**Journal of the Earth and Space Physics, Vol. 43, No. 4, Winter 2018, P. 73-81**

# Reconstruction of Data Gaps in Total-Ozone Records with a New Wavelet Technique

M. Farahani, M.[1]*, Pegahfar, N.[2] and Gharaylou, M.[3]

*1. Associate Professor, Department of Space Physics, Institute of Geophysics, University of Tehran, Iran*
*2. Assistant Professor, Atmospheric Sciences Research Center, Iranian National Institute for Oceanography and Atmospheric Science, Tehran, Iran*
*3. Assistant Professor, Department of Space Physics, Institute of Geophysics, University of Tehran, Iran*

*(Received: 29 May 2017, Accepted: 24 Oct 2017)*

## Abstract

This study introduces a new technique to fill and reconstruct daily observational of Total Ozone records containing void data for some days based on the wavelet theory as a linear time-frequency transformation, which has been considered in various fields of science, especially in the earth and space physics and observational data processing related to the Earth and space sciences. The initial corrupted records consist of six years of daily total Ozone measured by Dobson Photo-Spectrometer Instrument of Institute of Geophysics, University of Tehran. To verify the filled gaps resulted from this technique, the outputs of the proposed method are compared with the Total Ozone Mapping Spectrometer (TOMS) for the year 2005 and Ozone Monitoring Instrument (OMI) for the years 2006 – 2010 satellite data (hereafter used as TOMS/OMI data). The proposed technique consists of three steps: (1) quality control and denoising; (2) data-reconstruction based on Daubechies parent function (DB1); and (3) the combination of approximation and complementary coefficients using the Inverse Discrete Wavelet Transform (IDWT). Results show that this method was able to successfully reconstruct the missing data for gaps lasting no greater than 18 days. For gaps beyond this 18-day limit, however, this method was unable to reconstruct the voided data. As most instruments, including Dobson and Brewer Spectrometer, are working based on the optical interaction of stratospheric Ozone and sunshine, gaps in the Total-Ozone for more than 18 days should happen in atmospheric systems with longevity over 18 days in which overcast clouds persist longer than the 18-day limit. The proposed method could be applied with high efficiency.

**Keywords:** Total-Ozone, Discrete wavelet transform, Missing data, Interpolation, Signal reconstruction.

## 1. Introduction

Ozone is the most abundant photochemical oxidizing agent in the lower stratosphere. During its photochemical production process, the radiative heating of ozone results in short to medium range variations in dynamical properties of the stratosphere. Ozone is also a key factor in climatic changes (Werner, 2008). Quantitative properties of ozone that control the atmospheric variations are the trend of long-term variations in the amount of stratospheric ozone. It has fundamental role in dynamical and chemical processes in atmosphere, suggesting that it has a major contribution on the quality of the results of atmospheric modelling processes. In the lower stratosphere and upper troposphere, density of this gas experiences fluctuation due to atmospheric motions and chemical processes, proposing that ozone can be used as a tracer gas in detecting dynamical processes (Grytsai et al., 2005). As the major source of ozone (creation and destruction) is the Sun, variation of stratospheric ozone concentrations primarily depends on decadal, yearly and seasonal variations of the Sun's position relative to the Earth. In addition to these normal variations, irregular changes occur in time scales of a few days in response to short-time variations of atmospheric variables. For example, changes in ozone amounts due to synoptic systems (high- and low-pressure systems) are short-term (of the order of days); therefore, their dynamical lifecycle lasts up to ten days (Dobson et al., 1929). Duration of synoptic systems affects the trend of time series of ozone. However, it should be noted that in order to examine the non-steady and transient feature of ozone concentration, different continuous and discrete basic functions with finite energy must be used. This means that due to the variation of ozone density in different places

*Corresponding author: mazraeh@ut.ac.ir

(different wavelength), a single wavelength (Fourier transform) could not describe the ozone concentration. In this regard, instead of Fourier transform method that adopts sinusoidal stationary mode, a wavelet transform method with its special characteristics of ability to demonstrate non stationary properties of signals can be used.

Application of wavelet theory to data analysis in different scientific fields has been considerably increased. For example, Candes and Donoho (2002), Donoho and Johnstone (1998), Hall and Penev (2004) and Genovese and Wasserman (2005) have established the application of the wavelet method to signal analysis in various fields. This theory has also been introduced and applied in various meteorological studies. Among them, the time-series analysis of soil changes (Lark and Webster, 2001), examination of the relationship between rainfall and runoff (Labat et al., 2001), simulation of photochemical reactions (Heidarinasab et al., 2004), hurricane boundary layer studies (e.g. Zhu et al., 2010), mountain waves (Woods and Smith, 2010), Elniño Southern Oscillation (ENSO) model validation (Stevenson et al., 2010), continuous wavelet analysis for meteorological data (Wang and Lu, 2010) and drought forecasting (Ozger et al., 2012) can be named.

