

Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN  
CHIMICA

Ciclo 34

**Settore Concorsuale:** 03/A2 - MODELLI E METODOLOGIE PER LE SCIENZE CHIMICHE

**Settore Scientifico Disciplinare:** CHIM/02 - CHIMICA FISICA

NOISES: A NUISANCE AND A RESOURCE. DEVELOPMENT OF NEW  
DECOMPOSITION METHODS FOR NOISY DATA

**Presentata da:** Jacopo Fadanni

**Coordinatore Dottorato**

Luca Prodi

**Supervisore**

Francesco Zerbetto

**Co-supervisore**

Fabrizia Negri

**Esame finale anno 2022**

*A Katia,  
per l'instancabile supporto*

# Abstract

Noise is constant presence in measurements. Its origin is related to the microscopic properties of matter. Since the seminal work of Brown in 1828, the study of stochastic processes has gained an increasing interest with the development of new mathematical and analytical tools. In the last decades, the central role that noise plays in chemical and physiological processes has become recognized. The dual role of noise as nuisance/resource pushes towards the development of new decomposition techniques that divide a signal into its deterministic and stochastic components.

In this thesis I show how methods based on Singular Spectrum Analysis have the right properties to fulfil the previously mentioned requirement. During my work I applied SSA to different signals of interest in chemistry: I developed a novel iterative procedure for the denoising of powder X-ray diffractograms; I “denoised” bi-dimensional images from experiments of electrochemiluminescence imaging of micro-beads obtaining new insight on ECL mechanism. I also used Principal Component Analysis to investigate the relationship between brain electrophysiological signals and voice emission.



# Acknowledgements

This Ph.D. project has been supervised by Prof. Francesco Zerbetto. I am grateful to him for his steady scientific support and for the chance he gave me to work in his computational group. I would also thank all my colleagues in the lab for their endless support and for the profitable discussions. Finally, I would also like to thank professor Ralf Metzler and his group in Potsdam. The period in that lab gave me renewed energy to continue my work.



# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgement</b>	<b>iii</b>
<b>Introduction</b>	<b>1</b>
Bibliography . . . . .	5
<b>I Methods</b>	<b>9</b>
<b>1 Stochastic Time Series</b>	<b>11</b>
1.1 Preliminary Analysis . . . . .	11
1.1.1 Descriptive Statistics . . . . .	12
1.1.2 Fourier Analysis . . . . .	13
1.1.3 Autocorrelation Analysis . . . . .	14
1.2 Fractal Analysis . . . . .	15
1.2.1 Fractals . . . . .	15
1.2.2 Fractal dimensions . . . . .	16
1.2.3 Dispersional Analysis . . . . .	18
1.2.4 Rescaled Range Analysis . . . . .	19
1.2.5 Hurst exponent . . . . .	22
1.2.6 Detrended fluctuation analysis . . . . .	22
1.2.7 Multifractal Detrended fluctuation analysis . . . . .	24
1.2.8 Generalized Moments Method . . . . .	26
Bibliography . . . . .	29
<b>2 Coloured Noises and Fractional Noises</b>	<b>31</b>
2.1 White Noises . . . . .	31
2.2 Correlated Noises . . . . .	32
2.2.1 Red Noise . . . . .	32
2.2.2 Pink Noise . . . . .	33

2.3	Anti-correlated Noises . . . . .	33
2.3.1	Blue Noise . . . . .	33
2.3.2	Purple Noise . . . . .	33
2.4	Fractional Noises . . . . .	34
2.5	Langevin Equation and Coloured Noises . . . . .	35
2.6	Applications of Coloured Noises . . . . .	36
2.6.1	Noises in Molecular Dynamics . . . . .	37
2.6.2	Noises in Biology . . . . .	37
2.6.3	Noise in Physiology . . . . .	38
2.6.4	Noise in Electronics . . . . .	38
2.7	Synthesis of Coloured noises . . . . .	39
2.7.1	Spectral Processing . . . . .	39
2.7.2	Fractional Calculus . . . . .	39
	Bibliography . . . . .	41
<b>3</b>	<b>Singular Spectrum Analysis</b>	<b>45</b>
3.1	1D SSA . . . . .	46
3.1.1	Method . . . . .	46
3.1.2	Window length and decomposition . . . . .	49
3.1.3	Separability . . . . .	49
3.2	2D SSA . . . . .	50
3.2.1	Method . . . . .	51
3.3	Circulant SSA . . . . .	53
3.4	New SSA methods . . . . .	55
3.4.1	Windowed (Overlap) SSA . . . . .	55
3.4.2	Iterative Trend Extraction SSA (ITE-SSA) . . . . .	56
	Bibliography . . . . .	59
<b>4</b>	<b>Principal Component Analysis</b>	<b>61</b>
4.1	Basic PCA . . . . .	62
4.1.1	History . . . . .	62
4.1.2	Method . . . . .	62
4.1.3	PCA Reconstruction . . . . .	65
4.1.4	Implementation . . . . .	65
	Bibliography . . . . .	67
<b>II</b>	<b>Applications</b>	<b>69</b>
<b>5</b>	<b>SSA and XRD</b>	<b>71</b>
5.1	Powder X-ray diffraction and Structure Determination . . . . .	71

5.1.1	Diffraction . . . . .	71
5.1.2	X-ray Sources . . . . .	72
5.1.3	Rietveld method . . . . .	73
5.2	Method . . . . .	73
5.3	Results and Discussion . . . . .	75
5.3.1	First Set . . . . .	75
5.3.2	Second Set . . . . .	77
5.3.3	Third Set . . . . .	83
5.4	Conclusions . . . . .	86
	Bibliography . . . . .	87
<b>6</b>	<b>SSA and ECL Imaging</b>	<b>89</b>
6.1	ECL . . . . .	90
6.2	Method . . . . .	91
6.2.1	ECL Images . . . . .	91
6.2.2	Images Elaboration . . . . .	91
6.3	Results and Discussion . . . . .	92
6.3.1	Long integration time . . . . .	92
6.3.2	Transient ECL emission . . . . .	96
6.4	Conclusions . . . . .	104
	Bibliography . . . . .	105
<b>7</b>	<b>Electrocorticogram Analysis</b>	<b>107</b>
7.1	State of the art . . . . .	107
7.1.1	Speech Brain Computer Interface . . . . .	109
7.2	Project description . . . . .	109
7.3	Results . . . . .	110
7.4	Conclusions . . . . .	116
	Bibliography . . . . .	118
	<b>General Concluding Remarks</b>	<b>119</b>
	<b>List of Figures</b>	<b>121</b>
	<b>List of Tables</b>	<b>123</b>



# Introduction

Observing any measurement it is evident the presence of noise. Thermal fluctuations, fluctuations in the electrical current, effects of the electronic circuits, are some of the sources of noise in a measurement. Looking at noises some questions spontaneously arise. How to remove the noise from the data? What are its properties? Is it only a nuisance or does it contain information on the observed system? Or, better, are fluctuations only due to unwanted noises or are they intrinsic to the system?

In this thesis I tried to address the topics touched in the previous questions.

According to the Encyclopaedia Britannica noise is “any undesired sound, either one that is intrinsically objectionable or one that interferes with other sounds that are being listened to. In information theory, noise refers to those random, unpredictable, and undesirable signals, or changes in signals, that mask the desired information content”.<sup>1</sup> Also in this definition, noise is only a nuisance, something that interferes with the process of interest. One of the properties of noise that causes its bad reputation is randomness: noise values are unpredictable, it is impossible to write down an equation to track its behaviour in time. This contrasts with the predictability of deterministic processes whose values can be inferred solving equations.

Another name, less prone to prejudices, for noises is stochastic processes. A stochastic process refers to a family of random variables indexed against some other variable or set of variables.

