Autocorrelation Structure Analysis and Auto Regressive Prediction of the Time Series of Mean Monthly Total Ozone over Arosa, Switzerland

Goutami Bandyopadhyay, **goutami15@yahoo.co.in**

Surajit Chattopadhyay, **surajit_2008@yahoo.co.in**

1/19 Dover Place, Kolkata-700 019
West Bengal
India

Summary

The purpose of the present study is to look into the characteristics of the mean monthly total ozone time series over Arosa (46.8^0N/ 9.68^0E), Switzerland using statistical methodologies. In this paper, the time series pertains to the data between 1932 and 1971. The intrinsic deterministic patterns of the time series have been investigated through autocorrelation analysis. A second order Auto Regressive Model is tested for prediction potential.

Keywords: Arosa, Mean monthly total ozone, time series, Autocorrelation, Auto Regressive Model.

1.   Introduction

Ozone is a secondary pollutant and is formed in the atmosphere as a result of reactions between other pollutants emitted mostly by industries and automobiles. The ozone precursors are the oxides of nitrogen (NOx) and volatile organic components (VOC) like evaporative solvents and other hydrocarbons. In suitable ambient meteorological condition (e.g. warm, sunny/clear day) ultraviolet radiation (UV) causes the precursors to interact photochemically in a set of reactions that result in the formation of ozone (Comrie, 1997; Corani, 2005).

Ozone absorbs both incoming solar radiation in the UV and visible region, and terrestrially emitted infrared (IR) radiation. Stratospheric ozone absorbs about 12Wm$_{-2}$ of solar radiation and 8Wm$_{-2}$ of terrestrial IR radiation.

Total ozone is a measure of the number of ozone molecules between the ground and the top of the atmosphere. Total ozone is, mathematically, the integral of the ozone concentration with respect to height.

Since total ozone encompasses both the troposphere and the stratosphere, and depends upon weather variables, the corresponding time series is highly complex and non-linear. The present paper endeavors to recognize the intrinsic pattern of a long time series pertaining to mean monthly total ozone data made available from http://www.robhyndman.info/TSDL/monthly/arosa.dat. The data are measured in Dobson Units (DU).

2.   Purpose of the Research

Short-term total ozone changes at Northern mid-latitudes are to a large part caused by a redistribution of ozone in the lower stratosphere, where the lifetime of ozone is long (Bronniamann et al, 2000). As a consequence, total ozone is connected with planetary wave activity and embedded synoptic scale disturbances near the tropopause such as troughs or ridges. Since the total ozone time series over Arosa, Switzerland is the largest available time series of total ozone over mid latitude, this site is chosen for the present study. The study zone is presented in the following map.

The present study aims to view the available total ozone time series statistically. This statistical exploration of the time series would help us to develop predictive model for the

total ozone. For example, the deterministic pattern understood from this study would help to frame an Artificial Neural Net input matrix. Details of the implementation procedure are presented in the subsequent sections.

