SOLSTICE: An Electronic Journal of Geography and Mathematics.
(Major articles are refereed; full electronic archives available)

SOLSTICE, VOLUME XVIII, NUMBER 1; JUNE, 2007.

Front matter: June, 2007.
Editorial Board, Advice to Authors, Mission Statement.

Awards

ARTICLE
(reviewed)

Spatial Analysis through the Looking Glass
Peter Martin

ANNOUNCEMENT

3D Atlas of Ann Arbor, 3rd Edition

Sandra Lach Arlinghaus

NOTES

Update on the Varroa Mite Map
Diana Sammataro
[with Editorial Commentary]

Editorial Commentary on Essays of this Section

The Prediction of Indian Monsoon Rainfall: A Regression Approach
Goutami Bandyopadhyay

Autocorrelation Structure Analysis and Auto Regressive Prediction of the Time Series of Mean Monthly Total Ozone overArosa, Switzerland
Goutami Bandyopadhyay and Surajit Chattopadhyay

Mail

Solstice Archive

- *Solstice* was a Pirelli INTERNETional Award Semi-Finalist, 2001 (top 80 out of over 1000 entries worldwide)
- One article in Solstice was a Pirelli INTERNETional Award Semi-Finalist, 2003 (Spatial Synthesis Sampler).
- *Solstice* is listed in the Directory of Open Access Journals maintained by the University of Lund where it is maintained as a "searchable" journal.
- *Solstice* is listed on the journals section of the website of the American Mathematical Society, http://www.ams.org/
- Solstice is listed in Geoscience e-Journals
- *Solstice* is listed in the EBSCO database.
- IMaGe is listed on the website of the Numerical Cartography Lab of The Ohio State University: http://ncl.sbs.ohio-state.edu/4_homes.html

Congratulations to all *Solstice* contributors.

***Solstice: An Electronic Journal of Geography and Mathematics*,**
Volume XVII, Number 2
Institute of Mathematical Geography (IMaGe).

sarhaus@umich.edu
http://www.imagenet.org

1964 Boulder Drive,
Ann Arbor, MI 48104734.975.0246
image@imagenet.org

SOLSTICE: AN ELECTRONIC JOURNAL OF GEOGRAPHY AND MATHEMATICS
http://www.imagenet.org

June, 2007
VOLUME XVIII, NUMBER 1
ANN ARBOR, MICHIGAN

WebSite: http://www.imagenet.org

Electronic address: sarhaus@umich.edu

MISSION STATEMENT

The purpose of Solstice is to promote interaction between geography and mathematics. Articles in which elements of one discipline are used to shed light on the other are particularly sought. Also welcome are original contributions that are purely geographical or

purely
mathematical. These may be prefaced (by editor or author) with commentary suggesting directions that might lead toward the desired interactions.
Individuals wishing to submit articles or other material should contact an editor, or send e-mail directly to sarhaus@umich.edu.

SOLSTICE ARCHIVES

Back issues of Solstice are available on the WebSite of the Institute of Mathematical Geography, http://www.imagenet.org and at various sites
that can be found by searching under "Solstice" on the World Wide Web. Thanks to Bruce Long (Arizona State University, Department of Mathematics)
for taking an early initiative in archiving Solstice using GOPHER.

PUBLICATION INFORMATION

To cite the electronic copy, note the exact time of transmission from Ann Arbor, and cite all the transmission matter as facts of publication. Any copy that
does not superimpose precisely upon the original as transmitted from Ann Arbor should be presumed to be an altered, bogus copy of *Solstice*. The
oriental rug, with errors, serves as the model for creating this weaving of words and graphics.

Awards and Recognition

(See Press Clippings page for other.)

- Google 3D Warehouse, "Google Picks" then go to "Cities in Development" http://sketchup.google.com/3dwarehouse/ to see textured models of downtown Ann Arbor buildings.
- *3D Atlas of Ann Arbor*, *Version 2*. Google Earth Community, ranked a "Top 20 Rated Post" on Entrance page, December 8, 2006.
- *3D Atlas of Ann Arbor*, *Version 2*. Rated a 5 globe production (top score) in Google Earth Community, November 2006.
- Sandra L. Arlinghaus and William C. Arlinghaus, Spatial Synthesis Sampler, *Solstice*, Summer 2004. Semi-Finalist, Pirelli 2003 INTERNETional Award Competition.
- Sandra Lach Arlinghaus, recipient, The Presidentâ€™s Volunteer Service Award, March 11, 2004.
- Jeffrey A. Nystuen, won the 2003 Medwin Prize in Acoustical Oceanography given by the Acoustical Society of America. The citation was "for the innovative use of sound to measure rainfall rate and type at sea". It is awarded to a young/mid-career scientist whose work demonstrates the effective use of sound in the discovery and understanding of physical and biological parameters and processes in the sea.
- Sandra L. Arlinghaus, William C. Arlinghaus, and Frank Harary. *Graph Theory and Geography: an Interactive View (eBook)*, published by John Wiley and Sons, New York, April 2002. Finished as a Finalist in the 2002 Pirelli INTERNETional Award Competition (in the top 20 of over 1200 entries worldwide).
- *Solstice*, Semi-Finalist, Pirelli 2001 INTERNETional Award Competition in the Environmental Publishing category.
- *Solstice*, article about it by Ivars Peterson in *Science News*, 25 January, 1992..
- *Solstice*, article about it by Joe Palca, *Science* (AAAS), 29 November, 1991.

Solstice: An Electronic Journal of Geography and Mathematics, Institute of Mathematical Geography, Ann Arbor, Michigan.
Volume XVII, Number 2.
http://www.InstituteOfMathematicalGeography.org/

Spatial Analysis through the Looking Glass

Introduction

One of the delightful attractions that mathematics and geography have in common is that they offer fresh views of the world. Both fields of inquiry are circumspect almost by definition, and as soon as one looks around without prejudice, one discovers alternatives to one's provincial world view.

In mathematics, formalism and generalization are the seemingly sterile keys that open the gates to fertile fields of invention that, for its unexpected—even unexpectable—consequences, is indistinguishable from discovery. Formalism engenders the "what if" assumptions that lead to strange constructions, with their uncanny applicability to physics—non-Euclidean geometry, complex numbers, the Mandelbrot set. Generalization is the building of broader and broader analogies—the essence of cognition that Hofstadter has called 'chunking" (Hofstadter 2002). The joy of mathematics is that one gets more than what one bargained for; one discovers connections that one did not anticipate. The net effect is that the familiar is seen anew, as a "special case".

In geography, one discovers cultural symmetry, and cultural asymmetry. One realizes that relations are not absolute, but dependent on location in space and in time. Unless one is very well indoctrinated, one develops a sense of the relative, even "accidental", nature of one's point of view.

Science seeks to discover the laws which describe the rhythms and patterns of nature in time and space (Feynman 1967); in fact the essence of reality may be *relations*, or interactions, rather than any inherent objective attributes (Anandan 2003).

In this spirit, I'd like to offer here an example of a geographic-analytic point of view that I think is a refreshing departure from the customary provincial point of view in spatial analysis.

Case in point: Interpolation of spatial data

Interpolation of spatial data from irregular samples is a familiar problem in the geosciences. In the case of remotely-sensed data, the problem may be that of reconstruction of an image where data are missing; in the case of *in situ* data, the problem may be that of producing a data model on a regular spatial grid. Notwithstanding the prevailing view that "the location of events is considered an essential part of the observations" (Tobler 1966), it may be helpful to regard the data and the desired interpolation as objects in an abstract information space, not essentially dependent on the spatial coordinate system.

In this way, a data set $\mathbf{d}$ is thought of as a point in a space of as many dimensions (or coordinates) as the data have spatial locations. Let us say for specificity that there are n data points, and that the desired interpolation would have m data points, with $m > n$. The n components of the data vector $d_1 \ldots d_n$ are simply the n data values for each of the n coordinates (*i.e.* spatial locations). The desired interpolation $\hat{\mathbf{i}}$ would then be a point in a space of m dimensions (including $m - n$ additional dimensions for the missing data locations) and this interpolation would be determined by a *projection* of $\mathbf{d}$ from a space of n dimensions to one of m dimensions.

This way of looking at the problem is useful because of the importance of *context* to interpolation: any data set, like any signal, can be thought of as belonging to a class of similar data, which probably has a Gaussian distribution, in the m-dimensional information space, that is contained by a generalization of an ellipsoid (see, *e.g.*, (Goyal 2001)). The object of interpolation, then, is to project from $\mathbf{d}$ to $\hat{\mathbf{i}}$, where $\hat{\mathbf{i}}$ lies within the generalized m-dimensional ellipsoid of the expected distribution that represents *context.*

A linear *transform* of the data, such as a Fourier transform (or FFT, Fast Fourier Transform), Discrete Cosine Transform (DCT), or a Principal Component Transform (PCT), may be regarded simply as a change of the coordinate axes of the information space, not affecting the informational essence of the data, but affording a new view of the data in relation to its ellipsoid of context.

Band-limitation and its generalization

A class of data may be *band-limited*, which means that a limited number of the FFT or DCT components are significant in the representation of data sets from the class, ordinarily the longer-wavelength components. According to the Shannon-Whittaker-Kotel'nikov sampling theorem, accurate interpolation is possible if there are about two samples per shortest wavelength, not necessarily regularly-spaced samples (see, *e.g.* (Papoulis 1977, Unser 2000)). Given band-limitation, it may be much easier (and computationally more efficient) to carry out interpolation in the transform domain.