In atmospheric ozone studies, the results of study of Echer (2004) can be mentioned in which the Meyer wavelet method was employed as a band-pass filter for monthly ozone data to study the solar cycle variations and ENSO phenomena. Werner (2008) also applied the wavelet theory to study the ozone variability with latitudes. One of the problems that challenged scientists in the study of ozone layer is missing data in their records. Therefore, the main purpose of the present study is to introduce and examine an optimized method to reconstruct the missing data. The motivation comes from the filling gaps in ozone data used by Farahani et al. (2012b). They showed that the Spline interpolation method provides an acceptable degree of agreement between the interpolated and observed ozone data (it will be discussed later). They demonstrated that the Spline method has the biggest error in filling gaps that last for more than five days.

In this study, the six-year (2005–2010) Total-Ozone daily measured with Dobson spectrometer in the Institute of Geophysics, University of Tehran, is used as the raw data with many gaps in it and the gaps in it are filled by the proposed method. For validation of the suggested technique, the satellite data of World Ozone and Ultraviolet Radiation Data Centre (WOUDC) were used.

## 2. Wavelet Theory: Definitions and Properties

One of the substantial characteristics of the wavelet transform is its ability in local analysis of the non-stationary signals. Fourier transform lacks this feature for non-stationary signals. These advantages of specifying the location (temporal or spatial) of the discontinuity in a signal is done by the wavelet transform technique accurately. Therefore, the wavelet transform is considered as a prevailing tool in expressing the signals properties.

Theoretically, wavelet is a set of functions used to decompose a continuous signal into its frequencies components, where the resolution of each frequency is proportional to its scale. However, wavelet transform is decomposition of a function based on wavelet functions. The wavelets, known as daughter wavelets, are the transferred and scaled samples of a function (mother wavelet) with a finite length and a highly damped oscillation. Some examples of mother wavelet are Meyer, Morlet, Mexican Hat and Haar. The wavelet transform transfers a time series to frequency domain (using basis functions) and then represents the time series in different time and scales. The continuous transform of wavelet is defined as (Gencay et al. 2002):

$$w(u,s)=\int_{-\infty}^{\infty} x(t)\psi_{u,s}(t)\,\mathrm{d}t\,, \qquad (1)$$

where x(t) is the basis function and the mother wavelet $\psi_{u,s}$ is defined as:

$$\psi_{u,s}(t)=\frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right). \qquad (2)$$

Here, $\psi_{u,s}$ is a function of two continuous variables of u and s. Variable s is the wavelet size parameter, while u indicates the location of the wavelet. By changing these variables, the wavelet transform is able to decompose a time series in different time scales. When the time series is discrete, a discrete mother wavelet should be used (for example, Haar or Daubechies approach). In general, the discrete mother wavelet has the following form:

$$\psi_{j,k} = 2^{j/2}\psi(2^j - k), \quad j,k \in Z \tag{3}$$

where $k$ is the location of the wavelet and $j$ is its size. For example, if $j$ increases, the wavelet size decreases, while its scale and time resolution increases. In general, the coefficients of discrete wavelet transform (DWT; $d_{j,k}$) are obtained from internal multiplication of the time series $x(t)$ by the mother wavelet $\psi_{j,k}$:

$$d_{j,k} =< x(t), \psi_{j,k}(t) >= \int x(t)\psi_{j,k}(t)\,\mathrm{d}t. \tag{4}$$

Mallat (1989) outlined the wavelet theory based on multi-resolution decomposition of a time series. In this theory, a time series can be decomposed into two parts, one part is devoted to the approximations and the other involves the complementaries. In this case, in addition to mother wavelet, another basic function that can accommodate the approximations is required.

$$\phi_{j,k} = 2^{j/2}\phi(2^j - k). \quad j,k \in Z \tag{5}$$

The quadratically integrable function of $g(t) \in \mathrm{L}^2(\mathrm{R})$ can be shown using a parent wavelet and a scalar:

$$g(t) = \sum c_j(k)\phi_{c_j,k}(t) + \sum_k \sum_{j=j}^{\infty} c_j(k)\psi_{c_j,k}(t), \tag{6}$$

where $\sum c_j(k)\phi_{c_j,k}(t)$ is the approximation term or the flattened component of the time series, and $\sum_k \sum_{j=j}^{\infty} c_j(k)\psi_{c_j,k}(t)$ expresses the noise in the time series.

