian Sea level anomalies are compared with those obtained using an artificial neural network method with reasonable agreement found.

Key words: Caspian Sea, Satellite altimetry, Interpolation, Forecasting, Exponential smoothing

Citation: Imani, M., R. J. You, and C. Y. Kuo, 2013: Accurate forecasting of the satellite-derived seasonal Caspian Sea level anomaly using polynomial interpolation and Holt-Winters exponential smoothing. Terr. Atmos. Ocean. Sci., 24, 521-530, doi: 10.3319/TAO.2012.10.12.01(TibXS)

1. INTRODUCTION

Innovative application of historical geophysical data sets is the key to the progress of local and global climate change studies. Tide gauge measurements are one of the most reliable and helpful accessible data sets for monitoring sea level change (Parker 1992). In addition, the fluctuation of endotherm lakes, i.e., lakes in closed basins, is not only a critical parameter for understanding the water balance in drainage basins, but also a valuable indicator of climate change (Kostianoy et al. 2004). The Caspian Sea is the largest isolated water reservoir in the world with a surface area of 370000 km² (Fig. 1). Surrounded by five countries, Azerbaijan, Iran, Kazakhstan, Russia, and Turkmenistan, the Sea plays a key role in their interaction as well as being a vital natural resource for all (Stolberg et al. 2006).

Since lakes in closed basins are sensitive to climate change, the Caspian Sea observed traditionally by tide gauges exhibits large sea level variations. Scientists are interested in Caspian Sea variations because of its chronological oscillations in both surface area and depth causing complex geological and climatic evolution of the region (Rodionov 1994).

Sea level measurements covering the entire ocean basin could provide useful information on the water mass balance and inter-annual and decadal oscillations in response to climate change (Cazenave et al. 2002). Tide gauge records relative to the crustal surface at the local coast are considerably affected by a number of geographical and meteorological phenomena, including vertical crustal movements, changes in atmospheric pressure, wind, river discharge, water circulation, water density, and increscent water mass due to melting ice (Kostianoy and Kosarev 2005). In the last two decades, satellite-based sensors in particular satellite altimetry have offered a promising alternative for monitoring water surface heights with unprecedentedly high accuracy from inter-seasonal to inter-annual time scales and spatial coverage regardless of meteorological and geological constraints (Campos et al. 2001; Cheng et al. 2008). Satellite altimetry has been applied in a wide variety of studies (Mercier et al. 2002; Crétaux et al. 2011; Kuo and Kao 2011). An analysis of Caspian Sea level anomalies over a short time span (3.5 years) was conducted by Cazenave et al. (1997) to
estimate temporal sea level variations. Lebedev and Kos
tianoy (2008) investigated the variations of some meteor-
ological phenomena, including sea level, wind speed, and
wave height in different areas of the Caspian Sea, Kara-
Bogaz-Gol Bay, and the Volga River. However, their work
was only based on the spatially spare crossover points of
Topex/Poseidon (T/P) and Jason-1 (J-1) data sets.

Along-track altimetric products have some gaps along
the measurements due to technical noise or closeness to
coastal areas, these data gaps may affect the analysis of sea
level monitoring. In many instances, a careful interpolation
along the track of corrective parameters to fill the data gaps
is required to retrieve valid altimeter observations and the
reclamation of a significant number of good-quality altim-
metric measurements (Durand et al. 2008).

Interpolation is a process of estimating the values of
a certain variable of interest at un-sampled locations based
on measured values (Inggs and Lord 1996). Polynomial in-
terpolation is a commonly used method, in which for some
given points, a polynomial which goes exactly through
these points is found (Gerald and Wheatley 2008). Many
studies have shown the importance of interpolating values
of climate variables from stations to large areas (Kurtzman
and Kadmon 1999; Gigov and Nikolova 2002). A compara-
tive analysis of interpolation techniques, including Inverse
Distance Weighted, Polynomial, Splines, Ordinary Kriging,
and Universal Kriging, was conducted by Basistha et al.
(2008) to find the best interpolation method for the spatial
distribution of rainfall in the Indian Himalayas with a case
study of the Uttarakhand region.