The study zone for the present study (Monge Sanz and Marqu s, 2004)
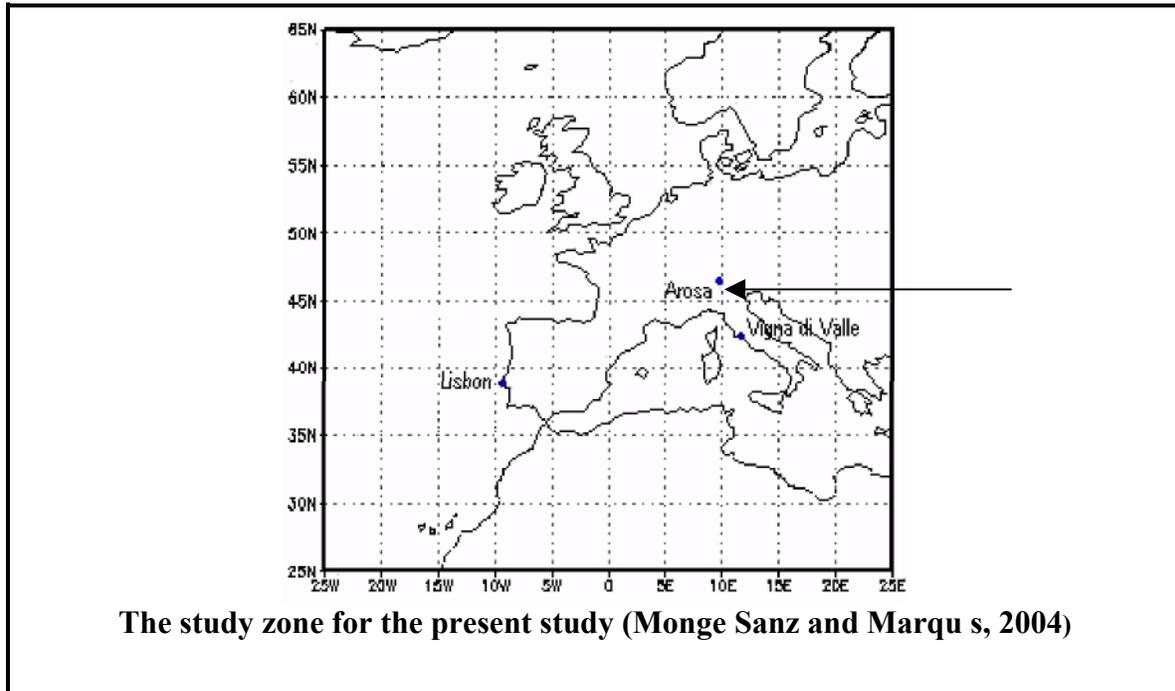
3.   Methodology

3.1   Testing the persistence of the time series

Mean monthly total ozone concentration over the study zone during the period under consideration contains 40 years' data. For each year there are time series of 12 continuous monthly data sets, producing a time series of 480 data sets.

As the first step, all the years are tested with respect to persistence in the mean monthly total ozone concentration. To measure persistence quantitatively, the autocorrelation coefficients for each year are computed as (Wilks, 1995)

r_k=[Covariance $(x(n-k), x\_(n-k)$)]/[Stdev$(x(n-k)$)][Stdev$(x\_(n-k)$)]   ...   (1)