One of the first observation of stochastic process has been done by Robert Brown in 1828.<sup>2</sup> The Scottish botanist in 1828 published a paper on his observation “*on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies*”.<sup>2</sup> In this work he describes the irregular motion of small particles on the surface of a water droplet. It is only after Einstein 1905 work on the theory of diffusion that the motion of those “molecules” was theorized.<sup>3</sup> Since Einstein, the study of stochastic processes gained new interest. It is applying Einstein’s results that Jean Baptiste Perrin, A. Westgren and Eugen Kappler experimentally

measured the Avogadro number with increasing accuracy.<sup>4-6</sup> The random walk, experimentally observable, represents a link between the microscopic dynamics of atoms and molecules bombarding a larger particle in suspension, and macroscopic observable like the Avogadro number and the diffusion coefficient.<sup>5</sup>

We can deduce that natural processes are intrinsically noisy with fluctuations that arise as a consequence of the complex system of interaction that is present in Nature. Even simple laws, like Maxwell's equations for electromagnetism or Schrödinger's equation for quantum mechanics, when applied billions of billions of times, produce an extremely complex system.<sup>7</sup> It is this complexity that is observed during measurement.

If nature is intrinsically noisy, then we can consider the noise in measurements as a fingerprint of the observed processes. In the last decades an increasing interest has been attracted by the constructive role of noise, as for example in stochastic resonance or Brownian ratchet,<sup>8</sup> in brain activity,<sup>9</sup> in ecology<sup>10</sup> or in reaction simulations.<sup>11</sup>

The recent interest on noises has been possible thanks to the work of many scientists from different fields of research that developed analysis methods for stochastic time series. A pioneer in the study of stochastic time series was been the hydrologist Harold Edwin Hurst. While making measurements for the building of the Aswan dam on the Nile river, he observed that records of flows or levels at the Roda gauge, near Cairo, were not random, but showed a certain correlation, with series of low-flow years and high-flow years. The "memory" or correlation between successive years created a serious problem: the dam would need to be much larger than it would if annual rainfalls and river flows were random. Observing other sets of data, he developed an approach to examining accumulations, or integrals, of natural stochastic time series.<sup>12</sup>

Another ground breaking scientist has been Benoit Mandelbrot, a Polish-born French American mathematician that developed the concept of fractal. In his book "*The Fractal Geometry of Nature*", published in its final version in 1983, he made a manifesto of the study of fractals.<sup>13</sup> Starting from the observation that "*Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line.*",<sup>13</sup> he developed a new field of math: *fractal analysis*. Using the concepts of self-similarity and scale-invariance, Mandelbrot developed a set of tools for the analysis of time series.<sup>13,14</sup> According to Bassingthwaite, for a statistical insight on a time series it is necessary to perform three analysis: autocorrelation analysis, spectral analysis and the Hurst exponent.<sup>15</sup> All this descriptors provide a brief idea on the properties of the given time series. In Chapter 1 I present a list of methods and tools for the analysis of a stochastic

time series.

The different analysis tools give information on the different properties of a time series. Looking at their properties, time series can be classified as correlated or anticorrelated, if the autocorrelation function is considered; the power exponent gives the color of noises; from the generalization of Brownian motion, it is possible to obtain Fractional Brownian motion and Fractional Gaussian noises.<sup>14</sup> In Chapter 2 I give a short introduction on colored noises and fractional noises.

The proposed analysis methods have been successfully applied in different fields of chemistry. Bakalis *et al.*<sup>16</sup> studied the diffusion of lipid in a membrane using molecular dynamics simulations. They observed a sub-diffusional motion in the case of a heterogeneous bi-layer. Liu *et al.*<sup>17</sup> combined multifractal detrended fluctuation analysis and neural networks into a novel technique for the determination of oleic acid content in rapeseed. Another application field of tools for the analysis of stochastic time series is in the study of biological system where it was observed an important contribution of  $1/f$  noises in membrane processes.<sup>18,19</sup> Also physiological data show stochastic behaviour as observed in Bakalis *et al.*<sup>20</sup> during the study of eye movement.

All the previous tools and classifications consider a time series as it contains mostly information on the system of interest and it is not contaminated by noise from other sources, like the environment or the instrument. When the contribution of external noise sources is relevant, the performances of the analysis tools drop producing biased results.

Typical denoising methods are mainly focused on the complete removal of all the stochastic components in a signals. In particular, they are based on frequency filters and Fourier transform.<sup>21,22</sup> Removing all the stochastic components from a signal precludes the possibility to extract information not only from the deterministic components but also from the noisy ones.

If we are interested in the study of the different components of a signal, instead of using filters, it is useful the use of signal decomposition techniques that separate a signal into a set of constituting components according to a given rule. Among them a remarkable role is played by Empirical Mode Decomposition (EMD) and Singular Spectrum Analysis (SSA).<sup>23</sup>

EMD decomposes a signal into a set of intrinsic mode functions based on the local characteristic time scale of the data.<sup>24</sup> Recently EMD was used for baseline correction for signals of analytical instruments improving the quality of signals from mass spectrometer or gas chromatography.<sup>25</sup>

SSA is a technique that decomposes a given signal into a set of interpretable components such as a trend, oscillatory components and noise.<sup>26</sup> SSA decomposition is based on the Karhunen–Loeve transformation that

decompose a given signal into a set of orthogonal components.

Restricted to the chemistry field SSA has not found large application. Among the applications, SSA was used to remove the solvent artefact in multidimensional NMR.<sup>27</sup> NMR spectra of the solvent distort the weak signal of the protein. Using SSA the effect of solvent was removed from the acquired data unveiling the weak signals. In this way NMR performances were improved. Habasaki applied SSA to the analysis of molecular dynamics of ionic liquids.<sup>28,29</sup> With SSA he was able to distinguish the different time scales of the ion diffusion improving the phase-space plot for glass transition.

Chapter 3 of this thesis is fully devoted to Singular Spectrum Analysis. SSA has been the method of choice for the projects presented in this thesis thanks to its flexibility. Being non-parametric and dividing a given signal into a set of additive components makes SSA a method of choice for automated signal denoising, giving the possibility to analyse the different components individually, obtaining information on all the processes involved in the measurement. In this thesis, SSA has been applied in denoising methods for powder X-ray diffractograms and ECL imaging.

Chapter 4 introduces Principal Components Analysis (PCA), a powerful technique for the reduction of dimensionality of large dataset with minimal loss of information<sup>30,31</sup> If a multivariate dataset is considered, it should be necessary to re-express the data in a more concise way making the interpretation of the analysis results easier. PCA express the dataset using a new set of orthogonal components obtained from the diagonalization of the covariance matrix of the dataset. Using only the components that weigh more in terms of eigenvalues, it is possible to reduce the dimensionality of the dataset preserving its information content.<sup>30</sup> PCA is widely applied in chemistry related fields whenever it is necessary to find the most important observable in a multivariate dataset. In a high cited review, Bro and Smilde<sup>32</sup> showed PCA applied to the study of wine properties. In chromatography, PCA was successfully applied in order to make a comparison between measured and calculated data for background extraction and for the reduction of the number of chromatographic systems, or analytes, to the minimum necessary for the solution of both practical and theoretical problems in chromatography.<sup>33</sup> PCA plays an important role also in organic chemistry where it was used in the selection of the best combination of ligand and solvent in challenging catalytic reactions.<sup>34</sup> In this thesis, PCA has been applied to the study of electrophysiological data as a tool to find a correlation between speech and neural signals.

# Bibliography

- [1] *Noise (acoustic)*, <https://www.britannica.com/science/noise-acoustics>.
- [2] R. Brown, *The Philosophical Magazine*, 1828, **4**, 161–173.
- [3] A. Einstein, *Annalen der Physik*, 1905, **17**, 549–560.
- [4] J. Perrin, *Ann. Chim. Phys*, 1909, **18**, 967.
- [5] R. Metzler and J. Klafter, *Physics Report*, 2000, **339**, 1–77.
- [6] R. Metzler and J. Klafter, *Journal of Physics A: Mathematical and General*, 2004, **37**, R161–R208.
- [7] N. Goldenfeld and L. P. Kadanoff, *Science*, 1999, **284**, 87–89.
- [8] P. V. McClintock, *Nature*, 1999, **401**, 23–25.
- [9] B. J. He, *Trends in Cognitive Sciences*, 2014, **18**, 480–487.
- [10] A. Fiasconaro, D. Valenti and B. Spagnolo, *Mathematical Biosciences and Engineering*, 2004, **1**, 185–211.
- [11] F. Lugli and F. Zerbetto, *Journal of Physical Chemistry C*, 2019, **123**, 825–835.
- [12] H. E. Hurst, *Transactions of the American Society of Civil Engineers*, 1951, **116**, 770–799.
- [13] B. B. Mandelbrot, *The fractal geometry of nature*, W. H. Freeman and Comp., New York, 3rd edn., 1983.
- [14] B. B. Mandelbrot and J. W. Van Ness, *SIAM Review*, 1968, **10**, 422–437.
- [15] J. B. Bassingthwaite, L. S. Liebovitch and B. J. West, *Fractal Physiology*, Springer New York, New York, NY, 1994.