The object of *principal component analysis* is to determine the major, semi-major, etc., axes of the m-dimensional dataset distribution ellipsoid in information space, and to define a new coordinate system aligned with those axes. There are typically few axes of significant extent, relative to the total number of dimensions of the information space, so members of the class can be distinguished by a minimum number of components. Likewise, by extension of the sampling theorem, reconstruction of data sets which are known to be members of the class can be based on very few point samples (as shown, for example, by (Everson & Siroich 1995), for the case of images of faces).

Many data classes, such as grayscale images, can be modeled as *autoregressive* sources, for which adjacent data values are related by a correlation coefficient ρ close to 1 plus a small noise term $\mathbf{z}$:

$$\mathbf{d}_k = \rho\mathbf{d}_{k-1} + \mathbf{z}_k \ .$$

For such data classes, the PCT can be approximated bt a DCT (Goyal 2001), which is the reason that the first JPEG image-compression standard was based on the DCT.

In consideration of various linear transforms by which the members of a data class might be represented by a limited number of components, a generalization of the concept of band-limitation suggests itself: if data to be interpolated belong to a class whose distribution ellipsoid can be aligned with the coordinate axes under some transform, then the data can be called "band-limited" in that transform space. More to the point, interpolation can then be carried out in a computationally efficient manner.

Interpolation in practice

Notwithstanding the ideal that interpolation should not introduce information not present in the given data (Briggs 1974), interpolation is necessarily a hypothesis about the character of the interpolated surface (Lam 1983). That is so because any interpolation is the addition of hypothetical information components to given information components, based on known constraints, as the data vector of given components is projected to an information space of higher dimension. The essence of information is context (Martin 2004). Therefore the problem of interpolation is to venture a least-biased projection of a limited vector of given data onto a more complete vector of interpolated data. With the best choice of transform, the additional components will have value zero, while the original data components (*i.e.* in the standard coordinates) will remain unchanged.

Common interpolation methods

Splines

For many purposes, a least-biased estimate is assumed to be that the interpolated surface should be smooth (or band-limited after FFT), so *minimum curvature* algorithms (employing splines) have been favored, and found to be the best for most scattered data sets (Rauth & Strohmer 1998). Briggs (Briggs 1974) advocated such a method for interpolation of geophysical fields, which was refined by Smith & Wessel's (Smith & Wessel 1990) method of "splines in tension" for gridding such data as that of seafloor topography, which may exhibit more abrupt changes of direction than do gravity or magnetic fields.

Equivalent source

The *equivalent source* field-interpolation technique of Dampney (Dampney 1969), extended by Cordell (Cordell 1992) is an algorithm employing geophysical inversion, solving for field sources consistent with given data, then calculating intervening field values from the posited sources. Its limitations are with respect to propriety of a field data model and with respect to computational efficiency.

Kriging

This is a statistically-based interpolation technique based on estimation of how data values vary as a function of distance (see, *e.g.*, (Journel 1989)). In this technique, hypotheses about the character of the surface are based strictly on the given data. It is explicitly a least-biased estimation technique, but it tends to be computationally intensive (Doucette & Beard 2000).

Iterative Fourier-based method

Computationally attractive iterative algorithms for reconstruction (or interpolation) of band-limited signals from irregular samples were devised by Marvasti (Marvasti 1991) and refined by Feichtinger, Gröchenig, Strohmer, and others (Feichtinger & Strohmer 1993, Strohmer 1997). In these algorithms, the error of a projection (referred back to the original basis of coordinates) is fed back into the algorithm, its projection added to the first approximation, and so on iteratively until a good "fit" to the given sample values is achieved.

Bandwidth in geoscientific data

The iterative Fourier-based methods are most effective for interpolation when the bandwidth of the data to be interpolated is known *a priori*. If the bandwidth is incorrectly assumed, the interpolation result will be unsatisfactory (Martin 2004): if the upper band limit is underestimated, the interpolation will be too "smooth", missing some of the sample points; if the upper band limit is overestimated, the interpolation will exhibit wild excursions even as it passes through all of the sample points (like curve-fitting with a polynomial of high degree).

Many data in geosciences belong to a class with $1/f$ spectral characteristic, where the magnitude of FFT components vary inversely with the "frequency", or wavenumber. For example, seafloor topography exhibits $1/f$ spectral characteristic over seven orders of magnitude (Bell 1975). In general, it is difficult to devise an appropriate spectral filter in order to implement the iterative Fourier-based method (Martin 2004).

The ideal case for interpolation is to have a large number of data sets, to which principal-component analysis might be applied, in order to arrive at a figure of generalized "bandwidth", in terms of number of significant principal components.

Computing principal components

Given a number s of data sets (data vectors $\mathbf{d}_k$) of the same presumed class, and of the same length n, principal component analysis may be carried out as follows using Matlab: first one constructs the s x n matrix $\mathbf{P}$ whose s rows are the s representative data sets, each of length n, from the data class[1]. The covariance matrix $\mathbf{C}$ is computed:

$$\mathbf{C} = \mathrm{cov}(\mathbf{P}) \ ,$$

and after this, a matrix $\mathbf{V}$ of eigenvectors and a diagonal matrix $\mathbf{D}$ of eigenvalues is computed:

$$[\mathbf{V}, \mathbf{D}] = \mathrm{eig}(\mathbf{C}) \ .$$

The columns of $\mathbf{V}$ are the principal component vectors, in reverse order of importance, which importance is indicated by the magnitude of the corresponding eigenvalues, on the diagonal of $\mathbf{D}$.

The columns of $\mathbf{V}$ thus derived can be reassembled to construct an inverse transform matrix for interpolation based on the hypothesis of generalized "band limitation", as described below.

Matrix computation for interpolation in a transform base

A linear transform $\mathbf{T}$, which is an operator used to go from the representation of a data set $\mathbf{i}$ in the standard base to the representation $\acute{\mathbf{i}}$ in the transform base, is a system of equations that can be expressed as the matrix multiplication

$$\mathbf{Ti} = \acute{\mathbf{i}} \ .$$

Likewise the inverse transform can be written

$$\mathbf{T^{-1}}\acute{\mathbf{i}} = \mathbf{i} \ .$$

Here we assume $\mathbf{T}$ to be an m x m matrix, where m is the the number of points in the desired interpolation.

If a set of samples $\mathbf{d}$ are taken at points $\mathbf{p}$ of $\mathbf{i}$, so that we can write

$$\mathbf{d} = \mathbf{i}(\mathbf{p}) \ ,$$

then to represent the data in the transform base, the general system of equations to solve can be written

$$\mathbf{T_p^{-1}}\acute{\mathbf{d}} = \mathbf{d} \ ,$$

[1]For spatial data of two or three dimensions, the data can be arranged into a one-dimensional vector, for example concatenating the rows of an image in alternating order.

where $\mathbf{T_p^{-1}}$ is $\mathbf{T^{-1}}$ having only the rows corresponding to $\mathbf{p}$. In Matlab notation,

$$\mathbf{T_p^{-1}} = \mathbf{T^{-1}}(\mathbf{p}, :) \ .$$

After a least-squares estimate of transform coefficients $\hat{\hat{\mathbf{d}}}$ is found (again in Matlab notation) by

$$\hat{\hat{\mathbf{d}}} = \mathbf{T_p^{-1}} \backslash \mathbf{d} \ , \tag{1}$$

the interpolation $\hat{\mathbf{i}}$ is computed as

$$\hat{\mathbf{i}} = \mathbf{T^{-1}} \hat{\hat{\mathbf{d}}} \ .$$

Equation 1 will not in general be well-determined since $m > n$. However, the key point is that a "band-limited" matrix computation can be formulated when solution is sought for a limited set of significant components. This is done by removing the columns of the inverse transform matrices $\mathbf{T_p^{-1}}$ and $\mathbf{T^{-1}}$ corresponding to components that are considered to be insignificant in the context of the class of data.

Non-iterative method proven with FFT

The above-described method of matrix computation was used in conjunction with FFT in a Matlab routine "FFTINTERP"[2], and found to run faster than an "Adaptive Weight, Conjugate Gradient" iterative routine (Martin 2004). It was also found to be more readily applicable to data characterized by $1/f$ spectra since *the source bandwidth need not be specified*. Instead, the bandwidth of interpolation is determined (by default) by the number of samples.

The routine "FFTINTERP" is not to be confused with the Matlab routine "INTERPFFT", which requires regularly-spaced samples. It does, however, converge to the same result for regular samples, and runs faster.

The "FFTINTERP" routine demonstrates the efficiency of interpolation in a transform space, as well as the idea of adaptive band limiting, based on number of samples. By substituting a different inverse transform matrix in the algorithm (such as a PCT), the idea of generalized "bandwidth" can be demonstrated and applied analogously.

Conclusion

A practical computation algorithm for interpolation of irregularly-spaced data on a transform basis has been presented, motivated by a conceptual understanding of data sets residing in information space. This has been offered as proof of the idea that even the customary spatial point of view of the geographer is not absolute.

The relation between generalized "bandwidth" and "context" of the data to be interpolated was discussed, leading to explicit computation formulas involving partial inverse transform matrices. This was offered as a substantial example of how a formal change in point of view affords one the tools to understand and solve problems from a fresh point of view.

References

Anandan, J. (2003). "Laws, symmetries, and reality" *International Journal of Theoretical Physics.* **42**(9): 1943–1955.

Bell, T. (1975). "Statistical features of sea floor topography" *Deep Sea Research.* **22**: 883–892.

Briggs, I. (1974). "Machine contouring using minimum curvature" *Geophysics.* **39**: 39–48.

Cordell, L. (1992). "A scattered equivalent-source method for interpolation and gridding of potential field data in three dimensions" *Geophysics.* **57**(4): 629–636.

Dampney, C. (1969). "The equivalent source technique" *Geophysics.* **34**(1): 39–53.

