TIME SERIES ANALYSIS

V. A. Gordin
Hydrometeorological Centre of Russia & Mathematical College of Independent Moscow University,Russia

Keywords: Windrose, Maunder's butterfly, $m-$ th order finite-difference ordinary equation, Fibonacci numbers, characteristic equation, Green function, boundary conditions, logistic equation, fixed point, stable point, basin of attraction, periodical points, Jacobi matrix, super-convergence, neutral case, Lotka – Volterra equation, Newton method, Julia set, fractal structure, Hausdorff dimension, weakened Newton method, Euler scheme, Runge – Kutta methods, Hoin scheme, Lebesgue constant, interpolational polynomial (Lagrange form and Newton form), basic interpolation polynomial, generalized Lagrange polynomial, difference quotients, Chebyshev polynomial, spline of degree n with the defect k, Schoenberg spline, rational spline, three-diagonal matrix with diagonal dominance, double-sweep method, local spline (Ryabenkii spline), Fourier transform, wavelet, method of least squares, Chebyshev (uniform) approximation, Weierstrass theorem, Chebyshev theorem, points of alternance

Contents

Summary

Different kind of natural time series as well as the methods (deterministic and stochastic) of their study are considered. Checking, interpolation, and approximation

can be useful for the experimental data assimilation. The equation series that are obtained as a result of calculations, e.g., as solutions of a finite – difference or as approximate solutions of a transcendental equation are considered, too.

1. Introduction

A set of numbers or vectors can be considered as a discrete times series if there is a "natural" ordering of the set: $< x_1, ..., x_N >$. The series can describe very different processes: historical, physical, chemical, financial, biological etc. They can be arising in a process of some recurrent (iterative) calculations. The series may be obtained both as a result of reproducible or irreproducible experiments. Sometimes one tries to minimize the number of expensive measurements. Sometimes the series is a result of "free" computer experiments.

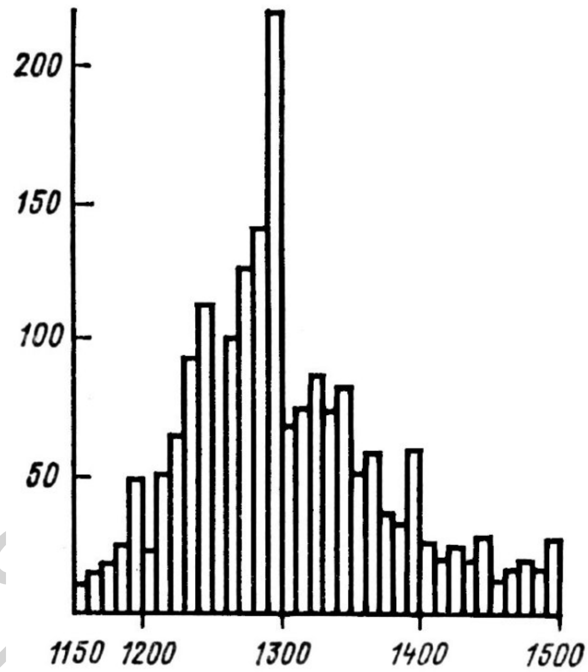

Figure 1. Distribution of towns' foundations years in Central Europe. Averaging by the period of 10 years. One can analyse the data, but a repetition of the experiment is impossible.

The digital times series will be considered here, when a corresponding ***range of values (a phase space***) is a finite-dimensional linear space, a manifold, or a grid. For instance, the phase space for the dynamics of a temperature and wind's direction at a given point, the number of animals in a population are the half-line $\mathbb{R}_+$, are the circumference $\mathbf{S}^1$ (or two-dimensional sphere $\mathbf{S}^2$, if the direction of three-dimensional wind is considered), and $\mathbb{Z}_+$.

There are more complex time series, when its elements x_j are some patterns (pictures, texts, speeches, music, etc). Usually one tries to imbed the ranged values into a suitable linear space. The impact of the imbedding is significant for the following recognition of the patterns.