Fig. 1. Map of Caspian Sea (adopted from http://eol.jsc.nasa.gov).
Hydrological and climatic conditions are key issues in the international arena because of significant impact on human activities, thus accurate prediction of hydro-climate extremes helps to mitigate disasters, reduce losses in infrastructure and productive activities and maximize the benefits of socioeconomic consequences of these effects. An optimal forecasting technique is adopted based on the availability of information and the time scale of the data. The most commonly used forecasting techniques within different time span from a few days to years include autoregressive-moving-average models, bilinear and smooth threshold autoregressive models, and artificial intelligence techniques (Sfetsos 2000). A statistical method named exponential smoothing can be used for not only forecasting and smoothing the analyzed function, but also presenting data conveniently and eliminating random errors. It is based on the intuitive application of movable averages where the mean of the last observations is employed to smooth the function (Taylor 2004a). The exponential smoothing method has been applied in various fields (Snyder et al. 2004; Taylor 2004b). The primary advantage of the method is its fast and efficient implementation together with descriptive and inferential statistics (Koçak 2008). Cadenas et al. (2010) analyzed and forecasted wind velocities in Chetumal, Quintana Roo, Mexico using the single-exponential smoothing method. Their results showed a very good accuracy of short-term forecasting.

The present study focuses on the analysis and forecast of Caspian Sea level anomalies from T/P and J-1 measurements along track 092 covering the years 1993 - 2008. The ground tracks of T/P and J-1 altimeters cover the optimal positions of the Caspian Sea surface (Fig. 2) with the proper orbital repeat period (~10 days) for effective study of inter-annual and seasonal sea level variations. As satellite altimetry coverage has data gaps; a least squares polynomial interpolation is employed to fill the gaps. Finally, 3-year forecast of Caspian Sea level anomalies is derived using Holt-Winters exponential smoothing (HWES) and then compared with that obtained using the artificial neural network (ANN) method.

2. DATA

T/P altimeter data in the form of sea surface height (SSH) was extracted from the Merged Geophysical Data Record (MGDR) provided by the NASA Physical Oceanography Distributed Active Archive Center (PODAAC) at the Jet Propulsion Laboratory (JPL) of the California Institute of Technology (Benada 1997). Similarly, SSHs of J-1 were retrieved from the Interim Geophysical Data Record (IGDR) and the Geophysical Data Record (GDR) provided by AVISO (Archivage, Validation et Interprétation des données des Satellites Océanographiques) and PODAAC (Picot et al. 2006). After all recommended corrections, including standard instrumental, media, and geophysical corrections were applied, sea level anomalies were computed by subtracting the geoid heights using the EGM96 geopotential model (Lemoine et al. 1998) from the corrected SSHs. In this study, the sea level anomalies of the pass 092 ground track of the Caspian Sea are used since it is the longest track (most data) and relatively free of strong winds and ice in the study area (Fig. 2).

3. METHODS

3.1 Least Squares Polynomial Interpolation

As shown in Fig. 3, altimetric observations contain some spatial and temporal gaps in sea level measurements, which lead to inaccurate analysis of Caspian Sea level anomalies. In order to compensate for the gap-filled time series of the satellite ground track and obtain an integrated surface covered by interpolated altimetric data, a least squares polynomial interpolation was used.

For the polynomial interpolation, values can be interpolated on regular grids using un-gridded observations by
polynomials, whose coefficients are determined by the least squares method. Accurate interpolation schemes can be used to increase the resolution available for computational analyses and thereby improve the accuracy of the results (Kozak and Krajnc 2007). The polynomial fitting technique was first introduced by Panofsky, Cressman, and Gilchrist based on Lagrange’s classic interpolation formula and then improved by Gilchrist and Cressman (1954) by introducing the concept of a region of influence. The concept indicates only those observations locating within the region of influence (i.e., in a neighborhood of an interpolated point) when interpolating at the point is performed (Lanciani and Salvati 2008).

In this study, the along-track T/P and J-1 data of all available cycles are used to form a local two-dimensional (2-D) coordinate system (or a 2-D matrix) with the X-axis as the cycle \((t)\) and the Y-axis as the distance \((d)\) (Daley 1991). There are \(K\) observations of sea level anomalies \([h_o(r_k), k = 1,..., K]\), where \(r_k = (t_k, d_k)\), so the interpolated field \(h_a(r)\) can be represented by a polynomial expansion as (Gerald and Wheatley 2008):

\[
h_a(r) = \sum \sum C_{mn} r^m d^n, \quad m + n \leq 3, \quad m, n \geq 0
\]  

where \(C_{mn}\) are the unknown coefficients that can be estimated using observations, \(h_o(r_k)\), and \(m + n\) within the order of the polynomial. The quadratic cost function, essentially the mean square error between the interpolation and the observations, can be then written as:

\[
J = \frac{1}{2} \sum_{k=1}^{K} [h_o(r_k) - h_a(r_k)]^2
\]  