Doucette, P. & Beard, K. (2000). "Exploring the capability of some GIS surface interpolators for DEM gap fill" *Photogrammetric Engineering and Remote Sensing.* **66**: 881–888.

[2]Available: http://martinandriener.com/Papers.

Everson, R. & Siroich, L. (1995). "Karhunen-Loève procedure for gappy data" *J.Opt. Soc. Am. A.* **12**(8): 1657–1664.

Feichtinger, H. & Strohmer, T. (1993). "Fast iterative reconstruction of band-limited images from irregular sampling values" In D. Chetverikov & W. Kropatsch (eds), *Proceedings, 5th International Conference, CAIP '93.* Budapest pp. 82–91.

Feynman, R. (1967). *The Character of Physical Law.* M.I.T. Press. Cambridge, MA.

Goyal, V. (2001). "Theoretical foundations of transform coding" *IEEE Signal Processing Magazine.* **18**(5): 9–21.

Hofstadter, D. (2002). "Analogy as the core of cognition" In D. Gentner, K. Holyoak & B. Kokinov (eds), *The Analogical Mind.* MIT Press. Cambridge, MA In "15",pp. 499–538.

Journel, A. (1989). *Fundamentals of Geostatistics in Five Lessons.* American Geophysical Union. Washington, DC.

Lam, N. (1983). "Spatial interpolation methods: A review" *American Cartographer.* **10**(2): 129–149.

Martin, P. (2004). *"Spatial Interpolation in Other Dimensions"* Master's thesis Oregon State University Corvallis.
URL: *http://martinandriener.com/Thesis*

Marvasti, F. (1991). "Recovery of signals from nonuniform samples using iterative methods" *IEEE Trans. Sig. Proc..* **39**: 872–877.

Papoulis, A. (1977). *Signal Analysis.* McGraw-Hill. New York.

Rauth, M. & Strohmer, T. (1998). "Smooth approximation of potential fields from noisy scattered data" *Geophysics.* **63**: 85–94.

Smith, W. & Wessel, P. (1990). "Gridding with continuous curvature splines in tension" *Geophysics.* **55**: 293–305.

Strohmer, T. (1997). "Computationally attractive reconstruction of bandlimited images from irregular samples" *IEEE Trans. Image Proc..* **6**(4): 540–548.

Tobler, W. (1966). "Spectral analysis of spatial series" *Proceedings, 1966 URISA Conference..* pp. 179–186.

Unser, M. (2000). "Sampling—50 years after Shannon" *Proc. IEEE.* **88**: 569–587.

***Announcement: 3D Atlas of Ann Arbor, 3rd Edition*;**

Sandra Lach Arlinghaus

(aka "Archimedes" in the Google 3D Warehouse environment)

The "Diag"--central quadrangle--The University of Michigan, Ann Arbor.

The current 3rd Edition of the *3D Atlas of Ann Arbor* (http://www.imagenet.org/) contains real-time 3d (virtual reality) models of the entire city of Ann Arbor. The files in that eBook cover just under 25,000 buildings in Ann Arbor. A number of files are fully textured; still others are partially textured; yet others are extruded from GIS software footprints according to correct heights; and, the remaining buildings outside the downtown and the university are extruded according to an arbitrary height. The geometry for all buildings is there. Recently, the 3D Warehouse featured the work of "Archimedes" in three of its four "Google Picks" categories: "Cities in Development"; "Featured Modelers"; and, "Help Model a City." The latter category is particularly exciting as it offers cities with full geometries, such as Ann Arbor, an opportunity to take advantage of free labor from those wishing exercises for students in texture application using Google SketchUpÂ® coupled with photos in PicasaÂ®, and so forth. Thus, a winning opportunity is once again created (as it was when municipal authorities began to embrace GIS software): the city that shares its files wins with free labor and an expanded spatial information base--university instructors win with fine real-world scenes to offer students as modeling exercises. Consider joining the growing group!

Updated Varroa Mite Map of Data of Diana Sammataro

One advantage of on-line publication is the capability to easily update files that depend on temporal data. *Solstice* author Diana Sammataro has been sending IMaGe current data for her Varroa Mite Map on a regular basis, since it first appeared in Volume IX, Number 1, 1998. The current form of the map is shown below.

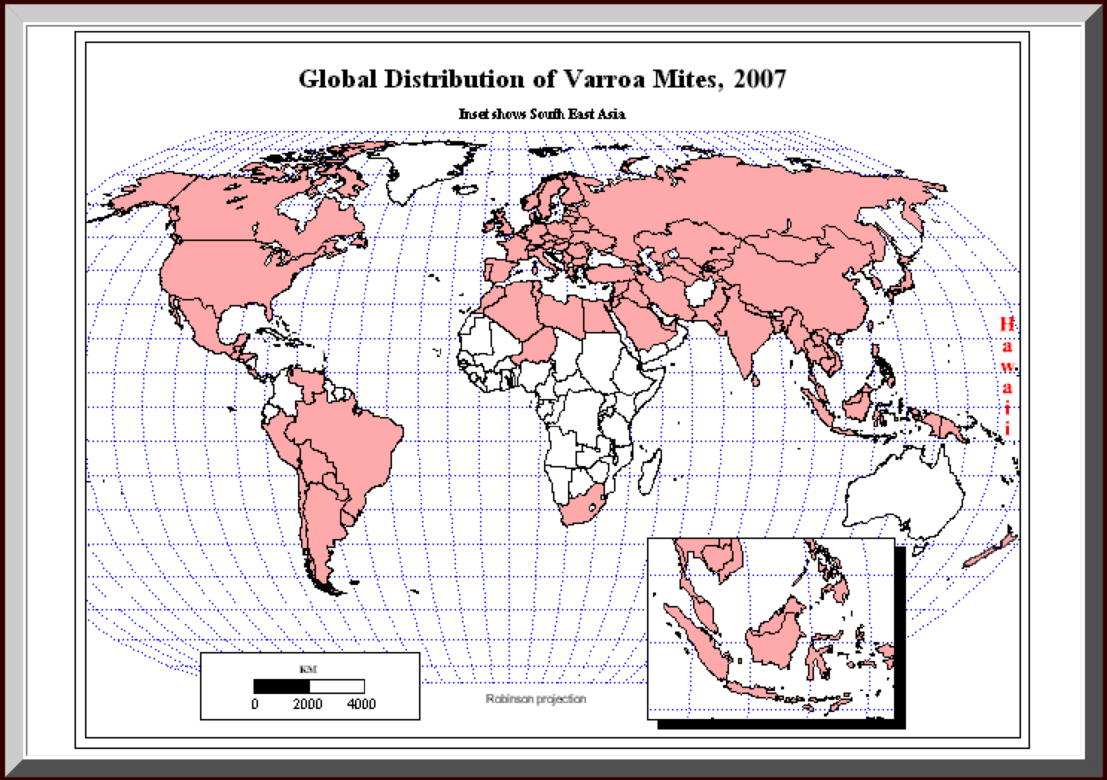
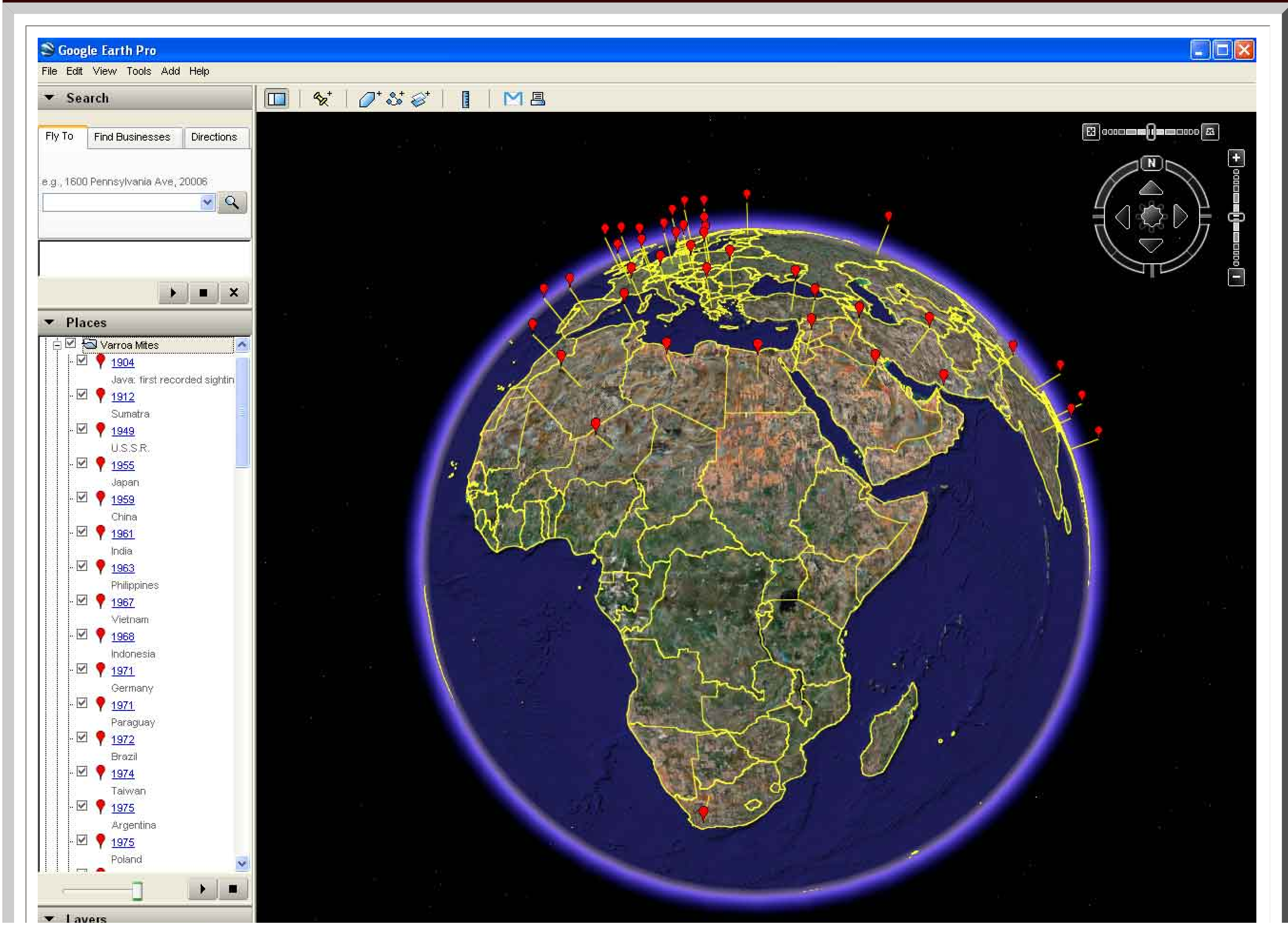
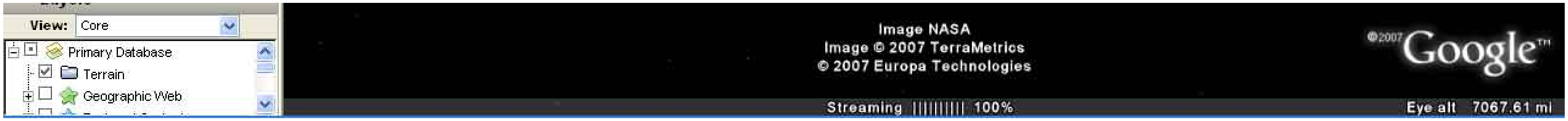


