



# Singular spectrum analysis, harmonic regression and El-Nino effect on total ozone (1979–1993) over India and surrounding regions

CHANDRAMADHAB PAL

*Department of Physics, Ramakrishna Mission Vidyamandira, Belurmath, Howrah 711 202, India.  
e-mail: cmpmagra@gmail.com*

MS received 29 December 2016; revised 8 June 2017; accepted 12 November 2017; published online 6 July 2018

The Total Ozone Mapping Spectrometer (TOMS) is a satellite instrument that records Total Column Ozone (TCO) concentration (in DB) of the atmosphere in the form of different overpass files. We have selected 23 sites over India (15), Pakistan (4), Bangladesh (1) and adjoining China (3) to investigate the TCO scenario over this region. About 114,000 historical records (1979M1–1992M12) from 23 overpass files were processed to generate 23 monthly mean time series (TS) of TCO. Statistical analysis followed by singular spectrum analysis (SSA), harmonic regression (HR) and spatial interpolation have been used to accomplish the investigation. Four dominant signals; quasi-biennial signal (QBI,  $T = 30.12$  months), quasi-annual signal (QAN,  $T = 19.69$  months), annual signal (ANN,  $T = 12$  months) and semi-annual signal (SAN,  $T = 6$  months) were discerned to explain the variability. Direct latitudinal effect on the TCO distribution was observed. The variance was limited between 80.53 and 90.13%; ANN contributes 65.93–93.22% followed by SAN 0.58–5.69%, QAN 0.33–5.48%, and QBI 0.06–5.94%. Peak values of the oscillations are estimated from phasor diagrams: QBI, March to May in the mid-latitude; QAN, April, and May; ANN, February to April; SAN, March to May. Incisive pictures of the average distribution and variability of four sinusoids were investigated from contour plots. Two ozone valley were discerned from Spatial interpolation; one over Deccan Plateau in low and other over Tibetan Plateau in high latitude. 179 outliers from  $23 \times 168$  observations have been identified after harmonic regression. The appearance of the outliers is highly consistent with extreme phases of multivariate ENSO Index and Dipole Mode Index.

**Keywords.** Total ozone; singular spectrum analysis; spatial interpolation; multivariate ENSO Index; Dipole Mode Index.

## 1. Introduction

The climatic portrayal of India and its surroundings is heterogeneous, ranging from tropical in the south to temperate and alpine in the Himalayan north, whereas the elevated region receives sustained winter snowfall. The climate is strongly influenced by Thar Desert, Himalayas, Tibetan Plateau, and 7515 km long coastal line around the

vast hydrosphere of the Bay of Bengal, the Indian Ocean, and the Arabian Sea. The Himalayas act as a barrier to the frigid katabatic winds flowing down from central Asia keeping the bulk of the Indian subcontinent warmer than most locations at similar latitudes. As such, the northern parts of the land experience a continental climate with alternate severe summer and winter (Attri and Tyagi 2010; Hassan 2014).

A tropical to sub-tropical climate prevails over India. In the lower western region from  $8^{\circ} \leq \lambda < 23^{\circ}\text{N}$ , is hot-humid whereas the eastern side is savanna with a steppe transition area in between. A far north-east part from  $23^{\circ} \leq \lambda < 37^{\circ}\text{N}$  is humid sub-tropical, but the northwest part has desert-like character. The composition of atmosphere and hydrosphere changes its behaviour due to periodic motion of the earth and different human activities. Therefore, there is a general need to study the nature of trend and periodic variation of all the ingredients. Detailed regression of periodic phenomena in biology and climatology is given by Bliss (1958). Phillips (1984) has reported the utility of transforming climate data from time domain to frequency domain using 20 years dataset of temperature and solar radiation over six cities in the USA. Existence of biennial, annual and semi-annual components in stratospheric ozone were detected by Maeda (1984) applying spherical harmonic. Fourier series was used by Bradshaw and Salazar (1985) to estimate the diurnal nature temperature at mountainous locations over the western United States. Analysis of solar radiation over seed in Oman (Dorvlo 2000) and precipitation climatology in Jordan (Tarawneh and Kadioglu 2003) proved the importance of harmonic analysis of climate data. In the present communication, the author is going to report about the trend and harmonic behavior in monthly time series (1979–1993) of ozone over the land surface of India and its surroundings. SSA followed by a statistical modeling of time series were used to analysis the data documented by 23 ozone monitoring sites over this region.

## 2. Study area

The analysis is limited to the land surface within the latitude (LAT)  $8.4^{\circ}$ – $36.2^{\circ}\text{N}$  and longitude (LON)  $66.9^{\circ}$ – $102.7^{\circ}\text{E}$ . Altitude (ALT) of the sites varies from 11 to 3816 m. Geophysical details of the data recording sites with corresponding Koppen classification (Gabler *et al.* 2007) are given in table 1. We note that extreme climate prevails over the northwest Himalayan region, which is arid with desert/steppe. The topography of the region and locations of the sites are presented in figure 1. The sites are unevenly poised; relatively closer in lower latitude and altitude than in higher latitude.

## 3. Data source

The TOMS instrument was built and launched by Orbital Sciences Corporation of Pomona, California. Of the five TOMS instruments which were built, four entered the orbit successfully. Nimbus-7 provided global measurements of total column ozone on a daily basis over 470 sites and together provided a complete dataset of daily ozone. The Level-2 and 3 data products were prepared by the Ozone Processing Team (OPT) of NASA/Goddard Space Flight Center and available from the NASA Goddard Space Flight Center Distributed Active Archive Center (NASA/GSFC/DAAC). In the present analysis, the data was obtained from <ftp://toms.gsfc.nasa.gov/pub/nimbus7/data/overpass/> (Nimbus-7 satellite, 31 October 1978–06 May 1993). About 114,000 historical records (1979M1–1992M12) from 23 overpass files were processed to generate 23 monthly mean time series (TS) of TCO. The HDF format of the records is arranged according to the Modified Julian Day from 43813.3 (1 November 1978, 07:12) to 49113.2 (4 May 1993, 04:48). On an average, 100 records are missing in each overpass file giving the concentration of atmospheric ozone over 23 selected sites. A program in Excel-2010 was written that directly generates 23 different monthly mean TS (1979M01–1992M12, 168 data) excluding the missing values.

MEI is regarded as the most comprehensive index for monitoring ENSO, since it combines analysis of multiple meteorological and oceanographic components. The index (MEI) is the SST anomaly over the Pacific Ocean was obtained from <http://www.esrl.noaa.gov/psd/enso/mei/table.html>.

Data of SST anomaly over the Indian Ocean called SST DMI (Dipole Mode Index) were derived from [http://www.jamstec.go.jp/frsgc/research/d1/iod/iod/dipole\\_mode\\_index.html](http://www.jamstec.go.jp/frsgc/research/d1/iod/iod/dipole_mode_index.html). It is the gradient of temperature between western equatorial Indian Ocean ( $50^{\circ}$ – $70^{\circ}\text{E}$  and  $10^{\circ}\text{S}$ – $10^{\circ}\text{N}$ ) and south-eastern equatorial Indian Ocean ( $90^{\circ}$ – $110^{\circ}\text{E}$  and  $10^{\circ}\text{S}$ – $0^{\circ}\text{N}$ ).

## 4. Methodology

Since the study of climate to a large extent is the study of the statistics of weather, analysis of the TS has been performed in four steps: statistical overview, singular spectrum analysis, regression diagnoses and spatial interpolations. Multi-Taper Method Toolkit (SSA-MTM) software (Ghil 1997)

Table 1. Geographical details of 23 ozone mapping sites.