- [16] E. Bakalis, S. Höfner, A. Venturini and F. Zerbetto, *Journal of Chemical Physics*, 2015, **142**, 26049526.
- [17] F. Liu, F. Wang, G. Liao, X. Lu and J. Yang, *Applied Sciences*, 2021, **11**, 5726.
- [18] E. Yamamoto, T. Akimoto, Y. Hirano, M. Yasui and K. Yasuoka, *Phys. Rev. E*, 2014, **89**, 022718.
- [19] E. Yamamoto, T. Akimoto, M. Yasui and K. Yasuoka, *Scientific Reports*, 2015, **5**, 8876.
- [20] E. Bakalis, H. Fujie, F. Zerbetto and Y. Tanaka, *Physica A: Statistical Mechanics and its Applications*, 2018, **512**, 945–953.
- [21] M. F. Wahab, F. Gritti and T. C. O’Haver, *TrAC Trends in Analytical Chemistry*, 2021, **143**, 116354.
- [22] B. Goyal, A. Dogra, S. Agrawal, B. Sohi and A. Sharma, *Information Fusion*, 2020, **55**, 220–244.
- [23] M. Zimoń, R. Prosser, D. Emerson, M. Borg, D. Bray, L. Grinberg and J. Reese, *Journal of Computational Physics*, 2016, **325**, 380–394.
- [24] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung and H. H. Liu, *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 1998, **454**, 903–995.
- [25] X. Xu, X. Huo, X. Qian, X. Lu, Q. Yu, K. Ni and X. Wang, *Analytica Chimica Acta*, 2021, **1157**, 338386.
- [26] N. Golyandina, V. Nekrutkin and A. Zhigljavsky, *Analysis of Time Series Structure SSA and related Techniques*, CHAPMAN & HALL/CRC, New York, 1st edn., 2001, p. 320.
- [27] W. M. Malloni, S. De Sanctis, A. M. Tomé, E. W. Lang, C. E. Munte, K. P. Neidig and H. R. Kalbitzer, *Journal of Biomolecular NMR*, 2010, **47**, 101–111.
- [28] J. Habasaki, K. L. Ngai and Y. Hiwatari, *The Journal of Chemical Physics*, 2005, **122**, 054507.
- [29] J. Habasaki and K. L. Ngai, *The Journal of Chemical Physics*, 2008, **129**, 194501.

- [30] I. T. Jolliffe, *Principal Component Analysis*, Springer New York, New York, NY, 2002.
- [31] I. T. Jolliffe and J. Cadima, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2016, **374**, 20150202.
- [32] R. Bro and A. K. Smilde, *Analytical Methods*, 2014, **6**, 2812–2831.
- [33] T. Cserháti, *Biomedical Chromatography*, 2010, **24**, 20–28.
- [34] J. D. Moseley and P. M. Murray, *Journal of Chemical Technology & Biotechnology*, 2014, **89**, 623–632.



# Part I

## Methods



# Chapter 1

## Stochastic Time Series Analysis

### Introduction

Stochastic fluctuations are driving processes in Nature. In order to understand Nature in terms of stochastic processes, a way to rationalize them is needed. Given a temporal process  $X(t)$ , it is possible to define a series of samples  $x_i$  for  $i = 1, \dots, N$  that are events of the process  $X(t)$ . This series  $X(i)$  is called time series. If the time interval between two consecutive samples is constant, then the time series is defined as equidistant with a sampling time  $\Delta t$  and sampling frequency  $f = 1/\Delta t$ .

Given a time series, it can be either deterministic, chaotic or stochastic. A deterministic time series can be expressed explicitly by an analytical expression, its past and future can be obtained given the analytical law. In a chaotic time series, strong nonlinear interactions between few factors shape the fluctuations. In a stochastic time series, a large number of weak influences is responsible for the fluctuations.<sup>1</sup> Looking at a stochastic time series, emerges that the values do not make any sense if taken individually, only looking at the sequence as a whole makes the description of the time series possible.

Many different approaches can be used to obtain insight from stochastic time series. In this chapter I present a workflow for the analysis of a stochastic time series. Each section of this chapter focuses on a step of the analysis

### 1.1 Preliminary Analysis

Given a time series  $X(t)$ , before performing the analysis, is necessary to check for a trend and deterministic components.

The removal of the trend can be done fitting the analytical law of the trend

or using decomposition techniques like SSA (see Chapter 3).

After detrending and removal of the deterministic components, the stochastic time series can be described using a set of parameters that belongs to the field of descriptive statistics. Descriptive statistics treats the elements of the time series as independent and the order is not important. Complementary to the use of descriptive statistics are the methods that consider the relationship between the elements of the time series.

### 1.1.1 Descriptive Statistics

A time series can be described in terms of its location and shape or in terms of its central tendency via mean, median and mode; and its variability using standard deviation (or variance), minimum and maximum values, skewness and kurtosis.

Mean, mode and median give information on the location of the distribution. The mean (eq. 1.1) is the most representative value of the set, mode is the value that appears most often in the set, median is the point that is located halfway between minimum and maximum.

The standard deviation (eq. 1.2) gives information on the amount of dispersion of the dataset.

Skewness (eq. 1.3) and kurtosis (eq. 1.4) give information on the shape of the distribution. Skewness describes the symmetry of the distribution and gives the directions of the tails, with positive skewed data tailed to the left and negative skewed data tailed to the right. Kurtosis describes how many data are far from the mean, it describes how fat are the tails.

$$\mu \equiv E[X] = \frac{1}{N} \sum_{i=1}^N x_i \quad (1.1)$$

$$\sigma \equiv \sqrt{E[(X - \mu)^2]} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu)^2} \quad (1.2)$$

$$\tilde{\mu}_3 = E \left[ \left( \frac{X - \mu}{\sigma} \right)^3 \right] = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^3}{\sigma^3} \quad (1.3)$$

$$\tilde{\mu}_4 = E \left[ \left( \frac{X - \mu}{\sigma} \right)^4 \right] = \frac{\mu_4}{\sigma^4} \quad (1.4)$$

$\tilde{\mu}_n$  standardized moment of order  $n$ ,  $\mu_n$  central moment of order  $n$

### 1.1.2 Fourier Analysis

Given a time series, it is possible to decompose it in a sum of sine and cosine functions. This is done using the Fourier transform (FT). For a continuous signal  $s(t)$  the Fourier transform reads

$$S(f) = \int_{-\infty}^{\infty} s(t) \cdot e^{-i2\pi ft} dt \quad (1.5)$$

with  $S(f)$  that is the transform of  $s(t)$  and is function of frequency. In the case of a discrete series the integral becomes a sum.

$$S_k = \sum_{n=0}^{N-1} s_n \cdot e^{-\frac{i2\pi}{N} kn} \quad (1.6)$$

The Fourier transform is usually performed using an effective computational technique called Fast Fourier Transform (FFT). FFT performs the discrete FT using an efficient algorithm that reduces the computational cost.

Fourier transform gives information on amplitude and phases of the frequencies forming the signal.

$$x_i = \sum_{n=0}^{N/2} A_n \cos[\omega_n t_i + \varphi_n] = \sum_{n=0}^{N/2} A_n \cos \left[ \frac{2\pi n}{N} i + \varphi_n \right] \quad (1.7)$$

were  $A_n$  is the amplitude,  $\varphi_n$  is the phase of the cosine component and  $\omega_n$  the angular frequency ( $f_n = \omega_n/2\pi$ ).  $A_n(f_n)$  and  $\phi_n(f_n)$  are called the amplitude, and phase spectrum of the signal, respectively. The square of the amplitudes  $A_n(f_n)^2$  is called the power spectrum or power spectral density of the signal (PSD). It gives information on the amount of energy contained at the different frequencies.

The form of the power spectral density gives information on the nature of the time series. In particular, if it presents a power law form, then the time series is fractal and can be analyzed using the tools I present in the next sections

$$A_n^2 = p\omega_n^{-\beta} \quad (1.8)$$

with  $\beta$  that is the spectral index.