Figure 1. Animated map. Map by Sandra L. Arlinghaus and John D. Nystuen.

[Ed.] Notice, however, that with the introduction of Hawaii into the picture in 2007, the visual geographic scale shift required to portray both Hawaii and the former Soviet Union on the same map is stretched to the limit. Shading of the Hawaiian Islands is not visible. One solution is simply to insert a visible word over the location of the islands (as above), interrupting the visual animation pattern (as in Figure 1). Another solution is to offer yet another inset map which might be distracting and cluttering. A third possibility is to recast the two dimensional animation as a three dimensional interactive model. Thus, Sammataro's data is recast in Google Earth®. To view it, first download a current version of the free Google Earth® and put it on your computer's desktop. Then, open the following file in Google Earth®: varroa.kmz . Screen captures from a couple of view of this model are shown in Figure 2 below (to get the full effect, however, the reader must load the .kmz file, varroa.kmz , in Google Earth®). Placemarks, as red balloons on sticks, mark the spatial/temporal appearance of a varroa recording. One drawback to this form of display is that the reader can see at most half the globe at once. The merit, however, of being able to control the globe, in orientation, scale, and other factors, generally outweights this single universal drawback to this sort of display and indeed this merit helps to overcome that drawback. Also, consider taking "tour" of the placemarks in Google Earth®: pull down "Tools" and then choose "Play Tour" (track the place names on the left side of the screen). To see the pattern in a single year (of more than one entry) or in a subset of years, click off all the other years individually (on the left side of the display). Or, make changes in the color, shape, height, opacity, or other cosmetic features of the placemarks; simple changes can emphasize differences in pattern. In the future, look to see further refinement of this data display, and of the content (added text, GPS coordinates of sightings, and so forth), that has now become possible by zooming in and out on the user-controlled globe!




View: Core
Primary Database
Terrain
Geographic Web
Image NASA
Image © 2007 TerraMetrics
© 2007 Europa Technologies
©2007 Google™
Streaming 100%
Eye alt 7067.61 mi

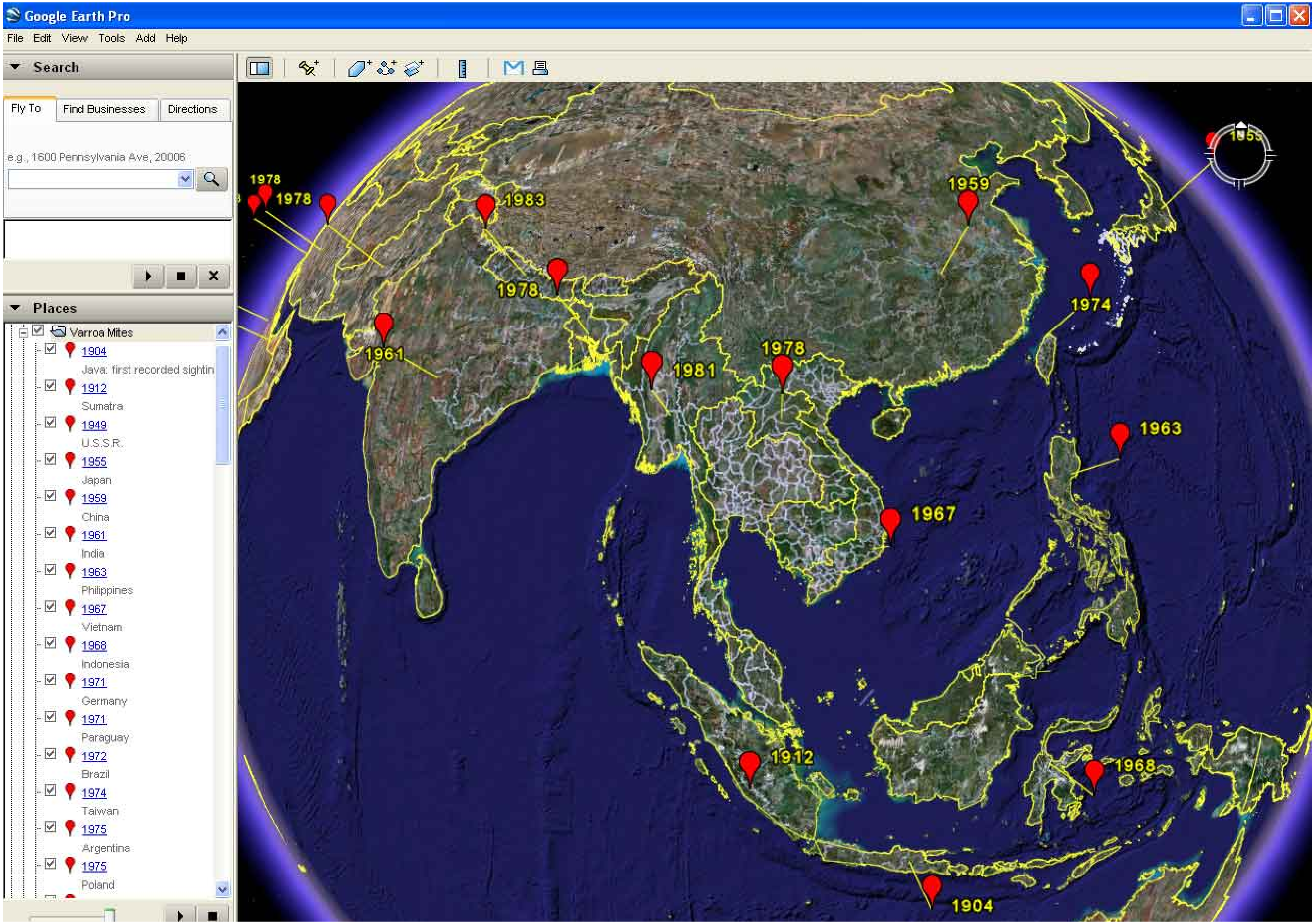

Google Earth Pro
File Edit View Tools Add Help
Search
Fly To
Find Businesses
Directions
e.g., 1600 Pennsylvania Ave, 20006
Places
Varroa Mites
1904
Java: first recorded sightin
1912
Sumatra
1949
U.S.S.R.
1955
Japan
1959
China
1961
India
1963
Philippines
1967
Vietnam
1968
Indonesia
1971
Germany
1971
Paraguay
1972
Brazil
1974
Taiwan
1975
Argentina
1975
Poland
1978
1978
1983
1959
1978
1974
1961
1981
1978
1963
1967
1912
1968
1904

Figure 2. Google Earth® screen captures. Model, varroa.kmz, by Sandra Lach Arlinghaus

The Prediction of Indian Monsoon Rainfall: A Regression Approach

Goutami Bandyopadhyay goutami15@yahoo.co.in

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

Abstract

The present paper analyses the monthly rainfall data of the Indian summer monsoon months between 1871-1999. Multiple linear regression is used to predict the average summer-monsoon rainfall using the previous years' data from the corresponding time period.

Keywords: summer-monsoon, India, rainfall, prediction, multiple linear regression

1. Introduction

India is basically an agricultural country and the success or failure of the harvest and water scarcity in any year is always considered with the greatest concern. These problems are closely linked with the behavior of the summer monsoon rains in India (Rajeevan, 2001). The term *monsoon* seems to have been derived either from the Arabic *mausin* or from the Malayan *monsin*. As first used it was applied to southern Asia and the adjacent waters, where it referred to the seasonal surface air streams which reverse their directions between winter and summer, southwest in summer and northeast in winter in this area. During the summer the continent is heated, leading to rising motion and lower pressure. This induces airflow from sea to land at low elevations.