Stations	Abbreviation	Latitude (Deg)	Longitude (Deg)	Altitude (m)	General climate	Köppen class	MxT	MnT	Hum
Trivandrum	TRI	8.48	76.95	61	Tropical savanna and monsoon	<i>Am</i>	31.3	23.5	78.1
Kodaikanal	KOD	10.23	77.47	2343	Subtropical highland	<i>Cfb/Cub</i>	17.2	10.5	65.7
Bangalore	BAN	12.98	77.58	921	Tropical savanna	<i>Aw</i>	29.4	19.0	66.2
Chennai	CHE	13.08	80.17	26	Tropical wet and dry, coastal	<i>Aw</i>	32.7	24.6	69.5
Hyderabad	HYD	17.38	78.47	542	Tropical wet and dry	<i>Aw</i>	32.0	20.2	59.3
Pune	PUN	18.53	73.85	559	Hot semi-arid, tropical wet and dry boundary	<i>BSh/Aw</i>	30.2	18.0	61.3
Mumbai	MUM	19.12	72.85	11	Tropical wet and dry	<i>Aw</i>	31.8	22.4	75.1
Nagpur	NAG	21.10	79.05	310	Tropical wet and dry	<i>Aw</i>	33.5	19.8	57.2
Kolkata	KOL	22.65	88.45	64	Tropical wet and dry	<i>Aw</i>	31.7	22.2	72.2
Ahmedabad	AHM	23.02	72.65	56	Hot and semi-arid	<i>BSh</i>	34.4	21.1	63.2
Dacca	DAC	23.72	90.42	17	Hot, wet, and humid tropical	<i>Cfb</i>	30.6	21.5	60.1
Mount Abu	MtA	24.60	72.72	1219	Subtropical highland	<i>BSh</i>	25.2	16.3	58.5
Karachi	KAR	24.87	67.05	6	Arid	<i>Buh</i>	30.0	22.2	61.7
Benares	BEN	25.00	83.00	228	Humid subtropical	<i>Cwa</i>	28.9	19.2	66.4
Kunming	KUN	25.02	102.68	1917	Subtropical highland	<i>Cub</i>	21.8	10.3	69.5
Varanasi	VAR	25.45	83.02	81	Humid subtropical	<i>Cwa</i>	27.9	19.5	64.3
New Delhi	NeD	28.67	77.22	239	Humid subtropical	<i>Cwa</i>	30.9	17.5	56.8
Lhasa	LHS	29.4	91.08	3640	Cool semi-arid	<i>BSk</i>	15.8	1.5	41.6
Quetta	QUE	30.18	66.95	1721	Warm temperate desert scrub	<i>BSk</i>	24.4	6.9	51.5
Lahore	LAH	31.58	74.30	166	Humid subtropical	<i>BSh</i>	30.8	17.8	59.2
Srinagar	SRI	34.08	74.83	1586	Humid subtropical	<i>BSh</i>	18.8	7.3	65.4
Chitral	CHI	35.85	71.83	2718	Cool temperate steppe	<i>BSk</i>	23.0	8.8	61.6
Mt. Waliguan	MtW	36.17	100.53	3816	Arid and semi-arid	<i>BWk</i>	17.0	8.0	49.3

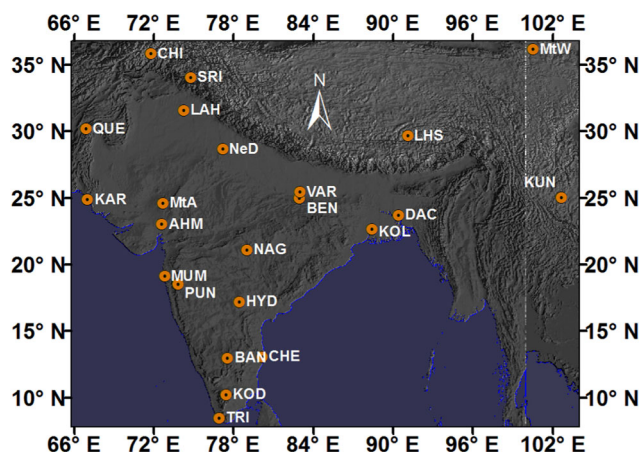


Figure 1. Locations of satellite data recording sites.

was employed to extract common dominant signal frequencies (sinusoids) in 23 different TS. E views, Minitab, and STATA-12 software were used to estimate the significant coefficients of sinusoids. The spatial interpolations of average and amplitude of sinusoids were generated from ArcGIS software.

#### 4.1 Statistical overview

Statistical methods are needed for a whole gamut of activities that contribute to the ultimate synthesis of climate knowledge, ranging from the collection of primary data to the interpretation and analysis of the resulting high-level datasets (Timmermann and Lohmann 2000; Zwiers and Storch 2004). Most atmospheric processes, in simple graphical plots, are visibly non-stationary and heterogeneous due to the inherent seasonal nature of climate (da Costa 2009; Derryberry and De Wayne 2014; Meko 2015). Statistical procedures include the identification of two basic components of TS; trend and seasonality. A synoptic outline of the statistics is given in table 2.

Regression needs the variables should satisfy the basic conditions, homogeneity and stationarity, prior to least squares fitting. The homogeneity of every raw TS was checked according to Standard Normal Homogeneity Test (SNHT). One-tailed  $p$ -value, computed using 10000 Monte Carlo simulations are given (second column) in table 2. As the computed  $p$ -values of the first 16, TS are greater than the significance level (0.05), one can accept the null hypothesis that the TS are homogeneous. However, substantial inhomogeneities exist in seven other TS at high latitude. Similar results were obtained from

Buishand's test and von Neumann's test. Therefore, we have to detrend the TS to improve the statistical picture. In general, most of the TS have inherent linear or slightly non-linear trends. Here cubic smoothing spline (equation 2) is used to detrend the raw TS. The improvement is very clear from comparative studies of  $p$ -values (SNHT) (table 2).

The mean, variance and autocorrelation structure of stationary TS must not change over time. For practical purposes, stationarity can usually be determined from a run sequence plot, a flat looking series. Here the stationarity of raw and detrend TS was checked using KPSS unit root test (Kwiatkowski–Phillips–Schmidt–Shin). The asymptotic critical value (one-sided LM statistics) at 1% level is 0.739, at 5% level is 0.463, and at 10% level it is 0.347 (Kwiatkowski *et al.* 1992). If the LM statistic is greater than the critical values, then the null hypothesis is rejected; that the series is non-stationary. We see that the results returned by the test (table 2) are well below the critical values. Detrend TS are now homogeneous as well as stationary and regression may be applied.

Box plots display variation in samples of a statistical population without making any assumptions about the underlying statistical distribution. Overall statistical characters of detrend TS may be obtained from the descriptive statistics (table 2) and Box–Whisker plot (figure 2). A series of records during Feb–Mar/1979, Feb–Mar/1982 and Dec/1987 are observed around the higher latitude. We note few observations (outliers) over KAR, VAR, NeD, QUE, and LAH, are not covered by the whiskers. Since they appear regularly, some global or local environmental change must be responsible for this temporal regularity. This statistics in advance is very important to continue the regression in a reliable direction. Interquartile ranges of the last six TS (NeD to CHI) are larger than rest of the sites. It appears that the ozone distribution from lowest to mid-latitude (TRI to BEN) is almost identical, but a significant difference is expected at higher latitude.

#### 4.2 Singular spectrum analysis

Singular spectrum analysis (SSA) is essentially a principal components analysis (PCA) in the time domain that extracts information from short and noisy TS (Schoellhamer 2001). Singular Spectrum Analysis–Multi Taper Method (SSA-MTM) Toolkit is a software program that provides insight into

Table 2. Descriptive statistics of raw and detrend ozone time series.

