An Introduction to Singular Spectrum Analysis with R

Dr. Hossein Hassani† and Dr. Mansi Ghodsi‡
†Senior Lecturer in Statistics and Econometrics
‡Lecturer in Statistics and Econometrics
The Statistical Research Centre
Bournemouth University, United Kingdom
April 6, 2014

1 Introduction

According to Google Scholar, the last decade has witnessed an exponential increase in the application of Singular Spectrum Analysis (SSA) for forecasting in various fields ranging from meteorology, biomedical science, and finance, to economics. SSA is a nonparametric time series analysis and forecasting technique which seeks to decompose a time series, filter the noise and reconstruct a less noisy series which can then be used to obtain forecasts. Moreover, a paper relating to SSA by Hassani, Heravi, and Zhigljavsky (2009)¹ has been recognized by the International Journal of Forecasting as one of the top cited papers, and further portrays the emergence of SSA as an important tool in the field of time series analysis and forecasting.

The aim of this three-hour workshop is to provide an introduction to SSA with R. The newly developed R code for SSA will be introduced step by step with examples which will help you to attain a sound understanding of the overall process underlying SSA, whilst learning to exploit the code correctly. You can bring your personal computer’s along with R installed, and you are most welcome to bring your own data sets for this workshop which will enable you to run the code on your own data during the session and get tips and advice on understanding any related complexities.

2 Motivation

In the real world we are faced with data that is affected increasing noise levels distorting the signal and intensifying the task of obtaining accurate forecasts. Shown in the following Figures are few examples of real world data which can benefit through the application of SSA in terms of extracting the noise, signal and harmonic components relating to these time series.

2.1 Economic Data

Here I present some output from a paper entitled Silva, Hassani and Webster (2014) that will be presented at ISF 2014. This application is related to the use of SSA for analysing and

forecasting U.S. Tourist arrivals. Figure 1 below shows the actual data and the extraction of trend and seasonal components for the in-sample data using SSA. Figure 2 presents the out-of-sample SSA forecasts for $h = 24$ months ahead in comparison to a variety of other forecasting techniques.

Figure 1: SSA extraction of trend and seasonal components in U.S. Tourist arrivals.

Figure 2: Out-of-sample forecast of monthly U.S. Tourist arrivals at 24 steps ahead.
2.2 Biomedical Data

In this section I present some recent work by Ghodsi and Hassani (2014) to be presented at ISF 2014 which is related to Biomedical Data. The objective is to use SSA for pattern recognition of gene expression. Figure 3 shows an image of a *Drosophila* who’s egg is analyzed using SSA on the left, SSA trend extraction when applied to the biomedical data. Figure 4 shows an illustration of the temporomandibular joint and its location\(^2\) on the right, and the extracted noise series for an individual with the disease on the left.

Figure 3: LEFT: An electron micrograph of a *Drosophila* adults head and wings. RIGHT: Temporal dynamics of bcd expression.

Figure 4: LEFT: An illustration of the temporomandibular joint and its location. RIGHT: Noise series for an individual with TMD.

2.3 Environmental Data

The application of SSA for analysing and forecasting rainfall in the French Guiana (shown in Figure 5) has yielded the following decompositions (Figure 6) which allows us to obtain an accurate reconstruction with filtered noise for forecasting future data points, and also to attain a better understanding of the seasonal rainfall patterns in French Guiana.

Figure 5: Actual monthly rainfall in French Guiana.

Figure 6: SSA decompositions for monthly rainfall in French Guiana.
Figure 7 below provides an application of Multivariate SSA (MSSA) in the field of Oceanic Climate Change and Sea Level, for the purpose of evaluating causality and forecasting\(^3\). In brief, the MSSA technique is much more flexible than the standard methods of modelling that involve at least one of the restrictive assumptions of linearity, normality and stationarity\(^4\). Whilst SSA is extracts the signal leaving the residual in a univariate series, MSSA extracts groups corresponding to the components of the signal from multiple series. A main advantages of MSSA is that it can consider time series with different lengths and different frequencies, thereby enabling one to model more complex problems in an appropriate form.

Figure 7: Merged MSLA rms (in cm) in the Mediterranean Sea over all the period 1993-2009.

\(^3\)http://sealevel.jpl.nasa.gov/science/ostsciencteam/scientistlinks/scientificinvestigations2013/kahlouche/