Eastern and southern Asia has the earth's largest and best developed monsoon circulations. The tropical monsoon circulation of southern Asia, including India-Pakistan and Southeast Asia, differs significantly from East Asia monsoon. The Indian monsoon is effectively separated from that of China by the Himalayan-Tibet system. In summer a deep and widespread surface pressure trough extends across northern India-Pakistan into Southeast Asia. This is part of the planetary intertropical convergence zone, which here reaches its maximum poleward displacement. To the south of the trough is a deep current of maritime tropical air called the southwest monsoon. This current appears to originate in the southeast trades of the Indian Ocean east of Africa. As this stream of air approaches and crosses the equator its direction becomes southerly and then southwesterly. Along the Somali coast of Africa the flow becomes especially strong, taking the form of a low-level jet. In crossing the Arabian Sea the southwesterly current gains considerable moisture and becomes less stable. This unstable southwesterly current crosses India, continues eastward over the Indochina peninsula, and then moves northward over much of eastern Asia. It is a great moisture source for most of southern Asia.

2. Indian monsoon rainfall- brief literature review

Guhathakurta (2006) implemented an Artificial Neural Network method in predicting monsoon-rainfall over Kerala, a southern state of India. This paper proved that a neural net approach could be applicable to predict rainfall over districts of Kerala up to 2003. But, a major drawback of this paper is that it did not analyze the autocorrelation structure of the rainfall time series and chose the input matrix quite arbitrarily. Gadgil et al. (2002) discussed various aspects of summer monsoon rainfall prediction. Rejeevan (2001) discussed various problems associated with prediction of Indian summer monsoon rainfall. Gadgil et al. (2005) discussed the reasons behind failure in prediction of Indian summer monsoon rainfall. Hasternrath (1988) discussed the usefulness of regression model in predicting Indian summer monsoon rainfall.

3. Data and analysis

Data used in the present study are collected from the website http://www.tropmet.res.in published by Indian Institute of Tropical Meteorology. In this study only four months' data (June-August) are explored because these three months are the Indian summer monsoon months.

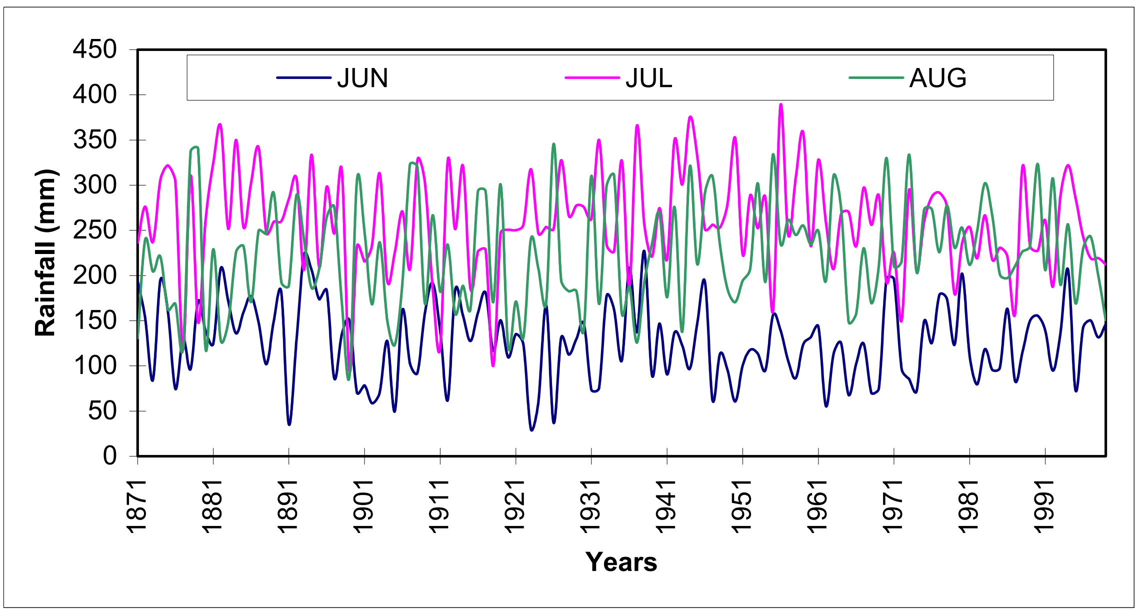


Fig.01- Monthly rainfall (mm) in the summer monsoon season over India during 1871-1999

The Pearson correlations between the data pertaining to different months are computed and are displayed in Table-01. The table shows that the months are not significantly correlated with respect to monthly rainfall of the summer monsoon season.

Months	Pearson correlation coefficients
Jun-Jul	-0.059362253
Jun-Aug	-0.013968157
Jul-Aug	-0.013968157

Table 01- Tabular presentation of the monthly rainfall amounts of different monsoon months

4. Multiple Linear Regression Model

In this research paper, a multiple linear regression (MLR) method is adopted to predict the average summer monsoon rainfall in a given year using the monthly rainfall data of the summer-monsoon of the previous year.

After computation, the MLR equation is set as

y=0.03x1+0.06x2+0.02x3+229

Where, x1= June rainfall of year Y

x2= July rainfall of year Y

x3= August rainfall of year Y

y= Average rainfall of year Y+1

[1]The actual and predicted average rainfalls are presented in Fig.02.

Readers wishing to see the detail of the derivation of the MLR equation are encouraged to contact the author directly.

[1] Paper submitted to *Solstice-An Electronic Journal of Geography and Mathematics*

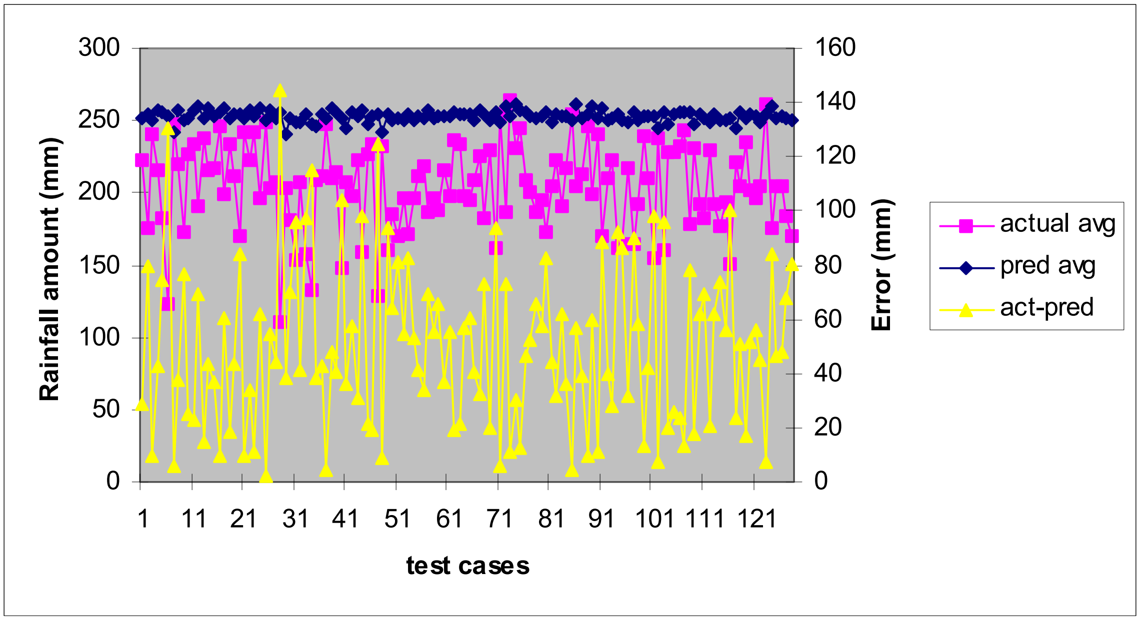


Fig.02- Schematic of the actual and predicted average monthly rainfall of Indian summer monsoon during 1872-1999

Overall prediction error is found to be 26.46%.

The t-statistics are computed from the MLR related components and are computed and tabular values are compared. The computed values are based upon the hypothesis that the input value is a good predictor of the predictand. The computed and critical tabular values are presented in Fig.03.

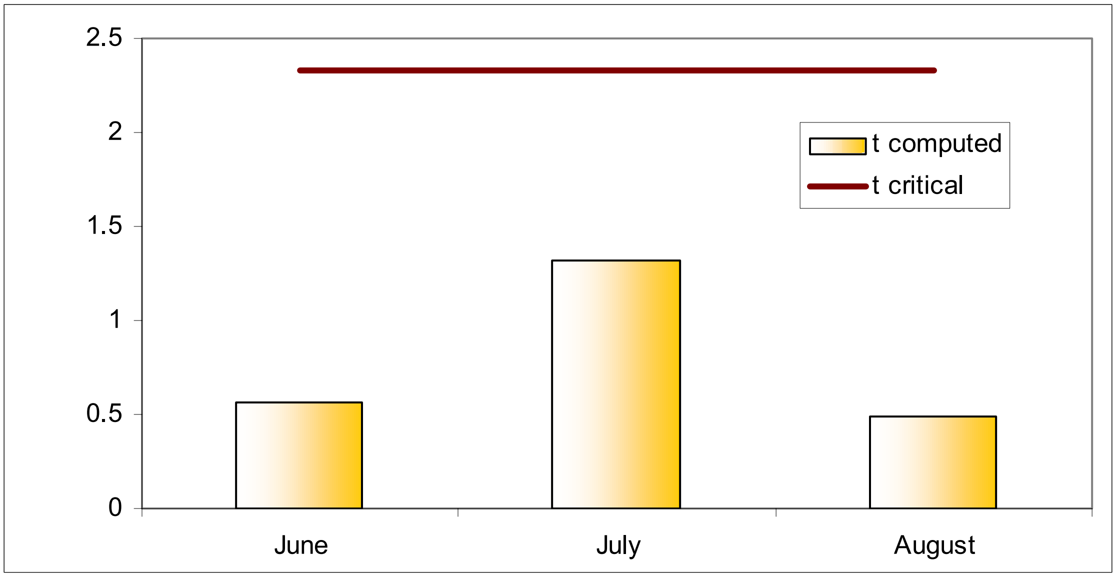


Fig.03- The computed and critical t-values

It is found that all the computed t-values fall below the tabular values of t. This illustrates, in the sample studied, that none of the months is a good predictor of average monsoon rainfall of a given year. The criticism (mentioned in section 2 above) of the neural net approach, that it did not analyze the autocorrelation structure of the rainfall time series, may be unfounded. The topic of monsoon-rainfall data series is highly complex; the role that multiple linear regression might play in this topic is one for future research—it appears, from the evidence here, not to be useful as a predictive model. Whether it might be useful for offering an approximate value of future monsoon rainfall remains to be seen.