Stations	SNHT test		KPSS test (LM)		Statistical details of detrend time series						Cons.
	Raw	Detrend	Raw	Detrend	Median	IQR	Lower WL	Upper WL	Skewness	Kurtosis	
TRI	0.60	0.71	0.04	0.02	268.57	21.81	14.97	18.78	-0.18	-0.97	0.06
KOD	0.72	0.76	0.04	0.03	268.93	22.69	18.93	13.83	-0.28	-1.05	0.04
BAN	0.89	0.92	0.04	0.03	270.31	26.88	24.27	20.55	-0.25	-1.03	0.05
CHE	0.76	0.84	0.05	0.02	269.38	22.95	24.11	10.33	-0.46	-0.90	0.05
HYD	0.74	0.96	0.03	0.02	269.61	27.78	24.14	11.09	-0.36	-1.06	0.03
PUN	0.67	0.97	0.03	0.02	270.37	31.42	23.11	13.52	-0.26	-1.15	0.03
MUM	0.53	0.92	0.03	0.01	272.56	28.85	22.82	13.39	-0.31	-1.04	0.04
NAG	0.39	0.80	0.04	0.01	272.22	25.97	25.56	13.40	-0.34	-0.88	0.04
KOL	0.32	0.61	0.05	0.01	273.03	24.63	24.92	16.67	-0.21	-0.69	0.06
AHM	0.16	0.55	0.04	0.01	274.67	26.93	29.71	13.80	-0.34	-0.82	0.07
DAC	0.22	0.54	0.05	0.01	273.53	23.81	27.16	19.06	-0.15	-0.59	0.08
MtA	0.10	0.40	0.05	0.01	277.27	25.06	33.97	13.33	-0.39	-0.59	0.10
KAR	0.06	0.35	0.05	0.02	277.77	22.25	33.42	13.66	-0.44	-0.50	0.11
BEN	0.11	0.46	0.06	0.02	275.65	21.20	30.99	17.75	-0.26	-0.44	0.10
KUN	0.23	0.37	0.06	0.01	269.28	25.63	29.89	20.46	-0.19	-0.78	0.08
VAR	0.08	0.45	0.06	0.02	277.24	21.00	31.99	17.42	-0.25	-0.35	0.11
NeD	0.01	0.30	0.09	0.02	282.78	21.09	38.81	26.90	-0.12	0.09	0.23
LHS	0.01	0.45	0.09	0.02	269.90	20.96	28.48	28.58	0.14	-0.22	0.26
QUE	0.00	0.07	0.10	0.03	280.64	18.96	38.09	34.51	0.25	0.42	0.33
LAH	0.00	0.07	0.08	0.03	290.49	21.30	38.41	38.93	0.37	0.45	0.30
SRI	0.01	0.06	0.05	0.02	289.16	31.46	30.50	44.09	0.58	-0.17	0.23
CHI	0.01	0.08	0.04	0.02	294.13	36.85	23.27	44.27	0.61	-0.45	0.18
MtW	0.02	0.19	0.05	0.02	295.99	39.08	20.14	37.92	0.37	-0.85	0.23

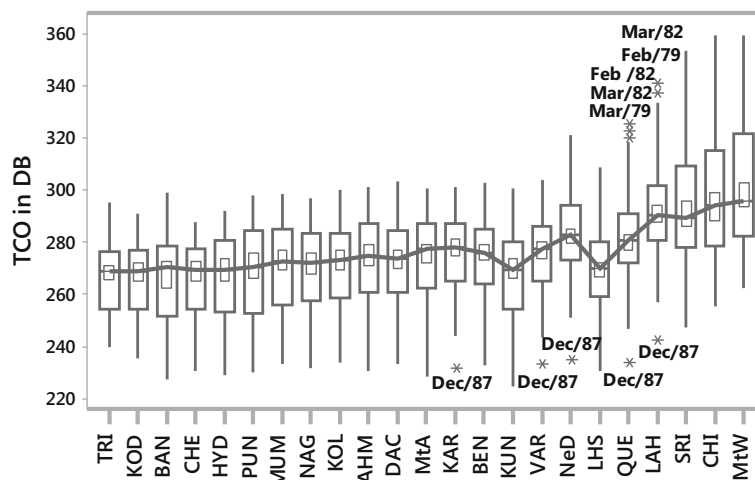


Figure 2. Box-Whisker plot of TCO in the order of latitude.

the unknown or only partially known dynamics of the underlying system which generated the series. The toolkit includes three kinds of power-spectrum estimation. These are the traditional Blackman–Tukey windowed correlogram, multi-taper method (MTM), and maximum-entropy method (MEM). Briefly, the time series is filtered by SSA so that only the statistically significant low-frequency components are retained. MTM attempts to reduce the variance of spectral estimates by using a small set of tapers (Thomson 1982; Percival and Walden 1993) rather than the unique data taper or spectral window used by Blackman–Tukey methods. Averaging over this small ensemble of spectra yields a better and more stable estimate than do single-taper methods. The harmonic analysis attempts to determine directly the (finite) amplitude of a (pure) line in the spectrum of a time series of finite length.

Basic SSA performs four steps: embedding, reconstruction, grouping, and diagonal averaging. At first (embedding), the one-dimensional series is represented as a multidimensional series whose dimension is called the window length. The sole parameter of this step is the window length and it should be proportional to the number of periods, for example, if we have a time series measure monthly we must consider  $L$  proportional to 12 (Vautard *et al.* 1992; Golyandina *et al.* 2001; Hassani 2007; Myung 2009). Evidence obtained using simulation experiments and real data sets indicate how, in practice, an optimal assignment for the window length can be made (Atikur and Poskitt 2013). In the present analysis  $N = 168$ , and best output was found when

$L = 72$ . It allows distinguishing clearly the oscillatory periods in the range  $(L/18, L)$ , and the leading pairs above the distinct break in the eigen value spectrum. Four leading and common oscillatory components with periods 12, 6, 19.69 and 30.12 months were discerned. They are defined here as ANN (annual,  $T = 12$  months), SAN (semi-annual,  $T = 6$  months), QAN (quasi-annual,  $T = 19.69$  months), and QBI (quasi-biennial,  $T = 31.12$  months). MTM time reconstruction with these four eigen triples explains, on average, 87% of the entire temporal scenario of TCO given by 23 TS.

### 4.3 Regression and diagnosis

Primary conditions that the TS must be stationary and homogeneous are satisfied after detrending. Ordinary least squares fitting of each TS with four sinusoids as predictor variables are used in the regression. Procedure and interpretation (Chatterjee and Hadi 2012) of the output are discussed.

#### 4.3.1 Theoretical background

The seasonal components  $Y_i(S)$  with four leading sinusoids and random residuals  $Y_i(R)$  may be expressed as (Bloomfield 2000; Myung 2009; Duchon and Hale 2010):

The oscillatory component  $Y_i(S)$  as the combination of different harmonic pairs is defined as



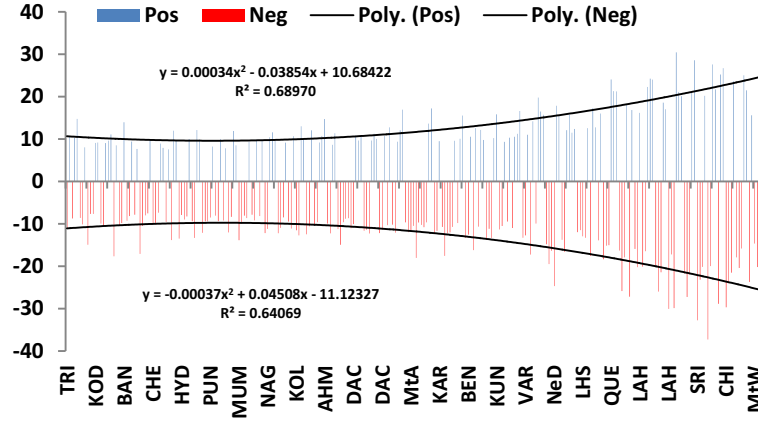


Figure 3. Spatial variation of residuals outside standard deviation (1.8).

$$Y_i(S) = Y_i(AVE) + \sum_{m=1}^4 A_m \cos \frac{2\pi t}{T_m} + \sum_{m=1}^4 B_m \sin \frac{2\pi t}{T_m} + Y_i(R). \quad (1)$$

The amplitude  $\alpha_m$  and the phase  $\phi_m$  of sinusoid are

$$\alpha_m = (A_m^2 + B_m^2)^{1/2}, \quad (2)$$

$$\phi_m = \tan^{-1} \frac{B_m}{A_m}. \quad (3)$$

Since regression coefficients  $A_m$  and  $B_m$  may be positive and negative, four different phasor diagrams are possible. A logical program in Excel-2010 was written to calculate the phase angle  $\phi_m$  directly. The peak month of each sinusoid may be evaluated from:

$$t_{\text{peak}} = \frac{T_m \phi_m}{2\pi}. \quad (4)$$

In general, it is a nine-parameter model (equation 1). But if any of the regression coefficients is statistically insignificant, the resulting model will be reduced to a lower dimension.

