Analysis of Time Series Structure

SSA and Related Techniques

NINA GOLYANDINA
VLADIMIR NEKRUTKIN
ANATOLY ZHIGLJAVSKY

CHAPMAN & HALL/CRC

Boca Raton London New York Washington, D.C.

CRACKED TRADING SOFTWARE

70+ DVD's FOR SALE & EXCHANGE

www.traders-software.com

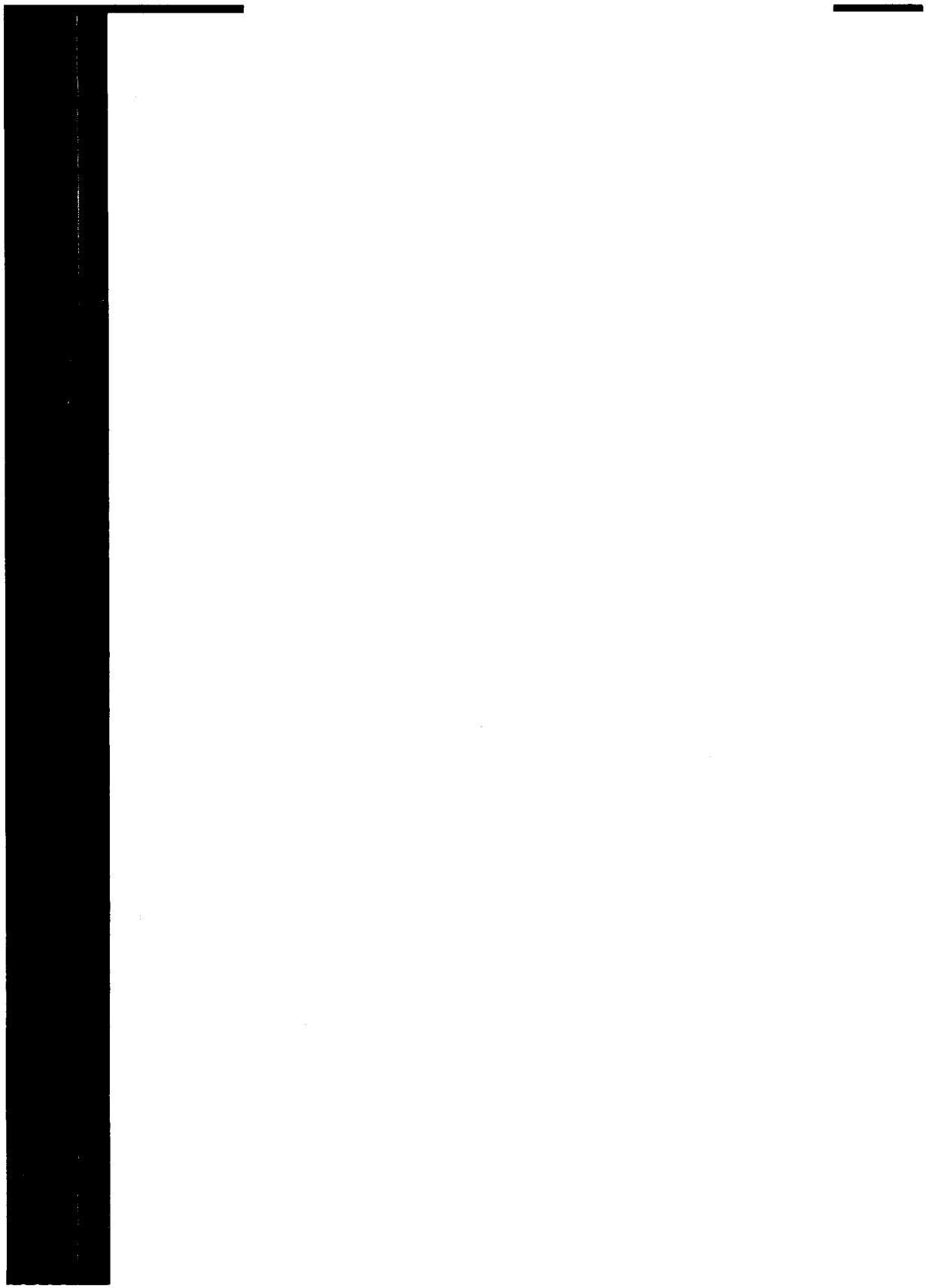
www.forex-warez.com

www.trading-software-collection.com

www.tradestation-download-free.com

Contacts

andreybbrv@gmail.com
andreybbrv@yandex.ru
Skype: andreybbrv



MONOGRAPHS ON STATISTICS AND APPLIED PROBABILITY

General Editors

D.R. Cox, V. Isham, N. Keiding, T. Louis, N. Reid, R. Tibshirani, and H. Tong

1 Stochastic Population Models in Ecology and Epidemiology *M.S. Barlett* (1960)
2 Queues *D.R. Cox and W.L. Smith* (1961)
3 Monte Carlo Methods *J.M. Hammersley and D.C. Handscomb* (1964)
4 The Statistical Analysis of Series of Events *D.R. Cox and P.A.W. Lewis* (1966)
5 Population Genetics *W.J. Ewens* (1969)
6 Probability, Statistics and Time *M.S. Barlett* (1975)
7 Statistical Inference *S.D. Silvey* (1975)
8 The Analysis of Contingency Tables *B.S. Everitt* (1977)
9 Multivariate Analysis in Behavioural Research *A.E. Maxwell* (1977)
10 Stochastic Abundance Models *S. Engen* (1978)
11 Some Basic Theory for Statistical Inference *E.J.G. Pitman* (1979)
12 Point Processes *D.R. Cox and V. Isham* (1980)
13 Identification of Outliers *D.M. Hawkins* (1980)
14 Optimal Design *S.D. Silvey* (1980)
15 Finite Mixture Distributions *B.S. Everitt and D.J. Hand* (1981)
16 Classification *A.D. Gordon* (1981)
17 Distribution-Free Statistical Methods, 2nd edition *J.S. Maritz* (1995)
18 Residuals and Influence in Regression *R.D. Cook and S. Weisberg* (1982)
19 Applications of Queueing Theory, 2nd edition *G.F. Newell* (1982)
20 Risk Theory, 3rd edition *R.E. Beard, T. Pentikäinen and E. Pesonen* (1984)
21 Analysis of Survival Data *D.R. Cox and D. Oakes* (1984)
22 An Introduction to Latent Variable Models *B.S. Everitt* (1984)
23 Bandit Problems *D.A. Berry and B. Fristedt* (1985)
24 Stochastic Modelling and Control *M.H.A. Davis and R. Vinter* (1985)
25 The Statistical Analysis of Composition Data *J. Aitchison* (1986)
26 Density Estimation for Statistics and Data Analysis *B.W. Silverman* (1986)
27 Regression Analysis with Applications *G.B. Wetherill* (1986)
28 Sequential Methods in Statistics, 3rd edition
G.B. Wetherill and K.D. Glazebrook (1986)
29 Tensor Methods in Statistics *P. McCullagh* (1987)
30 Transformation and Weighting in Regression
R.J. Carroll and D. Ruppert (1988)
31 Asymptotic Techniques for Use in Statistics
O.E. Bandorff-Nielsen and D.R. Cox (1989)
32 Analysis of Binary Data, 2nd edition *D.R. Cox and E.J. Snell* (1989)

33 Analysis of Infectious Disease Data *N.G. Becker* (1989)

34 Design and Analysis of Cross-Over Trials *B. Jones and M.G. Kenward* (1989)

35 Empirical Bayes Methods, 2nd edition *J.S. Maritz and T. Lwin* (1989)

36 Symmetric Multivariate and Related Distributions
K.T. Fang, S. Kotz and K.W. Ng (1990)

37 Generalized Linear Models, 2nd edition *P. McCullagh and J.A. Nelder* (1989)

38 Cyclic and Computer Generated Designs, 2nd edition
J.A. John and E.R. Williams (1995)

39 Analog Estimation Methods in Econometrics *C.F. Manski* (1988)

40 Subset Selection in Regression *A.J. Miller* (1990)

41 Analysis of Repeated Measures *M.J. Crowder and D.J. Hand* (1990)

42 Statistical Reasoning with Imprecise Probabilities *P. Walley* (1991)

43 Generalized Additive Models *T.J. Hastie and R.J. Tibshirani* (1990)

44 Inspection Errors for Attributes in Quality Control
N.L. Johnson, S. Kotz and X, Wu (1991)

45 The Analysis of Contingency Tables, 2nd edition *B.S. Everitt* (1992)

46 The Analysis of Quantal Response Data *B.J.T. Morgan* (1992)

47 Longitudinal Data with Serial Correlation—A state-space approach
R.H. Jones (1993)

48 Differential Geometry and Statistics *M.K. Murray and J.W. Rice* (1993)

49 Markov Models and Optimization *M.H.A. Davis* (1993)

50 Networks and Chaos—Statistical and probabilistic aspects
O.E. Barndorff-Nielsen, J.L. Jensen and W.S. Kendall (1993)

51 Number-Theoretic Methods in Statistics *K.-T. Fang and Y. Wang* (1994)

52 Inference and Asymptotics *O.E. Barndorff-Nielsen and D.R. Cox* (1994)

53 Practical Risk Theory for Actuaries
C.D. Daykin, T. Pentikäinen and M. Pesonen (1994)

54 Biplots *J.C. Gower and D.J. Hand* (1996)

55 Predictive Inference—An introduction *S. Geisser* (1993)

56 Model-Free Curve Estimation *M.E. Tarter and M.D. Lock* (1993)

57 An Introduction to the Bootstrap *B. Efron and R.J. Tibshirani* (1993)

58 Nonparametric Regression and Generalized Linear Models
P.J. Green and B.W. Silverman (1994)

59 Multidimensional Scaling *T.F. Cox and M.A.A. Cox* (1994)

60 Kernel Smoothing *M.P. Wand and M.C. Jones* (1995)

61 Statistics for Long Memory Processes *J. Beran* (1995)

62 Nonlinear Models for Repeated Measurement Data
M. Davidian and D.M. Giltinan (1995)

63 Measurement Error in Nonlinear Models
R.J. Carroll, D. Rupert and L.A. Stefanski (1995)

64 Analyzing and Modeling Rank Data *J.J. Marden* (1995)

65 Time Series Models—In econometrics, finance and other fields
D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen (1996)

66 Local Polynomial Modeling and its Applications *J. Fan and I. Gijbels* (1996)

67 Multivariate Dependencies—Models, analysis and interpretation
D.R. Cox and N. Wermuth (1996)

68 Statistical Inference—Based on the likelihood *A. Azzalini* (1996)

69 Bayes and Empirical Bayes Methods for Data Analysis
B.P. Carlin and T.A Louis (1996)

70 Hidden Markov and Other Models for Discrete-Valued Time Series
I.L. Macdonald and W. Zucchini (1997)

71 Statistical Evidence—A likelihood paradigm *R. Royall* (1997)

72 Analysis of Incomplete Multivariate Data *J.L. Schafer* (1997)

73 Multivariate Models and Dependence Concepts *H. Joe* (1997)

74 Theory of Sample Surveys *M.E. Thompson* (1997)

75 Retrial Queues *G. Falin and J.G.C. Templeton* (1997)

76 Theory of Dispersion Models *B. Jørgensen* (1997)

77 Mixed Poisson Processes *J. Grandell* (1997)

78 Variance Components Estimation—Mixed models, methodologies and applications
P.S.R.S. Rao (1997)

79 Bayesian Methods for Finite Population Sampling
G. Meeden and M. Ghosh (1997)

80 Stochastic Geometry—Likelihood and computation
O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout (1998)

81 Computer-Assisted Analysis of Mixtures and Applications—
Meta-analysis, disease mapping and others *D. Böhning* (1999)

82 Classification, 2nd edition *A.D. Gordon* (1999)

83 Semimartingales and their Statistical Inference *B.L.S. Prakasa Rao* (1999)

84 Statistical Aspects of BSE and vCJD—Models for Epidemics
C.A. Donnelly and N.M. Ferguson (1999)

85 Set-Indexed Martingales *G. Ivanoff and E. Merzbach* (2000)

86 The Theory of the Design of Experiments *D.R. Cox and N. Reid* (2000)

87 Complex Stochastic Systems
O.E. Barndorff-Nielsen, D.R. Cox and C. Klüppelberg (2001)

88 Multidimensional Scaling, 2nd edition *T.F. Cox and M.A.A. Cox* (2001)

89 Algebraic Statistics—Computational Commutative Algebra in Statistics
G. Pistone, E. Riccomagno and H.P. Wynn (2001)

90 Analysis of Time Series Structure—SSA and Related Techniques
N. Golyandina, V. Nekrutkin and A.A. Zhigljavsky (2001)

Analysis of Time Series Structure

SSA and Related Techniques

NINA GOLYANDINA
VLADIMIR NEKRUTKIN
ANATOLY ZHIGLJAVSKY

CHAPMAN & HALL/CRC

Boca Raton London New York Washington, D.C.