References

- Gadgil, S et al (2002) "On Forecasting the Indian summer Monsoon: the Intriguing Season of 2002", Available at: http://caos.iisc.ernet.in/formon_12aug.pdf
- Guhathakurta, P (2005) "Long-range monsoon rainfall prediction of 2005 for the districts and sub-division Kerala with artificial neural network", *Current Science*, 90, 773-779
- Rajeevan, M (2001) "Prediction of Indian summer monsoon: Status, problems and prospects", *Current Science*, 81, 1451-1457
- Gadgil, S et al (2005) "Monsoon prediction – Why yet another failure?", *Current Science*, 88, 1389-1400
- Hastenrath, S (1988) "Prediction of Indian Monsoon Rainfall: Further Exploration", *Journal of Climate*, 1, 298-304

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 (NO_X) 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.

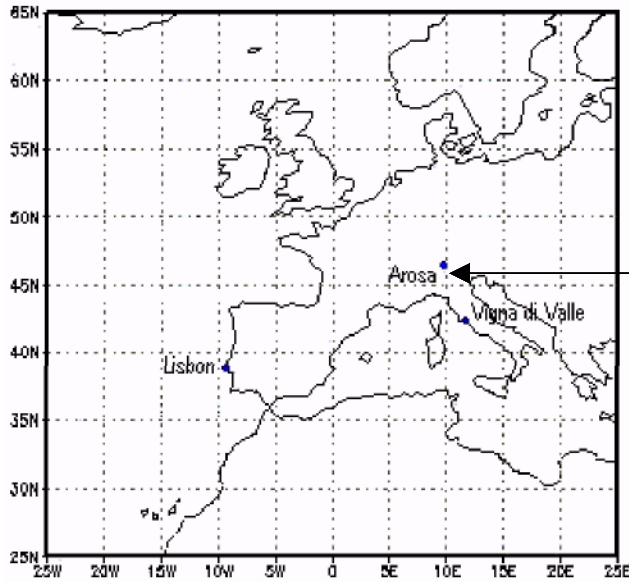
Total ozone is a measure of the number of ozone molecules between the ground and the top of the atmosphere. Total ozone is, mathematically, the integral of the ozone concentration with respect to height.

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)

$$r_k=[\text{Covariance }(x(n\text{-}k),x\_(n\text{-}k)]/[\text{Stdev}(x(n\text{-}k)][\text{Stdev}(x\_(n\text{-}k)] \quad \ldots \quad (1)$$

Where, r_k = autocorrelation of order k

$x(n\text{-}k)$ = 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.

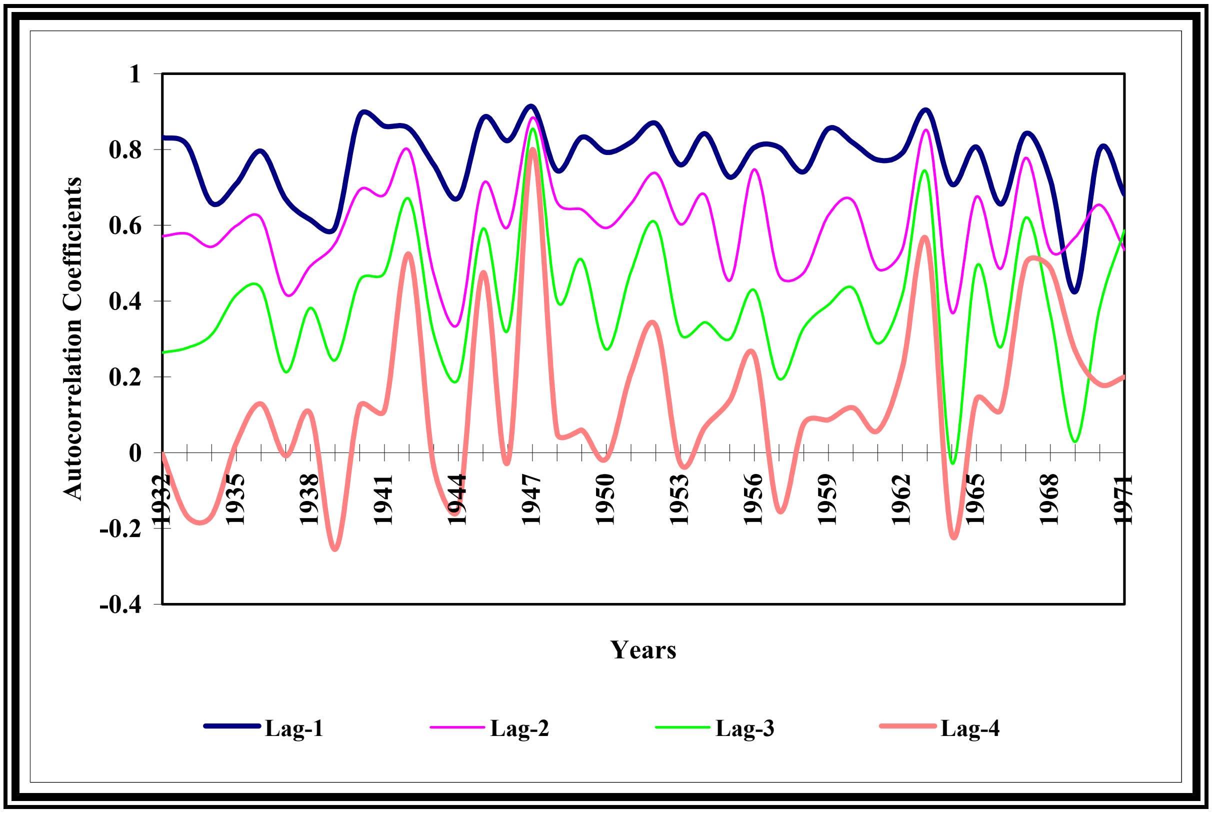


Fig.01-Autocorrelation Coefficients of different orders pertaining to mean monthly total ozone concentration time series in different years within study period.

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.

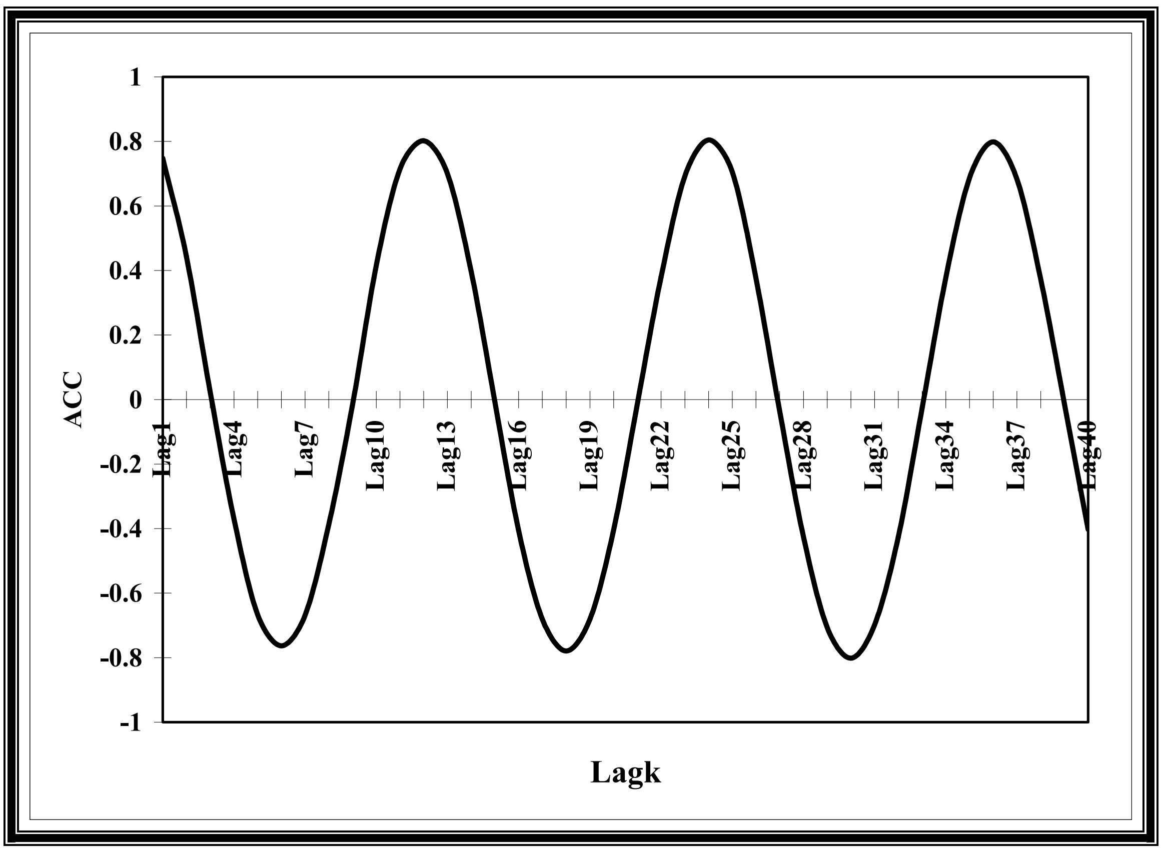


Fig.02-Autocorrelation Coefficients of different orders pertaining to mean monthly total ozone concentration time series with 480 data points.