Often the discrete time series is a restriction ($t = n\tau,\ \tau = const,\ n = 1,...,N)$) of a continuous process, i.e., $x = x(t), t \in \mathbb{R}$. Then the value τ is equal to the period between observations of the process. If one chooses a coarser step: $\theta = k\tau,\ M = N/k$, then the thinning out series $< y_1,..., y_M >$, where $y_i = x_j,\ j = ik$ or $y_i = (x_j + ... + x_l)/k,\ \ j = k(i-1)+1,\ \ l = ki$, can have lost details of the original process with short period (high frequency). For instance, a mean temperature for a given month cannot describe the concrete temperature of tomorrow – one needs a study with another temporal scale.

Sometimes an averaging of a time series can help to separate a useful signal from a significant noise.

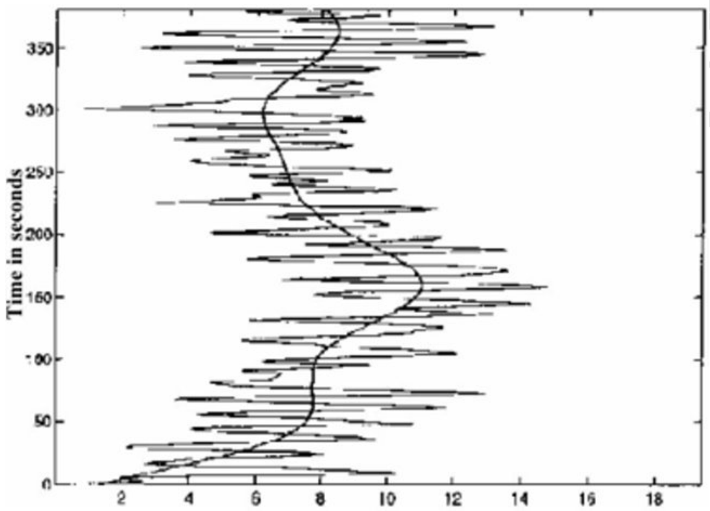

Figure 2. GPS (Global Position System) raw (the oscillations are the result of the "pendulum effect", i.e., oscillation of the sound under a balloon) and the averaged (with respect to time) wind speed. The vector function $\vec{v}$ is averaged here and its absolute value is shown (solid line).

The period $T = \tau(N-1)$ of measurements also limits the class of phenomena that are available for evaluations by the data.

Example: During several decades the level (mean annual) of the Caspian Sea decreased. The Soviet government began a huge work for its stabilization. In particular, some years ago the Kara-Bogaz-Gol Bay (Garabogazkol Aylagy) was separated from the sea by a dam to reduce the surface evaporation of the Sea. However, the level began to increase before the principal (and very expensive) works started, and the coastlanders had big problems with the flooding. Perhaps, the period of the sea level observations was not sufficient.

Another variant of a time series t_j is composed from instants of some events, e.g., hurricanes in an area. Then the interval $\left|t_{j+1}-t_j\right|$ strongly depends on j and they are subjects of investigations.

Let the times series $x(n)$ describe a process, where n is a discrete analogue of the time, x is a scalar variable or a vector, real or complex.

To describe shortly (and therefore, coarsely) the information from a long time series one can construct its mean value and the frequencies of various values.

Windroses cannot inform about possible scenario of the wind's dynamics and moreover about physical reasons of the statistics. However, it can be useful for a practical activity (e.g., for a building) in the corresponding region as well as for a verification of the models of the atmospheric circulation.

