

Use R!

Series editors

Robert Gentleman Kurt Hornik Giovanni Parmigiani

More information about this series at <http://www.springer.com/series/6991>

Nina Golyandina • Anton Korobeynikov •
Anatoly Zhigljavsky

Singular Spectrum Analysis with R

 Springer

Nina Golyandina
Faculty of Mathematics and Mechanics
Saint Petersburg State University
Saint Petersburg, Russia

Anton Korobeynikov
Faculty of Mathematics and Mechanics
Saint Petersburg State University
Saint Petersburg, Russia

Anatoly Zhigljavsky
School of Mathematics
Cardiff University
Cardiff, United Kingdom

ISSN 2197-5736

ISSN 2197-5744 (electronic)

Use R!

ISBN 978-3-662-57378-5

ISBN 978-3-662-57380-8 (eBook)

<https://doi.org/10.1007/978-3-662-57380-8>

Library of Congress Control Number: 2018940189

Mathematics Subject Classification (2010): 37M10, 68U10

© Springer-Verlag GmbH Germany, part of Springer Nature 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer-Verlag GmbH, DE part of Springer Nature.

The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Preface

Singular spectrum analysis (SSA) is a well-known methodology for analysis and forecasting of time series. Since quite recently, SSA was also used to analyze digital images and other objects that are not necessarily of planar or rectangular form and may contain gaps. SSA is multipurpose and naturally combines both model-free and parametric techniques; this makes it a very special and attractive methodology for solving a wide range of problems arising in diverse areas, most notably those associated with time series and digital images. An effective, comfortable, and accessible implementation of SSA is provided by the R-package *Rssa*, which is available from CRAN and reviewed in this book.

For time series, SSA can be used for many different purposes, from non-parametric time series decomposition and filtration to parameter estimation, forecasting, gap filling, and change-point detection; many of these techniques can be extended to digital images. An essential difference between one-dimensional SSA and the majority of methods that analyze time series with trend and/or periodicities lies in the fact that SSA does not require any model for trend and no prior knowledge is needed about the number of periodical components and their frequencies. Also, periodicities can be modulated in different ways, and therefore, the type of model, additive or multiplicative, is not necessary to be assumed and taken into consideration for applying SSA. In the process of analysis, SSA constructs a number of the so-called subspace-based models; an example of a time series which satisfies such a model is provided by a series that follows a linear recurrence relation in the presence of noise. The validity of any of these models, however, is not a prerequisite for SSA, which makes SSA a very flexible tool that could be applied to virtually any data in the form of a time series or a digital image.

There are three books devoted to SSA¹ as well as many papers related to the methodological and theoretical aspects of SSA and comparison with ARIMA and other methods but especially to various applications. This present book expands the SSA methodology in many different directions and unifies various approaches and modifications within the SSA framework. Those mostly interested in applications may consider this book as an RSSA textbook.

Any method needs effective, comfortable, and accessible implementation. The R-package RSSA² provides such an implementation for SSA. RSSA is well documented and contains many standard and nonstandard tools for time series analysis and forecasting and image processing; it also has many visual tools which are useful for making proper choice of SSA parameters and examination of results. RSSA is the only SSA implementation available from CRAN and is almost certainly the most efficient implementation of SSA. RSSA is well theoretically and methodologically supported as its main routines are based on the books mentioned above. Most of the new SSA developments, which are systemized in this book, are also implemented in RSSA.

This book has the following goals: (a) to present the up-to-date SSA methodology, including multidimensional extensions, in the language accessible to a large circle of users; (b) to interconnect a variety of versions of SSA into a single tool, (c) to show the diverse tasks that SSA can be used for; (d) to formally describe the main SSA methods and algorithms; and (e) to provide tutorials on the RSSA package and the use of SSA. The companion website for the book is <http://ssa-with-r-book.github.io>.