#### 4.3.2 Data fitting

Here, we have eight potential predictors which mean we can choose 256 possible models (subsets) in regression. The best subset is selected according to a well-established criterion; lowest Akaike Info Criterion (AIC), lowest Mallows's  $C_p$  and the smallest number of variables to get high  $R^2$  (adj), which should stay more than 60% for the rest of

the models. Initial fitting with the best subset for all the TS generates varied outputs as expected from Box–Whisker plot (figure 2). The variance of regression was more than 80% for 17 sites situated within low to mid-latitude  $8^\circ \leq \lambda < 26^\circ\text{N}$ , but the variance was much lower (67.5–77.1%) for six sites in higher latitude except MtW. The magnitude of residuals outside the standard deviation 1.8 obtained from case wise detection is given in figure 3. We may realize the difficulties in regression around high latitude. The data fitting in regression was repeated judiciously excluding the records giving high residuals to ‘streamline’ the process. Low values of Durbin–Watson (DW) statistics of the fittings (table 3) mean that errors are estimated with bias, and have strong to medium positive autocorrelation over time. Next autoregressive regression was tried. After first-order autoregression, the residuals ( $\varepsilon_t = \rho\varepsilon_{t-1} + e_t$ ,  $|\rho| < 1$ , white noise  $\{e_t\}$ ) were stationary. The expected highest correlation was noted in every correlogram (Lag-1). Jarque–Bera normality test, Breusch–Godfrey serial correlation LM (Lagrange multiplier =  $n \cdot R^2$ ) test and Breusch–Pagan–Godfrey heteroskedasticity test of the residuals are checked after each fitting for additional necessary guarantee. The influential cases or the cases with high leverage (outliers in statistics) were carefully excluded to improve the output. Residual statistics were improved substantially except a slight heteroskedasticity in NeD. So first order autoregression (AR1) should suffice the fitting.

Lowest variance inflation factor (VIF) of every regression confirmed the absence of redundancy of the independent variables. Since the observed F-statistics/prob. ( $10^{-5}$ ) are highly significant in every output, all the independent variables can

Table 3. Statistical details of residuals after least squares fitting.

Stations	DW stat without AC	1st auto corr. coeff.	No. iteration	No. outlier	DF	Normality		Heteroskedasticity	
						Jarque-Bera	Prob.	Prob. Chi <sup>2</sup>	DW stat.
TRI	0.85	0.818	4	6	7	3.96	0.137	0.092	2.14
KOD	0.89	0.816	6	5	6	0.94	0.624	0.115	2.07
BAN	0.98	0.825	4	7	7	1.44	0.485	0.127	2.10
CHE	1.14	0.800	4	5	7	0.99	0.609	0.076	2.09
HYD	1.32	0.813	3	6	8	0.09	0.958	0.151	1.88
PUN	1.23	0.829	4	3	7	2.39	0.303	0.177	1.97
MUM	1.24	0.824	4	2	7	1.16	0.561	0.128	2.03
NAG	1.19	0.815	4	7	7	4.47	0.107	0.068	1.96
KOL	1.36	0.785	5	4	6	1.32	0.517	0.923	1.93
AHM	1.16	0.813	4	7	7	2.38	0.305	0.087	1.90
DAC	1.36	0.788	4	7	7	1.52	0.469	0.070	2.01
MtA	1.19	0.804	6	6	5	0.12	0.941	0.217	1.98
KAR	1.05	0.805	6	7	6	5.79	0.055	0.102	1.95
BEN	1.32	0.782	3	6	8	0.94	0.624	0.057	1.96
KUN	1.28	0.803	4	6	7	2.36	0.306	0.082	2.01
VAR	1.24	0.776	3	7	8	2.01	0.365	0.095	1.85
NeD	0.96	0.746	3	13	8	2.87	0.237	<b>0.037</b>	<b>1.64</b>
LHS	1.14	0.757	5	11	7	0.31	0.857	0.048	1.76
QUE	1.07	0.727	3	16	8	3.90	0.147	0.043	1.72
LAH	1.27	0.716	3	14	8	0.04	0.982	0.047	2.03
SRI	1.34	0.719	3	12	8	4.05	0.132	0.049	2.03
CHI	1.42	0.740	6	14	5	5.57	0.059	0.061	1.95
MtW	1.49	0.790	6	6	5	0.28	0.871	0.094	2.11



influence the TS within the population. It implies that observed  $R^2$  are reliable, and are not a spurious result of oddities in the dataset. The outputs with few statistical parameters are presented in tables 4 and 5. We note that RHO values (inverse of AR-1 roots) are less than one. So the model is stationary and ergodic (Dunn 2007; Brooks 2008). The regression coefficients are statistically significant at 0.05% level of confidence. Since  $R^2$  values are very high ( $0.95 > R^2 > 0.81$ ), all the required criteria of regression are satisfied and infer that the four oscillatory component pairs can strongly influence the output within the sample. We conclude that in TCO the average ozone concentration (AVE) and the amplitude of sinusoids (equation 2) with respective peak month (equation 4) will provide a realistic portrayal of ozone distribution over India and surrounding regions.

#### 4.3.3 Discussion of regression output

The regression output is arranged in table 4 with increasing latitude of the sites. The variances of MTM reconstruction (MEM-RC var.), without auto-regression (initial var.) and with first order autoregression (final var.) are given in first four columns. Substantial improvement in the variance particularly around the high latitude is clearly observed. Tolson (1981) studied the TS (4M1970–4M1977) of total ozone field over the region 90°S–90°N using spherical harmonic model. Here he reported about three periodic components; annual semi-annual and quasi-biennial ( $23 < T < 26$  m). The amplitudes around 24°N were 20 DB (ANN), 3.2 DB (SAN) and 2.2 DB (QBI), which are highly consistent with present observation. We note that on average the annual cycle has the highest contribution (82.10%) to the overall variance of TCO. The rest three components contribute very weakly, but almost equally; QBI 2.83%, QAN 2.77%, and SAN 2.43%.

#### 4.4 Spatial interpolations

Since the sites are sparse and random over the large region with huge geological curvature, we have used statistical model; Ordinary Kriging from ArcGIS to generate contours. They are next piled up with a common layer; shaded relief in the grayscale 3D model for the topographic impression. The necessary digital elevation model (DEM) was obtained from US Geological Survey (<https://lta.cr.usgs.gov/GTOPO30>).

Every shaded relief should be illuminated by the Sun from its natural relative position over the region. It is a method of showing relief on a map by simulating shadows where oblique sunrays illuminate the surface so that slopes behind it are shaded to have a realistic impression of the area. The approximate position of the Sun over the study area is (112° and 42°) all the year round; this value is used to illuminate every image to get the natural topographic impression. Polygon shape file of the landform barrier of the study area was generated from the long coastal range in south-east Asia ([http://www.ngdc.noaa.gov/mgg\\_coastline/](http://www.ngdc.noaa.gov/mgg_coastline/)). It is used to extract only the land surface from the rectangular extent of spatial interpolation. Point shape file and polyline shape files of the average and amplitudes of four harmonic components are generated one after the next. Contour map shows valley, hill, and steepness in the slopes of the third variable. The contour interpolations against terrain land are given in figure 4(a–e). Since the surface is not homogenous throughout, the result cannot be claimed to be definitely realistic and reliable. However, we may guess how the average TCO and four oscillatory components vary over the region.