Library of Congress Cataloging-in-Publication Data

Golyandina, N. (Nina)
Analysis of time series structure : SSA and related techniques /
N. Golyandina, V. Nekrutkin, and A. Zhigljavsky.
p. cm.—(Monographs on statistics and applied probability)
Includes bibliographical references and index.
ISBN 1-58488-194-1 (alk. paper)
1. Time-series analysis. I. Nekrutkin, V.V. (Vladimir
Viktorovich) II. Zhigljavsky, A.A. (Anatoly Aleksandrovich) III.
Title. IV. Series.
QA280 .G65 2000
519.5'5—dc21 00-050442
CIP

International Standard Book Number 1-58488-194-1
Library of Congress Card Number 00-050442
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

Contents

Preface

This monograph is about a technique of time series analysis which is often called 'singular-spectrum analysis' (SSA). The basic SSA algorithm looks simple, but understanding of what it does and how it fits among the other time series analysis techniques is by no means simple. At least, it was difficult for us: we have spent a few years on this. This book is an account of what we have learned.

Spending so much time on just one technique should be somehow justified. For us, the justification is our belief in the capabilities of SSA: we are absolutely convinced that for a wide range of time series SSA can be extremely useful. More than that, we firmly believe that in the near future no statistical package will be sold without incorporating SSA facilities, and every time series analysis textbook will contain an SSA-related section.

Although not widely known among statisticians and econometrists, SSA has become a standard tool in meteorology and climatology; it is also a well-known technique in nonlinear physics and signal processing. We think that the lack of popularity of SSA among statisticians was mostly due to tradition and the lack of theory of SSA. We should also accept that the main methodological principle of SSA is not really statistical; SSA is more a technique of multivariate geometry than of statistics. In addition to statistics and multivariate geometry, the theory of SSA comprises the elements of signal processing, linear algebra, nonlinear dynamical systems, the theory of ordinary differential and finite-difference equations, and functional analysis. It is thus not surprising that it took a long time for us to achieve some level of understanding of what SSA is.

Despite the fact that the material of the book touches many different fields, a large part of the book is oriented towards a wide circle of readers who need or have an interest in time series analysis.

SSA is essentially a model-free technique; it is more an exploratory, model-building tool than a confirmatory procedure. It aims at a decomposition of the original series into a sum of a small number of interpretable components such as a slowly varying trend, oscillatory components and a 'structureless' noise. The main concept in studying the SSA properties is 'separability,' which characterizes how well different components can be separated from each other.

An important feature of SSA is that it can be used for analyzing relatively short series. On the other hand, asymptotic separation plays a very important role in the theory of SSA. There is no contradiction here because the asymptotic features (which hold as the length of the series N tends to infinity) are found to be met

for relatively small N. In practical applications, we typically deal with series of length varying from a few dozen to a few thousand.

Possible application areas of SSA are diverse: from mathematics and physics to economics and financial mathematics, from meteorology and oceanology to social science and market research. Any seemingly complex series with a potential structure could provide another example of a successful application of SSA.

There are a large number of examples in the book. Many of these examples are real-life series from different areas including medicine, physics, astronomy, economics, and finance. These examples are not the most exciting examples of application of SSA; they were not selected to impress the reader. The purpose of the selection was different: the examples serve only for illustrating the methodological and theoretical aspects discussed in the book. Also, each example illustrates a different feature of the method, so that the number of examples can hardly be reduced.

This book could not have been written had we not acquired a particular computer routine realizing SSA (see the Web site *http://vega.math.spbu.ru/caterpillar*). We were very lucky to have had in our team Kirill Braulov from St. Petersburg University who developed the software. We are very grateful to Kirill for his excellent work. We are also very grateful to our other collaborators and colleagues from the Faculty of Mathematics, St. Petersburg University, and especially to Sergei Ermakov, Vladislav Solntsev, Dmitrii Danilov and Alexander Bart, who have participated in a large number of seminars and discussions on the topic. These seminars and discussions were most useful, especially during the initial stage of the work. Also we are grateful to Dmitry Belov (Institute of Physiology, St. Petersburg University) for permission to use his EEG data for one of the examples in the book.

Our Cardiff University colleague, Gerald Gould, has carefully gone through the manuscript and improved the English where necessary; we are much obliged to him for a very important job. Comments from the Chapman & Hall editors have also helped very much in improving the manuscript; we are really thankful to them.

A part of this work has been done in accordance with the grant GR/M21713, "Multivariate methods in change-point detection problems" from the EPSRC. We are very grateful for this support. However, our main gratitude undoubtedly goes to the Procter & Gamble Company, which for many years has been extremely supportive of us. We have worked with a number of very bright and clever people from the company, but first of all we wish to acknowledge Phil Parker and Luigi Ciutti. Their interest in and support for our work have helped us tremendously.

Last but not least, we are very grateful to our families for their patience and understanding during the long period taken to write this book.

Nina Golyandina, Vladimir Nekrutkin, Anatoly Zhigljavsky

St. Petersburg – Cardiff, October 2000

Notation

SVD	singular value decomposition
LRF	linear recurrent formula
SSA	singular-spectrum analysis
c.d.f.	cumulative distribution function
F	time series
N	length of time series
$F_N = (f_0, \ldots, f_{N-1})$	time series of length N
$F_{i,j} = (f_{i-1}, \ldots, f_{j-1})$	subseries of a time series F_N
L	window length
$K = N - L + 1$	number of L-lagged vectors of F_N
X_i	ith L-lagged vector of time series
$\mathbf{X} = [X_1 : \ldots : X_K]$	trajectory matrix with columns X_i
$\mathbf{X}^{\mathrm{T}}$	transposed matrix $\mathbf{X}$
$\mathcal{M}_{L,K}$	linear space of $L \times K$ matrices
$\langle \mathbf{X}, \mathbf{Y} \rangle_{\mathcal{M}}$	inner product of matrices in $\mathcal{M}_{L,K}$
$\|\|\mathbf{X}\|\|_{\mathcal{M}}$	Frobenius matrix norm in $\mathcal{M}_{L,K}$
$\mathrm{rank}(\mathbf{X})$	rank of matrix $\mathbf{X}$
$\mathcal{H}$	Hankelization operator
λ_i	ith eigenvalue of the matrix $\mathbf{X}\mathbf{X}^{\mathrm{T}}$
$\mathbf{E}_{\mathrm{M}}$	identical $M \times M$ matrix
$\mathbf{0}_{\mathrm{LK}}$	zero $L \times K$ matrix
$\mathbf{0}_{\mathrm{M}}$	zero vector of dimension M
$\mathbf{1}_{\mathrm{M}}$	vector $(1, \ldots, 1)^{\mathrm{T}}$ of dimension M
$\mathbf{R}^M$	Euclidean space of dimension M
$\mathfrak{L}$	linear subspace of the Euclidean space
$\dim \mathfrak{L}$	dimension of a linear space $\mathfrak{L}$
$\mathfrak{L}_r$	linear space of dimension r
$\mathrm{span}(P_1, \ldots, P_n)$	linear space spanned by vectors $P_1, \ldots, P_n$
$\mathrm{span}(\mathbf{X})$	linear space spanned by the columns of $\mathbf{X}$
$\mathfrak{L}^{(L)} = \mathfrak{L}^{(L)}(F_N)$	L-trajectory space of a time series F_N
$\mathrm{dist}(X, \mathfrak{L})$	distance from a vector X to a linear space $\mathfrak{L}$
$\mathrm{fdim}(F_N)$	difference dimension of a time series F_N
$\mathrm{rank}_L(F_N)$	L-rank of a time series F_N
$\mathrm{rank}(F_N)$	rank of a time series F_N
U_i	ith eigenvector of the SVD of the matrix $\mathbf{X}$

V_i	ith factor vector of the SVD of the matrix $\mathbf{X}$
$\rho^{(L,M)}$	maximal cross-correlation of two series
$\rho_{12}^{(\omega)}$	weighted cross-correlation of two series
$\rho_{12}^{(\Pi)}$	spectral cross-correlation of two series
R_f	covariance function of a stationary series F
m_f	spectral measure of a stationary series F
p_f	spectral density of a stationary series F
Φ_f	spectral function of a stationary series F
Π_f^N	periodogram of a time series F_N
$g(F_1, F_2)$	heterogeneity index of time series F_1, F_2
$\mathbf{G} = \mathbf{G}_{B,T}$	heterogeneity matrix of a time series F
meas	Lebesque measure in $\mathbf{R}$

V_i	ith factor vector of the SVD of the matrix $\mathbf{X}$
$\rho^{(L,M)}$	maximal cross-correlation of two series
$\rho_{12}^{(\omega)}$	weighted cross-correlation of two series
$\rho_{12}^{(\Pi)}$	spectral cross-correlation of two series
R_f	covariance function of a stationary series F
m_f	spectral measure of a stationary series F
p_f	spectral density of a stationary series F
Φ_f	spectral function of a stationary series F
Π_f^N	periodogram of a time series F_N
$g(F_1, F_2)$	heterogeneity index of time series F_1, F_2
$\mathbf{G} = \mathbf{G}_{B,T}$	heterogeneity matrix of a time series F
meas	Lebesque measure in $\mathbf{R}$

V_i	ith factor vector of the SVD of the matrix $\mathbf{X}$
$\rho^{(L,M)}$	maximal cross-correlation of two series
$\rho_{12}^{(\omega)}$	weighted cross-correlation of two series
$\rho_{12}^{(\Pi)}$	spectral cross-correlation of two series
R_f	covariance function of a stationary series F
m_f	spectral measure of a stationary series F
p_f	spectral density of a stationary series F
Φ_f	spectral function of a stationary series F
Π_f^N	periodogram of a time series F_N
$g(F_1, F_2)$	heterogeneity index of time series F_1, F_2
$\mathbf{G} = \mathbf{G}_{B,T}$	heterogeneity matrix of a time series F
meas	Lebesque measure in $\mathbf{R}$

Introduction

SSA (singular-spectrum analysis) is a novel technique of time series analysis incorporating the elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems, and signal processing. Despite the fact that a lot of probabilistic and statistical elements are employed in the SSA-based methods (they relate to stationarity, ergodicity, principal component and bootstrap techniques), SSA is not a statistical method in terms of classical statistics. In particular, we typically do not make any statistical assumptions concerning either signal or noise while performing the analysis and investigating the properties of the algorithms.

The present book is fully devoted to the methodology and theory of SSA. The main topics are SSA analysis, SSA forecasting, and SSA detection of structural changes. Let us briefly consider these topics.

SSA analysis of time series

The birth of SSA is usually associated with publication of the papers by Broomhead and King (1986a, 1986b) and Broomhead *et al.* (1987). Since then, the technique has attracted a lot of attention. At present, the papers dealing with methodological aspects and applications of SSA number several hundred; see, for example, Vautard *et al.* (1992), Ghil and Taricco (1997), Allen and Smith (1996), Danilov and Zhigljavsky (1997), Yiou *et al.* (2000) and the references therein. An elementary introduction to the subject can be found in the recent book by Elsner and Tsonis (1996).

SSA has proved to be very successful, and has already become a standard tool in the analysis of climatic, meteorological and geophysical time series; see, for example, Vautard and Ghil (1989), Ghil and Vautard (1991), and Yiou *et al.* (1996). It is thus not surprising that among the main journals publishing SSA-related research papers are Journal of Climate, Journal of the Atmospheric Sciences, and Journal of Geophysical Research.

Let us turn to the description of SSA. The basic version of SSA consists of four steps, which are performed as follows. Let $F = (f_0, f_1, \ldots, f_{N-1})$ be a time series of length N, and L be an integer, which will be called the 'window length'. We set $K = N-L+1$ and define the L-lagged vectors $X_j = (f_{j-1}, \ldots, f_{j+L-2})^{\mathrm{T}}$, $j = 1, 2, \ldots, K$, and the trajectory matrix

$$\mathbf{X} = (f_{i+j-2})_{i,j=1}^{L,K} = [X_1 : \ldots : X_K].$$

Note that the trajectory matrix $\mathbf{X}$ is a Hankel matrix, which means that all the elements along the diagonal $i+j = \text{const}$ are equal. The construction of the trajectory matrix constitutes the first step of the algorithm.

The second step is the singular value decomposition (SVD) of the matrix $\mathbf{X}$, which can be obtained via eigenvalues and eigenvectors of the matrix $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$ of size $L \times L$. This provides us with a collection of L singular values, which are the square roots of the eigenvalues of the matrix $\mathbf{S}$, and the corresponding left and right singular vectors. (The left singular vectors of $\mathbf{X}$ are the orthonormal eigenvectors of $\mathbf{S}$; in SSA literature, they are often called 'empirical orthogonal functions' or simply EOFs. The right singular vectors can be regarded as the eigenvectors of the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$.) We thus obtain a representation of $\mathbf{X}$ as a sum of rank-one biorthogonal matrices $\mathbf{X}_i$ $(i = 1, \ldots, d)$, where d $(d \leq L)$ is the number of nonzero singular values of $\mathbf{X}$.