The authors have vast experience with SSA. Nina Golyandina (NG) and Anatoly Zhigljavsky (AZ) have worked on SSA for more than 30 years and wrote about 20 papers each on the SSA-related topics. They authored the books Golyandina et al. (2001) and Golyandina and Zhigljavsky (2013) and in doing so they have spent huge efforts on trying to develop and tidy up the SSA methodology. NG is the head of the “Caterpillar-SSA” project <http://gistatgroup.com/cat/> and a part of the team which has applied SSA to the analysis of gene expression; see Sect. 5.4.3. NG had supervised six PhD students (Eugene Osipov, Theodore Alexandrov, Konstantin Usevich, Alex Shlemov, Maxim Lomtev, and Nikita Zvonarev) whose projects were fully devoted to SSA. There is much material in this book which is based on recent papers of NG with coauthors; see Sect. 1.7.4 for details. AZ was behind the idea of using SSA for change-point detection and structural monitoring of time series, has participated in many projects on applications of SSA in economics, organized a number of conferences on SSA, and edited two recent SSA-oriented

¹Elsner JB, Tsonis AA (1996) Singular spectrum analysis: a new tool in time series analysis. Plenum; Golyandina N, Zhigljavsky A (2013) Singular spectrum analysis for time series. Springer briefs in statistics. Springer; Golyandina N, Nekrutkin V, Zhigljavsky A (2001) Analysis of time series structure: SSA and related techniques. Chapman & Hall/CRC.

²Korobeynikov A, Shlemov A, Usevich K, Golyandina N (2017) Rssa: a collection of methods for singular spectrum analysis. R package version 1.0. <http://CRAN.R-project.org/package=Rssa>.

special issues of the *Statistics and Its Interface* journal. Both NG and AZ have taught SSA to different audiences which enormously helped them with the books. Anton Korobeynikov (AK) specializes in data analysis, computational statistics, and programming. He has a vast experience in programming in C, C++, and R and contributed to many open-source projects. He stood behind the initial ideas toward fast SSA implementation.³ AK was the original author of RSSA; currently, he is its maintainer and one of the two major contributors to RSSA (the second major contributor is Alex Shlemov).

We believe that the book will be very useful to a very wide circle of readers including professional statisticians, specialists in signal and image processing, and specialists in numerous applied disciplines interested in using statistical methods of time series analysis, forecasting, and signal and image processing in their applications. The book is written on a level accessible to a very broad audience and contains a large number of examples; hence it can also be considered as a textbook for undergraduate and postgraduate courses on time series analysis and signal processing.

Acknowledgements The authors are very grateful to Vladimir Nekrutkin, their coauthor and one of the founders of the “Caterpillar-SSA” project, and to Alex Shlemov and Konstantin Usevich for their valuable contributions to RSSA and for their enormous contributions to the development of multidimensional extensions of SSA. Research of the third author was supported by the Russian Science Foundation, project No. 15-11-30022 “Global optimization, supercomputing computations, and application.”

Saint Petersburg, Russia
Saint Petersburg, Russia
Cardiff, UK

Nina Golyandina
Anton Korobeynikov
Anatoly Zhigljavsky

³Korobeynikov A (2010) Computation- and space-efficient implementation of SSA. *Stat Interface* 3(3):357–368.

Contents

1 Introduction: Overview	1
1.1 Generic Scheme of the SSA Family and the Main Concepts	2
1.1.1 SSA Methods	2
1.1.2 The Main Concepts	5
1.2 Different Versions of SSA	8
1.2.1 Decomposition of \mathbf{X} into a Sum of Rank-One Matrices	8
1.2.2 Versions of SSA Dealing with Different Forms of the Object	11
1.3 Separability in SSA	12
1.4 Forecasting, Interpolation, Low-Rank Approximation, and Parameter Estimation in SSA	13
1.5 The Package	14
1.5.1 SSA Packages	14
1.5.2 Tools for Visual Control and Choice of Parameters	15
1.5.3 Short Introduction to RSSA	16
1.5.4 Implementation Efficiency	17
1.6 Comparison of SSA with Other Methods	17
1.6.1 Fourier Transform, Filtering, Noise Reduction	18
1.6.2 Parametric Regression	19
1.6.3 ARIMA and ETS	20
1.7 Bibliographical Notes	21
1.7.1 Short History	21
1.7.2 Some Recent Applications of SSA	21
1.7.3 SSA for Preprocessing/Combination of Methods	22
1.7.4 Materials Used in This Book	23
1.8 Installation of RSSA and Description of the Data Used in the Book	24
1.8.1 Installation of RSSA and Usage Comments	24
1.8.2 Data Description	27
References	27