A power law relationship expresses the idea that doubling the frequency, the power changes by the same fraction ( $2^{-\beta}$ ) regardless of the chosen frequency.<sup>2</sup> Before applying FFT, the signal has to be preprocessed in order to improve the performances of the analysis. The main preprocessing steps are:

1. *Subtraction fo the mean;*

2. *Windowing* with a parabolic window. In this step each value of the time series is multiplied by a parabolic function given by

$$W(i) = 1 - \left( \frac{2j}{N+1} - 1 \right)^2 \quad \text{for } j = 1, \dots, N \quad (1.9)$$

applying this window reduces the effects of discontinuities at both ends of the data series;<sup>3</sup>

3. *End matching*. This step makes the first and last point of the time series to be equal. It works subtracting to the data the line connecting the first and last points of the series.

The preprocessing is extremely recommended for time series with  $\beta > 2$ .<sup>2</sup>

The fitting of the spectral index,  $\beta$ , is improved if the high-frequency part of the PSD is excluded from the fitting. For a good fitting the range  $1/8 < f < 1/2$  is excluded.

## Coloured Noises

As for light the colour is defined by its spectrum, also for stochastic processes it is possible to classify them according to their power spectrum defining a colour. Coloured noises show a power law scaling in their PSD and have a different color depending on the spectral index. White noises show flat PSD ( $\beta = 0$ ); anticorrelated noises have  $\beta < 0$ ; correlated noises have  $\beta > 0$  A detailed dissertation is contained in Chapter 2

### 1.1.3 Autocorrelation Analysis

The aim of the Autocorrelation Analysis is to determine to what extent the value at a given time  $s_t$  of the time series depends on its past values  $k$  lag apart. The  $k$ -lag autocorrelation coefficient defines the magnitude of the dependence. The  $k$ -lagged autocorrelation may be positive, zero or negative in the range  $[-1; 1]$ . Estimates of the autocorrelation for the lags  $k = 0, 1, \dots, N$  can be calculated as:

$$\hat{\rho}_k = \frac{\frac{1}{N-k-1} \sum_{i=k+1}^N (s_i - \mu)(s_{i-k} - \mu)}{\frac{1}{N-1} \sum_{i=1}^N (s_i - \mu)^2} \quad (1.10)$$

where  $\mu$  is the mean (eq. 1.1). A positive correlation indicates that  $s_i$  and  $s_{i-k}$  are on the same side of the distribution with respect to the mean while a negative correlation indicate that the two points are on the opposite sides of

the mean, they are anticorrelated. At any time, in a correlated signal a step in one direction is more likely followed by a step in the same direction, in an anticorrelated signal, a step in one direction is more likely followed by a step in the opposite direction. In an uncorrelated signal the steps are completely random and  $\hat{\rho}_k = 0 \forall k \geq 1$ .

## 1.2 Fractal Analysis

### 1.2.1 Fractals, definition and properties

Before introducing the techniques used to characterize a fractal signal we need to define what a fractal is and which are its properties.

*“Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line.”*

With this statement Mandelbrot opens his book on fractals,<sup>4</sup> introducing for the first time the concept of fractals and their properties. The idea of fractals dates back to the late 19th and early 20th century, when mathematicians started generating complex geometrical structures starting from simple objects (*the initiator*) applying a simple rule (*the generator*) an infinite number of iterative steps. In this way they obtained a complex structure that is equally rich in details at every scale of observation.<sup>5</sup> In his **“The fractal geometry of Nature”**,<sup>4</sup> Mandelbrot makes a manifesto of the study of fractals. It is in this book that he invented the word *fractal*, from the Latin word *fractus*, broken, fragmented, used in relation to the irregular shape of fractal objects.

#### Fractals properties

The two main properties that define a fractal are self-similarity and scaling, but other properties are also relevant for fractals.

**Self-similarity** Self-similarity means that fractals are equal to themselves at the different scales: a part of a fractal, if enlarged, is equal to the whole fractal. For mathematical fractals this is strictly true while natural fractal object are statistically self-similar. It means that the statistical properties of the pieces are proportional to the statistical properties of the whole, the pieces are like the whole but not identical.

**Scaling** The value measured for any property of an object depends on the characteristics of the object. If these characteristics depend on the mea-

surement resolution, then the measured value depends on the measurement resolution. There is no one true value for a measurement, the value changes with the scale.

**Scale invariance** Scale-invariance is strictly related to self-similarity, quantitative properties of the structure depends only on the ratio of scales and not on the absolute scale due to the uniform geometry.<sup>1</sup> The ratio of two estimates of  $q$  measures at two different scales,  $s_1$  and  $s_2$ ,  $q_1/q_2$  depends only on the ratio of scales  $s_1/s_2$  and not on the absolute scale. For a natural fractal, scale-invariance holds only for a restricted range of scales. It is possible to define a scaling range as the ratio between the biggest and the smallest scale at which scale-invariance holds.

$$SR = \log_{10}(s_{max}/s_{min}) \quad (1.11)$$

**Power law scaling** Self-similarity has effects on the scaling relationship between a quantitative property,  $q$ , and the scale,  $s$ .

$$q = f(s) \quad (1.12)$$

For a non fractal object,  $q$  converges to a single value when  $s$  becomes infinitely small. For fractals,  $q$  does not converge, but it exhibits a power law scaling relationship with  $s$ .

$$q = ps^\epsilon \quad (1.13)$$

where  $p$  is a proportionality factor and  $\epsilon$  is the scaling exponent ( $\epsilon < 0$ ).  $\epsilon$  can be determined from the log of eq.1.13

A useful relationship arising from eq.1.13 is between power and frequency in power spectral density with power increasing with frequency in anticorrelated time series and decreasing in correlated ones. For this reason it is possible to search for the power law in the power spectral density in order to define if a time series is fractal or not.

## 1.2.2 Fractal dimensions

Given a geometrical object it is possible to define the dimension of its embedding space, the *embedding dimension* or *Euclidean dimension*,  $E$ . The embedding dimension tells if the object is found on a line ( $E = 1$ ), on a plane ( $E = 2$ ) or in a three dimensional space ( $E = 3$ ).  $E$  gives the number of coordinates needed to address a point of the object in space.

The *topological dimension*  $D_T$  indicates the number of coordinates needed to determine the position of a point on the actual geometrical structure.

$D_T \leq E$ . The topological dimension describes the connectedness between the points in the fractal set.

For a classical geometrical object, embedding and topological dimensions define the object. For fractals a new dimension exists: the *fractal dimension*. It has values that usually are not integer and, accordingly to Mandelbrot<sup>4</sup>,  $D_T \leq D \leq E$ .  $D$  describes space-filling properties of the fractal set.  $D$  is related to the scaling exponent of a given attribute of the object. For an exact fractal  $D$  can be expressed as *self-similarity dimension* ( $D_{ss}$ ) and tells how many structural units of the object,  $N$ , are seen at a given resolution,  $R = 1/s$ .

$$D_{ss} = \log N / \log R \quad (1.14)$$

For real, non-exact, fractals it is not possible to determine  $D_{ss}$ . Thus, in order to estimate  $D_{ss}$  for a real fractal, a generalization of the self-similarity dimension is needed. The *capacity dimension*,  $D_{cap}$ , and the *box-counting dimension*,  $D_{box}$ , have been developed. The idea is to cover the fractal with ‘balls’ whose dimension equals  $E$  and then count the number of balls used.

For  $D_{cap}$  the fractal is covered by ‘balls’ whose dimension equals  $E$ . For  $E = 1$  the ball is a segment of length  $2r$ , for  $E = 2$  is a circle of radius  $r$  and for  $E = 3$  is a sphere of radius  $r$ . Given a radius  $r$ , the minimum number of balls of size  $r$ ,  $N(r)$ , is found, then  $r$  is decreased and  $N(r)$  is found again.<sup>6</sup>

$$D_{cap} = \lim_{r \rightarrow 0} \log N(r) / \log(1/r) \quad (1.15)$$

with one ball covering the object at  $r = 1$ .

The *box-counting dimension* uses balls that are contiguous nonoverlapping rectangular boxes.  $N(r)$  is the number of boxes containing at least one point of the object.