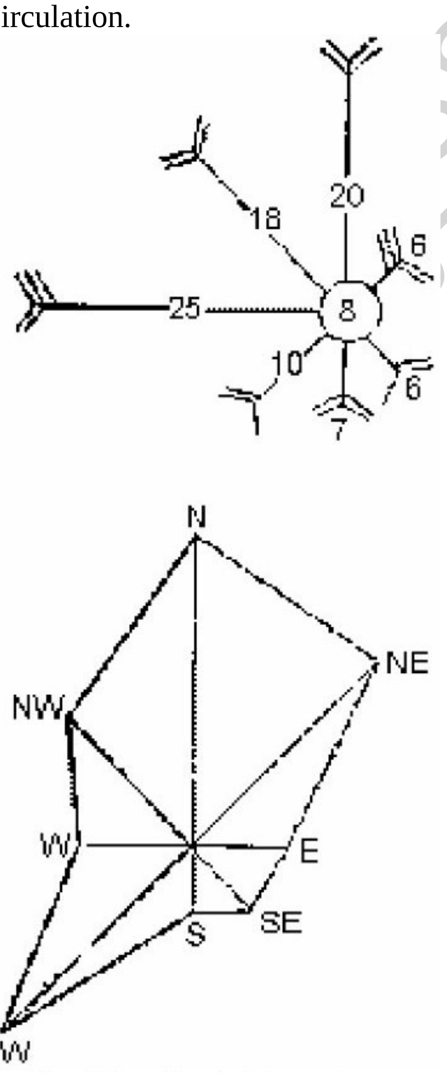

Figure 3. Two versions of a windrose are drawn. One can understand from the lower windrose relative frequencies of winds of any compass point (calm is excluded). On the upper windrose both the length of arrows and the values in the circles in every compass points (calm is taken into account here). The numbers of small lines mean the mean velocities (m/sec) for winds of a given compass point. Thus, one can determine, e.g.,

where and on what length shifts a particle of air during one hour over the region, season etc., for which the rose-wind was evaluated. For instance, according to the upper windrose, north-east is rarer than south-west, but stronger.

The Maunder's butterfly is a more complex statistical description. The color on Figure 4 describes the density of solar spots as a function of the time and the latitude. One can see here not only a periodicity with respect to time. In the beginning of a cycle the spots arise about 35° and then their maxima (north and south) move to the equator. Thus, the time series in the example takes its values in the space of distributions along latitude.

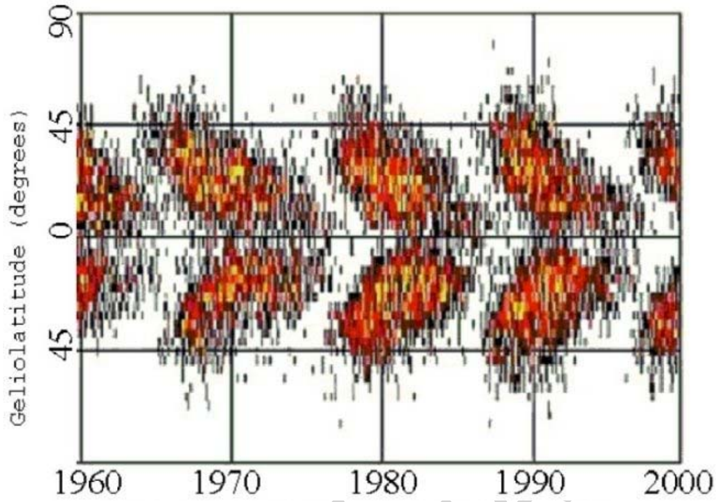

Figure 4. The Maunder's butterfly. The color characterizes the density and intensity of solar spots as a function of time and latitude.

One can verify modern and future magnetohydrodynamic models: are they generating the statistical picture or not ?

2. Finite-difference Equations

The processes $x(n)$ can informally be divided as: i) closed; ii) unclosed.

i) There is a function $f = f(n, z_1, ..., z_m)$ such that for any n

$$x(n) = f(n, x(n-1), ..., x(n-m)). \quad (1)$$

The formula (1) can be considered as a $m-$ ***th order finite-difference ordinary equation*** (or a ***system,*** if x and f are vectors). It is called ***autonomous*** if the function f does not depend on n, and ***linear***, if the function f is linear with respect to any arguments $z_1, ..., z_m$. The equation can be rewritten as a first order finite-difference m-th order system

$$\vec{y}(n) = \vec{F}(n, \vec{y}(n-1)); \quad (2)$$

where $\vec{y}(n) =< x(n), x(n-1),..., x(n-m+1) >$; $\vec{F} =< f, x(n-1),..., x(n-m+1) >$.

2.1 Linear FDE

Example: The ***Fibonacci*** (XIII century; his another name is Leonardo from Pisa***) numbers*** are obtained according to the formula

$$x(n) = x(n-1) + x(n-2), \qquad (3)$$

where $x(0) = 1, x(1) = 1$. Then $x(2) = 2, x(3) = 3, x(4) = 5, x(5) = 8$ etc. Eq. (3) is autonomous and linear, too. Here $m = 2$.

In the general case m given values $x(0), x(1),..., x(m-1)$ fix a unique solution of Eq. (1).

The general solution of (3) can be represented in the form $C_1\lambda_1^n + C_2\lambda_2^n$, where λ can be computed as roots of the ***characteristic*** (for Eq.(2)) quadratic ***equation*** $\lambda^2 = \lambda + 1$, i.e., $\lambda_{1,2} = 0.5 \pm \sqrt{5/4}$.