Where, r_k = autocorrelation of order k

$x(n-k)$ = First (n-k) data values

$x\_(n\text{-}k)$ = Last (n-k) data values

Covariance $(a,b) = (1/n)\sum[(a_i\text{-average}(a))(b_i\text{-average}(b))]$

Stdev $(a) = \sqrt{(1/n)\sum[(a_i\text{-average}(a))]^2}$

In the present problem, r_k are computed for k=1,2,3, and 4, i.e. Lag-1, Lag-2, Lag-3, and Lag-4 autocorrelation coefficients (ACC) are computed for each year using equation (1) and are plotted in Fig.01.
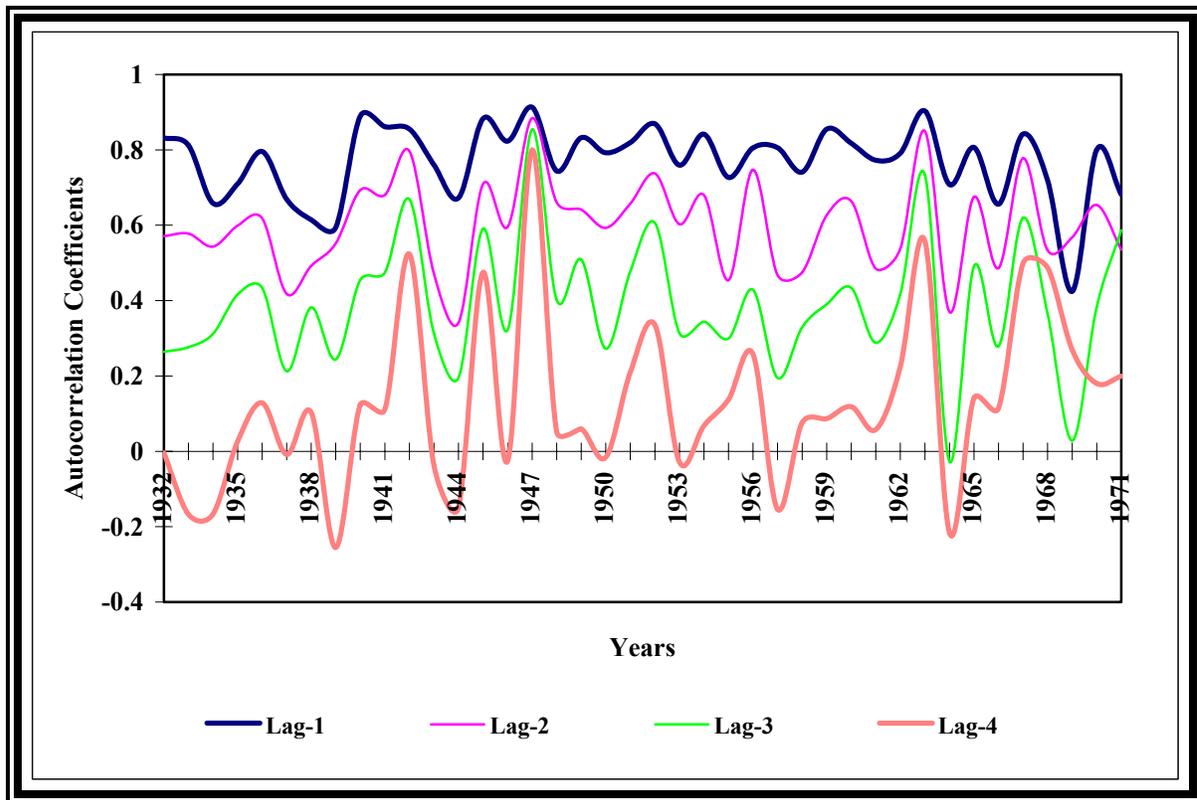

Fig.01-Autocorrelation Coefficients of different orders pertaining to mean monthly total ozone concentration time series in different years within study period.

It is evident from Fig.01 that in all years under study period Lag-1 ACC achieves the highest value. Furthermore, in 99% of cases Lag-1 ACC obtained positive values greater than 0.5. This observation suggests that there is a high serial correlation between data in month *n* and data in month (*n*+1). This further reflects that, there is a high degree of persistence in the data series. If we look into Lag-2 ACC, it becomes evident that the ACC values lie in the immediate neighborhood of 0.5. A linear association is therefore

found out between data in month *n* and in month (*n*+2). Thus, it appears that mean monthly total ozone over Arosa in month *n* maintains a pattern up to month (*n*+2).