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)

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

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.

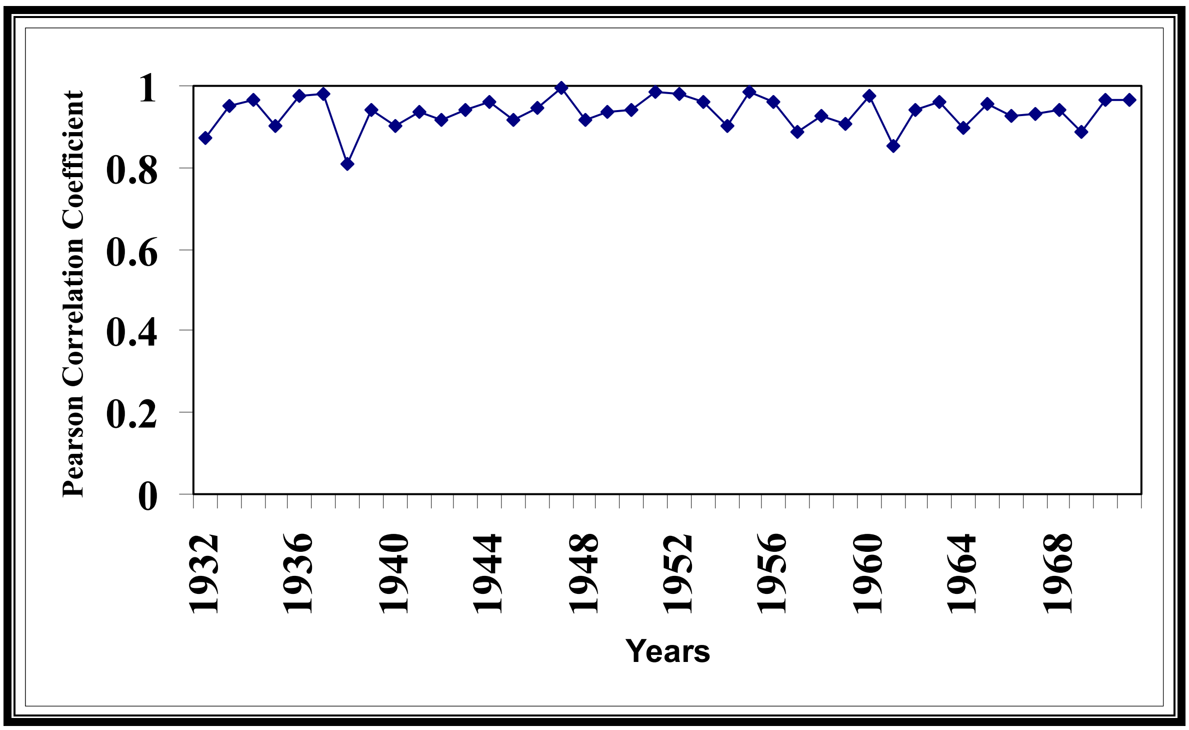


Fig.03- Schematic showing the linear association between actual monthly total ozone concentrations and those predicted through AR (2) model.

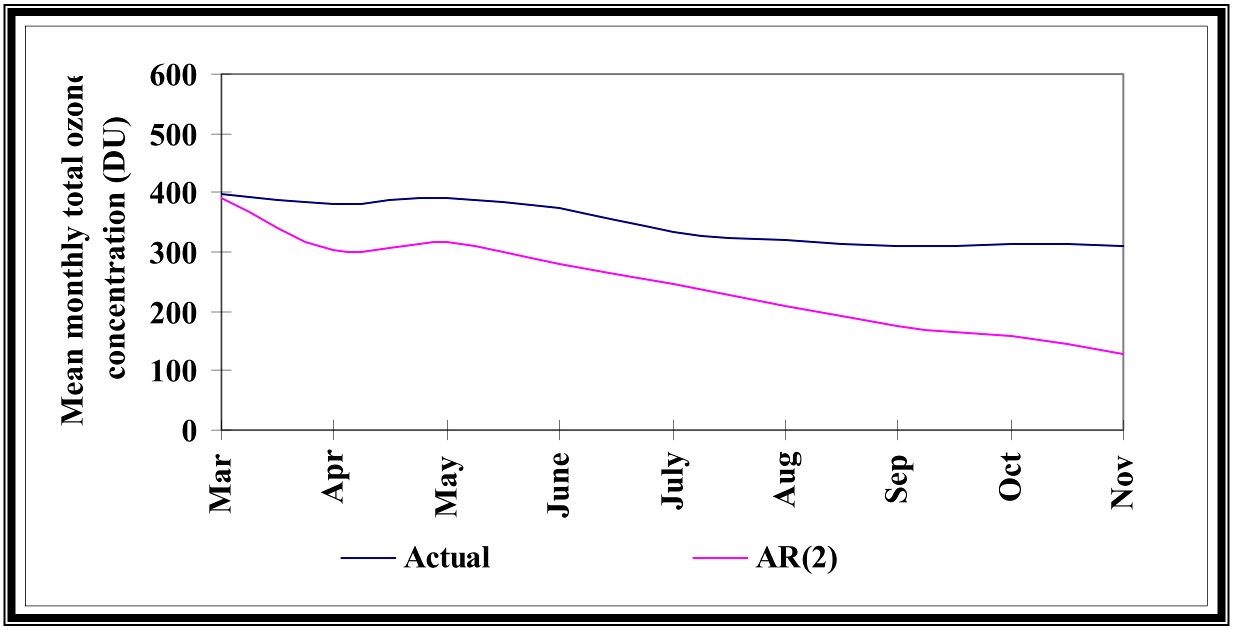


Fig.04- Schematic showing the actual and predicted (AR (2)) mean monthly total ozone concentration over Arosa in 1933

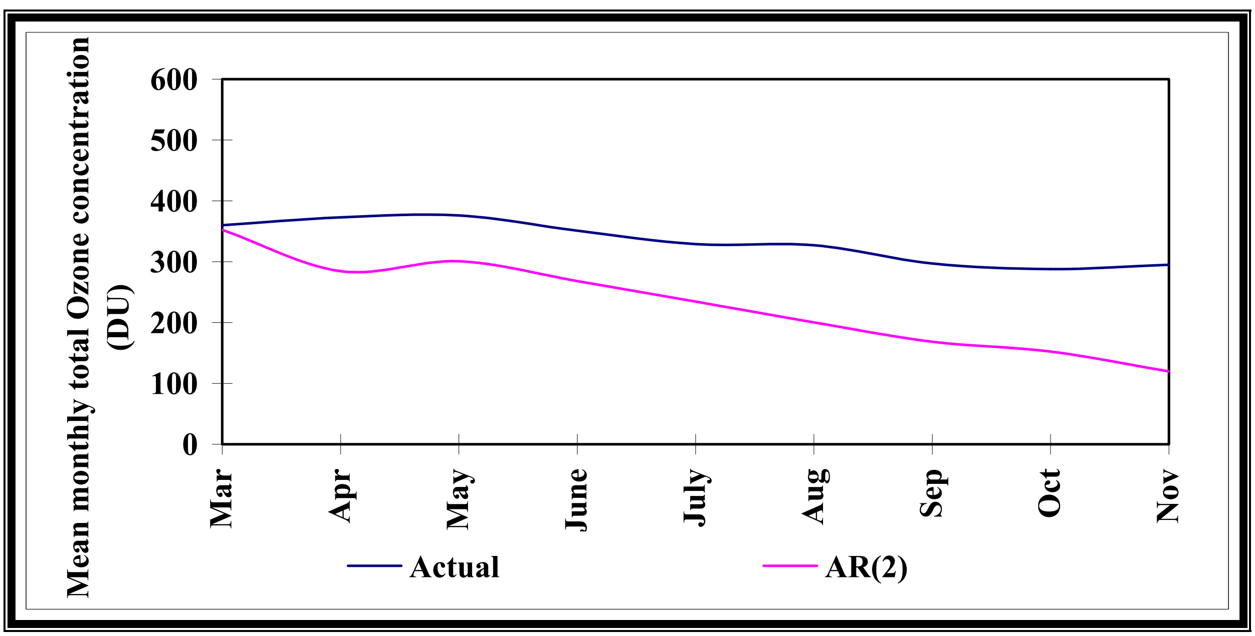


Fig.05- Schematic showing the actual and predicted (AR (2)) mean monthly total ozone concentration over Arosa in 1945