## 3. Methodology

In the present study, six years of daily column Total-Ozone data over the city of Tehran, measured with Dobson instrument located at the Institute of Geophysics (35º 44′ N, 51º 23′ E and 1418.6 meters above the sea level) were used as the raw data. The data set consists of average daily measured Total-Ozone of 2005 to 2010, while the OMI data were used after statistical quality control as our target objective or standard values. The time series and annual trends of the satellite and Dobson spectrometer data are plotted in Figure 1.

A nine-degree polynomial is overlapped to both data (observation and TOMS/OMI data) sets to illustrate the trend of the Total-Ozone variation from 2005 to 2010 (Figure 1). It also shows discontinuity in both Dobson measured and TOMS/OMI data sets with 2 major data lacks indicated with red circles. To expose more features of the data, monthly averaged values of Dobson and TOMS/OMI data sets are plotted in Figure 2. It shows (as expected) that the maximum concentrations of ozone occur in February, March and April, while the minimum values belong to October and November.

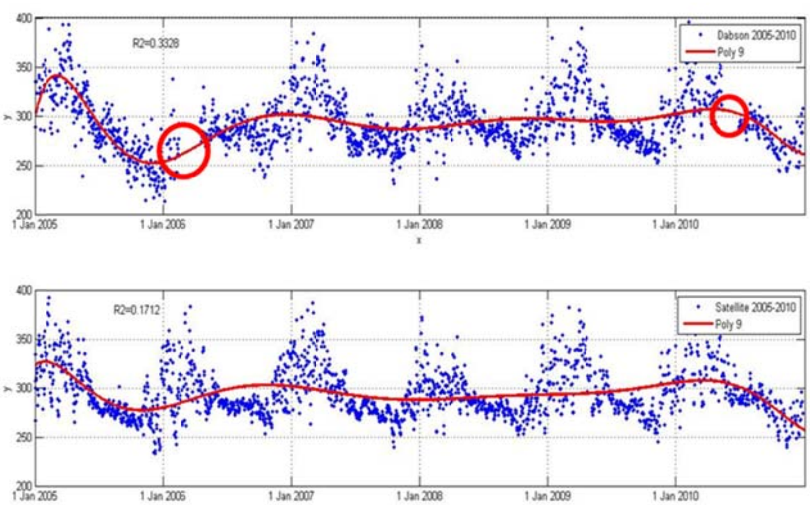


**Figure 1.** Time series of the Total-Ozone measured by Dobson instrument (top panel) and TOMS/OMI satellite (bottom panel) from 2005 to 2010. Thin red lines, show the trend line. The longest data gaps are marked by the red circles.