3.2   *Testing the persistence of the time series*

Now, the whole time series (i.e. 480 entries) is considered to investigate the overall autocorrelation structure. The ACC up to Lag-40 are computed and plotted in Fig.02.
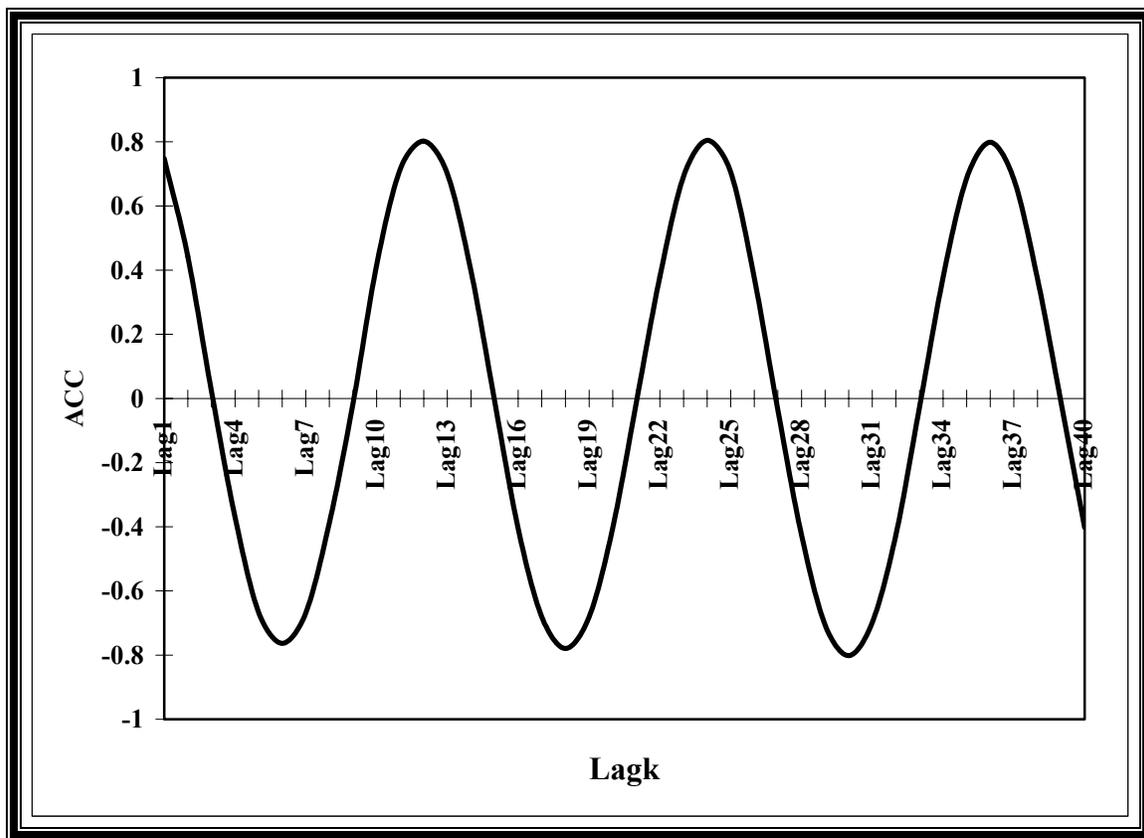

Fig.02-Autocorrelation Coefficients of different orders pertaining to mean monthly total ozone concentration time series with 480 data points.

Fig.02 shows that if 480 data points are considered and ACC are computed up to several lags, then a sinusoidal structure dominates the autocorrelation function. Here, peaks are there in every 6-time step gaps and similar pattern repeats in each 12 months periods. A yearly cyclic structure becomes apparent.

Fig.02 illustrates that ACCs are very high (-ve or +ve) at 6 months interval and demonstrates the existence of cyclic pattern in the data. Also, at each five months interval, specific pattern is repeated in the data series. For example, Lag-6 ACC is –0.78. This implies that, if total ozone concentration in January is very high, it would be very low in June. It would again be very high in December (Lag-12 ACC is very high and positive).

3.3 *Auto Regressive model construction*

In the previous section we found that the yearly time series of mean monthly total ozone are highly persistent up to order 2. Thus, we create an auto regressive predictive model for each year using the repetition of structure suggested in Fig. 02.

A second order auto regressive model (AR (2)) can be formulated as (Wilks, 1995)

$$x_{t+2} \ - \ \mu = \phi_1(x_{t+1} \ - \ \mu) + (x_t - \mu) \qquad \ldots \qquad \ldots \qquad (2)$$

Where, ϕ_1 and ϕ_2 are AR parameters.

Now, the AR parameters are computed for each year and separate AR (2) models are generated for each year. The models are all predictive and the predictions are compared with actual values via Pearson Correlation Coefficient (Chattopadhyay, 2002).

The results are presented in Fig.03. This figure suggests that in all the years, there is high linear association between actual and predicted values. For further clarification, some sample figures (Figs.04-07) are drawn. It is understandable from Figures 04-06 that AR (2) model fits better during Spring-Summer than winter. From the raw data series it is found that total ozone falls significantly with the advancement of winter. Thus from the less goodness of fit of AR (2) model in winter it is inferred that the winter data series are more chaotic than Spring-Summer.

4. Conclusion

The study leads to us conclude the following:

- Mean monthly total ozone time series over Arosa, Switzerland is highly periodic with periodicity of twelve months.
- Every year, mean monthly total ozone time series over Arosa, Switzerland exhibit high degree of persistence up to Lag 2.