At the third step, we split the set of indices $I = \{1, \ldots, d\}$ into several groups $I_1, \ldots, I_m$ and sum the matrices $\mathbf{X}_i$ within each group. The result of the step is the representation

$$\mathbf{X} = \sum_{k=1}^{m} \mathbf{X}_{I_k}, \quad \text{where } \mathbf{X}_{I_k} = \sum_{i \in I_k} \mathbf{X}_i .$$

At the fourth step, averaging over the diagonals $i+j = \text{const}$ of the matrices $\mathbf{X}_{I_k}$ is performed. This gives us an SSA decomposition; that is, a decomposition of the original series F into a sum of series

$$f_n = \sum_{k=1}^{m} f_n^{(k)}, \quad n = 0, \ldots, N-1, \tag{I.1}$$

where for each k the series $f_n^{(k)}$ is the result of diagonal averaging of the matrix $\mathbf{X}_{I_k}$.

The basic scheme of SSA for analysis of time series and some modifications of this scheme are known in the SSA literature cited above. Note that SSA is usually regarded as a method of identifying and extracting oscillatory components from the original series; see, for example, Yiou *et al.* (1996), Ghil and Taricco (1997), Fowler and Kember (1998). The standard SSA literature, however, does not pay enough attention to theoretical aspects which are very important for understanding how to select the SSA parameters and, first of all, the window length L for the different classes of time series. The concept of separability and related methodological aspects and theoretical results provide us with this understanding. It is the study of separability which makes the biggest distinction between our research on SSA analysis and the standard approach to SSA.

The choice of parameters in performing the SSA decomposition (they are the window length L and the way of grouping the matrices $\mathbf{X}_i$) must depend on the properties of the original series and the purpose of the analysis.

The general purpose of the SSA analysis is the decomposition (I.1) with additive components $f_n^{(k)}$ that are 'independent' and 'identifiable' time series; this is

what we mean when we talk about analyzing the structure of time series by SSA. Sometimes, one can also be interested in particular tasks, such as 'extraction of signal from noise,' 'extraction of oscillatory components' and 'smoothing'.

For a properly made SSA decomposition, a component $f_n^{(k)}$ in (I.1) can be identified as a trend of the original series, an oscillatory series (for example, seasonality) or noise. An oscillatory series is a periodic or quasi-periodic series which can be either pure or amplitude-modulated. Noise is any aperiodic series. The trend of the series is, roughly speaking, a slowly varying additive component of the series with all the oscillations removed.

Note that no parametric model for the components in (I.1) is fixed and these components are produced by the series itself. Thus, when analyzing real-life series with the help of SSA one can hardly hope to obtain the components in the decomposition (I.1) as exact harmonics or linear trend, for example, even if these harmonics or linear trend are indeed present in the series (by a harmonic we mean any sine series with some amplitude, frequency and phase). This is an influence of noise and a consequence of the non-parametric nature of the method. In many cases, however, we can get a good approximation to these series.

In the ideal situation the components in (I.1) must be 'independent'. Achieving 'independence' (or 'separability') of the components in the SSA decomposition (I.1) is of prime importance in SSA. From the authors' viewpoint, separability of components in this decomposition is the main theoretical problem in SSA research and the main target in the selection of SSA parameters. Separability of components is the central problem in the book; it is touched upon in virtually every section.

There are different notions of separability (more precisely, L-separability, since the fact of separability depends on the window length L). The most important is weak separability, defined as follows. Provided that the original time series f_n is a sum of m series $f_n^{(k)}$ $(k = 1, \ldots, m)$, for a fixed window length L, weak L-separability means that any subseries of length L of the kth series $f_n^{(k)}$ is orthogonal to any subseries of length L of the lth series $f_n^{(l)}$ with $l \neq k$, and the same holds for their subseries of length $K = N - L + 1$. This is equivalent to the fact that there is a way of constructing the SVD of the trajectory matrix $\mathbf{X}$ and grouping the matrices $\mathbf{X}_j$ so that for each k the matrix $\mathbf{X}_{I_k}$ is the trajectory matrix of the series $f_n^{(k)}$.

The demand of exact separability of components is a strict requirement which rarely holds in practice. The notion of approximate separability is more important (and much less restrictive) than the exact one. For a relatively long series, approximate separability of the components is often achieved due to the theoretical concept of asymptotic separability which holds for a rather wide class of components.

To measure the degree of 'separability' of the components in (I.1) we use a number of different characteristics, such as 'spectral correlation coefficient' or 'weighted correlation coefficient'.

Weak separability may not be sufficient to guarantee that a particular SSA decomposition properly reflects the structure of the original time series. Indeed, in the case when two or more of the singular values of the trajectory matrices $\mathbf{X}^{(k)}$ and $\mathbf{X}^{(l)}$ corresponding to two different components $f_n^{(k)}$ and $f_n^{(l)}$ of the original series are equal (in practice, if the singular values are close), then the SVD is not uniquely defined and the two series $f_n^{(k)}$ and $f_n^{(l)}$ are mixed up, so that an additional analysis (such as rotations in the L-dimensional space of the lagged vectors) is required to separate the two series. If there is (approximate) weak separability and all eigenvalues corresponding to different components in (I.1) are sufficiently isolated from each other, then we have (approximate) strong separability, which means that for a proper grouping the SSA decomposition (approximately) coincides with the one assumed.

The absence of approximate strong separability is often observed for series with complex structure. For these series and series of special structure, there are different ways of modifying SSA. Several modifications of the basic SSA technique can be of interest, such as SSA with single and double centring, Toeplitz SSA, and sequential SSA (when the basic scheme is applied several times with different parameters to the residuals from the previous analysis). SSA with centring and Toeplitz SSA are based on particular non-optimal decompositions of the trajectory matrices; they may be useful in analysis of time series of special structure, such as series with linear-like tendencies and stationary-like series.

Toeplitz SSA was suggested in Vautard and Ghill (1989); it is a well known modification of the basic SSA method. By contrast, SSA with double centring of the trajectory matrix is a new version of SSA.

SSA forecasting of time series

The principles of SSA forecasting developed in this book are new with respect to the main-stream SSA approach. Let us now briefly consider the methodological aspects of SSA forecasting.

An important property of the SSA decomposition is the fact that, if the original series f_n satisfies a linear recurrent formula (LRF)

$$f_n = a_1 f_{n-1} + \ldots + a_d f_{n-d} \tag{I.2}$$

of some dimension d with some coefficients $a_1, \ldots, a_d$, then for any N and L there are at most d nonzero singular values in the SVD of the trajectory matrix $\mathbf{X}$; therefore, even if the window length L and $K = N - L + 1$ are larger than d, we only need at most d matrices $\mathbf{X}_i$ to reconstruct the series.

The fact that the series f_n satisfies an LRF (I.2) is equivalent to its representability as a sum of products of exponentials, polynomials and harmonics, that is as

$$f_n = \sum_{k=1}^{q} \alpha_k(n) e^{\mu_k n} \sin(2\pi\omega_k n + \varphi_k)\,. \tag{I.3}$$

Here $a_k(n)$ are polynomials, μ_k, ω_k and φ_k are arbitrary parameters. The number of linearly independent terms q in (I.3) is smaller than or equal to d.

SSA forecasting is based on a fact which, roughly speaking, states the following: if the number of terms r in the SVD of the trajectory matrix $\mathbf{X}$ is smaller than the window length L, then the series satisfies some LRF of some dimension $d \leq r$. Certainly, this assertion must not be understood *ad litteram*. However, for infinite series a similar fact can be found in Gantmacher (1998, Chapter XVI, Section 10, Theorem 7). The theorem due to Buchstaber (1994) amplifies these considerations for finite time series; this theorem says that under the above-mentioned conditions the series (with the possible exception of the last few terms) satisfies some LRF. This assertion, however, does not directly lead to a forecasting algorithm, since the last terms of the series are very important for forecasting.

An essential result for SSA forecasting was obtained in Danilov (1997a, 1997b). It can be formulated as follows: if the dimension r of the linear space $\mathfrak{L}_r$ spanned by the columns of the trajectory matrix is less than the window length L and this space is not a vertical space, then the series satisfies a natural LRF of dimension $L-1$. (If $e_L \notin \mathfrak{L}_r$, where $e_L = (0, 0, \ldots, 0, 1)^{\mathrm{T}} \in \mathbb{R}^L$, then we say that $\mathfrak{L}_r$ is a 'non-vertical' space.)

If we have a series satisfying an LRF (I.2), then we can obviously continue it for an arbitrary number of steps using the same LRF. It is important that any LRF governing a given series provides the same continuation, and thus we do not necessarily need the LRF with the minimal value of d. Thus, we now know how to continue time series with non-vertical spaces and small ranks of trajectory matrices.

Of course, when we are dealing with real-life time series we can hardly hope to have a time series that is governed by an LRF of small dimension (in terms of SVD, a 'real-life' trajectory matrix with $L \leq K$ has, as a rule, rank L). However, the class of series that can be approximated by the series governed by the LRFs of the form (I.2) or, equivalently, by the (deterministic) time series of the form (I.3) with a small number of terms, is very broad and we can attempt forecasting of these series using an SSA-based forecasting method. We may also be interested in continuing (forecasting) some periodic (perhaps, amplitude-modulated) components of the original series and in forecasting the trend, ignoring noise and all oscillatory components of the series.

The idea of SSA forecasting of a certain time series component is as follows. The selection of a group of $r < \operatorname{rank} \mathbf{X}$ rank-one matrices $\mathbf{X}_i$ on the third step of the basic SSA algorithm implies the selection of an r-dimensional space $\mathfrak{L}_r \subset \mathbb{R}^L$ spanned by the corresponding left singular vectors.

If the space $\mathfrak{L}_r$ is non-vertical, then, as was mentioned previously, this space produces the appropriate LRF, which can be used for forecasting (called recurrent forecasting) of the series component, corresponding to the chosen rank-one matrices.

As in the basic SSA, the separability characteristics help in selection of both the window length L and the space $\mathfrak{L}_r$. Moreover, separability is directly related

to LRFs: roughly speaking, if two series are separable, then they satisfy certain LRFs.

The SSA recurrent forecasting algorithm can be modified in several ways. For example, we can base our forecast on the Toeplitz SSA or SSA with centring rather than on the basic SSA (the $\mathfrak{L}_r$ is then spanned by the corresponding versions of left singular vectors); in some cases, we can also base the forecast on the LRF of minimal order. Perhaps the most important modification is the so-called SSA vector forecasting algorithm developed in Nekrutkin (1999). The idea of this method is as follows.

For any group of indices I selected at the grouping stage, the application of SSA gives us $K = N - L + 1$ vectors $\widehat{X}_1, \ldots, \widehat{X}_K$ that lie in an r-dimensional subspace $\mathfrak{L}_r$ of $\mathbf{R}^L$. Here r is the number of elements in I, for each j the $\widehat{X}_j$ is the projection of the L-lagged vector X_j onto the subspace $\mathfrak{L}_r$, and the subspace $\mathfrak{L}_r$ is spanned by the r left eigenvectors of the trajectory matrix $\mathbf{X}$ with the indices in the group I. We then continue the vectors $\widehat{X}_1, \ldots, \widehat{X}_K$ for M steps in such a way that (i) the continuation vectors Z_m $(K < m \leq K + M)$ belong to the space $\mathfrak{L}_r$ and (ii) the matrix $[\widehat{X}_1 : \ldots : \widehat{X}_K : Z_{K+1} : \ldots : Z_{K+M}]$ is approximately a Hankel matrix. The forecasting series is then obtained by means of diagonal averaging of this matrix.

While the recurrent forecasting algorithm performs the straightforward recurrent continuation of a one-dimensional series (with the help of the LRF so constructed), the vector forecasting method makes the continuation of the vectors in an r-dimensional space and only then returns to the time-series representation. Examples show that vector forecasting appears to be more stable than the recurrent one, especially for long-term forecasting.

Confidence intervals for the forecasts can be very useful in assessing the quality of the forecasts. However, unlike the SSA forecasts themselves (their construction does not formally require any preliminary information about the time series), for constructing confidence bounds we need some assumptions to be imposed on the series and the residual component, which we associate with noise.

We consider two types of confidence bounds; the first one is for the values of the series itself at some future point $N + M$, and the second one is for the values of the signal at this future point (under the assumption that the original series consists of a signal and additive noise). These two types of confidence intervals are constructed in different ways: in the first case, we use the information about forecast errors obtained during the analysis of the series; the second one uses the bootstrap technology.

To build the confidence intervals for the forecast of the entire initial series, we construct the forecasting LRF of dimension $L - 1$ (in the case of the recurrent forecast) and repeatedly apply it to all subseries of the same dimension within the observation period $[0, N-1]$. Then we compare the results with the corresponding values of the series. Under the assumption that the residual series is stationary and

ergodic, we can estimate the quantiles of the related marginal distribution, and therefore build the confidence bounds.