Given this definition we can conclude that, in a broad sense, the topological dimension tells us about the type of object the fractal is, and the fractional dimension tells us how wiggly it is.<sup>6</sup>

### Fractal Dimension from Scaling Properties

Self-similarity is a core property of fractals. The power law scaling is a result of it, while the fractal dimension is based on self-similarity. Using this observation we can find a direct way to determine the fractal dimension starting from the power law scaling. Using eq. 1.13 and modifying eq. 1.15 as:

$$N(r) = Br^{-D} \quad (1.16)$$

where  $B$  is a constant, we can determine the fractal dimension  $D$  from the scaling exponent  $\epsilon$  if we know how the measured property  $q$  depends on the number of pieces  $N(r)$ .

### 1.2.3 Dispersional Analysis

Dispersional analysis is one of the statistical approaches used to determine the fractal dimension. It was introduced by Bassingthwaighte in 1988 and it involves the measurement of variance or standard deviation of a signal at a succession of different levels of resolution. The resolution is changed grouping adjacent data points and using the mean to replace them: increasing the size of the groups reduces the resolution. Bassingthwaighte<sup>7</sup> observed that the relationship between the log of the variance, or of the Relative Dispersion ( $SD/\mu$ ), and the log of the size of the observed unit is linear

$$RD(m)/RD(m_0) = (m/m_0)^{1-D} \quad (1.17)$$

and the fractal dimension is

$$D = 1 - \frac{\log[RD(m)/RD(m_0)]}{\log(m/m_0)} \quad (1.18)$$

with  $m$  the element size used to calculate  $RD$  and  $m_0$  the reference size. From this analysis it is possible to obtain the Hurst coefficient  $H = E + 1 - D$  for any euclidean dimension  $E$ .

#### Procedure

The typical sequence used to estimate the Hurst exponent of a time series using dispersional analysis is presented here.

1. Given a time series  $x_i$ ,  $i = 1, \dots, N$ , obtained from a signal sampled at even time intervals  $\Delta t$ , we can define a resolution of the observation  $m$  as the number of sample points grouped in the observation interval. For each resolution  $n = N/m$  intervals are defined;
2. Using the highest resolution, calculate the mean  $\bar{x}$  of the whole time series

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (1.19)$$

3. Calculate the standard deviation of the set at the resolution  $m$

$$SD(m) = \frac{1}{n} \sqrt{n \sum x_i^2 - \left( \sum x_i \right)^2} \quad (1.20)$$

4. Calculate the relative deviation as

$$RD(m) = SD(m)/\bar{x} \quad (1.21)$$

5. Increase  $m$  decreasing the resolution. Define a new time series grouping adjacent points according to  $m$  and substituting the values with the mean on the interval

$$y_{i,m} = \frac{1}{m} \sum_{j=mi}^{mi+m} x_j \quad i = 0, \dots, n \quad (1.22)$$

6. For the new time series calculate standard deviation and relative dispersion;
7. Repeat steps 5 and 6;
8. Plot the logarithm of  $RD$  versus the logarithm of  $m$  and determine the slope of the relationship. Using the fitted value determine the Hurst exponent  $H = \text{slope} + 1$ ;

### Strengths and Limitations

According to an exhaustive study by Bassingthwaighte and Raymond<sup>8</sup>, Dispersional analysis is more robust than Hurst Rescaled Range method. In particular, it performs well also for short time series and the Hurst coefficient suffers of low bias.

A weakness of dispersional analysis is the susceptibility to nonstationarity of the signal, with the trend that will increase the variance at all levels of resolution. This weakness limits dispersional analysis to stationary time series.

#### 1.2.4 Rescaled Range Analysis

Rescaled range analysis was invented by Hurst while studying for the construction of the Aswan dam.<sup>9</sup> In his work, Hurst proposed an empirical descriptor to describe temporal signals of natural phenomena. The idea had been developed in order to determine the minimum height of the proposed Aswan dam that would have allowed a control on the flow of the Nile downstream. The descriptor proposed by Hurst is now known as *Hurst exponent* (or parameter),  $H$ , and the method proposed to determine it is called *Rescaled Range Analysis* (R/S)

Rescaled Range Analysis is an approach to examining accumulations or integrals of naturally fluctuating events. The rescaled range  $R/S$  is the range  $R$  of cumulative deviations from the mean, divided by the standard deviation  $S$ , where  $R/S$  is estimated for subsets of duration  $\tau$  for many different values of  $\tau$ .

Given a time process  $X(t)$ , for each interval, the time series of cumulative deviations from the mean  $V(u)$  is defined as

$$V(u) = \int_0^u (X(\lambda) - \bar{X}(\tau)) d\lambda; \quad 0 < u \leq \tau \quad (1.23)$$

where  $\bar{X}(\tau)$  is the mean value on the interval of duration  $\tau$  and  $\lambda$  is a dummy variable for integration.

The range of  $V(u)$  over the period  $\tau$  is defined

$$R(t, \tau) = V_{max} - V_{min} = \max(V(u)) - \min(V(u)) \quad \text{for } 0 < u \leq \tau \quad (1.24)$$

To normalize the range relative to the fluctuations, the standard deviation of  $X(t)$  over the same period is used

$$S(t, \tau) = \left[ \frac{1}{\tau} \int_0^\tau (X(t) - \bar{X}(\tau))^2 dt \right]^{1/2} \quad (1.25)$$

thus several estimates of the “rescaled range”,  $R/S$  have been obtained for each  $\tau$ .

$$R/S = R(t, \tau)/S(t, \tau) \quad (1.26)$$

Hurst observed that the relationship between  $R/S$  and the length of observation  $\tau$  shows a power law scaling.

$$(R/S)_\tau = p\tau^H \quad (1.27)$$

where  $p$  is a prefactor and  $H$  is the Hurst exponent (see Section 1.2.5).

$H$  is related to the fractal dimension with the relationship

$$H = E + 1 - D \quad (1.28)$$

## Procedure

Here we present rescaled range analysis for a time series  $X_i$  consisting of equidistant samples of the time process  $X(t)$ , sampled at intervals  $\Delta t$ .

1. Define the total duration as  $\tau_{max} = N\Delta t$  and calculate the mean over the whole dataset

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^N x_i \quad (1.29)$$

2. Sum the differences from the mean to get an accumulation time series using the discrete form of equation (1.23)

$$V(n, k) = \sum_{j=0}^k (x_j - \bar{x}_N); \quad 0 < k \leq n \quad (1.30)$$

3. Find the maximum,  $V_{max}$ , and the minimum,  $V_{min}$  of  $V(N, k)$  for  $0 < k \leq n$  and calculate the range

$$R(n) = V_{max} - V_{min} \quad (1.31)$$

4. Calculate the standard deviation of the observation over the period  $\tau = n \Delta t$

$$S(n) = \left[ \frac{1}{n} \sum_{i=0}^n (x_i - \bar{x}_N)^2 \right]^{1/2} \quad (1.32)$$

where  $\bar{x}_N$  is the mean over the whole time series.

5. Calculate  $R/S = R(n)/S(n)$
6. Decrease the dimension of the intervals,  $n$ , from  $N$  to  $n_{min}$ , and repeat the entire procedure.
7. Plot the logarithm of  $R/S$  versus the logarithm of  $n$  and obtain the Hurst exponent as the slope of the regression line.

$$H = \frac{\log(R/S)_n - \log p}{\log n} \quad (1.33)$$

For each  $n$ ,  $N/n$  estimates of  $R/S$  are produced. In order to fit the slope, two different procedure have been proposed: i) averaging the values of  $R/S$  for each  $n$ , ii) using 10 values of  $R/S$  for each  $n$  without averaging. The second method was proposed by Mandelbrot and avoids overweighting the values of  $R/S$  for small  $n$ .<sup>10</sup>

### Strengths and Limitations

Bassingthwaighte and Raymond<sup>11</sup> in 1994 evaluated rescaled range method. From this study emerges that the linear regression on the logarithms produces an estimate of  $H$  that is indistinguishable from that obtained using the nonlinear optimization (fitting directly the exponential law). For  $H < 0.5$  the method overestimates  $H$  for all lengths. For  $H > 0.5$  a systematic error

in the results is present. A suggested correction for this error is to remove the trend of the signal. However, the method gives biased estimates converging to the true value for a signal of infinite length. Series with positive correlation leads to better estimates of  $H$  using  $R/S$ . For series with  $H < 0.5$  alternative methods should be used instead.