- Second order Auto Regressive model can be considered as suitable predictive model for mean monthly total ozone time series over Arosa, Switzerland.
- During winter mean monthly total ozone time series over Arosa, Switzerland exhibits more chaotic nature than spring-summer.

Scope of future research

In the present work, a cyclic pattern of the time series has been revealed. Furthermore, an auto regressive predictive equation has been generated. This equation has some success in predicting a long time series. On the basis of this sort of deterministic pattern, we suggest that an Artificial Neural Net model can be developed in future to attain more accuracy in prediction.
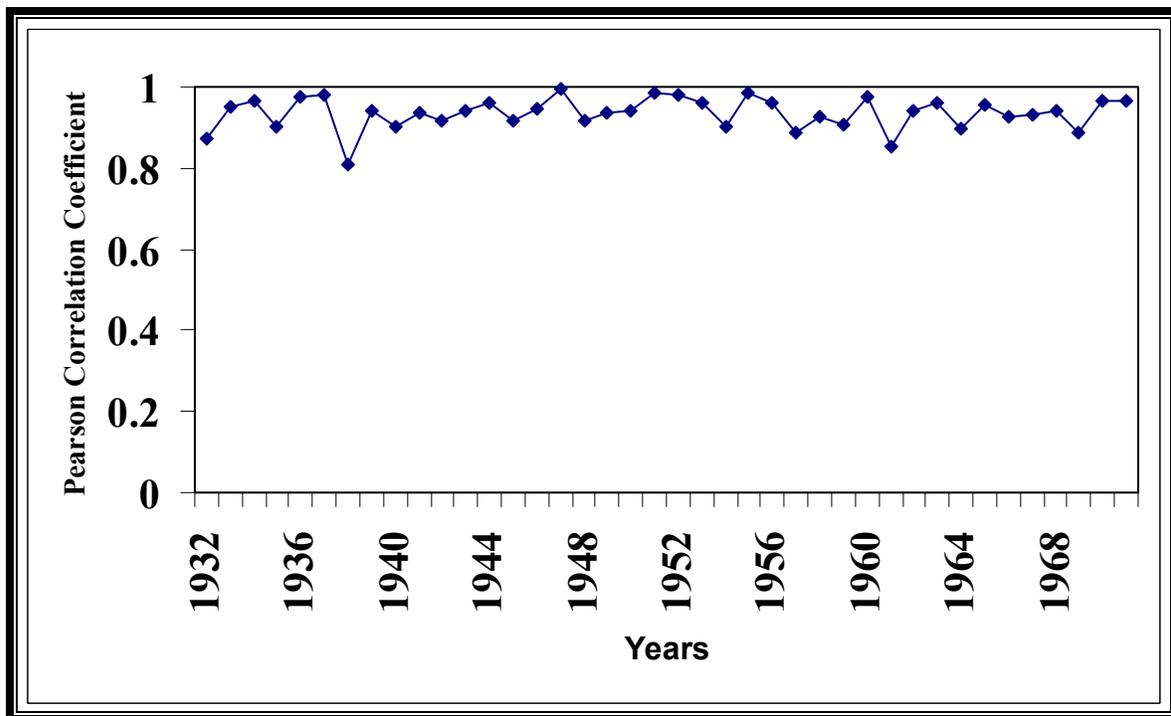