The bootstrap technique is useful for constructing confidence intervals for the signal $F^{(1)}$ at some future time $N + M$ under the assumption that the series $F_N = (f_0, \ldots, f_{N-1})$ is a sum of a signal $F_N^{(1)}$ and noise $F_N^{(2)} = F_N - F_N^{(1)}$. To do that, we first obtain the SSA decomposition $F_N = \widetilde{F}_N^{(1)} + \widetilde{F}_N^{(2)}$, where $\widetilde{F}_N^{(1)}$ (the reconstructed series) approximates $F_N^{(1)}$, and $\widetilde{F}_N^{(2)}$ is the residual series. Assuming that we have a (stochastic) model for the residuals $\widetilde{F}_N^{(2)}$, we then simulate some number S of independent copies $\widetilde{F}_{N,i}^{(2)}$ of the series $F_N^{(2)}$, obtain S series $\widetilde{F}_N^{(1)} + \widetilde{F}_{N,i}^{(2)}$ and get S forecasting results $\widetilde{f}_{N+M-1,i}^{(1)}$. Having obtained the sample $\widetilde{f}_{N+M-1,i}^{(1)}$ $(1 \le i \le S)$ of the forecasting results, we use it to calculate the empirical lower and upper quantiles of fixed level γ and construct the corresponding confidence interval for the forecast.

Note that the bootstrap confidence bounds can be constructed not only for the SSA forecasts but also for the terms of the SSA decomposition when we are dealing with separation of a signal from noise.

SSA detection of structural changes in time series

We call a time series F_N homogeneous if it is governed by an LRF of order d that is small relative to the length of the series N.

Assume now that the series is homogeneous until some time $Q < N$, but then it stops following the original LRF (this may be caused by a perturbation of the series). However, after a certain time period, it again becomes governed by an LRF. In this case, we have a structural change (heterogeneity) in the series. We may have either a permanent heterogeneity (in this case the new LRF is different from the original one) or a temporary heterogeneity, when both LRFs coincide. Note that even in the latter case, the behaviour of the series after the change is different from the behaviour of the homogeneous (unperturbed) series; for example, the initial conditions for the LRF after the perturbation can be different from the unperturbed initial conditions.

The main idea of employing SSA for detecting different types of heterogeneity is as follows. The results of Section 5.2 imply that for sufficiently large values of the window length L the L-lagged vectors of a homogeneous series span the same linear space $\mathfrak{L}^{(L)}$ independently of N, as soon as N is sufficiently large. Therefore, violations in homogeneity of the series can be described in terms of the corresponding lagged vectors: the perturbations force the lagged vectors to leave the space $\mathfrak{L}^{(L)}$. The corresponding discrepancies are defined in terms of the distances between the lagged vectors and the space $\mathfrak{L}^{(L)}$, which can be determined for different subseries of the original series.

Since, in practice, the series are described by LRFs only approximately, the problem of approximate construction of the spaces $\mathfrak{L}^{(L)}$ arises again. Analogous to the problems of forecasting, the SVD of the trajectory matrices is used for

this purpose. As everywhere in the book, the concept of separability plays a very important role when we are interested in detecting changes in components of the series (for example, in the signal, under the presence of additive noise). Unlike the forecasting problems, for studying structural changes in time series, the properties of the SVDs of subseries of the initial series F become of prime importance.

We consider two subseries (say F' and F'') of the series F; we call them 'base subseries' and 'test subseries'. Assume that the lengths of these subseries are fixed and equal to B and T, respectively. Suppose that $B > L$ and $T \geq L$, where L is the window length. Let us make an SVD of the trajectory matrix of the base subseries, select a group of $r < L$ left singular vectors, consider the linear space $\mathfrak{L}'_r$ spanned by these vectors and compute the sum of the squared distances between the space $\mathfrak{L}'_r$ and the L-lagged vectors corresponding to the test subseries. If we normalize this sum by the sum of the squared norms of the L-lagged vectors of the test subseries, then we obtain the so-called heterogeneity index $g = g(F', F'')$ formally defined in Section 3.1. The heterogeneity index $g(F', F'')$ measures the discrepancy between F' and F'' by computing the relative error of the optimal approximation of the L-lagged vectors of the time series F'' by vectors from the space $\mathfrak{L}'_r$.

The main tool used to study structural changes (heterogeneities) in time series is the 'heterogeneity matrix' of size $(N-B+1) \times (N-T+1)$. The entries of this matrix are the values of the heterogeneity index $g = g(F', F'')$, where F' and F'' run over all possible subseries of the series F of fixed lengths B and T, respectively.

The columns, rows and some diagonals of the heterogeneity matrix constitute the 'heterogeneity functions'. Change in the indexation system gives us the 'detection functions'; they are more convenient for the purpose of change detection.

We also consider three groups of supplementary detection characteristics. The first group is obtained when we use a different normalization in the expression for the heterogeneity index (rather than using the sum of the squared norms of the L-lagged vectors of the test subseries, we use the sum of the squared terms of the whole series). This renormalization of the heterogeneity index often helps when we monitor changes in monotone series and their components.

The second group of characteristics relates to the series of the roots of the characteristic polynomials of the LRFs that correspond to the SSA decomposition of the base subseries F'. The roots of the characteristic polynomials monitor the dynamics of the linear spaces $\mathfrak{L}'_r$. In particular, this monitoring can be very useful for distinguishing the changes that actually happen in the series from spurious changes that are caused by the fact that abrupt changes in the dynamics of the linear spaces $\mathfrak{L}'_r$ may be related to the changes in the order of the singular values.

The third group of characteristics is basically the moving periodograms of the original series; this group is used to monitor the spectral structure of the original series.

Composition of the book

The book has two parts; they are devoted to the methodology and theory of SSA, respectively. The methodological principles of SSA are thoroughly considered in Part I of the book. This part consists of three chapters, which deal with SSA analysis, SSA forecasting and SSA detection of structural changes, respectively.

SSA analysis of time series is dealt with in Chapter 1. In Section 1.1, the basic algorithm is described. In Section 1.2, the steps of this algorithm are explained and commented on. In Section 1.3, the main capabilities of the basic algorithm are illustrated by a number of real-life examples. In Section 1.4, the major tasks that can be attempted by SSA are formulated and discussed. In Section 1.5, the concept of separability is considered in detail. These considerations play a very important role in the selection of the parameters of SSA, the problem which is dealt with in Section 1.6. In Section 1.7, supplementary SSA techniques, such as SSA with centring and Toeplitz SSA, are considered.

Chapter 2 is devoted to SSA forecasting methodology. In Section 2.1, we formally describe the SSA recurrent forecasting algorithm. In Section 2.2, the principles of SSA forecasting and links with LRFs are discussed. Several modifications of the basic SSA recurrent forecasting algorithm are formulated and discussed in Section 2.3. The construction of confidence intervals for the forecasts is made in Section 2.4. In Section 2.5, we summarize the material of the chapter, and in Section 2.6 we provide several examples illustrating different aspects of SSA forecasting.

The methodology of SSA detection of structural changes in time series is considered in Chapter 3. In Section 3.1, we introduce and discuss the main concepts. In Section 3.2, we consider various violations of homogeneity in time series and the resulting shapes of the heterogeneity matrices and detection functions. In Section 3.3, we generalize the results of Section 3.2 to the case when we are detecting heterogeneities in one of the components of the original series rather than in the series itself (this includes the case when the series of interest is observed with noise). The problem of the choice of detection parameters is dealt with in Section 3.4. In Section 3.5, we consider several additional detection characteristics, and in Section 3.6 we provide a number of examples.

Chapters 4, 5 and 6 constitute the second (theoretical) part of the book, where all the statements of Part I are properly formulated and proved (with the exception of some well-known results where the appropriate references are given).

Chapter 4 considers the singular value decomposition (SVD) of real matrices, which is the main mathematical tool in the SSA method. The existence and uniqueness of SVDs is dealt with in Section 4.1. In Section 4.2, we discuss the structure and properties of the SVD matrices with special attention paid to such features of SVD as orthogonality, biorthogonality, and minimality. In Section 4.3, we consider optimal features of the SVD from the viewpoints of multivariate geometry and approximation of matrices by matrices of lower rank. A number of

results on optimality of the standard SVD are generalized in Section 4.4 to the SVD with single and double centring.

Chapter 5 provides a formal mathematical treatment of time series of finite rank; the L-trajectory matrices of these series have rank less than $\min(L, K)$ for all sufficiently large L and K. General properties of such series are considered in Section 5.1. As discussed above, the series of finite rank are related to the series governed by the LRFs; these relations are studied in Section 5.2. The results concerning the continuation procedures are derived in Section 5.3.

In Chapter 6, we make a formal mathematical study of four topics that are highly important for the SSA methodology. Specifically, in Section 6.1 we study weak separability of time series, in Section 6.2 diagonal averaging (Hankelization) of matrices is considered, while centring in SSA is studied in Section 6.3, and specific features of SSA for deterministic stationary sequences are discussed in Section 6.4.

Other SSA and SSA-related topics

On the whole, this book considers many important issues relating to the implementation, analysis and practical application of SSA. There are, however, several other topics which are not covered here. Let us mention some of them.

1. *Multichannel SSA.* Multichannel SSA is an extension of the standard SSA to the case of multivariate time series (see Broomhead and King, 1986b). It can be described as follows. Assume that we have an l-variate time series $f_n = (f_n^{(1)}, \ldots, f_n^{(l)})$, where $n = 0, 1, \ldots, N-1$ (for simplicity we assume that the time domain is the same for all the components of the series). Then for a fixed window length L we can define the trajectory matrices $\mathbf{X}^{(i)}$ $(i = 1, \ldots, l)$ of the one-dimensional time series $f_n^{(i)}$. The trajectory matrix $\mathbf{X}$ can then be defined as

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \cdots \\ \mathbf{X}^{(l)} \end{pmatrix}. \tag{I.4}$$

The other stages of the multichannel SSA procedure are identical to the one-dimensional procedure discussed above with obvious modification that the diagonal averaging should be applied to each of the l components separately. (Multichannel SSA can be generalized even further, for analyzing discrete time random fields and image processing problems; see Danilov and Zhigljavsky, 1997.)

There are numerous examples of successful application of the multichannel SSA (see, for example, Plaut and Vautard, 1994; Danilov and Zhigljavsky, 1997), but the theory of multichannel SSA is yet to be developed. The absence of a theory is the reason why, in the present book, we have confined ourselves to the univariate case only. This case is already difficult enough, and multichannel SSA has additional peculiarities.

Construction of the trajectory matrix in multichannel SSA is not obvious; there are several alternatives to (I.4). The matrix (I.4) seems to be the natural candidate

for the trajectory matrix of a multivariate series, but its advantages are not clear. Note also that there is a version of SSA that deals with complex-valued series; it can be considered as a version of multichannel SSA as well. It is, however, not clear how to compare the two-channel SSA with the one-channel complex SSA.

2. *Continuous time SSA.* The basic SSA scheme and most of its variations can be modified for the case of continuous time. There are many significant changes (with respect to the material of the book) that would to be made if one were to try to analyze the corresponding procedure: instead of sums we get integrals, instead of matrices we have linear operators, the SVD becomes the Schmidt decomposition in the corresponding Hilbert space, LRFs become ordinary differential equations, and so on. Note that the theory of generalized continuous time SSA includes the standard discrete time SSA as a particular case. In addition, such a generalization allows us to consider not only embeddings of Hankel type but also many other mappings which transfer functions of one variable to the functions of two variables. Those interested in this approach can find a lot of related material in Nekrutkin (1997).

3. *Use of different window lengths.* The use of different values of the window length is discussed in Section 1.7 in relation to the so-called 'Sequential SSA'. There are some other suggestions in the literature, such as selecting the window length at random (see Varadi *et al.*, 1999) or keeping the ratio L'/N' fixed, where L' is the window length for the subseries of the original series of length $N' = N/k$ which is obtained by sieving the original series (see Yiou *et al.*, 2000). Both methods are suggested for analyzing long series; the latter one is shown to have some similarity with the wavelet analysis of time series.

4. *SSA for sequential detection of structural changes.* The methodology of Chapter 3 aims at a nonsequential (posterior) detection of structural changes in time series. Some of these algorithms can be modified for the more standard change-point problem of sequential detection of change-points. This approach is implemented in Moskvina and Zhigljavsky (2000), where some of the detection algorithms are analyzed as proper statistical procedures. The Web site *http://www.cf.ac.uk/maths/stats/changepoint/* contains more information on the subject and a link to the software that can be downloaded.

Let us mention some other areas related to SSA.

During the last forty years, a variety of techniques of time series analysis and signal processing have been suggested that use SVDs of certain matrices; for surveys see, for example, Marple (1987) or Bouvet and Clergeot (1988). Most of these techniques are based on the assumption that the original series is random and stationary; they include some techniques that are famous in signal processing, such as Karhunen-Loève decomposition and the MUSIC algorithm (for the signal processing references, see, for example, Madisetti and Williams, 1998). Some statistical aspects of the SVD-based methodology for stationary series are considered, for example, in Brillinger (1975, Chapter 9), Subba Rao (1976) and Subba Rao and Gabr (1984).

The analysis of periodograms is an important part of the process of identifying the components in the SSA decomposition (I.1). For example, noise is modeled by aperiodic (chaotic) series whose spectral measures do not have atoms (white noise has constant spectral density). A comparison of the observed spectrum of the residual component in the SSA decomposition with the spectrum of some common time series (these can be found, for example, in Priestley, 1991 and Wei, 1990, Chapter 11) can help in understanding the nature of the residuals and formulation of a proper statistical hypothesis concerning the noise. However, a single realization of a noise series can have a spectrum that significantly differs from the theoretical one. Several simulation-based tests for testing the white noise zero hypothesis against the 'red noise' alternative (i.e., an autoregressive process of the first order) have been devised; the approach is called 'Monte Carlo SSA', see Allen and Smith (1996). This approach has attracted a lot of attention of researchers; for its extension and enhancement see, for example, Paluš and Novotna (1998).