##### 4.4.1 Average variation

Contour plot of the average distribution of TCO is given in figure 4(a). We see an upward convex slope in the distribution over the north-east corner which converges to an ‘ozone hill’ near Chitral (CHI). The distribution over the southeast region is reversed gradually to a downward concave slope that finally creates an ‘ozone valley’ over Tibetan Plateau (TP) near Lhasa (LHS). The deep over TP was reported earlier (Zhou *et al.* 1995; Bian *et al.* 2011). Contour lines are straight and dispersed over the central India. The steepness in distribution is reduced gradually to almost uniform nature without any slope. Towards the lower latitude a concave slope starts. This ultimately forms another ‘ozone valley’ over the lower flank of Deccan Plateau (DP) near Bangalore (BAN). The weak deep near the converging zone of Western and Eastern Ghat mountain range was also reported earlier (Pal 2010). In general, the average TCO follows a parabolic relation with LAT ( $\lambda$ )

$$\begin{aligned} Oz_{ave} &= 0.0914\lambda^2 - 2.4714\lambda + 281.0, \\ R^2 &= 0.93. \end{aligned} \quad (5)$$

Table 4. Result of least squares fitting.

Station	MEM-RC Var. (%)	Initial Var. (%)	Final Var. (%)	LAT	LON	ALT	AVE	QBI	QAN	ANN	SAN	RHO	n
TRI	90.2	86.5	90.9	8.48	76.95	61	263.64	3.36	1.11	16.11	3.04	0.51	6
KOD	92.1	88.7	92.1	10.23	77.47	2343	265.20	1.72	1.08	17.74	3.02	0.52	5
BAN	94.4	91.1	94.1	12.98	77.58	921	264.41	0.49	2.27	20.17	2.48	0.44	7
CHE	92.3	88.4	92.2	13.08	80.17	26	266.57	1.12	2.57	16.51	4.25	0.39	5
HYD	92.9	91.7	94.5	17.2	78.47	542	266.25	2.85	3.91	20.07	3.17	0.28	6
PUN	94.1	92.7	94.9	18.53	73.85	559	267.38	3.22	3.58	22.02	1.76	0.36	3
MUM	93.4	91.6	94.2	19.12	72.85	11	270.45	3.63	3.80	20.33	2.71	0.36	2
NAG	92.2	89.9	92.7	21.1	79.05	310	270.10	3.24	3.78	18.81	2.95	0.35	7
KOL	88.1	86.7	89.8	22.65	88.45	64	272.49	3.64	4.01	17.89	3.97	0.26	4
AHM	90.7	88.6	91.9	23.02	72.65	56	273.92	3.75	3.91	18.34	2.17	0.34	7
DAC	88	86.3	89.3	23.72	90.42	17	274.56	3.52	3.55	17.65	3.58	0.30	7
MtA	88.8	86.2	90.5	24.6	72.72	1219	275.88	3.80	3.67	16.97	2.01	0.36	6
KAR	87.1	84.2	89.1	24.87	67.05	6	277.98	3.61	3.46	16.09	2.89	0.41	7
BEN	85.9	83.6	88.7	25	83	228	276.46	3.67	3.42	15.74	3.02	0.31	6
KUN	90.2	87.1	91.2	25.03	102.68	1917	268.71	3.46	3.25	19.04	2.33	0.33	6
VAR	84.8	81.7	87.3	25.45	83.02	81	277.52	3.82	3.47	15.47	3.09	0.33	7
NeD	77.8	<b>72.1</b>	80.7	28.67	77.22	239	286.78	2.98	2.25	14.28	2.14	0.34	13
LHS	81.2	<b>76.3</b>	84.2	29.67	91.08	3640	274.04	3.66	3.62	15.77	2.87	0.35	11
QUE	71.8	<b>67.8</b>	80.2	30.18	66.95	1721	290.78	4.27	4.10	14.21	3.68	0.27	16
LAH	72.1	<b>67.5</b>	80.6	31.58	74.3	166	300.00	3.63	2.91	15.99	3.22	0.29	16
SRI	76.8	<b>73.3</b>	84.5	34.06	74.83	1586	306.00	4.63	4.75	25.30	4.62	0.30	12
CHI	81.7	<b>77.1</b>	86.9	35.83	71.83	2718	309.83	3.66	3.88	28.21	5.01	0.21	14
MtW	89.0	82.9	89.8	36.17	100.53	3816	311.44	1.34	2.71	27.70	2.61	0.26	6

Table 5. Variance and phase angle with peak month of sinusoids.

Stations	QBI			QAN			ANN			SAN		
	Var. (%)	Phase	Peak month	Var. (%)	Phase	Peak month	Var. (%)	Phase	Peak month	Var. (%)	Phase	Peak month
TRI	1.87	6.44	Jul	0.25	4.92	May	78.62	3.73	Apr	2.62	4.28	May
KOD	0.30	10.59	Nov	0.53	2.77	Mar	85.83	3.79	Apr	2.81	6.28	Jul
BAN	0.49	0.00	Jan	0.72	3.51	Apr	88.32	3.70	Apr	1.52	4.34	May
CHE	0.60	0.00	Jan	1.45	3.82	Apr	78.81	3.63	Apr	5.04	4.50	May
HYD	2.36	2.36	Mar	2.49	4.18	May	83.73	3.51	Apr	2.27	4.35	May
PUN	2.51	2.52	Mar	1.98	4.92	May	86.64	3.49	Apr	0.66	4.20	May
MUM	3.40	2.85	Mar	2.46	4.92	May	82.97	3.42	Apr	1.57	4.30	May
NAG	4.15	3.15	Apr	3.62	4.92	May	78.33	3.31	Apr	1.86	4.09	May
KOL	3.69	3.82	Apr	3.19	4.92	May	74.01	3.07	Apr	3.88	4.10	May
AHM	4.82	3.93	Apr	3.38	4.92	May	76.49	3.24	Apr	1.44	3.87	Apr
DAC	3.53	4.11	May	3.12	4.92	May	73.95	2.99	Mar	3.60	4.17	May
MtA	4.57	4.26	May	3.00	4.12	May	69.43	3.14	Apr	1.69	3.51	Apr
KAR	5.53	4.31	May	3.52	4.13	May	66.91	3.06	Apr	2.50	3.14	Apr
BEN	5.42	4.08	May	3.78	3.93	Apr	68.17	2.94	Mar	2.72	3.88	Apr
KUN	2.90	4.52	May	2.28	3.99	Apr	74.38	3.08	Apr	1.33	3.95	Apr
VAR	4.34	4.29	May	2.97	3.81	Apr	66.30	2.89	Mar	2.94	3.84	Apr
NeD	5.02	4.81	May	4.05	3.36	Apr	57.74	2.53	Mar	2.71	3.14	Apr
LHA	2.47	4.98	May	1.94	3.63	Apr	65.55	2.24	Mar	2.97	3.14	Apr
QUE	5.54	4.77	May	3.27	3.78	Apr	56.09	2.08	Mar	6.22	2.45	Mar
LAH	4.32	5.10	Jun	2.98	3.06	Apr	61.25	1.95	Feb	4.70	2.59	Mar
SRI	2.33	4.82	May	1.53	3.34	Apr	76.06	1.66	Feb	3.86	2.74	Mar
CHI	1.74	4.53	May	1.20	4.92	May	83.59	1.57	Feb	3.50	3.14	Apr
MtW	0.20	10.59	Nov	0.81	4.14	May	85.81	1.57	Feb	1.92	2.63	Mar