Fig.03- Schematic showing the linear association between actual monthly total ozone concentrations and those predicted through AR (2) model.
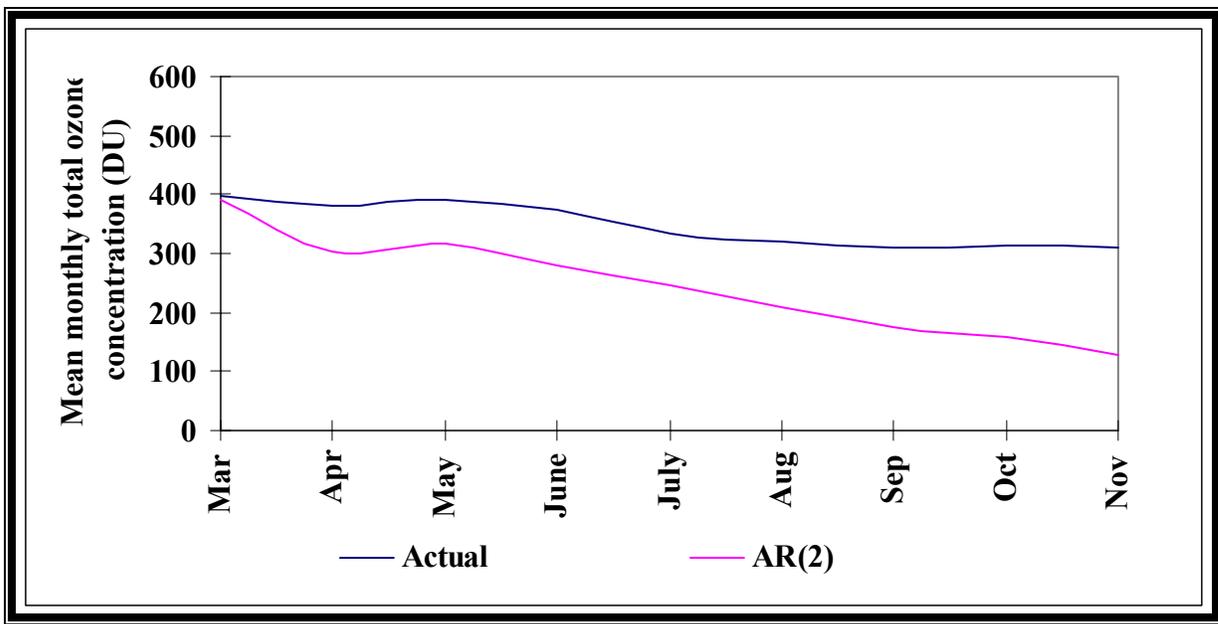
Fig.04- Schematic showing the actual and predicted (AR (2)) mean monthly total ozone concentration over Arosa in 1933
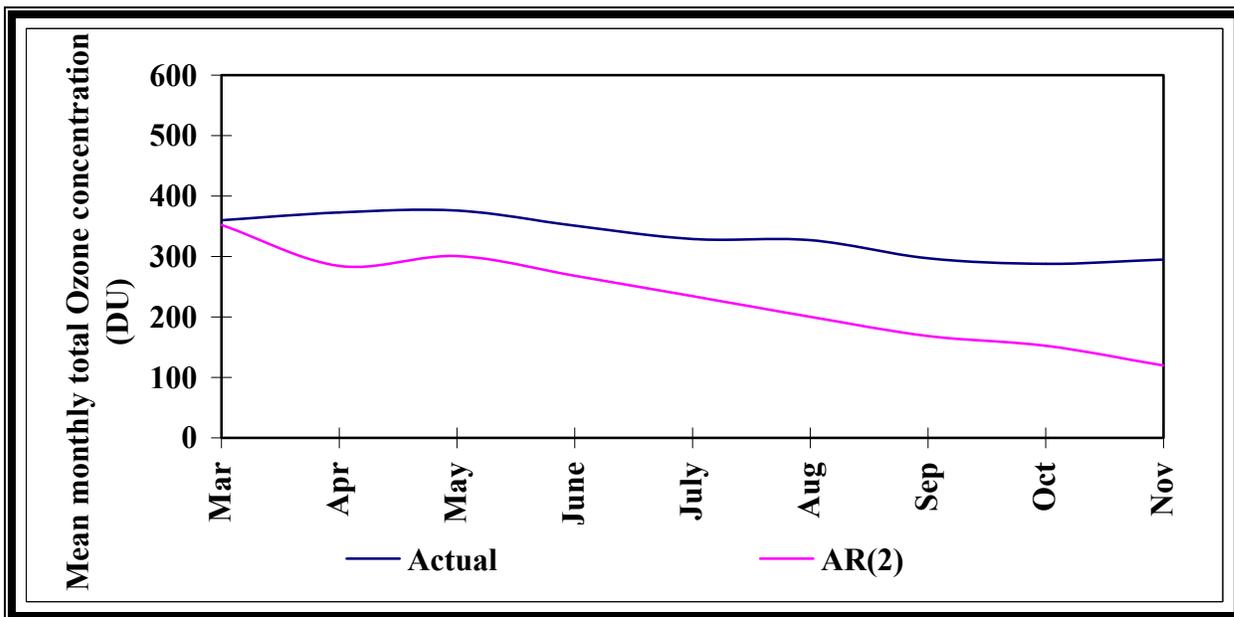