Another area which SSA is related to is nonlinear (deterministic) time series analysis. It is a fashionable area of rapidly growing popularity; see the recent books by Cutler and Kaplan (1997), Kantz and Schreiber (1997), Abarbanel (1996), Tong (1993), and Weigend and Gershenfeld (1993). Note that the specialists in nonlinear time series analysis (as well as statisticians) do not always consider SSA as a technique that could compete with more standard methods; see, for example, Kantz and Schreiber (1997, Section 9.3.2).

It is impossible to discuss all the fields related to SSA. In a certain wide sense, one can consider SSA as a method of approximating the original series (or its component) with the other series governed by an LRF. Then we can consider a long list of publications on the theme, starting with Prony (1795).

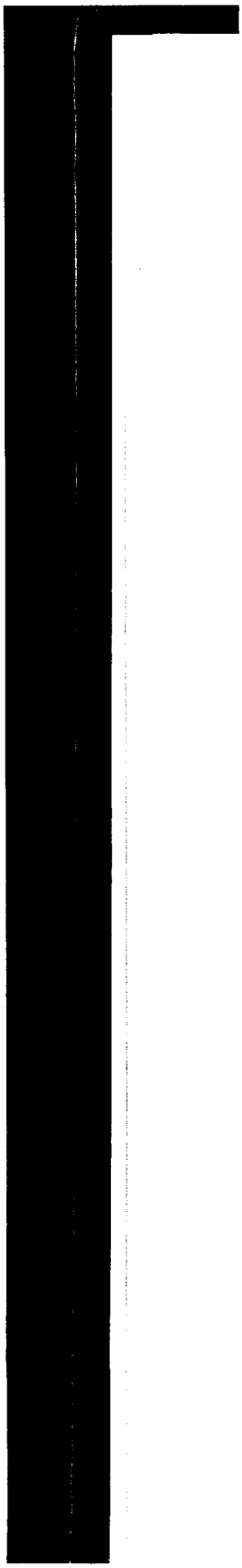
On the other hand, the essential feature of SSA is the choice of the optimal basis consisting of the left singular vectors. If we do not restrict ourselves to strong optimality (see the discussion on Toeplitz and centring SSA), then we arrive at a wide class of methods dealing with different bases (including, for example, the wavelet bases) that can be used for the decomposition of the lagged vectors.

As has already been mentioned, in signal processing, nonlinear physics and some other fields, a number of methods are in use that are based on SVDs of the trajectory matrices (as well as other matrices calculated through the terms of time series); these methods are used for different purposes.

Thus, the area of SSA-related methods is very wide. This is one of the reasons why we are confident that the ideas and methodology of SSA described in this book will be useful for a wide circle of scientists in different fields for many years to come.

PART I

SSA: Methodology



CHAPTER 1

Basic SSA

This chapter deals with the basic scheme of SSA and several modifications of it. Only the problem of analysis of the structure of a one-dimensional real-valued time series is considered. Some refined generalizations of the basic scheme adapted to the problems of time series forecasting and homogeneity analysis (including the change-point detection problem) are considered in the subsequent chapters.

Briefly, in this chapter we consider Basic SSA as a model-free tool for time series structure recognition and identification. We do not want to specify the notion 'structure' at the moment but mention that the goal of Basic SSA is a decomposition of the series of interest into several additive components that typically can be interpreted as 'trend' components (that is, smooth and slowly varying parts of the series), various 'oscillatory' components (perhaps with varying amplitudes), and 'noise' components.

In this chapter we do not assign any stochastic meaning to the term 'noise': the concept of a deterministic stationary 'noise' series is generally more convenient for SSA since it deals with a single trajectory of a time series rather than with a sample of such trajectories. Also, it may occur that we are not interested in certain components of the series and can therefore subsume them under the noise components.

Basic SSA performs four steps. At the first step (called the *embedding step*), the one-dimensional series is represented as a multidimensional series whose dimension is called the *window length*. The multidimensional time series (which is a sequence of vectors) forms the *trajectory matrix*. The sole (and very important) parameter of this step is the window length.

The second step, *SVD step*, is the singular value decomposition of the trajectory matrix into a sum of rank-one bi-orthogonal matrices. The first two steps together are considered as the *decomposition stage* of Basic SSA.

The next two steps form the *reconstruction stage*. The *grouping step* corresponds to splitting the matrices, computed at the SVD step, into several groups and summing the matrices within each group. The result of the step is a representation of the trajectory matrix as a sum of several *resultant matrices*.

The last step transfers each resultant matrix into a time series, which is an additive component of the initial series. The corresponding operation is called *diagonal averaging*. It is a linear operation and maps the trajectory matrix of the initial series into the initial series itself. In this way we obtain a decomposition of the initial series into several additive components.

Let us describe these steps formally and discuss their meaning and features.

1.1 Basic SSA: description

Let $N > 2$. Consider a real-valued time series $F = (f_0, \ldots, f_{N-1})$ of length N. Assume that F is a nonzero series; that is, there exists at least one i such that $f_i \neq 0$. Though one can usually assume that $f_i = f(i\Delta)$ for a certain function of time $f(t)$ and a certain time interval Δ, this does not play any specific role in our considerations.

Moreover, the numbers $0, \ldots, N-1$ can be interpreted not only as discrete time moments but also as labels of any other linearly ordered structure. The numbering of the time series values starts at $i = 0$ rather than at the more standard $i = 1$; this is only for convenience of notation.

As was already mentioned, Basic SSA consists of two complementary stages: decomposition and reconstruction.

1.1.1 First stage: decomposition

1st step: Embedding

The *embedding* procedure maps the original time series to a sequence of multidimensional lagged vectors.

Let L be an integer (*window length*), $1 < L < N$. The embedding procedure forms $K = N - L + 1$ *lagged vectors*

$$X_i = (f_{i-1}, \ldots, f_{i+L-2})^{\mathrm{T}}, \quad 1 \leq i \leq K,$$

which have dimension L. If we need to emphasize the dimension of the X_i, then we shall call them *L-lagged vectors*.

The *L-trajectory matrix* (or simply *trajectory matrix*) of the series F:

$$\mathbf{X} = [X_1 : \ldots : X_K]$$

has lagged vectors as its columns. In other words, the trajectory matrix is

$$\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} f_0 & f_1 & f_2 & \cdots & f_{K-1} \\ f_1 & f_2 & f_3 & \cdots & f_K \\ f_2 & f_3 & f_4 & \cdots & f_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & f_{L+1} & \cdots & f_{N-1} \end{pmatrix}. \tag{1.1}$$

Obviously $x_{ij} = f_{i+j-2}$ and the matrix $\mathbf{X}$ has equal elements on the 'diagonals' $i + j = \text{const}$. (Thus, the trajectory matrix is a *Hankel matrix*.) Certainly if N and L are fixed, then there is a one-to-one correspondence between the trajectory matrices and the time series.

2nd step: Singular value decomposition

The result of this step is the singular value decomposition (SVD) of the trajectory matrix. Let $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$. Denote by $\lambda_1, \ldots, \lambda_L$ the *eigenvalues* of $\mathbf{S}$ taken in the decreasing order of magnitude ($\lambda_1 \geq \ldots \geq \lambda_L \geq 0$) and by $U_1, \ldots, U_L$ the orthonormal system of the *eigenvectors* of the matrix $\mathbf{S}$ corresponding to these eigenvalues. Let $d = \max\{i, \text{ such that } \lambda_i > 0\}$.

If we denote $V_i = \mathbf{X}^{\mathrm{T}} U_i / \sqrt{\lambda_i}$ $(i = 1, \ldots, d)$, then the SVD of the trajectory matrix $\mathbf{X}$ can be written as

$$\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_d, \tag{1.2}$$

where $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^{\mathrm{T}}$. The matrices $\mathbf{X}_i$ have rank 1; therefore they are *elementary matrices*. The collection $(\sqrt{\lambda_i}, U_i, V_i)$ will be called ith *eigentriple* of the SVD (1.2).

1.1.2 Second stage: reconstruction

3rd step. Grouping

Once the expansion (1.2) has been obtained, the grouping procedure partitions the set of indices $\{1, \ldots, d\}$ into m disjoint subsets $I_1, \ldots, I_m$.

Let $I = \{i_1, \ldots, i_p\}$. Then the *resultant matrix* $\mathbf{X}_I$ corresponding to the group I is defined as $\mathbf{X}_I = \mathbf{X}_{i_1} + \ldots + \mathbf{X}_{i_p}$. These matrices are computed for $I = I_1, \ldots, I_m$ and the expansion (1.2) leads to the decomposition

$$\mathbf{X} = \mathbf{X}_{I_1} + \ldots + \mathbf{X}_{I_m}. \tag{1.3}$$

The procedure of choosing the sets $I_1, \ldots, I_m$ is called the *eigentriple grouping*.

4th step: Diagonal averaging

The last step in Basic SSA transforms each matrix of the grouped decomposition (1.3) into a new series of length N.

Let $\mathbf{Y}$ be an $L \times K$ matrix with elements y_{ij}, $1 \leq i \leq L$, $1 \leq j \leq K$. We set $L^* = \min(L, K)$, $K^* = \max(L, K)$ and $N = L + K - 1$. Let $y^*_{ij} = y_{ij}$ if $L < K$ and $y^*_{ij} = y_{ji}$ otherwise.

Diagonal averaging transfers the matrix $\mathbf{Y}$ to the series $g_0, \ldots, g_{N-1}$ by the formula:

$$g_k = \begin{cases} \dfrac{1}{k+1} \displaystyle\sum_{m=1}^{k+1} y^*_{m,k-m+2} & \text{for } 0 \leq k < L^* - 1, \\ \dfrac{1}{L^*} \displaystyle\sum_{m=1}^{L^*} y^*_{m,k-m+2} & \text{for } L^* - 1 \leq k < K^*, \\ \dfrac{1}{N-k} \displaystyle\sum_{m=k-K^*+2}^{N-K^*+1} y^*_{m,k-m+2} & \text{for } K^* \leq k < N. \end{cases} \tag{1.4}$$

The expression (1.4) corresponds to averaging of the matrix elements over the 'diagonals' $i + j = k + 2$: the choice $k = 0$ gives $g_0 = y_{11}$, for $k = 1$ we have

$g_1 = (y_{12} + y_{21})/2$, and so on. Note that if the matrix $\mathbf{Y}$ is the trajectory matrix of some series $(h_0, \ldots, h_{N-1})$ (in other words, if $\mathbf{Y}$ is the Hankel matrix), then $g_i = h_i$ for all i.

Diagonal averaging (1.4) applied to a resultant matrix $\mathbf{X}_{I_k}$ produces the series $\widetilde{F}^{(k)} = (\widetilde{f}_0^{(k)}, \ldots, \widetilde{f}_{N-1}^{(k)})$ and therefore the initial series $f_0, \ldots, f_{N-1}$ is decomposed into the sum of m series:

$$f_n = \sum_{k=1}^{m} \widetilde{f}_n^{(k)}. \tag{1.5}$$

1.2 Steps in Basic SSA: comments

The formal description of the steps in Basic SSA requires some elucidation. In this section we briefly discuss the meaning of the procedures involved.

1.2.1 Embedding

Embedding can be regarded as a mapping that transfers a one-dimensional time series $F = (f_0, \ldots, f_{N-1})$ to the multidimensional series $X_1, \ldots, X_K$ with vectors $X_i = (f_{i-1}, \ldots, f_{i+L-2})^{\mathrm{T}} \in \mathbf{R}^L$, where $K = N - L + 1$. Vectors X_i are called *L-lagged vectors* (or, simply, *lagged vectors*).

The single parameter of the embedding is the *window length* L, an integer such that $2 \le L \le N - 1$.

Embedding is a standard procedure in time series analysis. With the embedding being performed, further development depends on the purpose of the investigation.

For specialists in dynamical systems, a common technique is to obtain the empirical distribution of all the pairwise distances between the lagged vectors X_i and X_j and then calculate the so-called correlation dimension of the series. This dimension is related to the fractal dimension of the attractor of the dynamical system that generates the time series. (See Takens, 1981; Sauer, Yorke and Casdagli, 1991, for the theory and Nicolis and Prigogine, 1989, Appendix IV, for the corresponding algorithm.) Note that in this approach, L must be relatively small and K must be very large (formally, $K \to \infty$).

If L is sufficiently large, then one can consider each L-lagged vector X_i as a separate series and investigate the dynamics of certain characteristics for this collection of series. The simplest example of this approach is the well-known 'moving average' method, where the averages of the lagged vectors are computed. There are also much more sophisticated algorithms.

For example, if the initial series can be considered as a locally stationary process, then we can expand each lagged vector X_i with respect to any fixed basis (for instance, the Fourier basis or a certain wavelet basis) and study the dynamics of such an expansion. These ideas correspond to the dynamical Fourier analysis. Evidently, other bases can be applied as well.