In order to estimate \(C_{mn}\), \(J\) has to be minimized. The minimization can be done by differentiating \(J\) with respect to \(C_{mn}\) and setting it equal to 0. The linear system can then be solved using standard algebraic techniques (Gerald and Wheatley 2008). If the error variances of the observations, \(S_i^2\), are known, \(J\) can be rewritten as:

\[
J = \frac{1}{2} \sum_{i=1}^{K} \left[ \frac{h_o(r_i) - h_a(r_i)}{S_i^2} \right]^2
\]  

Fig. 3. (a) T/P tracks over Caspian Sea in January, 1994. (b) J-1 tracks over Caspian Sea in June, 2005.
that is, each increment is weighted by the corresponding variance (Holton 1992; Gerald and Wheatley 2008).

### 3.2 Exponential Smoothing Approach

Exponential smoothing techniques are widely used for the time series analysis because of their simplicity and robustness as automatic forecasting procedures (Corberán-Vallet et al. 2011). Referring to Bowerman et al. (2005), the exponential smoothing approach can be effectively used to forecast stochastic time series. It constructs a time trend forecast with the ability to easily adjust for past errors and prepare follow-on forecasts. Single, double, and triple exponential smoothing techniques are available depending on whether no linear or quadratic trend is specified. Among different approaches, the Holt’s Linear Exponential Smoothing Technique is used to handle non-seasonal series by introducing smoothing parameters (Oyatoye and Fabson 2011).

Holt-Winters exponential smoothing technique extended from Holt’s method can be used to forecast data containing both trend and seasonality. The method has two versions, additive and multiplicative, depending on the characteristics of the particular time series (Chatfield and Yar 1988). In the multiplicative case, when a new observation, $X_t$, becomes available, the local mean level, trend, and seasonal index at time $t$, namely $L_t$, $T_t$, and $I_t$, are respectively updated using:

\[
L_t = \alpha \left( \frac{X_t}{I_{p}} \right) + (1 - \alpha)(L_{t-1} + T_{t-1})
\]

\[
T_t = \gamma(L_t - L_{t-1}) + (1 - \gamma)T_{t-1}
\]

\[
I_t = \delta\left( \frac{X_t}{L_t} \right) + (1 - \delta)I_{t-p}
\]

where $\alpha$, $\gamma$, and $\delta$ denote the smoothing parameters for updating the mean level, trend, and seasonal index, respectively, and $p$ denotes the number of observations per seasonal cycle. The new forecast, $\hat{X}_t$, at time $t$ of the value $k$ periods ahead is given by (Hanke and Wichern 2003):

\[
\hat{X}_t(k) = (L_t + kT_t)I_{t-p+k}, \text{ for } k = 1, 2, ..., p
\]

Exponential smoothing is recommended for time series data with increasing or decreasing trends. The HWES technique is here applied to the 1993 - 2005 (140 months) monthly Caspian Sea level anomalies of the pass 092 ground track. 37-month (2005 - 2008) altimetric sea level anomalies are used for verification of the model prediction results.

### 4. RESULTS AND DISCUSSION

As shown in Fig. 4, Caspian Sea level changes show a nonlinear trend over the considered time span with considerable annual amplitude of sea level anomalies. The maximum value of the Caspian Sea level was -26.80 m (summer 1995) and the minimum was -27.57 m (winter 2002).

By mid-1996, the Caspian Sea level started to decrease abruptly at an average rate of -23 cm yr$^{-1}$. This drop, initiated in July 1995, can be observed until 2001. At the end of this continuous falling, the sea level reached its lowest value among the studied years (Fig. 4). Two phases of rising and falling Caspian Sea levels were observed during the years

![Fig. 4. Caspian Sea level anomalies of pass 092 ground track from 1993 - 2008.](image-url)
2002–2008 where hydrometeorological phenomena (including river discharge and precipitation/evaporation) and water usage strategies could be responsible for the nonlinear trend of annual Caspian Sea level anomalies (Table 1) (Kostianoy and Kosarev 2005).

The Caspian Sea is usually divided into three distinct physical regions, namely the northern, middle, and southern Caspian (Fig. 2) (Amirahmadi 2008). As Table 1 shows, sea level changes in the northern (in the area of the Volga delta), middle, and southern Caspian have different rates over different time spans although correlation results (Pearson correlation coefficient) show similar patterns in sea level anomalies ($r > 0.7$; P-value < 0.001).