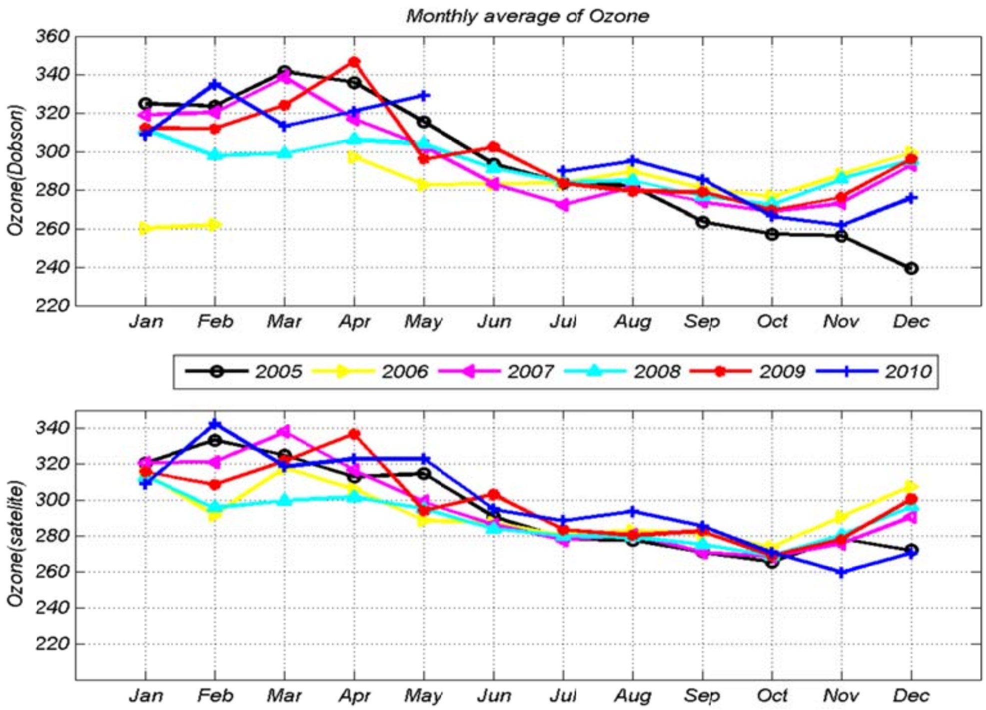


**Figure 2.** Monthly averaged values of Dobson (top panel) and TOMS/OMI satellite data (bottom panel).

Here we reconstruct the Dobson missing Total-Ozone data including two major parts (48 days in 2010 and 31 days in 2006) along with other gaps of lesser duration and compare the results with the satellite data as our reference. These long gaps are due to the calibration/intercomparison and technical problems of Dobson instrument during those years.

As the reconstructed values ultimately will be compared with the TOMS/OMI satellite data, the time overlap between Dobson and TOMS/OMI satellite data should be specified at the first step. Therefore, calendars of the missing data for both data sets were tabulated.

It would be informative to have the knowledge of the correlation between the satellite and Dobson data. To do so, determination coefficients have been calculated for each season over the whole period of study based on the R-squared values (Table 1). Table 1 reveals that the maximum $R^2$ is during winter, while the minimum value is in summer. The maximum value is obtained for autumn 2010 ($R^2 = 0.9623$) and the minimum $R^2$ value is found in 2005 ($R^2 = 0.0596$).

**Table 1.** Determination coefficient ($R^2$) between Dobson and TOMS/OMI for seasons of 2005-2010.

| Year / Season | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|
| Autumn | 0.0596 | 0.6608 | 0.7877 | 0.9505 | 0.9169 | 0.9623 |
| Winter | 0.6071 | 0.1949 | 0.8932 | 0.8861 | 0.9538 | 0.9384 |
| Spring | 0.5312 | 0.7273 | 0.8938 | 0.7586 | 0.9365 | 0.8630 |
| Summer | 0.3692 | 0.1018 | 0.2057 | 0.7127 | 0.6116 | 0.5868 |

In order to have some primary information about our data and general idea about nature of records including the missing data, the efficiency of some ordinary interpolation methods to fill the data gaps are evaluated. These methods include "Nearest Neighbor", "Linear", "Spline" and "Piecewise Cubic Hermite Interpolating Polynomial (PCHIP)". For concluding and evaluating the productivity quality with regards to the accuracy of each method, the index of agreement for each year has been computed. The index of agreement is defined as:

$$IOA = 1 - \frac{\sum_i (O_i - P_i)^2}{\sum (|O_i - \bar{O}| + |P_i - \bar{O}|)^2}, \qquad (7)$$

where $P$ is the interpolated value, $O$ is the observed value and subscript $i$ stands for sequential sampling index. $\bar{O}$ represents the average value of observations. The greater the values of this index relative to 0.5, the better are the results achieved for $P$ by the applied method (Willmott, 1981).

Farahani et al. (2012a) have examined the poor performance of each interpolation method over the longer-duration gaps. They concluded that PCHIP method produces the highest index of agreement.

## 4. Results

The process of computation consists of two steps: 1- Data denoising, and 2- data reconstruction. The details of these two steps are presented as follows.

### 4.1. Data Denoising

Results of the first step (section 3) indicate that both original data sets from Dobson instrument and TOMS/OMI satellite have acceptable degree of scattering. Therefore, the comparison of reconstructed data with the original satellite data may impact the errors. In many cases, some pre-processing operations on the original data set are required before the main interpolation process. Denoising of data is one of those preprocessing calculation. Results of examining different methods show that the Gaussian Function produces the best line fit to the available data. The best degree of Gaussian fitted functions to denoise the Dobson and satellite Total-Ozone data for each year are listed in Table 2. The best fitted Gaussian functions to the signals along with the data sets from Dobson and TOMS/OMI satellite are plotted in Figure 3 (for 2005 to 2007) and Figure 4 (for 2008 to 2010).