For the given initial data $x(0) = 1, x(1) = 1$ the constants are $C_1 = \dfrac{1-\lambda_2}{\lambda_1 - \lambda_2}$, $C_2 = \dfrac{\lambda_1 - 1}{\lambda_1 - \lambda_2}$.

The general solution of a linear m-th order homogeneous finite-difference equation with constant coefficients

$$a_0 x(n) + a_1 x(n-1) + ... + a_m x(n-m) = 0, \qquad a_0, a_m \neq 0, \qquad (4)$$

can be represented in the form of a linear combination of ***basic solutions*** of (4):

$\sum_{j=1}^{m} C_j \lambda_j^n$, if the roots λ_j of the algebraic ***characteristic equation*** for (4):

$$a_0\lambda^m + a_1\lambda^{m-1} + ... + a_m = 0$$

are not multiple, and in the form $\sum_{j=1}^{d} \sum_{i=1}^{p_j} C_{ji} n^{(i-1)} \lambda_j^n$, otherwise. Here p_j is the multiplicity of the root λ_j, C_j and C_{ji} are arbitrary constants. The constants can be determined, if initial values $x(0),..., x(m-1)$ are given.

ii) A non-homogeneous linear finite-difference equation

$$a_0 x(n) + a_1 x(n-1) + ... + a_m x(n-m) = \varphi(n) \qquad (5)$$

gives us a simplest example of a unclosed process; $\varphi(n)$ is a forcing at $t = n\tau$. The general solution of Eq. (5) is $X(n) + \sum_{j=1}^{d} \sum_{i=1}^{p_j} C_{ji} n^{(i-1)} \lambda_j^n$, where $X(n)$ is a partial solution of (5). It can be represented via a ***Green function:***

$$X(n) = \sum_{j=-\infty}^{\infty} K(n, j)\varphi(j),$$

where $K(n, j)$ can be represented explicitly via independent solutions of Eq.(4); if the coefficients a_k are not depending on n, then $K = K(|n - j|)$.

2.2 Boundary Value Problems

If instead of initial conditions ($< x(0),..., x(m-1) >$ for Eq. (1) or $\vec{y}(0)$ for (2)) some ***boundary conditions*** are given, e.g., $< x(0),..., x(k), x(N + k), ..., x(N + m - 2) >$, N>1, then the algebraic linear system on the unknown coefficients C_{ji}, may be singular. However, in some important cases its non-singularity can be proved.

-
-
-

Bibliography

Ahlberg J.H., Nilson E.N., Walsh J.L.(1967) *The Theory of Splines and Their Applications.* Acad. Press, New York and London, 284pp. [This is a handbook on splines, estimations of interpolation errors, algorithms etc.]

Anderson T.W. (1971) *The Statistical Analysis of Time Series*. John Wiley & Sons, New York, 718pp. [This includes a review of methods of evaluation of covariances, spectral densities, covariance functions etc.]

Arnold V.I. *Ordinary Differential Equations*. “Nauka", Moscow, 1984, 271pp. [Russian], MIT Press, Cambridge (Massachusetts), London, 1988, 280pp. [English], *Equations differentielles ordinaires*. Moscou, “Mir", 1988, 334pp. [French]. [Classical handbook on ODE, including Lotka – Volterra eq.]

Babenko K.I. (1986). *Foundation of Numerical Analysis*. [Russian], Moscow, “Nauka”, 744pp. [This considers the interpolation and approximation problems.]

Daubechies I. (1992). *Ten Lectures on Wavelets*. SIAM, 357pp. [Popular and strong explanation of the wavelet theory.]

Digital Pattern Recognition. (1976) Ed. R.S.Fu. Berlin-Heidelberg-New York, Springer, 206pp. [The reduction of various patterns into digital pattern recognition is considered, e.g., syntactic (linguistic) patterns and speech ones; resemblance and dissemblance; clustering algorithms; speech understanding.]

Freidlin M. (1996) *Markov Processes and Differential Equations.* Birkhauser, 148pp. [Differential equations with a stochastic forcing are considered here.]

Gelfand I.M., Vilenkin N.Ia. (1964) *Some Applications of the Harmonic Analysis. The Framed Spaces*. London, Acad. Press, 384pp. [This describes the applications of the generalized functions (distributions) to the theory of stationary (and generalized stationary) random processes.]