### 1.2.5 Hurst exponent

Introduced by Hurst in 1951, the Hurst exponent,  $H$ , is a parameter that gives information on the *smoothness* of a fractal object. In other terms it is a measure of the correlation between elements in a time series.  $H$  is related to the fractal dimension and Euclidean dimension with the following relationship

$$H = E + 1 - D \quad (1.34)$$

in this way  $0 \leq H \leq 1$ .

A signal with  $H$  close to 0 is one with rough structuring, frequently changing sign with respect to the mean; a signal with  $H$  close to 1 is a smooth one, with gentle changes of sign. In the case of  $H = 0.5$  the signal is completely random; for  $H \neq 0.5$  the signal is not completely random, for  $H < 0.5$  the signal is anticorrelated while for  $H > 0.5$  the signal is correlated.

Two equivalent ways of defining  $H$  are common in literature. The two definitions are based on two different stochastic processes. One is based on fractional Gaussian noise ( $fGn$ ), a stationary random process that shows a correlation given by eq. (1.35) that represent a process that is second-order stationary.<sup>12</sup> The other is based on fractional Brownian motion ( $fBm$ ), a non-stationary self-similar process with interdependent increments.<sup>13</sup> The two processes are strictly related, a  $fGn$  is often defined as the differentiation of a  $fBm$ .<sup>12,13</sup>

$$\rho_n = 0.5 \{ |n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H} \} \quad (1.35)$$

$H$  is also related to the spectral index  $\beta$  with a relationship depending on class of the process. For a Fractional Gaussian noise  $H = \frac{\beta+1}{2}$ , for a Fractional Brownian motion  $H = \frac{\beta-1}{2}$  In Chapter 2 I will give a more detailed view on  $fBm$  and  $fGn$ .

### 1.2.6 Detrended Fluctuation Analysis (DFA)

Detrended fluctuation analysis (DFA) was introduced by Peng *et al.* in 1994 with the aim to detect long-range correlations embedded in a patchy landscape.<sup>14</sup> After that publication, DFA became a widely used method for the

analysis of noisy data in the presence of trends of unknown origin.<sup>15</sup> Nonstationarity, due to a trend, introduces artifacts in long-range correlation making difficult to extract a scaling property. Generally, an artificial crossover between scaling regimes appears in the presence of a trend.<sup>15,16</sup> Detrending fluctuation analysis permits the detection of long-range correlation in nonstationary time series removing a “local trend” that is responsible for an excess of short-range correlation.

The DFA method applies as follow:

Given a sequence of length  $N$  and elements  $x_i$ ,  $i = 1, \dots, N$ :

1. Sum the signal and subtract the mean

$$y_j = \sum_{i=1}^j x_i - \hat{\mu} \quad (1.36)$$

The subtraction of the mean is not compulsory, it would be eliminated by later detrending.

2. Divide the summed signal  $y$  into  $N_l = N/l$  nonoverlapping boxes of size  $l$ , and define the “local trend”,  $y_{j,l}$ , as the least-squares fit for the signal in that box. If the signal length,  $N$ , is not a multiple of the considered box size, a short part of the signal will remain out of the boxes. In order to not disregard this part of the record, the same procedure is repeated starting from the other end of the signal. In this way  $2N_l$  segments are obtained;
3. Define the detrended signal as the difference between the original signal and the local trend. Calculate the variance about the detrended signal for each box, and calculate the average of these variances over all the boxes of size  $l$ ,  $F_l$ ;

$$F_l = \sqrt{\frac{1}{N} \sum_{j=1}^N (y_j - y_{j,l})^2} \quad (1.37)$$

$F_l$  is the root mean square fluctuation

4. Repeat steps 2 and 3 for different sizes  $l$ .

The scaling properties of the fluctuations arises between  $l$  and the variances.

$$F_l \sim l^\alpha \quad (1.38)$$

where  $\alpha$  is the scaling exponent.

The function used for detrending can be a polynomial of different order,  $l$ , and the method names DFA- $l$  (DFA-1 if  $l = 1$ , DFA-2 if  $l = 2$ ).<sup>15,16</sup> Different polynomial degrees differ in the capability of eliminating trends in the data. In  $n$ -th order DFA, trends of order  $n$  in the summed signal (of order  $n - 1$  in the original data) are removed. Thus, a comparison of results for DFA- $l$  of different order allows to estimate the strength of the trends in the time series.<sup>15</sup> Changing the order of detrending allows for a “correct” identification of the trend for the subsequent analysis.

For a fGn signal,  $y_j$  will be an fBm and  $F_l = pl^H$  with  $\alpha = H$ . For an fBm,  $y_j$  becomes a summed fBm and  $F_l = pl^{H+1}$ , Eke *et al.*<sup>1</sup>

### 1.2.7 Multifractal Detrended Fluctuation Analysis

Many signals do not exhibit a simple monofractal scaling behavior, since they do not have a single scaling exponent. The different scaling properties can be distributed in different parts of the series, requiring a scaling exponent for each of them or, the scaling changes with the scale of the observation. In this case there exist a crossover scale,  $s_\times$ , separating regimes with different scaling exponents.<sup>15-17</sup> The existence of more than one scaling exponent defines the series as multifractal.

In general, it is possible to distinguish two different types of multifractality in time series:

- (i) Multifractality due to a broad probability density function for the values of the time series. This type of multifractality can not be removed by shuffling the time series.
- (ii) Multifractality due to different long-range correlations of the small and large fluctuations. In this case, shuffling removes multifractality with the shuffled series that is nonmultifractal.

If both kinds of multifractality are present, the shuffled series will exhibit a weaker multifractality than the original one.

Given a multifractal time series, many analysis methods are available, examples are reported in Salat *et al.*<sup>18</sup> and Kantelhardt *et al.*<sup>17</sup>. Among them, the method developed by Kantelhardt *et al.*, *Multifractal detrended fluctuation analysis* (MF DFA), is widely used. The method develops as a generalization of the DFA method (Subsec. 1.2.6). MF DFA consists of five steps, the first three of them are identical to those in DFA.

We can rewrite step 3 in DFA as follow:

- determine the variance

$$F^2(i, l) = \frac{1}{l} \sum_{j=1}^l \{y[(i-1)l + j] - p_i(j)\}^2 \quad (1.39)$$

for each segment  $i$ ,  $i = 1, \dots, N_l$

- determine the variance

$$F^2(i, l) = \frac{1}{l} \sum_{j=1}^l \{y[N - (i - N_l)l + j] - p_i(j)\}^2 \quad (1.40)$$

for segments  $i = N_l + 1, \dots, 2N_l$ .

$p_i(j)$  is the fitting polynomial in segment  $i$ . Equation (1.40) is necessary in order to consider the fact that a short part of the signal will remain out of the boxes if the signal length,  $N$ , is not a multiple of the box size  $l$ .

After the first three steps MDFDA has the following steps:

1. Average over all segments to obtain the fluctuation function of order  $q$

$$F_q(l) = \left\{ \frac{1}{2} N_l \sum_{i=1}^{2N_l} [F^2(i, l)]^{q/2} \right\}^{1/q} \quad (1.41)$$

where the order  $q$  can take any real value. For  $q = 0$  eq. (1.41) becomes

$$F_0(l) = \exp \left\{ \frac{1}{4N_l} \sum_{i=1}^{2N_l} \ln [F^2(i, l)] \right\} \sim l^{h(0)} \quad (1.42)$$

2. Determine the scaling behavior of the fluctuation functions for each value of  $q$  analyzing log-log plots  $f_q(l)$  versus  $l$

$$F_q(l) \sim l^{h(q)} \quad (1.43)$$

where  $h(q)$  is a generalized Hurst exponent, for  $q = 2$  the classical Hurst exponent is retrieved.

For very large scales,  $l > N/4$ , the number of segments becomes very small making  $F_q(l)$  statistically unreliable. Usually, only scales  $l < N/4$  are used.

Another range with  $f_q(l)$  that deviates from eq. (1.43) is for  $l < 10$ . For this region Kantelhardt *et al.*<sup>17</sup> proposed a correction: instead of using  $f_q(l)$  they proposed to use  $F_q(l)/F_q^{shuf}(l)$  where  $F_q^{shuf}(l)$  is the fluctuation function for the shuffled time series. In order to obtain a more robust estimation,

$F_q^{shuf}(l)$  can be obtained averaging over a large number of randomly shuffled time series.<sup>17</sup>

For monofractal time series  $h(q)$  is independent on  $q$  and is equal to  $H$ . Only when small and large fluctuations scale differently there is a dependence of  $h(q)$  on  $q$ . The presence of different scaling exponents defines multifractality.