Fig.05- Schematic showing the actual and predicted (AR (2)) mean monthly total ozone concentration over Arosa in 1945
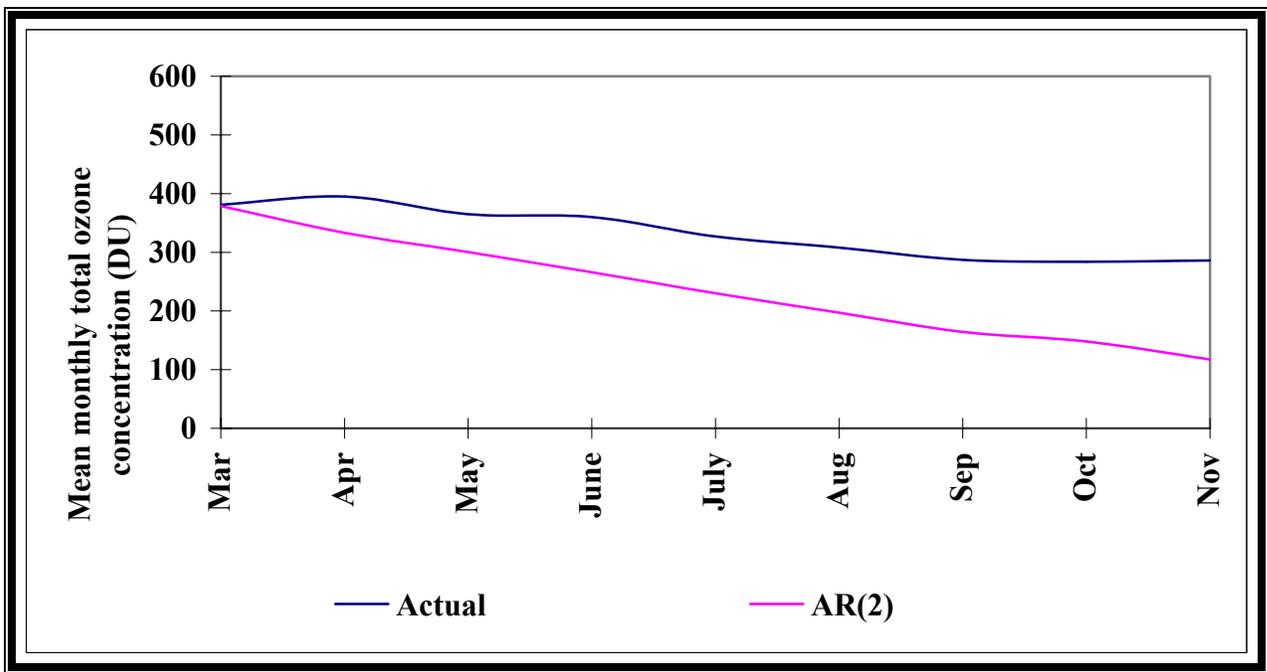
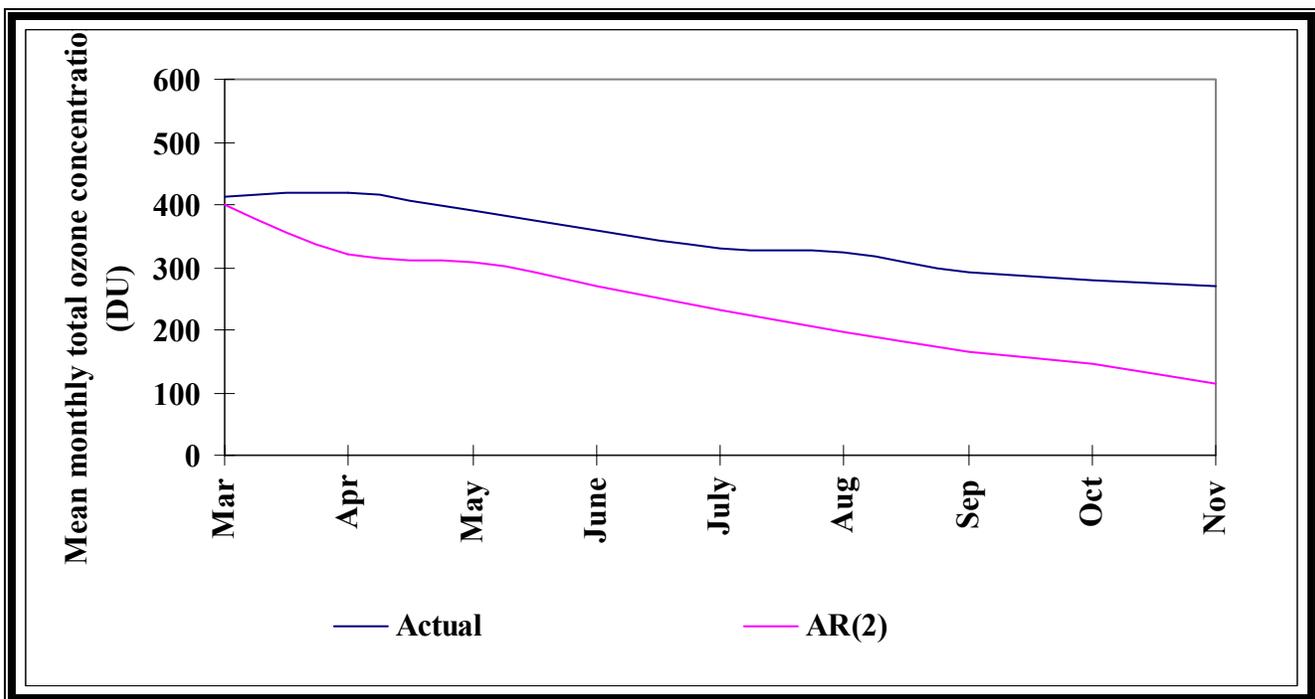**Fig.06- Schematic showing the actual and predicted (AR (2)) mean monthly total ozone concentration over Arosa in 1956**