### 4.1 Smoothing and Forecasting

In conventional forecasting methods, the mean of the past $n$ observations is used for future forecasting. This implies an equal weight (equal to $1/n$) for all $n$ data points. However, through forecasting, the most recent observations are usually the best guide for predicting the future. Furthermore, a weighting scheme that has decreasing weights as the observations get older can provide better results. If the weights decrease exponentially when the observations become older, exponential smoothing methods can be very useful for forecasting.

In order to quantitatively determine the best model, the following statistical error measurements were employed to evaluate performance of the models: mean absolute error (MAE), root mean square error (RMSE), mean square deviation (MSD), and mean absolute percentage error (MAPE). The best smoothing parameters were determined as $\alpha = 0.2$, $\gamma = 0.01$, and $\delta = 0.2$, which lead to the least errors for the selected model (Table 2). Figure 5 shows the results of HWES modeling (blue) compared with the observed time series (black). Although the modeling misses some peaks, such as those over 1997–1999, there is still satisfactory agreement between the two time series. The computed Pearson correlation coefficient between the observed and modeled value is high ($r > 0.91$; P-value < 0.001), which means that the selected model is excellent. The correlation coefficient between the forecasted value from the model (red) and observed data is 0.86, which is satisfactory in common model applications. In addition, as shown in Fig. 5, not only the annual signal of sea level anomalies but also the negative trend over 2005–2008 compared to the large positive trend of observed sea level over 2001–2005 can be accurately predicted.

In order to compare the results obtained by the HWES model with other forecasting methods, the ANN model proposed by Cadenas and Rivera (2007) was employed. Figure 6 compares the HWES and the ANN methods against the observed data. Both methods adjust quite well to the real data; however, it is not possible to assess the models in detail based on these figures, so the statistical error measurements for the two models were thus evaluated.

Statistical error measurements for the HWES and ANN methods compared with the observed values (Table 2) show that there is a satisfactory agreement among the models. HWES is thus a useful tool for short-time forecasting of Caspian Sea level anomalies. The main advantage of HWES is its relative simplicity compared to that of ANN. Due to the widespread familiarity of the standard Holt-Winters method, the formulation can be accepted. Furthermore, the method is straightforward to implement (Taylor 2010) and shows satisfactory accuracy based on statistical error measurements.

### 5. CONCLUSION

Sea level anomalies in the Caspian Sea were analyzed and forecasted using HWES. In order to achieve information

| Time Span | Region   | Rate (m yr$^{-1}$) | Time Span | Region   | Rate (m yr$^{-1}$) |
|-----------|----------|-------------------|-----------|----------|-------------------|
|           | Northern |                   | 1993-1995 | +0.17 ± 0.02 | +0.15 ± 0.02     |
|           | Middle   |                   | 1995-1996 | -0.22 ± 0.01 | -0.19 ± 0.01     |
|           | Southern |                   | 1996-2001 | -0.05 ± 0.02 | -0.06 ± 0.02     |
|           | Whole Sea|                   | 2002-2005 | +0.08 ± 0.01 | +0.09 ± 0.01     |
|           |         |                   | 2005-2008 | -0.09 ± 0.01 | -0.08 ± 0.01     |

| Model     | RMSE | MAE | MSD   | MAPE % |
|-----------|------|-----|-------|--------|
| HWES      | 0.07 | 0.05| 0.005 | 22.9   |
| ANN       | 0.06 | 0.05| 0.006 | 20.8   |

Table 1. Rates of Caspian Sea level anomalies in different regions and time span.

Table 2. Statistical error measurements (m) of the forecast methods compared with observed values during the years 2005 - 2008.
continuity over the considered track, a least squares polynomial interpolation was used to fill the gaps and acquired an integrated time series. Significantly higher correlations were found between sea level anomalies in northern, middle and southern regions of the Caspian Sea, which may indicate the same explanation (sounds Volga river discharge) for the observed variations through whole area. HWES provides satisfactory precision and accuracy for supporting water reservoir management plans. HWES is a robust, easy-to-use projection procedure which generally works quite well in practice. It is accentuated significantly due to its simplicity and exactness (based on the analysis of statistical error measurements) which has performed consistently well in more recent forecasting scenarios, when compared with other complicated projection techniques; for example, ANN which has greater computational burden. HWES is thus a useful tool for forecasting sea levels in water bodies with similar stochastic behavior.