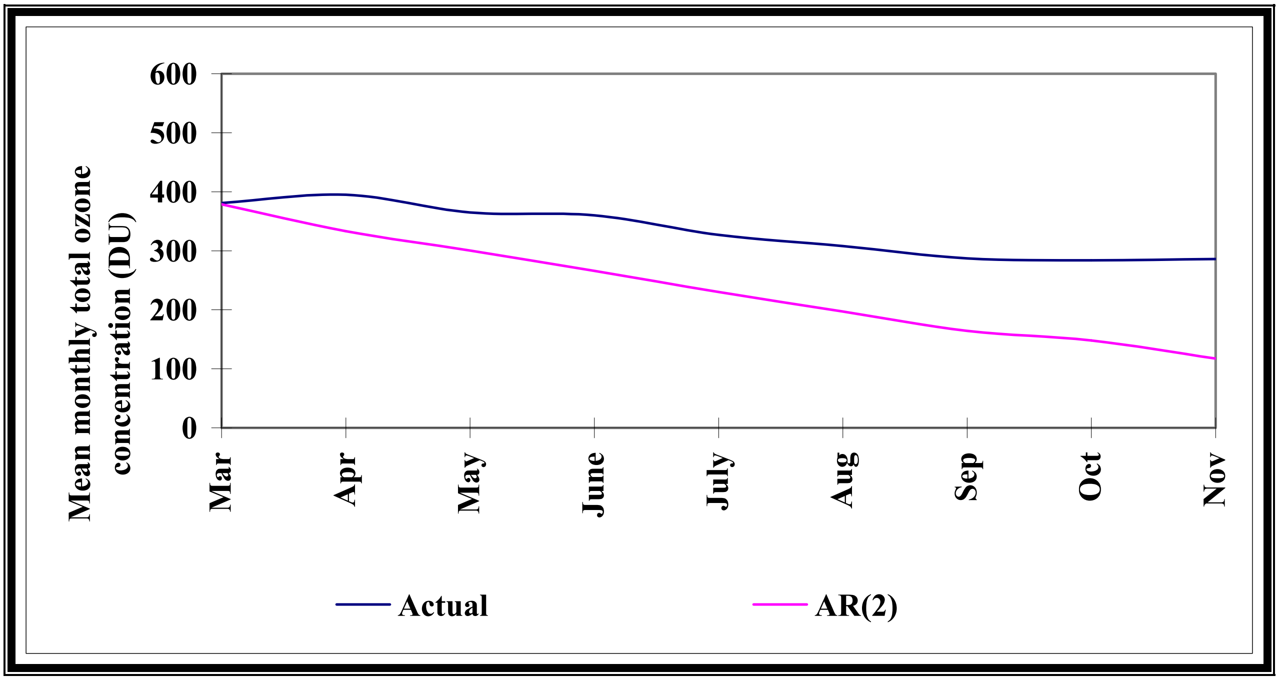


Fig.06- Schematic showing the actual and predicted (AR (2)) mean monthly total ozone concentration over Arosa in 1956

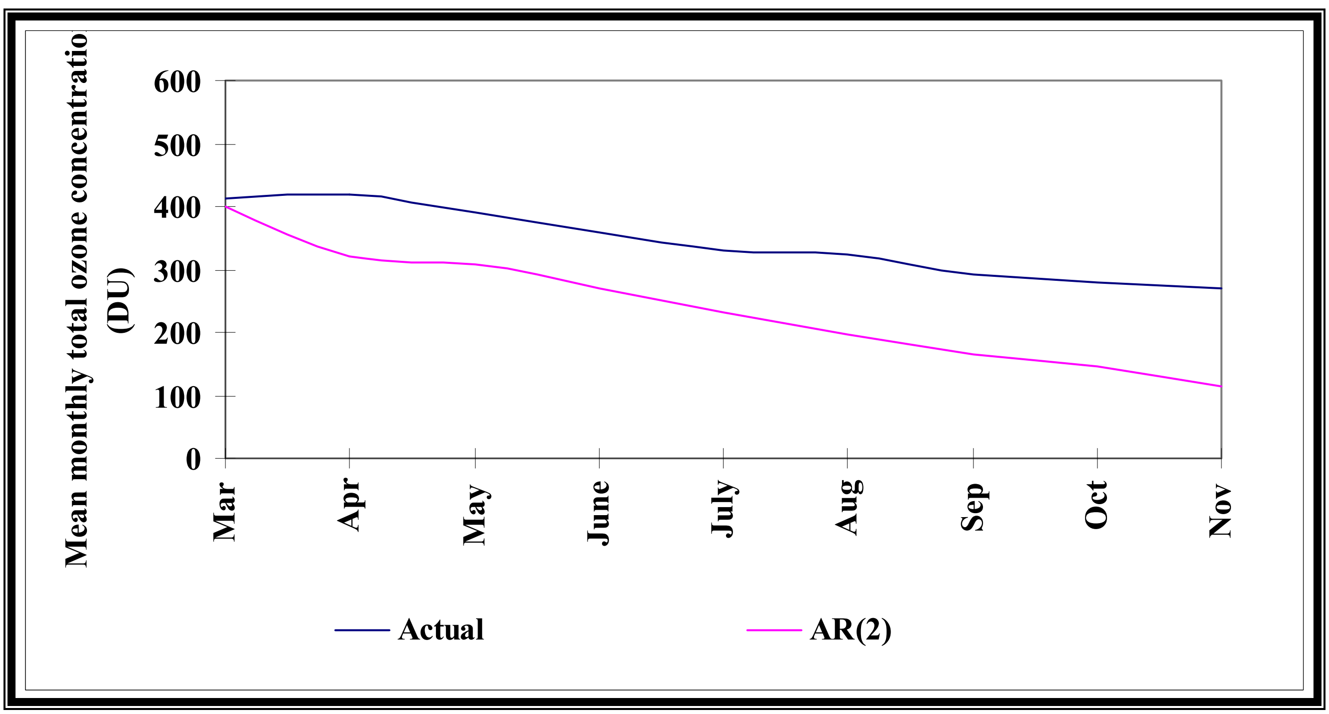
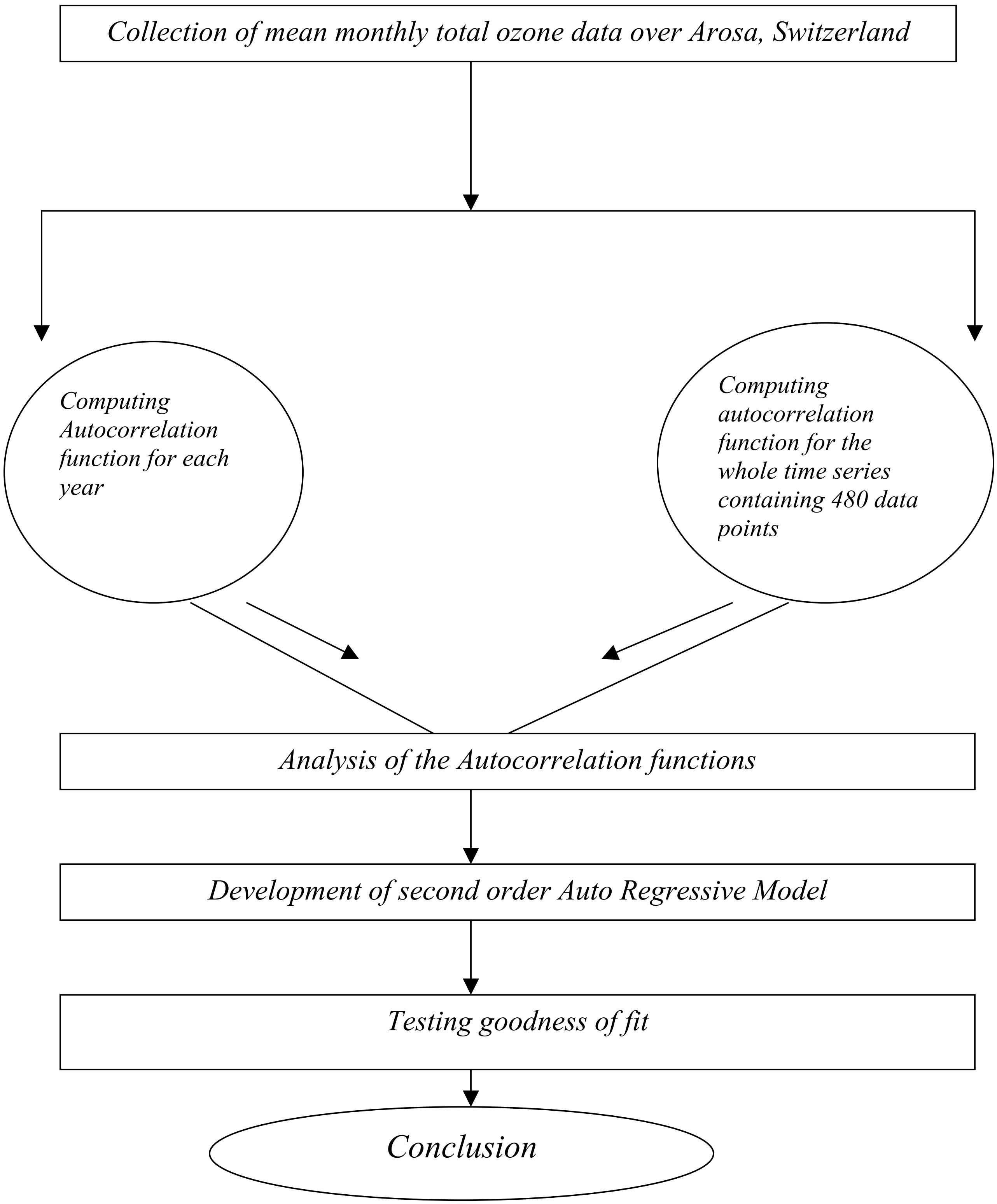


Fig.07- Schematic showing the actual and predicted (AR (2)) mean monthly total ozone concentration over Arosa in 1970

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



Mail
List of Individuals who sent e-mail subsequent to the last issue.

Cathy Antonakos
Kim O'Brien Chalmers
Urska Demsar
Nick Garaycochea
Senan Gorman
Bruno Granier
James Greene
Joan Gerard
Joseph Kerski
Harold Moellering
Caroline Mohai
Ellen Offen
Bethany Osborne
Lynn Scott
Carlo Werlen