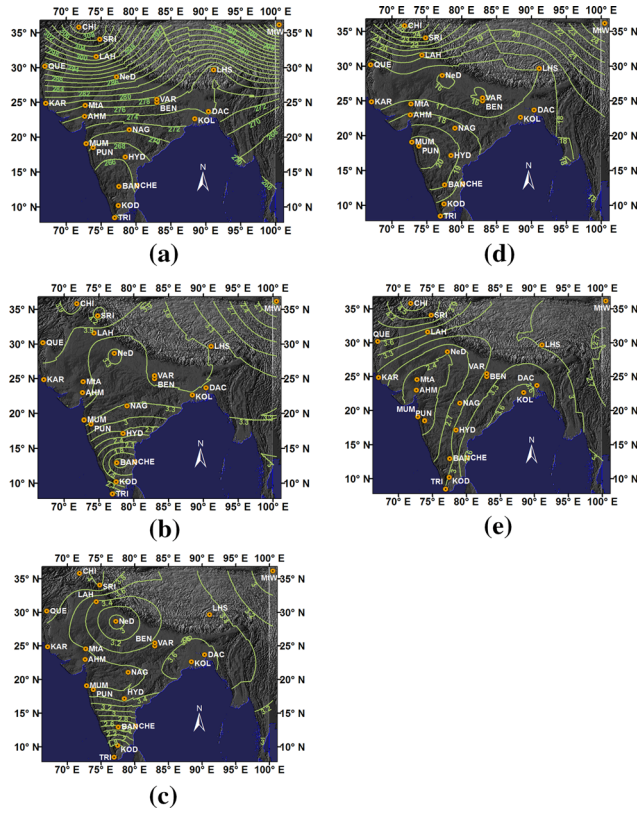


Figure 4. (a) Spatial distribution of average value in TCO. (b) Spatial distribution of quasi-biennial amplitude in TCO. (c) Spatial distribution of quasi-annual amplitude in TCO. (d) Spatial distribution of annual amplitude in TCO. (e) Spatial distribution of semi-annual amplitude in TCO.

#### 4.4.2 Sinusoidal variations

Quasi-biennial variation of ozone (figure 4b). The contours here are seen to converge within the familiar downward-pointing triangle of the Indian subcontinent's coastline. However, no clear pattern is seen elsewhere. The amplitude of QBI approximately follows cubic relation

$$Oz_{qbi} = -0.0009\lambda^3 + 0.0581\lambda^2 - 1.0123\lambda + 7.3246, \quad R^2 = 0.53. \quad (6)$$

This QBI cycle attains peak values between March and May except at the lowest and highest latitudes (tables 3 and 4). Gray and Dunkerton (1990) and Baldwin *et al.* (2001) reported that QBI variation in TCO is due to quasi-biennial oscillation ( $T = 28 \pm 4$  months) in the mean zonal winds around the equatorial lower stratosphere.

Quasi-annual variation of ozone (figure 4c): Contours of the QAN amplitude are curved with a positive gradient from Nilgiri range to entire DP. A weak 'sink' is apparent over NeD. The amplitude changes in a parabolic manner

$$Oz_{qan} = 0.0006\lambda^3 - 0.0492\lambda^2 + 0.3857\lambda - 1.2627, \quad R^2 = 0.66. \quad (7)$$

The peak values of QAN swings between two months, April and May all over the region.

Annual variation of ozone (figure 4d): The annual amplitude has a strong peak over western DP. After a minimum over the entire central region, the intensity is boosted except over southern TP. Intensity of ANN follows a cubic relation with LAT

$$Oz_{ann} = 0.0061\lambda^3 - 0.3898\lambda^2 + 7.6229\lambda - 26.054, \quad R^2 = 0.83. \quad (8)$$

Annual cycles attain maximum in February to April, starting from higher to lower latitude. The peak period of ANN around the mountainous region leads about two months that of around the aquatic region.

Semi-annual variation of ozone (figure 4e): Semi-annual part of the contour plot represents a 'minimax point' response surface. Here slopes of the orthogonal components defining the surface become zero (stationary point), but this is not a local extremum on both axes. The stationary point (also called saddle point) is observed over TP. Two eyes are also created, one over CHI and other over KOL. In general SAN oscillation attains peak values within March to May.

## 5. Outliers and global influences

We have observed that 242 data out of 3864 to remain outside the standard deviation of 1.8 after initial fitting. However, the number is reduced to 179 (table 4, last column) after best fitting and they are outliers. The outliers are highly influential; may pull regression relationships away from their true best fit to bias the regression coefficients (Cook 1977; Stevens 1984). They may be due to: (i) instrumental defects, (ii) sudden changes like volcanic eruption, earthquake, forest fire (Kita *et al.* 2000), ENSO, NAO, etc., in the environment. Transient changes will definitely modify the chemistry in atmosphere and hydrosphere including many others. The signatures will be carried by the historical record which will appear as 'outliers' in regression.

We define a number  $n$  as the frequency of outliers to understand the influence of outliers on the ozone distribution. Since in the present analysis, we

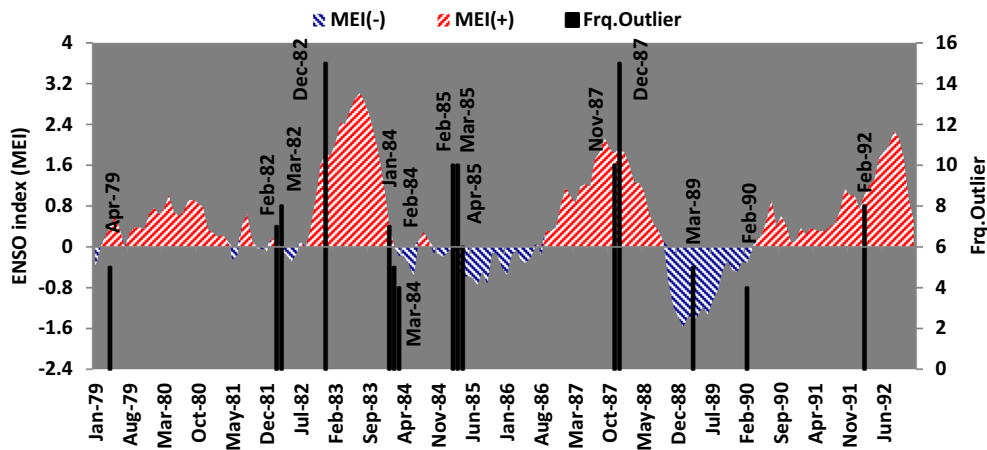


Figure 5. Temporal appearance of outliers against MEI.

have 23 sites the frequency in any particular date is limited as  $0 \leq n \leq 23$ .

The direct effect of SLP and SST gradients on ozone concentration is a well-established environmental phenomenon (Ziemke and Chandra 2003; Zeng and Pyle 2005; Doherty *et al.* 2006; Ganguly and Iyer 2006). Maruyama and Tsuneoka (1988) observed that the rapid descent of QBO westerly was accompanied with the El-Nino; longer westerly in the years 1965, 1968, 1976, 1979, 1982/83, 1986/87 and 1991/92, and shorter easterly in the years 1982/83 and 1987. Recently Sahu *et al.* (2013, 2014, 2016) have reported the impact of tropical convection and ENSO on the seasonal and inter-annual variability of TCO over the study area.

Therefore, the systematic appearance of outliers must be linked with some anomaly in nature which exactly happened here. Temporal appearance of the outliers was critically investigated with SLP and SST anomalies to find any teleconnection. Multivariate ENSO Index (MEI) is favoured over the conventional indices since it combines the significant features of all the observed surface fields in the Tropical Pacific. MEI is determined as the first principal component of six main observed variables over the Tropical Pacific (Wotler and Timlin 1998).