2	SSA Analysis of One-Dimensional Time Series	31
2.1	Basic SSA	32
2.1.1	Method	32
2.1.2	Appropriate Time Series	34
2.1.3	Separability and Choice of Parameters	36
2.1.4	Algorithm	38
2.1.5	Basic SSA in RSSA	39
2.2	Toeplitz SSA	46
2.2.1	Method	46
2.2.2	Algorithm	48
2.2.3	Toeplitz SSA in RSSA	48
2.3	SSA with Projection	51
2.3.1	Method	51
2.3.2	Appropriate Time Series	54
2.3.3	Separability	55
2.3.4	Algorithm	56
2.3.5	SSA with Projection in RSSA	57
2.4	Iterative Oblique SSA	63
2.4.1	Method	63
2.4.2	Separability	69
2.4.3	Algorithms	71
2.4.4	Iterative O-SSA in RSSA	72
2.5	Filter-Adjusted O-SSA and SSA with Derivatives	79
2.5.1	Method	79
2.5.2	Separability	81
2.5.3	Algorithm	82
2.5.4	Filter-Adjusted O-SSA in RSSA	84
2.6	Shaped 1D-SSA	88
2.6.1	Method	88
2.6.2	Separability	89
2.6.3	Algorithm	89
2.6.4	Shaped SSA in RSSA	90
2.7	Automatic Grouping in SSA	92
2.7.1	Methods	92
2.7.2	Algorithm	96
2.7.3	Automatic Grouping in RSSA	97
2.8	Case Studies	103
2.8.1	Extraction of Trend and Oscillations by Frequency Ranges	103
2.8.2	Trends in Short Series	104
2.8.3	Trend and Seasonality of Complex Form	106
2.8.4	Finding Noise Envelope	108
2.8.5	Elimination of Edge Effects	109

- 2.8.6 Extraction of Linear Trends 111
- 2.8.7 Automatic Decomposition..... 114
- 2.8.8 Log-Transformation 118
- References 120
- 3 Parameter Estimation, Forecasting, Gap Filling 121**
 - 3.1 Parameter Estimation 122
 - 3.1.1 Method 122
 - 3.1.2 Algorithms 124
 - 3.1.3 Estimation in RSSA 126
 - 3.2 Forecasting 129
 - 3.2.1 Method 131
 - 3.2.2 Algorithms 134
 - 3.2.3 Forecasting in RSSA 136
 - 3.3 Gap Filling 139
 - 3.3.1 Method 141
 - 3.3.2 Algorithms 143
 - 3.3.3 Gap-Filling in RSSA 145
 - 3.4 Structured Low-Rank Approximation 151
 - 3.4.1 Cadzow Iterations 151
 - 3.4.2 Algorithms 153
 - 3.4.3 Structured Low-Rank Approximation in RSSA 154
 - 3.5 Case Studies 157
 - 3.5.1 Forecasting of Complex Trend and Seasonality 157
 - 3.5.2 Different Methods of Forecasting 159
 - 3.5.3 Choice of Parameters and Confidence Intervals..... 162
 - 3.5.4 Gap Filling 167
 - 3.5.5 Parameter Estimation and Low-Rank Approximation 171
 - 3.5.6 Subspace Tracking 176
 - 3.5.7 Automatic Choice of Parameters for Forecasting 178
 - 3.5.8 Comparison of SSA, ARIMA, and ETS 182
 - References 186
- 4 SSA for Multivariate Time Series 189**
 - 4.1 Complex SSA 190
 - 4.1.1 Method 190
 - 4.1.2 Separability 190
 - 4.1.3 Algorithm 191
 - 4.1.4 Complex SSA in RSSA 192
 - 4.2 MSSA Analysis 194
 - 4.2.1 Method 194
 - 4.2.2 Multi-Dimensional Time Series and LRRs 197
 - 4.2.3 Separability 200
 - 4.2.4 Comments on 1D-SSA, MSSA and Complex SSA 201
 - 4.2.5 Algorithm 204
 - 4.2.6 MSSA Analysis in RSSA 205