**Table 2.** Degrees of Gaussian function to denoise Dobson and TOMS/OMI records.

| Year | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|
| Original signal of Dobson | | | | | | |
| Degree | 5 | 3 | 6 | 5 | 6 | 5 |
| TOMS/OMI | | | | | | |
| Degree | 5 | 5 | 5 | 5 | 7 | 6 |
| Reconstructed signal after using wavelet theory | | | | | | |
| Degree | 8 | 5 | 6 | 5 | 7 | 6 |

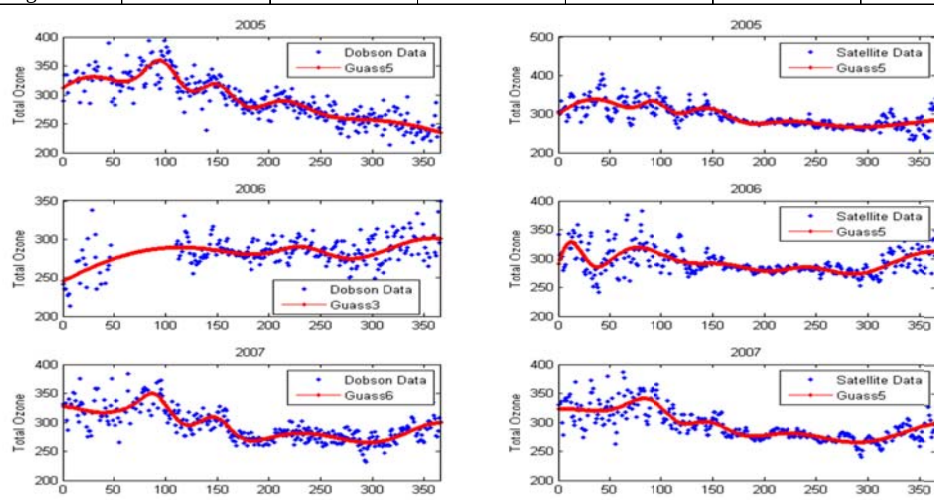


**Figure 3.** The best fitted Gaussian function to denoise the Dobson signal (left) and the satellite signal (right) during 2005 (top), 2006 (middle) and 2007 (bottom). The x axis displays the days of the year.

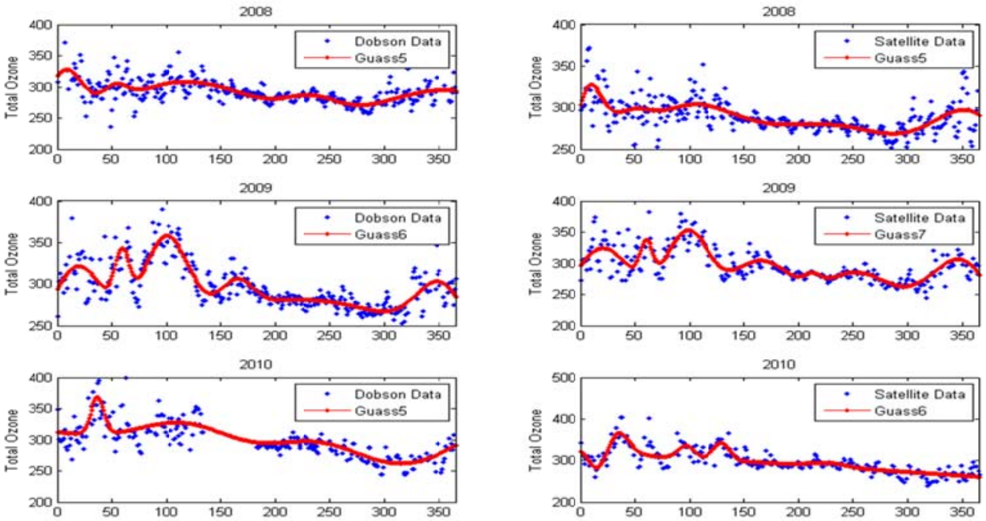


**Figure 4.** The same as Figure 3, but for 2008 (top panel), 2009 (middle) and 2010 (bottom).

## 4.2. Data Reconstruction Using Wavelet Transform Technique

Discrete Wavelet Transform of a signal results in transforming the signal into a set of approximation and complementary terms. The derived coefficients approximately represent the most of the features of the signals but with noises modulated on the main signals. Therefore, a threshold value is defined to filter out the noises imbedded in the coefficients. This filtering process leads to a new set of coefficients. Then an Inverse Discrete Wavelet Transform (hereafter DWT) is performed using the new wavelet coefficients and the previous approximation coefficients to get the same number of samples.