Fig.07- Schematic showing the actual and predicted (AR (2)) mean monthly total ozone concentration over Arosa in 1970

References

- Comrie, A.C., Comparing Neural Networks and Regression Models for ozone forecasting, *J. Air & Waste Manage. Assoc.*, **47**, 653-663 (1997).
- Corani, G., Air quality prediction in Milan: feed-forward neural networks, pruned neural networks and lazy learning, *Ecological Modelling*, **185**, 513-529 (2005)
- Wilks, D.S., *Statistical Methods in Atmospheric Sciences* (Academic Press, USA, 1995)
- Chattopadhyay, S., Predicting pre-monsoon thunderstorms-A statistical view through prepositional logic, 2002, *Solstice: Journal of Geographical Mathematics*, 13(2), http://www.imagenet.org/
- Monge Sanz B. M. and Medrano Marqu s N. J., Total ozone time series analysis: a neural network model approach, *Nonlinear Processes in Geophysics*, **11**, 683–689(2004)
- Bronnimann, S et al, Variability of total ozone at Arosa, Switzerland, since 1931 related to atmospheric circulation indices, *Geophysical Research Letters*, **27**, 2213-2216, 2000

Diagrammatic presentation of the research

Collection of mean monthly total ozone data over Arosa, Switzerland

Computing Autocorrelation function for each year

Computing autocorrelation function for the whole time series containing 480 data points

Analysis of the Autocorrelation functions

Development of second order Auto Regressive Model

Testing goodness of fit

Conclusion