In statistics, cross-correlation is a measure of similarity of two series as a function of the displacement of one relative to the other. The cross-correlation coefficient between the seasonal variation of MEI data and outliers increases from 0.021 reaching up to 0.312 when MEI time series leads to two to months. In figure 5, the bar plot (black vertical lines with date) of outlier-frequency ( $n \geq 4$ ) is coupled with MEI variation. Strong peaks of outliers are observed in the years 1982

(Feb, Mar, Dec), 1984 (Jan), 1985 (Feb, Mar, Apr), 1987 (Nov, Dec), 1989 (Mar), and 1992 (Feb). Consistency between the appearance of outliers and ENSO events is established.

The large hydrosphere of Indian Ocean through Arabian Sea and Bay of Bengal is connected with the concerned land surface. The Indian Ocean Dipole (IOD) phenomenon has been catalogued as an important manifestation of the tropical air-sea interaction (Behera *et al.* 1999; Saji *et al.* 1999; Webster *et al.* 1999). The SST gradient of Indian Ocean between the western equatorial ( $50^{\circ}$ – $70^{\circ}$ E and  $10^{\circ}$ S– $10^{\circ}$ N) and the southeastern equatorial Indian Ocean ( $90^{\circ}$ – $110^{\circ}$ E and  $10^{\circ}$ S– $0^{\circ}$ N) is specified by Dipole Mode Index (SST DMI). It is also investigated whether TCO variation over the concerned region is independent of IOD events or not. The cross-correlation coefficient between seasonally stratified DMI data and outliers increases from 0.036 reaching up to 0.373 if DMI time series are ahead of three months of the stratified outliers (figure 6).

Therefore, we are apt to conclude that TCO vitiation over this region lags two months behind ENSO and three months behind IOD events. The high (+) and low (–) phase of MEI and DMI episodes are highly consistent with the temporal appearance of outliers.

Delayed influences of SLP and SST of the vast hydrosphere are established directly from this new and easy approach. We observe an outbreak of comparatively maximum outliers (13 on average) around the north-west region; QUE, LAH, SRI, NeD. Our atmosphere is a dynamical system. Processes like deep convective outflow, shallow convection, stratosphere–troposphere



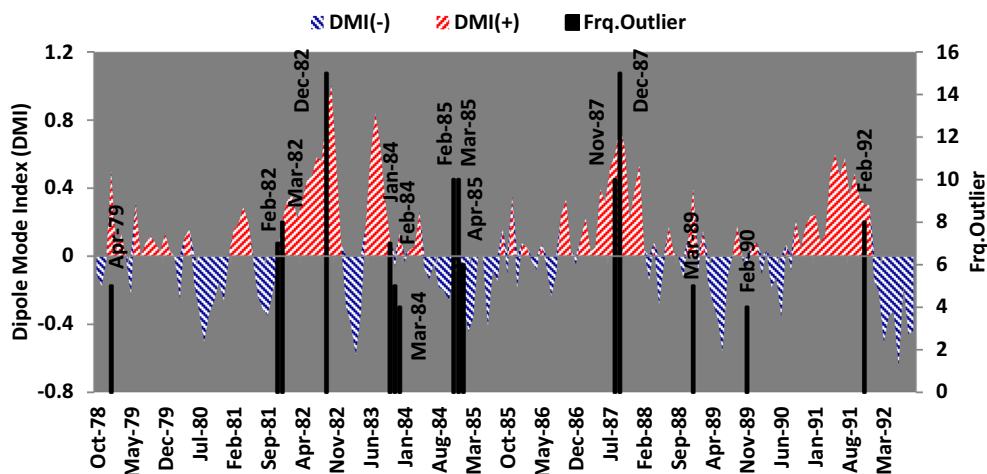


Figure 6. Temporal appearance of outliers against DMI.

exchange, convective downdraft, convective overshooting, etc., play important roles in the distribution of ozone (Sherwood and Dessler 2003; Sahu *et al.* 2014). The extreme climate (table 1) over this northwest Himalayan region, different dynamical processes and huge loading of chemical pollutants (CO, CO<sub>2</sub>, NO<sub>x</sub>, SO<sub>2</sub>, VOC, burning of crude oil) due to the intermittent military and militant interaction may be correlated with the anomalous distribution of TCO.

## 6. Conclusions

The effective harmonics in the monthly time series of TCO over India, Pakistan, Bangladesh, and a part of adjoining China is determined using singular spectrum analysis. This is definitely a new idea to find the important sinusoids in the periodic variation of TCO over this region. Four common sinusoids are discerned and next used as independent variables in regression to obtain statistically significant output; about 90.13% of the variation in TCO is explained. Here we recall that the contributions from other relatively weak signals ( $T = 21.34, 31.28, 37.31, \dots$  month) observed in SSA-MEM, are rejected in the present analysis. Besides dynamical process like thermally driven turbulent due to convergence and divergence in the atmosphere, mid-level subsidence, overshoots, etc., should have a substantial influence on the assumed static TCO distribution. In particular, one of the most important influential dynamical processes is Quasi-Biennial Oscillation in different geopotential height (Maruyama and Tsuneoka 1988; Rigozo *et al.* 2012 and references

therein). The rest of the variance may be explained accordingly.

The direct impact of SST, SLP anomaly on ozone variation over this region is established following this new approach; critical analysis of the regular appearance of outliers in TCO distribution. The effects nicely and clearly reflected in the bar diagrams; the regular appearance of outliers exactly during extreme phases. The method is very simple and straight forward.

Annual oscillation (ANN) has maximum loading 65.93%, but the aggregate contribution from other three oscillations is only 2.68%. The lowest annual contribution is observed over Quetta, but here we also find highest loadings from QBI (5.94%), QAN (5.48%) and SAN (4.41%).

Overall phasors are 18.71,  $\angle 5.56$  (ANN); 3.26,  $\angle 4.56$  (QAN); 3.18,  $\angle 5.79$  (QBI); 3.07,  $\angle 3.57$  (SAN) having peaks around June, May, June, and April, respectively. The average ozone concentration in TCO possesses a strong latitudinal dependence with two valleys; over the Deccan and Tibetan Plateau. The harmonic analysis is then a useful tool to characterize different climate regimes with transition regions.

## Acknowledgements

The author is grateful to ‘Ozone Processing Team (OPT) of NASA/Goddard Space Flight Center’ for the historical records of ozone data. He wishes to thank his son Mr. Subhradeep Pal, Ph.D. scholar in IIT Patna, India and Br. Subrata, Department of Statistics, Ramakrishna Mission Vidyamandira, Belurmath, Howrah, India for their active support and valuable discussion.