Beginning with DWT of the original signal of year 2005 (Figure 5-a) and based on Daubechies mother function (Figure 5-b) the approximation and complementary coefficients are calculated. Results of the decomposition process of the original signal for year 2005 leading to approximation and complementary coefficients are plotted in **Figure**s 5-c and 5-d.

As the achieved coefficients in this way are representatives of the original signal, every operation on them is similar to the operation on the original signal. Innovation of the present study is from here on, where according to the signals and the calculated coefficients we are facing sectors in them which lack definite values for coefficients. To this end, the approximation and complementary coefficients have been denoised by applying a threshold value based on the largest measured values of the Total-Ozone in the corresponding signal. The approximation and complementary coefficients were interpolated, following the work of Farahani et al. (2012a). They obtained the best ordinary interpolation method to fill gaps in Total-Ozone data set. According to their results for the years 2005–2010, the best interpolation method was found to be Spline, and PCHIP, respectively.

Farahani et al. (2012a) used some methods of interpolation for filling the gaps of total ozone for the same period as in this study. The poor quality of results urged us to apply a rather complicated interpolation method in approximation and complementary coefficients for the considered period. Ultimately, the revised original signals were derived based on the reconstructed coefficients using Inverse Discrete Wavelet Transform. The outcome of this modified signal of 2005 is plotted in Figure 5-e. For comparison, the observed Total-Ozone signals measured by TOMS/OMI satellite are superimposed in Fig 5-e. Similar computations were carried out for years 2006–2010. To indicate the consistency between reconstructed signals and reference signal (measured signals by TOMS/OMI satellite), the index of agreement is evaluated. Resulted values of this index for two signals are presented in Table 3.

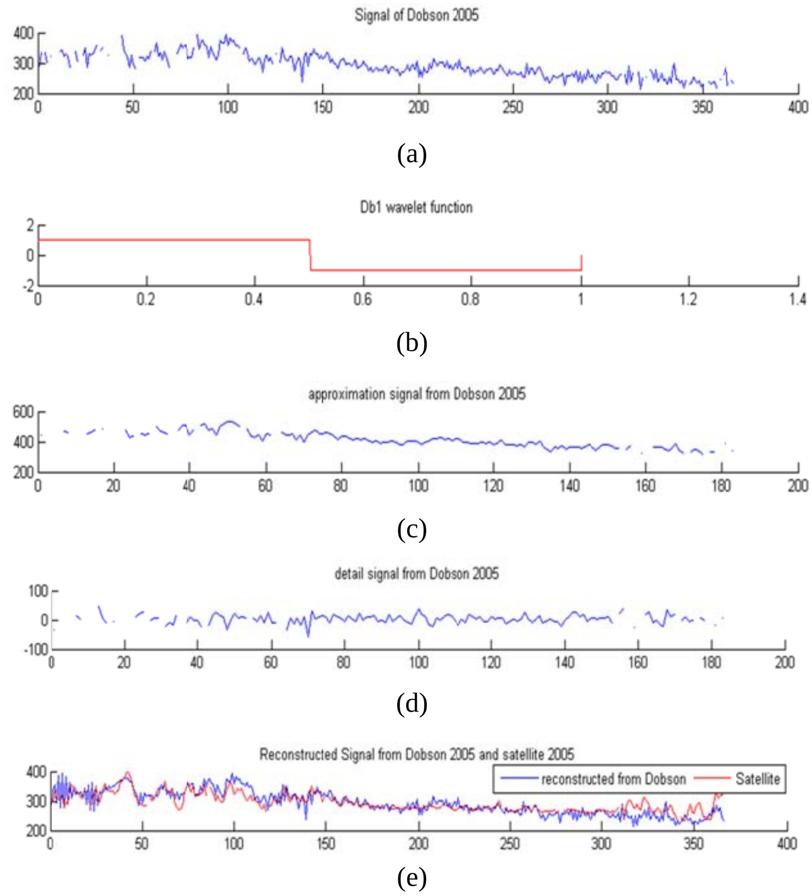


**Figure 5.** Different steps of application of wavelet transform and inverse wavelet transform. (a) Original Dobson signal for 2005, (b) mother wavelet (Db1), (c) approximation coefficients, (d) complementary coefficients and (e) reconstructed signal (after interpolation), as well as the signal measured by the satellite at the same time for year 2005.