- 4.3 MSSA Forecasting 210
 - 4.3.1 Method 211
 - 4.3.2 Algorithms 214
 - 4.3.3 MSSA Forecasting in RSSA 216
 - 4.3.4 Other Subspace-Based MSSA Extensions..... 221
- 4.4 Case Studies 222
 - 4.4.1 Analysis of Series in Different Scales (Normalization)..... 222
 - 4.4.2 Forecasting of Series with Different Lengths
and Filling-In 224
 - 4.4.3 Simultaneous Decomposition of Many Series..... 226
- References 229
- 5 Image Processing..... 231**
 - 5.1 2D-SSA..... 232
 - 5.1.1 Method 232
 - 5.1.2 Elements of 2D-SSA Theory 235
 - 5.1.3 Algorithm 236
 - 5.1.4 2D-SSA in RSSA..... 237
 - 5.2 Shaped 2D-SSA 241
 - 5.2.1 Method 242
 - 5.2.2 Rank of Shaped Arrays 245
 - 5.2.3 Algorithm 246
 - 5.2.4 Shaped 2D-SSA in RSSA..... 247
 - 5.2.5 Comments on nD Extensions 252
 - 5.3 Shaped 2D ESPRIT 253
 - 5.3.1 Method 253
 - 5.3.2 Theory: Conditions of the Algorithm Correctness 255
 - 5.3.3 Algorithm 256
 - 5.3.4 2D-ESPRIT in RSSA 257
 - 5.4 Case Studies 260
 - 5.4.1 Extraction of Texture from Non-Rectangle Images 260
 - 5.4.2 Adaptive Smoothing 261
 - 5.4.3 Analysis of Data Given on a Cylinder 263
 - 5.4.4 Analysis of nD Objects: Decomposition of a Color Image ... 266
 - References 269
- Index..... 271**

Common Symbols and Acronyms

SVD	Singular value decomposition
LRR	Linear recurrence relation
SSA	Singular spectrum analysis
\mathbb{X} or \mathbb{X}_N	Time series or ordered collection of numbers
N	Length or size of \mathbb{X}
L	Window length or size
K	The number of L -lagged vectors obtained from \mathbb{X} ; in the 1D case, $K = N - L + 1$
\mathcal{T}	Embedding operator
$\mathbf{X} = \mathcal{T}(\mathbb{X})$	The trajectory matrix of size $L \times K$ associated with \mathbb{X}
$\ \mathbf{X}\ _F$	Frobenius matrix norm
$\text{rank } \mathbf{X}$	Rank of \mathbf{X}
$\mathcal{X} = \mathcal{X}^{(L)}(\mathbb{X})$	L -trajectory space of \mathbb{X}
$\text{rank}_L(\mathbb{X})$	L -rank of \mathbb{X}
$\Pi_{\mathcal{H}}$	Hankelization operator
λ_i	i th eigenvalue of the matrix $\mathbf{X}\mathbf{X}^T$
U_i	i th eigenvector of the matrix $\mathbf{X}\mathbf{X}^T$
$V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$	i th factor vector of the matrix \mathbf{X}
$(\sqrt{\lambda_i}, U_i, V_i)$	i th eigentriple of the SVD of the matrix \mathbf{X}
\mathbf{I}_M	Identity $M \times M$ matrix
\mathbb{R}^L	Euclidean space of dimension L
$\mathbb{R}^{L \times K}$	Space of $L \times K$ matrices
$\mathcal{M}_{L,K}^{(H)}$	Set of trajectory matrices
$\text{span}(P_1, \dots, P_n)$	Linear subspace spanned by vectors P_1, \dots, P_n
\mathbf{A}^\dagger	Pseudo-inverse of \mathbf{A}