The approximation of a stationary series with the help of the autoregression models can also be expressed in terms of embedding: if we deal with the model

$$f_{i+L-1} = a_{L-1} f_{i+L-2} + a_1 f_i + \varepsilon_{i+L-1}, \quad i \geq 0, \tag{1.6}$$

then we search for a vector $A = (a_1, \ldots, a_{L-1}, -1)^{\mathrm{T}}$ such that the inner products (X_i, A) are described in terms of a certain noise series.

Note that these (and many other) techniques that use the embedding can be divided into two large parts, which may be called 'global' and 'dynamical'. The global methods treat the X_i as L-dimensional vectors and do not use their ordering.

For instance, if we calculate the empirical distribution of the pairwise distances between the lagged vectors, then the result does not depend on the order in which these vectors appear. A similar situation occurs for the autoregression model (1.6) if the coefficients a_i are calculated through the whole collection of the lagged vectors (for example, by the least squares method).

This invariance under permutation of the lagged vectors is not surprising since both models deal with stationary-like series and are intended for finding global characteristics of the whole series. The number of lagged vectors K plays the role of the 'sample size' in these considerations, and therefore it has to be rather large. Theoretically, in these approaches L must be fixed and $K \to \infty$.

The situation is different when we deal with the dynamical Fourier analysis and similar methods, and even with the moving averages. Here the order of the lagged vectors is important and describes the dynamics of interest. Therefore, the nonstationary scenario is the main application area for these approaches. As for L and K, their relationship can generally be arbitrary and should depend on the concrete data and the concrete problem.

At any rate, the window length L should be sufficiently large. The value of L has to be large enough so that each L-lagged vector incorporates an essential part of the behaviour of the initial series $F = (f_0, \ldots, f_{N-1})$.

In accordance with the formal description of the embedding step (see Section 1.1.1), the result of this step is a *trajectory matrix*

$$\mathbf{X} = [X_1 : \ldots : X_K]$$

rather than just a collection of the lagged vectors X_i. This means that generally we are interested in the dynamical effects (though some characteristics that are invariant under permutations of the lagged vectors will be important as well).

The trajectory matrix (1.1) possesses an obvious symmetry property: the transposed matrix $\mathbf{X}^{\mathrm{T}}$ is the trajectory matrix of the same series $f_0, \ldots, f_{N-1}$ with window length equal to K rather than L.

1.2.2 Singular value decomposition

Singular value decomposition (SVD) of the trajectory matrix (1.1) is the second step in Basic SSA. SVD can be described in different terms and be used for dif-

ferent purposes. (See Chapter 4 for the mathematical results.) Most SVD features are valid for general $L \times K$ matrices, but the Hankel structure of the trajectory matrix adds a number of specific features. Let us start with general properties of the SVD important for the SSA.

As was already mentioned, the SVD of an arbitrary nonzero $L \times K$ matrix $\mathbf{X} = [X_1 : \ldots : X_K]$ is a decomposition of $\mathbf{X}$ in the form

$$\mathbf{X} = \sum_{i=1}^{d} \sqrt{\lambda_i} U_i V_i^{\mathrm{T}}, \tag{1.7}$$

where λ_i $(i = 1, \ldots, L)$ are eigenvalues of the matrix $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$ arranged in decreasing order of magnitudes,

$$d = \max\{i, \text{ such that } \lambda_i > 0\} = \operatorname{rank} \mathbf{X},$$

$\{U_1, \ldots, U_d\}$ is the corresponding orthonormal system of the eigenvectors of the matrix $\mathbf{S}$, and $V_i = \mathbf{X}^{\mathrm{T}} U_i / \sqrt{\lambda_i}$.

Standard SVD terminology calls $\sqrt{\lambda_i}$ the *singular values*; the U_i and V_i are the *left* and *right singular vectors* of the matrix $\mathbf{X}$, respectively. The collection $(\sqrt{\lambda_i}, U_i, V_i)$ is called *ith eigentriple* of the matrix $\mathbf{X}$. If we define $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^{\mathrm{T}}$, then the representation (1.7) can be rewritten in the form (1.2), i.e. as the representation of $\mathbf{X}$ as a sum of the elementary matrices $\mathbf{X}_i$.

If all the eigenvalues have multiplicity one, then the expansion (1.2) is uniquely defined. Otherwise, if there is at least one eigenvalue with multiplicity larger than 1, then there is a freedom in the choice of the corresponding eigenvectors. We shall assume that the eigenvectors are somehow chosen and the choice is fixed.

Since SVD deals with the whole matrix $\mathbf{X}$, it is not invariant under permutation of its columns $X_1, \ldots, X_K$. Moreover, the equality (1.7) shows that the SVD possesses the following property of symmetry: $V_1, \ldots, V_d$ form an orthonormal system of eigenvectors for the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ corresponding to the same eigenvalues λ_i. Note that the rows and columns of the trajectory matrix are subseries of the original time series. Therefore, the left and right singular vectors also have a temporal structure and hence can also be regarded as time series.

SVD (1.2) possesses a number of optimal features. One of these properties is as follows: among all the matrices $\mathbf{X}^{(r)}$ of rank $r < d$, the matrix $\sum_{i=1}^{r} \mathbf{X}_i$ provides the best approximation to the trajectory matrix $\mathbf{X}$, so that $||\mathbf{X} - \mathbf{X}^{(r)}||_{\mathcal{M}}$ is minimum.

Here and below the (*Frobenius*) *norm* of a matrix $\mathbf{Y}$ is $\sqrt{\langle \mathbf{Y}, \mathbf{Y} \rangle_{\mathcal{M}}}$, where the *inner product* of two matrices $\mathbf{Y} = (y_{ij})_{i,j=1}^{q,s}$ and $\mathbf{Z} = (z_{ij})_{i,j=1}^{q,s}$ is defined as

$$\langle \mathbf{Y}, \mathbf{Z} \rangle_{\mathcal{M}} = \sum_{i,j=1}^{q,s} y_{ij} z_{ij}.$$

Note that $||\mathbf{X}||_{\mathcal{M}}^2 = \sum_{i=1}^{d} \lambda_i$ and $\lambda_i = ||\mathbf{X}_i||_{\mathcal{M}}^2$ for $i = 1, \ldots, d$. Thus, we shall consider the ratio $\lambda_i / ||\mathbf{X}||_{\mathcal{M}}^2$ as the characteristic of the contribution of the

matrix $\mathbf{X}_i$ in the expansion (1.2) to the whole trajectory matrix $\mathbf{X}$. Consequently, $\sum_{i=1}^{r} \lambda_i / ||\mathbf{X}||^2_{\mathcal{M}}$, the sum of the first r ratios, is the characteristic of the optimal approximation of the trajectory matrix by the matrices of rank r.

Let us now consider the trajectory matrix $\mathbf{X}$ as a sequence of L-lagged vectors. Denote by $\mathfrak{L}^{(L)} \subset \mathbf{R}^L$ the linear space spanned by the vectors $X_1, \ldots, X_K$. We shall call this space the *L-trajectory space* (or, simply, *trajectory space*) of the series F. To emphasize the role of the series F, we use notation $\mathfrak{L}^{(L)}(F)$ rather than $\mathfrak{L}^{(L)}$. The equality (1.7) shows that $\mathcal{U} = (U_1, \ldots, U_d)$ is an orthonormal basis in the d-dimensional trajectory space $\mathfrak{L}^{(L)}$.

Setting $Z_i = \sqrt{\lambda_i} V_i$, $i = 1, \ldots, d$, we can rewrite the expansion (1.7) in the form

$$\mathbf{X} = \sum_{i=1}^{d} U_i Z_i^{\mathrm{T}}, \tag{1.8}$$

and for the lagged vectors X_j we have

$$X_j = \sum_{i=1}^{d} z_{ji} U_i, \tag{1.9}$$

where the z_{ji} are the components of the vector Z_i.

By (1.9), z_{ji} is the ith component of the vector X_j, represented in the basis $\mathcal{U}$. In other words, the vector Z_i is composed of the ith components of lagged vectors represented in the basis $\mathcal{U}$.

Let us now consider the transposed trajectory matrix $\mathbf{X}^{\mathrm{T}}$. Introducing $Y_i = \sqrt{\lambda_i} U_i$ we obtain the expansion

$$\mathbf{X}^{\mathrm{T}} = \sum_{i=1}^{d} V_i Y_i^{\mathrm{T}},$$

which corresponds to the representation of the sequence of K-lagged vectors in the orthonormal basis $V_1, \ldots, V_d$. Thus, the SVD gives rise to two dual geometrical descriptions of the trajectory matrix $\mathbf{X}$.

The optimal feature of the SVD considered above may be reformulated in the language of multivariate geometry for the L-lagged vectors as follows. Let $r < d$. Then among all r-dimensional subspaces $\mathfrak{L}_r$ of $\mathbf{R}^L$, the subspace $\mathfrak{L}_r^{(0)} \stackrel{\text{def}}{=} \mathfrak{L}(U_1, \ldots, U_r)$, spanned by $U_1, \ldots, U_r$, approximates these vectors in the best way; that is, the minimum of $\sum_{i=1}^{K} \mathrm{dist}^2(X_i, \mathfrak{L}_r)$ is attained at $\mathfrak{L}_r^{(0)}$. The ratio $\sum_{i=1}^{r} \lambda_i / \sum_{i=1}^{d} \lambda_i$ is the characteristic of the best r-dimensional approximation of the lagged vectors.

Another optimal feature relates to the properties of the directions determined by the eigenvectors $U_1, \ldots, U_d$. Specifically, the first eigenvector U_1 determines the direction such that the variation of the projections of the lagged vectors onto this direction is maximum.

Every subsequent eigenvector determines a direction that is orthogonal to all previous directions, and the variation of the projections of the lagged vectors onto this direction is also maximum. Therefore, it is natural to call the direction of ith eigenvector U_i the *ith principal direction*. Note that the elementary matrices $\mathbf{X}_i = U_i Z_i^{\mathrm{T}}$ are built up from the projections of the lagged vectors onto ith directions.

This view on the SVD of the trajectory matrix composed of L-lagged vectors and an appeal to associations with *principal component analysis* lead to the following terminology. We shall call the vector U_i the ith (principal) *eigenvector*, the vector V_i will be called the *ith factor vector*, and the vector Z_i the *vector of ith principal components*.

1.2.3 Grouping

Let us now comment on the grouping step, which is the procedure of arranging the matrix terms $\mathbf{X}_i$ in (1.2). Assume that $m = 2$, $I_1 = I = \{i_1 \ldots, i_r\}$ and $I_2 = \{1, \ldots, d\} \setminus I$, where $1 \le i_1 < \ldots < i_r \le d$.

The purpose of the grouping step is separation of the additive components of time series. Let us discuss the very important concept of separability in detail. Suppose that the time series F is a sum of two time series $F^{(1)}$ and $F^{(2)}$; that is, $f_i = f_i^{(1)} + f_i^{(2)}$ for $i = 0, \ldots, N-1$. Let us fix the window length L and denote by $\mathbf{X}$, $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ the L-trajectory matrices of the series F, $F^{(1)}$ and $F^{(2)}$, respectively.

Consider an SVD (1.2) of the trajectory matrix $\mathbf{X}$. (Recall that if all the eigenvalues have multiplicity one, then this expansion is unique.) We shall say that the series $F^{(1)}$ and $F^{(2)}$ are (weakly) *separable by the decomposition* (1.2), if there exists a collection of indices $I \subset \{1, \ldots, d\}$ such that $\mathbf{X}^{(1)} = \sum_{i \in I} \mathbf{X}_i$ and consequently $\mathbf{X}^{(2)} = \sum_{i \notin I} \mathbf{X}_i$.

In the case of separability, the contribution of $\mathbf{X}^{(1)}$, the first component in the expansion $\mathbf{X} = \mathbf{X}^{(1)} + \mathbf{X}^{(2)}$, is naturally to measure by the share of the corresponding eigenvalues: $\sum_{i \in I} \lambda_i \Big/ \sum_{i=1}^{L} \lambda_i$.

The separation of the series by the decomposition (1.2) can be looked at from different perspectives. Let us fix the set of indices $I = I_1$ and consider the corresponding resultant matrix $\mathbf{X}_{I_1}$. If this matrix, and therefore $\mathbf{X}_{I_2} = \mathbf{X} - \mathbf{X}_{I_1}$, are Hankel matrices, then they are necessarily the trajectory matrices of certain time series that are separable by the expansion (1.2).

Moreover, if the matrices $\mathbf{X}_{I_1}$ and $\mathbf{X}_{I_2}$ are close to some Hankel matrices, then there exist series $F^{(1)}$ and $F^{(2)}$ such that $F = F^{(1)} + F^{(2)}$ and the trajectory matrices of these series are close to $\mathbf{X}_{I_1}$ and $\mathbf{X}_{I_2}$, respectively (the problem of finding these series is discussed below). In this case we shall say that the series are *approximately separable*.

Therefore, the purpose of the grouping step (that is the procedure of arranging the indices $1, \ldots, d$ into groups) is to find several groups $I_1, \ldots, I_m$ such that the matrices $\mathbf{X}_{I_1}, \ldots, \mathbf{X}_{I_m}$ satisfy (1.3) and are close to certain Hankel matrices.