## 1.2.8 Generalized Moments Method

Another method for the analysis of fractal time series is the Generalized Moments Method (GMM). GMM is one of the most robust methods for the analysis of nonstationary time series.<sup>19</sup> It uses the scaling of statistical moments of various orders and gives information on whether the time series is monofractal or multifractal.

### Procedure

Given a time series  $x_i$  of length  $N$  sampled at time interval  $\tau = \tau_{min}$ , its length in time is  $T = N\tau$ .

1. Given  $x_i$  define a lag time  $\Delta = n\tau$  and compute the time series of the increments for the given lag time as

$$y_i(\Delta) = |x_{i+\Delta} - x_i| \quad (1.44)$$

with  $i = 1, \dots, (T - \Delta)\tau$

2. The statistical moments of  $y_\Delta$  are estimated as

$$K_q(\Delta) = \frac{1}{T - \Delta} \sum_{i=1}^{T-\Delta} y_i(\Delta)^q \quad (1.45)$$

where the values of  $q$  are selected in the range  $0 < q \leq 2$  as in Bakalis *et al.*<sup>20</sup>

3. steps 1 and 2 are repeated changing the lag time,  $\Delta$ , from the range  $\Delta = \tau_{min}, \dots, \tau_{max}$ , usually  $\tau_{max} = N/10\tau$

the relationship between  $K_q(\Delta)$  and  $\Delta$  is a power law

$$K_q(\Delta) \approx \Delta^{qH(q)} \quad (1.46)$$

where  $H(q)$  is a generalized form of the well known Hurst exponent<sup>21</sup> and  $qH(q) = z(q)$  is also known as structure function. The shape of  $z(q)$  gives

information on the stochastic mechanism governing the process. If  $z(q)$  is linear with respect to  $q$  then the process is monofractal,  $H(q)$  is constant for all  $q$  values while, if the structure function has a convex shape, the process is multifractal.<sup>22,23</sup> Moments of different order,  $q$ , provide information on different portion of the PDF of the given time series. Small values of  $q$  sample the central part of the distribution while, large values of  $q$  sample the tails.<sup>24</sup>

### Generalized Hurst exponent

The Hurst exponent is related to the scaling properties of a time series, in particular, it is associated with the scaling of the absolute spread in the increments or to the variance. Using moments of different order  $q$ , a set of scaling exponents  $H(q)$  is produced with each  $q$  that is associated with different characterizations of the scaling complexity of the signal.<sup>25</sup> For  $q = 1$ ,  $H(q)$  is the classical Hurst exponent, while for  $q = 2$ ,  $H(q)$  is proportional to the autocorrelation function.<sup>21</sup>  $H(2)$  is also related to the power spectrum  $\beta = 1 + 2H(2)$  with  $\beta$  spectral index. Only for fBm  $H(2) = H$  too, lets remember that, for fBm,  $H = \frac{\beta-1}{2}$ .

In order to use GMM to compute the Hurst exponent, the time series has to be of an fBm process, if not, the accumulation time series can be used.



# Bibliography

- [1] A. Eke, P. Herman, L. Kocsis and L. R. Kozak, *Physiological Measurement*, 2002, **23**, R1–38.
- [2] A. Eke, P. Herman, J. B. Bassingthwaighte, G. M. Raymond, D. B. Percival, M. Cannon, I. Balla and C. Ikrényi, *Pflügers Archiv : European journal of physiology*, 2000, **439**, 403–15.
- [3] P. F. Fougere, *Journal of Geophysical Research: Space Physics*, 1985, **90**, 4355–4366.
- [4] B. B. Mandelbrot, *The fractal geometry of nature*, W. H. Freeman and Comp., New York, 3rd edn., 1983.
- [5] H.-O. Peitgen, H. Jürgens and D. Saupe, *Chaos and Fractals. New Frontiers of Science*, Springer, New York, 2nd edn., 2004.
- [6] J. B. Bassingthwaighte, L. S. Liebovitch and B. J. West, *Fractal Physiology*, Springer New York, New York, NY, 1994.
- [7] J. B. Bassingthwaighte, *Physiology*, 1988, **3**, 5–10.
- [8] J. B. Bassingthwaighte and G. M. Raymond, *Annals of Biomedical Engineering*, 1995, **23**, 491–505.
- [9] H. E. Hurst, *Transactions of the American Society of Civil Engineers*, 1951, **116**, 770–799.
- [10] B. B. Mandelbrot and J. R. Wallis, *Water Resources Research*, 1969, **5**, 242–259.
- [11] J. B. Bassingthwaighte and G. M. Raymond, *Annals of Biomedical Engineering*, 1994, **22**, 432–444.
- [12] M. J. Cannon, D. B. Percival, D. C. Caccia, G. M. Raymond and J. B. Bassingthwaighte, *Physica A: Statistical Mechanics and its Applications*, 1997, **241**, 606–626.

- [13] B. B. Mandelbrot and J. W. Van Ness, *SIAM Review*, 1968, **10**, 422–437.
- [14] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley and A. L. Goldberger, *Physical Review E*, 1994, **49**, 1685–1689.
- [15] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. Rego, S. Havlin and A. Bunde, *Physica A: Statistical Mechanics and its Applications*, 2001, **295**, 441–454.
- [16] K. Hu, P. C. Ivanov, Z. Chen, P. Carpena and H. E. Stanley, *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 2001, **64**, 19.
- [17] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde and H. E. Stanley, *Physica A: Statistical Mechanics and its Applications*, 2002, **316**, 87–114.
- [18] H. Salat, R. Murcio and E. Arcaute, *Multifractal methodology*, 2017.
- [19] J. Barunik and L. Kristoufek, *Physica A: Statistical Mechanics and its Applications*, 2010, **389**, 3844–3855.
- [20] E. Bakalis, V. Gavriil, A. C. Cefalas, Z. Kollia, F. Zerbetto and E. Sarantopoulou, *Journal of Physical Chemistry B*, 2021, **125**, 10883–10892.
- [21] T. Di Matteo, T. Aste and M. M. Dacorogna, *Journal of Banking & Finance*, 2005, **29**, 827–851.
- [22] E. Bakalis, H. Fujie, F. Zerbetto and Y. Tanaka, *Physica A: Statistical Mechanics and its Applications*, 2018, **512**, 945–953.
- [23] E. Bakalis, L. R. Parent, M. Vratsanos, C. Park, N. C. Gianneschi and F. Zerbetto, *The Journal of Physical Chemistry C*, 2020, **124**, 14881–14890.
- [24] R. Ferrari, A. J. Manfroi and W. R. Young, *Physica D: Nonlinear Phenomena*, 2001, **154**, 111–137.
- [25] T. Di Matteo, *Quantitative Finance*, 2007, **7**, 21–36.

# Chapter 2

## Coloured Noises and Fractional Noises

### Introduction

Noise is a lot more than a nuisance playing a central role in many physical, chemical and biological processes.<sup>1</sup> White noise is connected to thermal fluctuations, Brownian noise is related to the motion of molecules and particles, flicker noise to cooperativity.

As light color is defined according to its frequency spectrum, also for noises exists a color code classification based on their frequency content. Depending on the scaling law of the power spectral density(PSD), noises can have different “colors”.

In this chapter I present the main properties of the different coloured noises classified according to their PSD. I also introduce a second classification method that develops from the extension of Brownian motion to processes that are not Brownian in order to define two classes of processes that are one the integral (derivative) of the other.<sup>2</sup> All the stochastic processes considered in this chapter belongs to monofractal processes. They are random processes with a unique time scale.

For all the considered noises, the typical relationship between power and frequency is of the form

$$PSD(f) \propto 1/f^\beta \quad (2.1)$$

### 2.1 White Noises

White noise is widely used as a descriptor for fluctuation in physical, chemical and biological systems. White noise contains no information, it is completely

random. At any time, a step in one direction is as likely as a step in the opposite direction. A white noise is an uncorrelated stochastic process that is stationary. Its values are independent and identically distributed (i.i.d). Its mean value is constant in time and its autocorrelation function is a delta function with correlation vanishing even for the smallest lag time. Looking at the power spectrum of a white noise, it results flat with the power equally distributed on all frequencies. A white noise is considered a good representation for thermal fluctuations. In fact, at equilibrium, thermal fluctuations are not directional nor changing in time. In a gas or a liquid molecules move randomly in all directions.