**Table 3.** Index of agreement between the reconstructed and TOMS/OMI records.

| Year | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|
| IOA Before using wavelet | 0.8214 | 0.3755 | 0.7718 | 0.3928 | 0.7996 | 0.8566 |
| IOA After using wavelet | 0.9832 | 0.9128 | 0.9518 | 0.9602 | 0.9859 | 0.9533 |

The degree of the Gaussian functions fitted to the original signal (with gaps) and the reconstructed one (resulted by applying the wavelet transform to the original signal) are similar for 2007, 2008 and 2010, although they have different Gaussian coefficients (Table 2). To illustrate this issue, the two Gaussian functions fitted on both signals (original and reconstructed) for each year are plotted in Figure 6. This figure shows that there is an acceptable agreement between two trends during 2007, 2008 and 2009 and for the others are reasonable. As we remember, in 2006 there is the biggest gap of data between all records (48 days), and reconstructing the record of this year with IOA of 0.9128 shows the ability of this method to make up the gaps. Reminding the readers that IOA of this year before filling the gaps is about 0.3 and changing this value to about 0.9 is remarkable. For year 2010 with second biggest gap (31 days) in it, the situation is a little different, although the IOA after reconstruction is quite good (about 0.9), but before filling the gaps, the IOA had an acceptable value too. Variation of IOA for considered years before and after applying the wavelet is listed in Table 3. The values Larger than 0.5 point to the fact that the observed difference between the trends fitted to the original and reconstructed signals in each figure is negligible.

## 5. Discussion and Conclusions

To retrieve the information from a time-dependent variable in a record and to reveal the underlying dynamics that corresponds to the recorded signal, a different but proper signal processing technique is required. Generally, the signal processing process consists of transforming a time varying variable record, signal, into another suitable function with different domain based on the purpose of the process. The purpose of signal processing is to extract the characteristic information embedded within the time series that is not readily observable in its original form. The measured daily mean Total-Ozone in long time, make it as a signal with all unclear characteristic of Total-Ozone especially the time variation of it. In most of signals obtained through measurement in laboratory, continuity is almost guaranteed but in such a large laboratory as the Earth, measurements frequently contain gaps, and discontinuity in signals is a very common feature. Discontinuity of a signal makes retrieving the underlying feature of that signal more worthwhile. In the previous study conducted by Farahani (2012a), they tried to indicate that the ordinary methods of interpolation such as Nearest Neighbor, Linear, Spline and PCHIP adequately retrieve the missing data for gaps that last less than five days, and that they are not applicable for longer gaps. Following that study, exposing the shortcoming of ordinary interpolation methods motivated us to do the current study to fill the Total-Ozone data gaps by utilizing a complex method known as the wavelet method (even for longer than 5-day gaps). The proposed method based on the application of wavelet transform proved that this method is appropriate to fill the gaps of the Total-Ozone data for the periods of longer than five days and shorter than 18 days. This latest limitation of applying this method for gaps shorter than 18 days has been found through the computation and mathematical conception, and readers should not conflict it with two larger gaps (31 and 48 days) that we filled with the current technique. Results indicate that there is a good agreement (based on IOA) between the TOMS/OMI satellite signals as our reference and reconstructed Dobson signals (using wavelet technique to fill the gaps). Ultimately, the fitted Gaussian functions on the original and reconstructed Dobson signals showed an acceptable consistency. We suggest that this proposed method based on wavelet theory is applicable in filling gaps for up to 18 missed data.