Let us now look at the grouping step from the viewpoint of multivariate geometry. Let $\mathbf{X} = [X_1 : \ldots : X_K]$ be the trajectory matrix of a time series F, $F = F^{(1)} + F^{(2)}$, and the series $F^{(1)}$ and $F^{(2)}$ are separable by the decomposition (1.2), which corresponds to splitting the index set $\{1, \ldots, d\}$ into I and $\{1, \ldots, d\} \setminus I$.

The expansion (1.3) with $m = 2$ means that $U_1, \ldots, U_d$, the basis in the trajectory space $\mathfrak{L}^{(L)}$, splits into two groups of basis vectors. This corresponds to the representation of $\mathfrak{L}^{(L)}$ as a product of two orthogonal subspaces (*eigenspaces*) $\mathfrak{L}^{(L,1)} = \mathfrak{L}(U_i, i \in I)$ and $\mathfrak{L}^{(L,2)} = \mathfrak{L}(U_i, i \notin I)$ spanned by $U_i, i \in I$, and $U_i, i \notin I$, respectively.

Separability of two series $F^{(1)}$ and $F^{(2)}$ means that the matrix $\mathbf{X}_I$, whose columns are the projections of the lagged vectors $X_1, \ldots, X_K$ onto the eigenspace $\mathfrak{L}^{(L,1)}$, is exactly the trajectory matrix of the series $F^{(1)}$.

Despite the fact that several formal criteria for separability will be introduced, the whole procedure of splitting the terms into groups (i.e., the grouping step) is difficult to formalize completely. This procedure is based on the analysis of the singular vectors U_i, V_i and the eigenvalues λ_i in the SVD expansions (1.2) and (1.7). The principles and methods of identifying the SVD components for their inclusion into different groups are described in Section 1.6.

Since each matrix component of the SVD is completely determined by the corresponding eigentriple, we shall talk about grouping of the eigentriples rather than grouping of the elementary matrices $\mathbf{X}_i$.

Note also that the case of two series components ($m = 2$) considered above is often more sensibly regarded as the problem of separating out a single component (for example, as a noise reduction) rather than the problem of separation of two terms. In this case, we are interested in only one group of indices, namely I.

1.2.4 Diagonal averaging

If the components of the series are separable and the indices are being split up accordingly, then all the matrices in the expansion (1.3) are Hankel matrices. We thus immediately obtain the decomposition (1.5) of the original series: for every k and n, $\widetilde{f}_n^{(k)}$ is equal to all the entries $x_{ij}^{(k)}$ along the secondary diagonal

$$\{(i, j), \text{ such that } i + j = n + 2\}$$

of the matrix $\mathbf{X}_{I_k}$.

In practice, however, this situation is not realistic. In the general case, no secondary diagonal consists of equal elements. We thus need a formal procedure of transforming an arbitrary matrix into a Hankel matrix and therefore into a series. As such, we shall consider the procedure of *diagonal averaging,* which defines

the values of the time series $\widetilde{F}^{(k)}$ as averages of the corresponding diagonals of the matrices $\mathbf{X}_{I_k}$.

It is convenient to represent the diagonal averaging step with the help of the *Hankelization* operator $\mathcal{H}$.

The operator $\mathcal{H}$ acts on an arbitrary $(L \times K)$-matrix $\mathbf{Y} = (y_{ij})$ in the following way (assume for definiteness that $L \le K$): for $i + j = s$ and $N = L + K - 1$ the element $\widetilde{y}_{ij}$ of the matrix $\mathcal{H}\mathbf{Y}$ is

$$\widetilde{y}_{ij} = \begin{cases} \dfrac{1}{s-1}\displaystyle\sum_{l=1}^{s-1} y_{l,s-l} & \text{for } 2 \le s \le L-1\,, \\ \dfrac{1}{L}\displaystyle\sum_{l=1}^{L} y_{l,s-l} & \text{for } L \le s \le K+1\,, \\ \dfrac{1}{K+L-s+1}\displaystyle\sum_{l=s-K}^{L} y_{l,s-l} & \text{for } K+2 \le s \le K+L\,. \end{cases} \tag{1.10}$$

For $L > K$ the expression for the elements of the matrix $\mathcal{H}\mathbf{Y}$ is analogous, the changes are the substitution $L \leftrightarrow K$ and the use of the transposition of the original matrix $\mathbf{Y}$.

Note that the Hankelization is an optimal procedure in the sense that the matrix $\mathcal{H}\mathbf{Y}$ is closest to $\mathbf{Y}$ (with respect to the matrix norm) among all Hankel matrices of the corresponding size (see Section 6.2). In its turn, the Hankel matrix $\mathcal{H}\mathbf{Y}$ defines the series uniquely by relating the values in the diagonals to the values in the series.

By applying the Hankelization procedure to all matrix components of (1.3), we obtain another expansion:

$$\mathbf{X} = \widetilde{\mathbf{X}}_{I_1} + \ldots + \widetilde{\mathbf{X}}_{I_m}, \tag{1.11}$$

where $\widetilde{\mathbf{X}}_{I_l} = \mathcal{H}\mathbf{X}_{I_l}$.

A sensible grouping leads to the decomposition (1.3) where the resultant matrices $\mathbf{X}_{I_k}$ are almost Hankel ones. This corresponds to approximate separability and implies that the pairwise inner products of different matrices $\widetilde{\mathbf{X}}_{I_k}$ in (1.11) are small.

Since all the matrices on the right-hand side of the expansion (1.11) are Hankel matrices, each matrix uniquely determines the time series $\widetilde{F}^{(k)}$ and we thus obtain (1.5), the decomposition of the original time series.

The procedure of computing the time series $\widetilde{F}^{(k)}$ (that is, building up the group I_k plus diagonal averaging of the matrix $\mathbf{X}_{I_k}$) will be called *reconstruction of a series component* $\widetilde{F}^{(k)}$ *by the eigentriples* with indices in I_k.

1.3 Basic SSA: basic capabilities

In this section we start discussing examples that illustrate basic capabilities of Basic SSA. Note that terms such as 'trend', 'smoothing', 'signal', and 'noise' are

used here in their informal, common-sense meaning and will be commented on later.

1.3.1 Trends of different resolution

The example 'Production' (crude oil, lease condensate, and natural gas plant liquids production, monthly data from January 1973 to September 1999) shows the capabilities of SSA in extraction of trends that have different resolutions. Though the series has a seasonal component (and the corresponding component can be extracted together with the trend component), for the moment we do not pay attention to periodicities.

Taking the window length $L = 120$ we see that the eigentriples 1-3 correspond to the trend. Choosing these eigentriples in different combinations we can find different trend components.

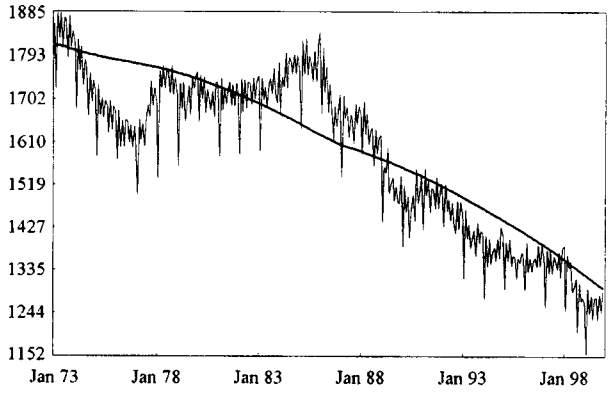


Figure 1.1 *Production: general tendency (rough trend).*

Figs. 1.1 and 1.2 demonstrate two alternatives in the trend resolution. The leading eigentriple gives a general tendency of the series (Fig. 1.1). The three leading eigentriples describe the behaviour of the data more accurately (Fig. 1.2) and show not only the general decrease of production, but also its growth from the middle 70s to the middle 80s.

1.3.2 Smoothing

The series 'Tree rings' (tree ring indices, Douglas fir, Snake river basin, U.S., annual, from 1282 to 1950), is described in Hipel and McLeod (1994, Chapter 10) with the help of a (3,0)-order ARIMA model. If the ARIMA-type model is accepted, then it is generally meaningless to look for any trend or periodicities. However, we can smooth the series with the help of Basic SSA.

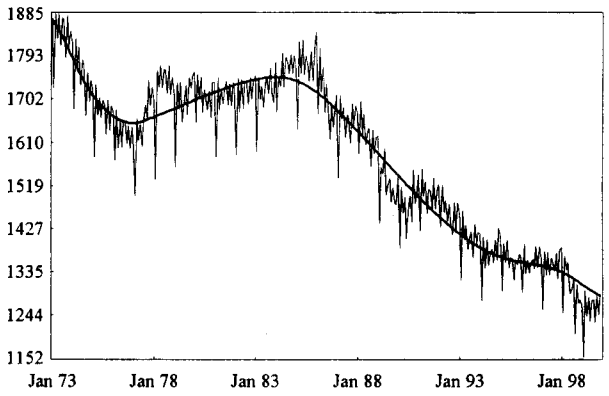


Figure 1.2 *Production: accurate trend.*

Fig. 1.3 shows the initial series and the result of its SSA smoothing, which is obtained by the leading 7 eigentriples with window length 120. Fig. 1.4 depicts the residuals.

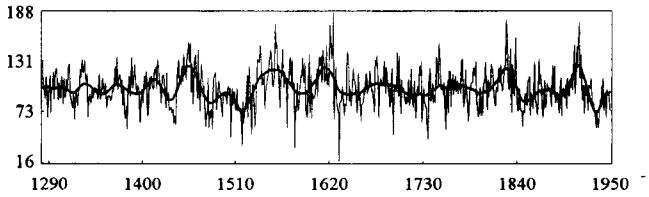


Figure 1.3 *Tree rings: smoothing result.*

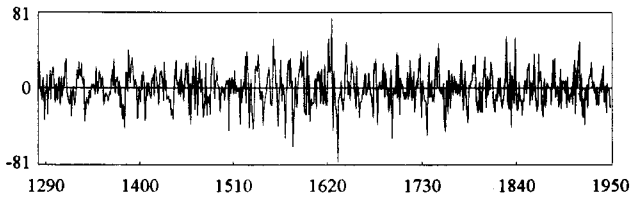


Figure 1.4 *Tree rings: residuals.*

Another example demonstrating SSA as a smoothing technique uses the 'White dwarf' data, which contains 618 point measurements of the time variation of the intensity of the white dwarf star PG1159-035 during March 1989. The data is

discussed in Clemens (1994). The whole series can be described as a smooth quasi-periodic curve with a noise component.

Using Basic SSA with window length $L = 100$ and choosing the leading 11 eigentriples for the reconstruction, we obtain the smooth curve of Fig. 1.5 (thick line). The residuals (Fig. 1.6) seem to have no evident structure (to simplify the visualization of the results; these figures present only a part of the series).

Further analysis shows that the residual series can be regarded as a Gaussian white noise, though it does not contain very low frequencies (see the discussion in Section 1.6.1).

Thus, we can assume that in this case the smoothing procedure leads to noise reduction and the smooth curve in Fig. 1.5 describes the signal.

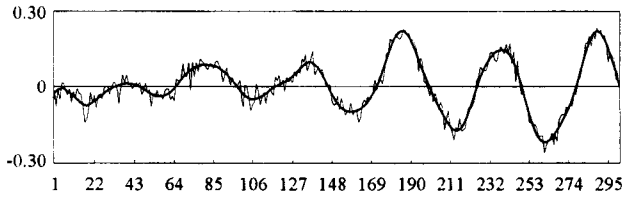


Figure 1.5 *White dwarf: smoothed series.*

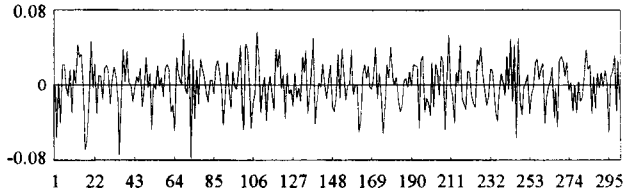


Figure 1.6 *White dwarf: residuals.*

1.3.3 Extraction of seasonality components

The 'Eggs' data (eggs for a laying hen, monthly, U.S., from January 1938 to December 1940, Kendall and Stuart, 1976, Chapter 45) has a rather simple structure: it is the sum of an explicit annual oscillation (though not a harmonic one) and the trend, which is almost constant.

The choice $L = 12$ allows us to extract simultaneously all seasonal components (12, 6, 4, 3, 2.4, and 2-months harmonics) as well as the trend.

The graph in Fig. 1.7 depicts the initial series and its trend (thick line), which is reconstructed from the first eigentriple.