Acknowledgements This research is supported by the grants from Tarbiat Modares University (Iran), National Cheng Kung University (Taiwan), and the National Science Council of Taiwan (NSC 100-2221-E-006-234). The co-author, Chung-Yen Kuo, is partially supported by a grant from the National Science Council of Taiwan (NSC 100-2221-E-006-233). Altimeter data products are from AVISO (Archivage, Validation et Interprétation des données des Satellites Océanographiques). We thank anonymous reviewers for their constructive comments. The figures are prepared using the GMT graphics package (Wessel and Smith, EOS 1991).

Fig. 5. (a) Comparison of HWES model values with observed time series, and (b) observed and forecasted values.
REFERENCES

Amirahmadi, H., 2008: The Caspian Region at a Crossroad: Challenges of a New Frontier of Energy and Development (Hardcover). St. Martin’s Press, Retrieved, 112 pp.

Basistha, A., D. S. Arya, and N. K. Goel, 2008: Spatial distribution of rainfall in Indian Himalayas - A case study of Uttarakhand Region. Water Resour. Manag., 22, 1325-1346, doi: 10.1007/s11269-007-9228-2. [Link]

Benada, R. J., 1997: Merged GDR (TOPEX/POSEIDON). Generation B Users Handbook, Version 2.0. Physical Oceanography Distributed Active Archive Center (PODAAC), Jet Propulsion Laboratory, Pasadena, JPL D-11007, 131 pp.

Bowerman, B. L., R. T. O’Connell, and A. B. Koehler, 2005: Forecasting, Time series, and Regression. Duxbury applied series, Thomson Brooks/Cole published.

Cadenas, E. and W. Rivera, 2007: Wind speed forecasting in the south coast of Oaxaca, México. Renew. Energy, 32, 2116-2128, doi: 10.1016/j.renene.2006.10.005. [Link]

Cadenas, E., O. A. Jaramillo, and W. Rivera, 2010: Analysis and forecasting of wind velocity in chetumal, quintana roo, using the single exponential smoothing method. Renew. Energy, 35, 925-930, doi: 10.1016/j.renene.2009.10.037. [Link]

Campos, I. O., F. Mercier, C. Maheu, G. Cochonneau, P. Kosuth, D. Blitzkow, and A. Cazenave, 2001: Temporal variations of river basin waters from Topex/Poseidon satellite altimetry. Application to the Amazon ba-

Fig. 6. (a) Comparison of HWES and ANN model values with observed time series, and (b) observed and forecasted values of HWES and ANN.
Parker, B. B., 1992: Sea level as an indicator of climate and global change. *Mar. Technol. Soc. J.*, 25, 13-24.

Picot, N., K. Case, S. Desai, and P. Vincent, 2006: AVISO and PODAAC User Handbook. IGDR and GDR Jason Products, SMM-MU-M5-OP-13184-CN (AVISO), JPL D-21352 (PODAAC), 3rd Edition, 112 pp.

Rodionov, S. N., 1994: Global and Regional Climate Interaction: The Caspian Sea Experience. Water Science and Technology Library vol. 11, Kluwer Academic Publishers, 241 pp.

Sfetsos, A., 2000: A comparison of various forecasting techniques applied to mean hourly wind speed time series. *Renew. Energy*, 21, 23-35, doi: 10.1016/S0960-1481(99)00125-1. [Link]

Snyder, R. D., A. B. Koehler, R. J. Hyndman, and J. K. Ord, 2004: Exponential smoothing models: Means and variances for lead-time demand. *Eur. J. Oper. Res.*, 158, 444-455, doi: 10.1016/S0377-2217(03)00360-6. [Link]

Stolberg, F., O. Borysova, I. Mitrofanov, V. Barannik, and P. Eghtesadi, 2006: Global International Waters Assessment: Caspian Sea - GIWA Regional assessment 23. Published by the University of Kalmar on behalf of United Nations Environment Programme, 71 pp.

Taylor, J. W., 2004a: Smooth transition exponential smoothing. *J. Forecast.*, 23, 385-404, doi: 10.1002/for.918. [Link]

Taylor, J. W., 2004b: Volatility forecasting with smooth transition exponential smoothing. *Int. J. Forecast.*, 20, 273-286, doi: 0.1016/j.ijforecast.2003.09.010. [Link]

Taylor, J. W., 2010: Exponentially weighted methods for forecasting intraday time series with multiple seasonal cycles. *Int. J. Forecast.*, 26, 627-646, doi: 10.1016/j.ijforecast.2010.02.009. [Link]