## References

- Atikur Rahman Khan M and Poskitt D S 2013 A note on window length selection in singular spectrum analysis; *Austr. N. Z. J. Stat.* **55**(2) 87–108.
- Attri S D and Tyagi A 2010 *Climate Profile of India*; India Meteorological Department, Lodi Road, New Delhi, India.
- Baldwin M P *et al.* 2001 The quasi-biennial oscillation; *Rev. Geophys.* **39**(2) 179–229.
- Behera S K, Krishnan R and Yamagata T 1999 Unusual ocean–atmosphere conditions in the tropical Indian Ocean during 1994; *Geophys. Res. Lett.* **26** 3001–3004.
- Bian J, Yan R, Chen H, Daren L U and Steven T M 2011 Formation of the summertime ozone valley over the Tibetan Plateau: The Asian summer monsoon; *Adv. Atmos. Sci.* **28**(6) 1318–1325.
- Bliss C I 1958 *Periodic regression in biology and climatology*; Connecticut Agricultural Experiment Station, New Haven.
- Bloomfield P 2000 *Fourier Analysis of Time Series: An Introduction*; 2nd edn, John Wiley and Sons Inc.
- Bradshaw L S and Salazar L A 1985 On using a Fourier series model for estimating diurnal temperatures at mountainous locations in the western United States; *J. Appl. Meteorol. Climatol.* **24** 1104–1106.
- Brooks C 2008 *Introductory Econometrics For Finance*; 2nd edn, Cambridge University Press, Cambridge, UK.
- Chatterjee S and Hadi A S 2012 *Regression Analysis by Example*; 4th edn, John Wiley and Sons Inc.
- Cook R 1977 Detection of influential observations in linear regression; *Technometrics* **19** 15–18.
- da Costa A C M 2009 Stochastic space-time models for the characterization of precipitation extreme values; PhD thesis, Universidade Técnica de Lisboa, Universidade, Nova de Lisboa (UNL).
- Derryberry and De Wayne R 2014 *Basic data analysis for time series with R*; John Wiley and Sons Inc., Hoboken, New Jersey.
- Doherty R M, Stevenson D S, Johnson C E, Collins W J and Sanderson M G 2006 Tropospheric ozone and El Niño–Southern Oscillation: Influence of atmospheric dynamics, biomass burning emissions, and future climate change; *J. Geophys. Res. Atmos.* **111**(D19) 2156–2202.
- Dorvlo A S S 2000 Fourier analysis of meteorological data for seed; *Energy Conserv. Manag.* **41** 1283–1291.
- Duchon C and Hale R 2010 *Time Series Analysis in Meteorology and Climatology*; 1st edn, John Wiley and Sons.
- Dunn P 2007 *Statistics for Climate Research*; University of Southern Queensland Toowoomba Queensland, Australia.
- Gabler R E, Petersen J F and Trapasso L M 2007 *Essentials of Physical Geography*; 8th edn, Thomson Higher Education, 10 Davis Drive, Belmont, CA 94002-3098 USA, 206p.
- Ganguly N D and Iyer K N 2006 Effect of El Niño southern oscillation on the ozone concentration of Indian cities; *Indian J. Physics* **80**(3) 229–234.
- Ghil M 1997 The SSA-MTM Toolkit: Applications to analysis and prediction of time series; In: *Proc. SPIE* 3165, *Applications of Soft Computing* **216**, <https://doi.org/10.1117/12.279594>.
- Golyandina N, Nekrutkin V and Zhigljavsky A 2001 *Analysis of Time Series Structure: SSA and Related Techniques*; Chapman and Hall.
- Gray L J and Dunkerton T J 1990 The role of seasonal cycle in the Quasi-biennial oscillation of ozone; *J. Atmos. Sci.* **47**(20) 2429–2451.
- Hassan M 2014 *Geography of India*; 5th edn, McGraw-Hill Education Private Limited, India.
- Hassani H 2007 Singular spectrum analysis: Methodology and comparison; *J. Data Sci.* **5**(2) 239–257.
- Kita K, Fujiwara M and Kawakami S 2000 Total ozone increase associated with forest fires over the Indonesian region and its relation to the El Niño–Southern Oscillation; *Atmos. Environ.* **34** 2681–2690.
- Kwiatkowski D, Phillips P C B, Schmidt P and Shin Y 1992 Testing the null of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *J. Econom.* **54**(1–3) 159–178.
- Maeda K 1984 Semiannual oscillation of stratospheric ozone; *Geophys. Res. Lett.* **11**(6) 583–586.
- Maruyama T and Tsuneoka Y 1988 Anomalous short duration of the easterly wind phase of the QBO at 50 hPa in 1987 and its relationship to an El-Niño event; *J. Meteor. Soc. Japan* **66** 629–634.
- Meko D M 2015 Applied Time Series Analysis, University of Arizona, Tucson, <http://www.ltrr.arizona.edu/~dmeko/geos585a.html>.
- Myung N K 2009 Singular Spectrum Analysis; M.S. thesis, University of California, USA.
- Pal C 2010 Variability of total ozone over India and its adjoining regions during 1997–2008; *Atmos. Environ.* **44** 1927–1936.
- Percival D B and Walden A T 1993 *Spectral Analysis for Physical Applications*; Cambridge University Press, Cambridge UK, 583p.
- Phillips W E 1984 Harmonic analysis of climate data; *Sol. Energy* **32**(3) 319–328.
- Rigozo N R *et al.* 2012 Reconstruction and searching ozone data periodicities in southern Brazil (29°S, 53°W); *Revista Brasileira de Meteorologia* **27**(2) 243–252.
- Saji N H, Goswami B N, Vinayachandran P N and Yamagata T 1999 A dipole mode in the tropical Indian Ocean; *Nature* **401** 360–363.
- Sahu L K, Sheel V, Kajino M, Gunthe S S, Thouret V, Nedelec P and Smit H G 2013 Characteristics of tropospheric ozone variability over an urban site in southeast Asia: A study based on MOZAIC and MOZART vertical profiles; *J. Geophys. Res. Atmos.* **118** 8729–8747.
- Sahu L K, Sheel V, Kajino M, Deushi M, Gunthe S S, Sinha P R, Sauvage B, Valérie T and Smit H G 2014 Seasonal and interannual variability of tropospheric ozone over an urban site in India: A study based on MOZAIC and CCM vertical profiles over Hyderabad; *J. Geophys. Res. Atmos.* **119** 3615–3641.
- Sahu L K, Sheel V, Kajino M, Deushi M, Gunthe S S, Sinha P R, Yadav R, Pal D, Nedelec P, Valérie T and Smit H G 2016 Impact of tropical convection and ENSO variability in vertical distributions of CO and O<sub>3</sub> over an urban site of India; *Clim. Dyn.*, <https://doi.org/10.1007/s00382-016-3353-7>.

- Schoellhamer D H 2001 Singular spectrum analysis for time series with missing data; *Geophys. Res. Lett.* **28**(16) 3187–3190.
- Sherwood S C and Dessler A E 2003 Convective mixing near the tropical tropopause: Insights from seasonal variations; *J. Atmos. Sci.* **60** 2674–2685.
- Stevens J P 1984 Outliers and influential data points in regression analysis; *Psychol. Bull.* **95**(2) 334–344.
- Tarawneh Q and Kadioglu M 2003 An analysis of precipitation climatology in Jordan; *Theor. Appl. Climatol.* **74**(1–2) 123–136.
- Thomson D J 1982 Spectrum estimation and harmonic analysis; *Proc. IEEE* **70**(9) 1055–1096.
- Timmermann A and Lohmann G 2000 Noise-induced transitions in a simplified model of the thermohaline circulation; *J. Phys. Ocean.* **30** 1891–1900.
- Tolson R H 1981 Spatial and temporal variations of monthly mean total columnar ozone derived from 7 years of BUV data; *J. Geophys. Res. Oceans* **86**(C8) 7312–7330.
- Vautard R, Yiou P and Ghil M 1992 Singular spectrum analysis: A toolkit for short noisy chaotic signals; *Physica D* **58** 95–126.
- Webster P J, Moore A M, Loschnigg J P and Leben R R 1999 The great Indian Ocean warming of 1997–98: Evidence of coupled atmospheric instabilities; *Nature* **401** 356–360.
- Zeng G and Pyle J A 2005 Influence of El-Nino Southern Oscillation on stratosphere/troposphere exchange and the global tropospheric ozone budget; *Geophys. Res. Lett.* **32**(1), <https://doi.org/10.1029/2004GL021353>.
- Zhou X, Leo C, Li W and Shi J 1995 Variation of the total ozone concentration over China and the minimum center over the Tibetan Plateau; *Chinese Sci. Bull.* **40** 1396–1398.
- Ziemke J R and Chandra S 2003 La Nina and El Nino-induced variabilities of ozone in the tropical lower atmosphere during 1970–2001; *Geophys. Res. Lett.* **30**(3), <https://doi.org/10.1029/2002GL016387>.
- Zwiers F W and Storch H V 2004 On the role of statistics in climate research; *Int. J. Climatol.* **24** 665–680.

Corresponding editor: KAVIRAJAN RAJENDRAN