Gelfond A.O. (1967). *The Finite-Difference Calculus*. "Nauka", Moscow, 375pp. [Russian]. [Handbook on finite-difference equations.]

Gordin V.A. (2000). *Mathematical Problems and Methods in Hydrodynamic Weather Forecasting*. Gordon & Breach Publ. House, 842pp. [This includes a description and applications of the mentioned (and contiguous) approaches.]

Hannan E.J. (1970). *Multiple Time Series*. John Wiley & Sons, Inc. New York etc. 536pp. [This is a view on time series analysis from the point of probability theory.]

Khasminskii R.Z. (1980) *Stochastic Stability of Differential Equations.* Sijthoff & Noordhoff, Alphen aan der Rijn, 344pp. [Differential equations with a stochastic forcing are considered here.]

Kiepenheuer K.O. (1953) Ch.6. *Solar Activity*. pp. 322-465. In *The Solar System*. v.1, *The Sun*. Ed. Kuiper G.P. Univ. Chicago Press, Chicago, Illinois. [Maunder's butterfly and other information about solar activity.]

Mandelbrot B.B. (1983). *The Fractal Geometry of Nature*. Freeman, San Francisco, 477pp. [Theory and examples of fractal phenomena.]

Peitgen H.-O., Richter P.H. (1986). *The Beauty of Fractals*. Springer-Verlag, 211pp. [Theory and examples of fractal phenomena, including a lot of beautiful computer pictures.]

Percival D.B., Walden A.T. (2000). *Wavelet Methods for Time Series Analysis.* Cambridge Univ. Press, 594pp. [Fundamental source of the wavelet approach: theory, applications, comparison with Fourier' one.]

Ruelle D. (1987) *Chaotic Evolution and Strange Attractors: The Statistical Analysis of Time Series for Deterministic Non-linear Systems*. Cambridge Univ. Press, 107pp. [In particular, evaluation of Lyapunov's exponents.]

Sanz-Serna J.M., Calvo M.P. (1994) *Numerical Hamiltonian Problems*. Chapman & Hall, London, 224pp. [The special methods for Hamiltonian systems are specially useful for long integration periods.]

Torrence C., Compo G.P. (1998) *A Practical Guide to Wavelet Analysis.* Bull. Am. Meteor. Soc., v.79 (1), pp. 61-78. [Geophysical examples of wavelet applications; source of Fig.11, 12.]

Biographical Sketch

Vladimir Alexander Gordin, was born on May 18, 1949, Leningrad, USSR. He left special mathematical Moscow school N 2 in 1966, and he graduated in 1972 in Moscow Institute of Electronic Machines as M. Sci. in Applied Mathematics. 1972- : collaboration in Hydrometeorological Center of USSR (later of Russia); presently leader fellow scientist. 1999- teaching at Independent Moscow University his own "Applied Mathematics" course. Also he taught his own mathematical courses in special Moscow schools N2 and N1313 as well as in mathematical groups and summer school for students and undergraduates. PhD in Physics and Mathematics in 1979: Hydrometeorological Center of USSR; title of thesis: "The Study of the Finite Difference Approximations and the Boundary Conditions for Systems of Forecasting Equations". Dr.Sci. degree in Physics and Mathematics in 2000: Moscow Institute of Physics and Technology (MFTI). The title of the Dr.Sci. thesis: "Mathematical Problems and Methods in Hydrodynamical Weather Forecasting". He published the following books: "Mathematical Problems of the Hydrodynamical Weather Forecasting. Analytical Aspects". Gidrometeoizdat, Leningrad, 256p. (1987, Russian); "Mathematical Problems of the Hydrodynamical Weather Forecasting. Numerical Aspects". ibid, 264p. (1987, Russian); Mathematics, Computer, Weather Forecasting. V.A.Gordin. ibid, 224p. (1991, Russian); Mathematical Problems and Methods in Hydrodynamical Weather Forecasting. V.A.Gordin. Gordon & Breach, 2000, 842p. (English); and about 70 articles. Member of the Moscow and American Mathematical Society. For his taking part in August 1991 events he was decorated with the medal "Defender of Free Russia". B.A. in Jewish Sciences (Touro College, Moscow, 1996).