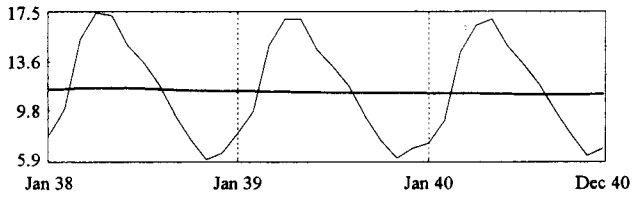


Figure 1.7 *Eggs: initial series and its trend.*

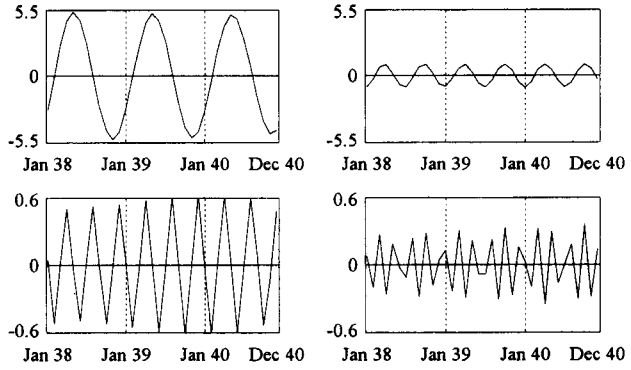


Figure 1.8 *Eggs: four leading seasonal harmonics.*

The four leading seasonal *harmonic components* (briefly, *harmonics*) of the series are depicted in Fig. 1.8; they are: 12-months, 6-months (presented in the same scale), 4-months and 2.4-months harmonics (also in the same scale). The corresponding pairs of the eigentriples are 2-3; 4-5; 6-7, and 8-9. The two weakest harmonics, 3-months and 2-months (10-11 and 12 eigentriples, respectively), are not shown.

1.3.4 Extraction of cycles with small and large periods

The series 'Births' (number of daily births, Quebec, Canada, from January 1, 1977 to December 31, 1990) is discussed in Hipel and McLeod (1994). It shows, in addition to a smooth trend, two cycles of different ranges: the one-year periodicity and the one-week periodicity.

Both periodicities (as well as the trend) can be simultaneously extracted by Basic SSA with window length $L = 365$. Fig. 1.9 shows the one-year cycle of the series added to the trend (white line) on the background of the 'Births' series from 1981 to 1990. Note that the form of this cycle varies in time, though the main two peaks (spring and autumn) remain stable. The trend corresponds to the

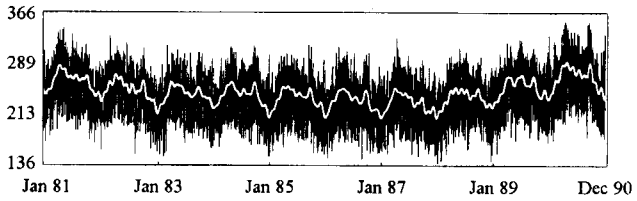


Figure 1.9 *Births: initial time series and its annual periodicity.*

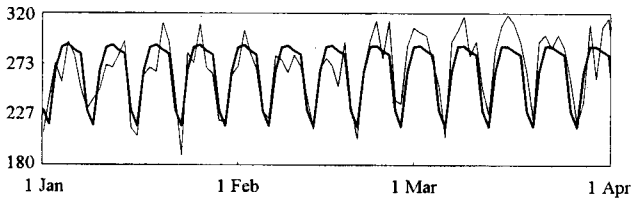


Figure 1.10 *Births: one-week periodicity.*

leading eigentriple, while the one-year periodic component is reconstructed from the eigentriples 6-9 and 12-19.

Fig. 1.10 demonstrates the one-week cycle on the background of the initial series for approximately the first three months of 1977. This cycle corresponds to the eigentriples 2-5 and 10-11.

1.3.5 Extraction of periodicities with varying amplitudes

The capability of SSA in extracting an oscillating signal with a varying amplitude can be illustrated by the example of the 'Drunkenness' series (monthly public drunkenness intakes, Minneapolis, U.S., from January 1966 to July 1978, McCleary and Hay, 1980). The initial series is depicted in Fig. 1.11 (thin line).

Taking $L = 60$ in Basic SSA and reconstructing the series from the fourth and fifth eigentriples, we see (bottom line in Fig. 1.11) an almost pure 12-months periodic component. The amplitude of this annual periodic component approximately equals 120 at the beginning of the observation time. The amplitude then gradually decreases and almost disappears at the end. The amplitude is reduced by a factor of about 10, but the trend in the data is diminished only by a factor of three to four.

1.3.6 Complex trends and periodicities

The 'Unemployment' series (West Germany, monthly, from April 1950 to December 1980, Rao and Gabr, 1984) serves as an example of SSA capability of

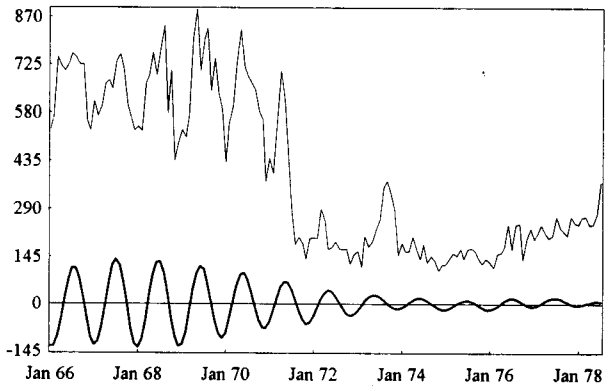


Figure 1.11 *Drunkenness: varying amplitudes.*

extracting complex trends simultaneously with the amplitude-modulated periodicities.

The result of extraction is presented in Fig. 1.12 (the initial series and the reconstructed trend) and in Fig. 1.13 (seasonality).

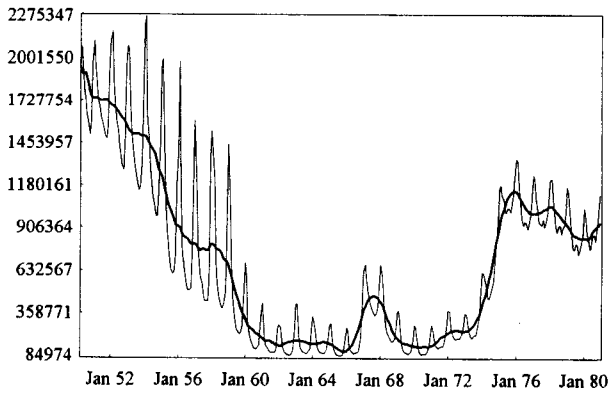


Figure 1.12 *Unemployment: trend.*

The window length was taken as $L = 180$. Since both the trend and the seasonality are complex, many eigentriples are required to reconstruct them. The trend is reconstructed from the eigentriples 1, 2, 5-7, 10, 11, 14, 15, 20, 21, 24, 27, 30, and 33, while the eigentriples with numbers 3, 4, 8, 9, 12, 13, 16-19, 22, 23, 25, 26, 34, 35, 43, 44, 71, and 72 describe the seasonality.

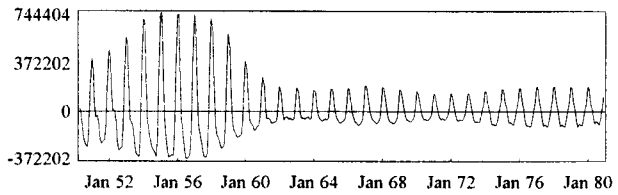


Figure 1.13 *Unemployment: seasonality.*

If we were to take a smaller number of eigentriples for the trend, then we would obtain a less refined description of a smooth, slowly varying component of the series corresponding to a more general tendency in the series.

1.3.7 Finding structure in short time series

The series 'War' (U.S. combat deaths in the Indochina war, monthly, from 1966 to 1971, Janowitz and Schweizer, 1989, Table 10) is chosen to demonstrate the capabilities of SSA in finding a structure in short time series.

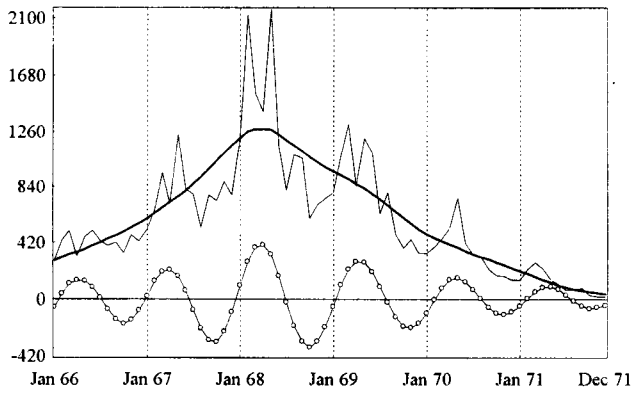


Figure 1.14 *War: trend and annual periodicity.*

Selecting a window length $L = 18$, we can see (Fig. 1.14) that the two leading eigentriples perfectly describe the trend of the series (thick line on the background of the initial data). This trend relates to the overall involvement of U.S. troops in the war.

The third (bottom) plot of Fig. 1.14 shows the component of the initial series reconstructed from the eigentriples 3 and 4. There is little doubt that this is an annual oscillation modulated by the war intensity. This oscillation has its origin

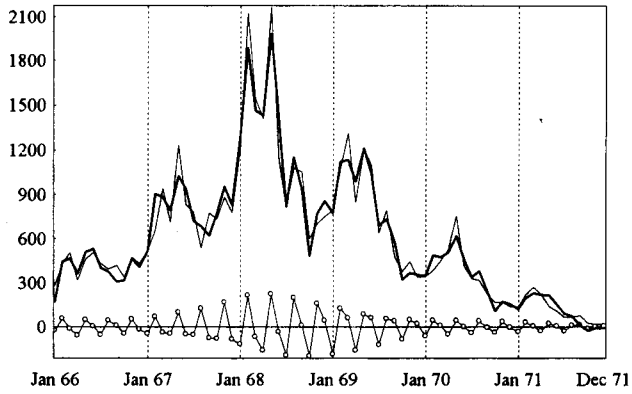


Figure 1.15 *War: quarter periodicity and series approximation.*

in the climatic conditions of South-East Asia: the summer season is much more difficult for any activity than the winter one.

Two other series components, namely that of the quarterly cycle corresponding to the eigentriples 5 and 6 (depicted at the bottom of Fig. 1.15) and the omitted 4-months cycle, which can be reconstructed from the eigentriples 7 and 8, are both modulated by the war intensity and both are less clear for interpretation. Nevertheless, if we add all these effects together (that is, reconstruct the series component corresponding to the eight leading eigentriples), a perfect agreement between the result and the initial series becomes apparent: see Fig. 1.15, top two plots, with the thick line corresponding to the reconstruction.

1.4 Time series and SSA tasks

In the previous section the terms 'trend', 'smoothing', 'amplitude modulation' and 'noise' were used without any explanation of their meaning. In this section we shall provide the related definitions and corresponding discussions. We shall also describe the major tasks that can be attempted by Basic SSA. Examples of application of Basic SSA for solving these tasks have been considered in Section 1.3.

1.4.1 Models of time series and the periodograms

Formally, SSA can be applied to an arbitrary time series. However, a theoretical study of its properties requires specific considerations for different classes of series. Moreover, different classes assume different choices of parameters and expected results. We thus start this section with a description of several classes of time series, which are natural for the SSA treatment, and use these classes

to discuss the important concept of (approximate) separability defined earlier in Section 1.2.3. (For the theoretical aspects of separability see Section 6.1.)

Since the main purpose of SSA is a decomposition of the series into additive components, we always implicitly assume that this series is a sum of several simpler series. These 'simple' series are the objects of the discussion below. Note also that here we only consider deterministic time series, including those that can be regarded as 'noise'. Stochastic models of the noise series, in their relation to the separability problem, are discussed in Sections 6.1.3 and 6.3.

(a) Stationary series

The concept of a deterministic stationary time series is asymptotic (rigorous definitions and results on the subject are given in Section 6.4, here we stick to a looser style). Specifically, an infinite series $F = (f_0, f_1, \dots, f_n, \dots)$ is called *stationary* if for all nonnegative integers k, m the following convergence takes place:

$$\frac{1}{N}\sum_{j=0}^{N-1} f_{j+k}f_{j+m} \underset{N\to\infty}{\longrightarrow} R_f(k-m), \tag{1.12}$$

where the (even) function $R_f(n)$ is called the *covariance function* of the series F. The covariance function can be represented as

$$R_f(n) = \int\limits_{(-1/2,1/2]} e^{i2\pi n\omega} m_f(d\omega),$$

where m_f is a measure called the *spectral measure* of the series F.

The form of the spectral measure determines properties of the corresponding stationary series in many respects. For example, the convergence (1.12) implies, loosely speaking, the convergence of the averages

$$\frac{1}{N}\sum_{j=0}^{N-1} f_{j+k} \underset{N\to\infty}{\longrightarrow} 0 \tag{1.13}$$

for any k if and only if m_f does not have an atom at zero.

Thus, the definition of stationarity is related to the ergodicity not only of the second order, but also of the first order as well. Below, when discussing stationarity, we shall always assume that (1.13) holds, which is the zero-mean assumption for the original series.

If the measure m_f is discrete, then, roughly speaking, we can assume that the stationary series F has the form

$$f_n \sim \sum_k a_k \cos(2\pi\omega_k n) + \sum_k b_k \sin(2\pi\omega_k n), \quad \omega_k \in (0, 1/2], \tag{1.14}$$

where $a_k = a(\omega_k), b_k = b(\omega_k), b(1/2) = 0$ and the sum $\sum_k (a_k^2 + b_k^2)$ converges. (Note that $a(1/2) \neq 0$ if one of the ω_k is exactly $1/2$.)