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

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 SSA[1] 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 RSSA[2] 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

[1]Elsner 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.

[2]Korobeynikov 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	Nina Golyandina
Saint Petersburg, Russia	Anton Korobeynikov
Cardiff, UK	Anatoly Zhigljavsky

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

Contents

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 $\mathbb{X}$
$\Vert\mathbf{X}\Vert_{\mathrm{F}}$	Frobenius matrix norm
rank $\mathbf{X}$	Rank of $\mathbf{X}$
$\mathcal{X} = \mathcal{X}^{(L)}(\mathbb{X})$	L-trajectory space of $\mathbb{X}$
$\mathrm{rank}_L(\mathbb{X})$	L-rank of $\mathbb{X}$
$\Pi_{\mathcal{H}}$	Hankelization operator
λ_i	ith eigenvalue of the matrix $\mathbf{X}\mathbf{X}^{\mathrm{T}}$
U_i	ith eigenvector of the matrix $\mathbf{X}\mathbf{X}^{\mathrm{T}}$
$V_i = \mathbf{X}^{\mathrm{T}} U_i / \sqrt{\lambda_i}$	ith factor vector of the matrix $\mathbf{X}$
$(\sqrt{\lambda_i}, U_i, V_i)$	ith eigentriple of the SVD of the matrix $\mathbf{X}$
$\mathbf{I}_M$	Identity $M \times M$ matrix
R^L	Euclidean space of dimension L
$\mathsf{R}^{L \times K}$	Space of $L \times K$ matrices
$\mathcal{M}^{(\mathrm{H})}_{L,K}$	Set of trajectory matrices
$\mathrm{span}(P_1, \ldots, P_n)$	Linear subspace spanned by vectors $P_1, \ldots, P_n$
$\mathbf{A}^{\dagger}$	Pseudo-inverse of $\mathbf{A}$