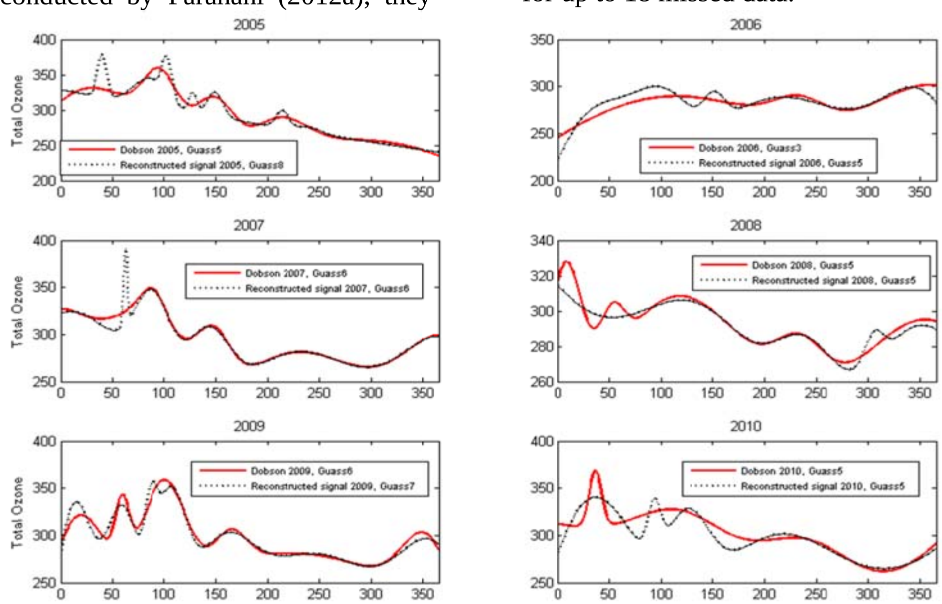


**Figure 6.** The best Gaussian functions fitted to the original and reconstructed signals during 2005–2010.

## References


Candes, E. J. and Donoho, D. L., 2002, Recovering edges in ill-posed inverse problems: optimality of curvelet frames. Ann. Statist. 30, 784–842.

Dobson, G. M., Harrison, D. N. and Lawrence, J., 1929, Measurements of the amount of ozone in the Earth's atmosphere and its reaction to other geophysical conditions, Part III, Proceedings of the Royal Society, London, Series A 122, 456–486.

Donoho, D. L. and Johnstone, I. M., 1998, Minimax estimation via wavelet shrinkage. Ann. Statist. 26, 879–921.

Echer, E., 2004, Multi-resolution analysis of global total ozone column during 1979–1992 Nimbus-7 TOMS period. Ann. Geophys., 22, 1487–1493.

Farahani, M. M., Pegahfar, N. and Gharaylou, M., 2012a, Application of various methods in reconstruction of gaps of total ozone data. 15th Iranian Geophysics Conference, Tehran, Iran.

Farahani, M. M., Pegahfar, N. and Gharaylou, M., 2012b, New Method for filling data Gap of total Ozone. 39th COSPAR Scientific Assembly - Mysore, India.

Genovese, C. R. and Wasserman, L., 2005, Confidence sets for nonparametric wavelet regression. Ann. Statist. 33, 698–729.

Grytsai, A., Grytsai, Z., Evtushevsky, A., Milinevsky, G. and Leonov, N., 2005, Zonal wave number 1–5 in planetary waves from the TOMS total ozone at 65°S, Ann. Geophys., 23 (5), 1565–1573.

Hall, P. and Penev, S., 2004, Wavelet-based estimation with multiple sampling rates. Ann. Statist. 32, 1933–1956.

Heidarinasab, A., Dabir, B. and Sahimi, M., 2004, Multi-resolution wavelet-based simulation of transport and photochemical reactions in the atmosphere, Atmos. Environ., 38, 6381-6397.

Labat, D., Ababou, R. and Mangin, A., 2001, Introduction of wavelet analyses to rainfall/runoffs relationship for a karstic basin: the case of Licq–Atherey Karstic system (France). Ground Water, 39, 605–615.

Lark, R. M. and Webster, R., 2001, Changes in variance and correlation of soil properties with scale and location: analysis using an adapted maximal overlap discrete wavelet transform. European Journal of Soil Science, 52, 547–562.

Mallat, S. A., 1989, theory for multi-resolution signal decomposition: The wavelet representation. IEEE Trans. Pat. Anal. Mach. Intell., 11, 674-693.

Ozger, M., Mishra, A. K. and Singh, V. P., 2012, Long Lead Time Drought Forecasting Using a Wavelet and Fuzzy Logic Combination Model: A Case Study in Texas. J. Hydrometeor, 13(1), 284-297.

Stevenson, S., Fox-Kemper, B., Jochum, M., Rajagopalan, B. and Yeager, S.G., 2010, S.G. ENSO Model Validation Using Wavelet Probability Analysis. J. Climate., 23(20), 5540-5547.

Wang, K. N. and Lu, C., 2010, Two-Dimensional Continuous Wavelet Analysis and Its Application to Meteorological Data. J. Atmos. Oceanic Technol., 27(4), 652-666.

Werner, R., 2008, The Latitudinal Ozone Variability Study using Wavelet Analysis, Journal of Atmospheric and Solar-Terrestrial Physics, 70(2-4), 261-267.

Willmott, C. J., 1981, On the validation of models. Physical Geography, 2, 184-194.

Woods, B. K. and Smith, R. B., 2010, Energy Flux and Wavelet Diagnostics of Secondary Mountain Waves. J. Atmos. Sci., 67 (11), 3721-3738.

Zhu, P., Zhang, J. A. and Masters, F. J., 2010, Wavelet Analyses of Turbulence in the Hurricane Surface Layer during Landfalls. J. Atmos. Sci., 67(12), 3793-3805.