According to the central limit theorem, typical white noises are with a Gaussian probability distribution, however white noises with other PDFs can be observed.

## 2.2 Correlated Noises

In a correlated noise, at any time, a step in one direction is more likely followed by a step in the same direction. In this way the autocorrelation function is no longer a delta function but is positive also for lags larger than the minimum. Correlated noises have the power in PSD that decreases at the increasing of the frequency (see eq. 2.1) with the power exponent,  $\beta$ , that is positive.

### 2.2.1 Red Noise

Red noise is the most famous coloured noise. It arises from the temporal integration of a white noise and has a power exponent  $\beta = 2$ . Red noise is also known as Brownian noise as an extension of the Brownian motion to stochastic processes other than motion.

Brownian motion has been observed and described in details by the Scottish botanist Robert Brown in 1827.<sup>3</sup> He observed that, when a pollen grain is placed on the surface of a liquid starts moving randomly. Physically Brownian motion (BM) arises as the combined effect of all the collision between liquid molecules and the particle. It is only after Einstein publication in 1905 that a unified theoretical explanation of Brownian motion was established.<sup>4</sup>

Autocorrelation function (ACF) for a red noise is exponential in linear lag-time, as  $ACF(t) \sim \exp(-t/\tau)$  and  $\tau$  is the characteristic time.

### 2.2.2 Pink Noise

The power exponent for a pink noise is  $\beta = 1$ . It can be considered the integration of order  $\frac{1}{2}$  of a white noise. Sometimes pink noise is also known as flicker noise, which actually is any noise with a power exponent  $0.5 < \beta < 1.5$ .

Pink noises are commonly found in nature<sup>2,5</sup> with a proven presence in human cognition processes. The wide occurrence of pink noises is also due the fact that, several red (Brownian) noises with different relaxation times sums to a pink noise like time series.<sup>6</sup> Caution has to be paid when analysing signals in order to distinguish pink noises from similar processes.

The Hurst exponent shows a discontinuity for  $\beta = 1$ .  $H = 1$  if computed with Dispersional analysis and considering the process as the half-integration of a white noise  $H = (\beta + 1)/2$ .  $H = 0$  if pink is considered as half-differentiation of red noise and using the relation  $H = (\beta - 1)/2$ . In section 2.4 a wider explanation is given.

## 2.3 Anti-correlated Noises

In an anti-correlated noise, at any time, a step in one direction is more likely followed by a step in the opposite direction. It is not completely random. The autocorrelation function shows a negative peak for the minimum lag time. Anticorrelated noises have a PSD with power that increases proportionally to the frequency with power exponent  $\beta < 1$ .

### 2.3.1 Blue Noise

A blue noise is characterized by a power exponent  $\beta = -1$ . It can be considered the fractional derivative of order  $\frac{1}{2}$  of a white noise.

Blue noise characterizes image processing and vision with retina cells arranged in a blue-noise pattern.<sup>7</sup> For blue noises the Hurst exponent  $H = 0$ ,  $H = (\beta + 1)/2$ .

### 2.3.2 Purple Noise

Purple (or violet) noise has a power exponent  $\beta = -2$ , that is the power density increases per octave with increasing frequency over a finite frequency range. It can be considered as the derivative of a white noise. It has been suggested that acoustic thermal noise of water in the Oceans has a violet spectrum at high frequencies.<sup>8</sup> Anticorrelation in purple noises is so strong that for lag time  $\tau = 2$  the ACF shows a positive value before setting to zero for larger lag times.

## 2.4 Fractional Noises

Brownian motion can be described as the integer integral, in time, of a thermal (white) noise. For a BM the average displacement,  $\sqrt{\langle X^2(t) \rangle}$  are proportional to  $t^H$  with  $H = 1/2$ . Measurements of single particle trajectories in crowded environment show that motion is not Brownian with  $H \neq 1/2$ . One of the way to describe anomalous diffusion is to use an extension of Brownian motion in order to consider the slower(faster) motion of the particles.

Mandelbrot and Van Ness(MVN)<sup>2</sup> proposed a family of Gaussian function that modifies ordinary Brownian motion calling it “fractional Brownian motion”, *fBm*. Fractional Brownian motion can be defined as follow:

- define a Brownian motion  $B(t)$  and a parameter (Hurst)  $H$ ,  $0 < H < 1$ ;
- Its increments  $dB(t)$  are from a Gaussian white noise  $W(t)$ ;
- fBm of exponent  $H$  is a moving average of  $W(t)$  with the past increments of  $B(t)$  weighted by a kernel  $(t - s)^{H-1/2}$

For an ordinary BM kernel exponent vanishes and no memory is retained. In fBm increments retain memory either persistent  $H > 1/2$  or antipersistent  $H < 1/2$  with a memory that can be infinite.

The equation of such process can be written as

$$B_H(t) - B_H(0) = \frac{1}{\Gamma(H + \frac{1}{2})} \left\{ \int_{-\infty}^0 [(t - s)^{H-1/2} - (-s)^{H-1/2}] W_0(s) ds + \int_0^t (t - s)^{H-1/2} W_0(s) ds \right\} \quad (2.2)$$

with  $B_H(0) = b_0$  that is the starting value. The second integral is the Holmgren-Riemann-Liouville fractional integral, already proposed by Lévy for fBm while the first integral is the Weyl integral that MVN proposed to adjust the effects of the origin.<sup>2,9</sup>

For the process generated by equation (2.2) the increments  $B_H(t + \tau) - B_H(t)$  are stationary: the mean of increments is constant. The Hurst exponent for fBm defines three families with different properties:

- For  $0 < H < \frac{1}{2}$  the process is antipersistent with increments that are anticorrelated;
- For  $\frac{1}{2} < H < 1$  the process is persistent with increments that are correlated;

- $H = \frac{1}{2}$  the process is ordinary Brownian motion and increments are uncorrelated.

Persistency or antipersistency give information on the nature of the memory of the process.<sup>9</sup>

Looking at the power spectrum of fBm, it shows a power exponent  $\beta > 1$  indicating a strong positive correlation.  $\beta$  is related to the Hurst exponent in the form

$$H_{fBm} = \frac{\beta - 1}{2} \quad (2.3)$$

Fractional Brownian motion, as ordinary Brownian motion, is not differentiable, so an approximate method is used. As for ordinary Brownian motion its “derivative” is a Gaussian white noise, for fBm the derivative may be called fractional Gaussian noise, *fGn*.

fGn has spectral density equal to  $f^{1-2H}$  with  $\beta_{fGn} = \beta_{fBm} - 1$ . As a consequence, the Hurst exponent becomes

$$H_{fGn} = \frac{\beta + 1}{2} \quad (2.4)$$

The differentiation/integration relationship between fBm and fGn implies the facts that a fBm and its fGn have the same Hurst exponent and  $\beta_{fBm} = \beta_{fGn} + 2$ .

In the analysis of a signal the methods used to estimate  $H$  differs whether the signal is fBm or fGn. For fBm, usually Rescaled range analysis or generalized Hurst are used while, for fGn Dispersional analysis is the method of choice.<sup>10</sup>

## 2.5 Langevin Equation and Coloured Noises

The Langevin equation is the simplest equation of motion that one can use to reproduce a Brownian motion. It directly originates from Newton’s equation of motion

$$m \frac{dv}{dt} = F_{tot}(t) \quad (2.5)$$

where  $m$  is the mass of a spherical particle,  $v$  is the velocity and  $F_{tot}(t)$  is the instantaneous force experienced by the particle. If the initial position and all the terms contributing to the force are known, then the motion of the particle is predicted in time. An exact form for  $F_{tot}$  is impractically complicated. Other than the friction forces slowing down the particle velocities, the collision between the particle and the surrounding fluid need to be considered in

the equation of motion. This originates what is known as Langevin equation

$$m \frac{dv}{dt} = -\zeta v + \delta F(t) \quad (2.6)$$

with  $-\zeta v$  friction force and  $\delta F(t)$  a fluctuating force. The introduction of the fluctuating forces in eq. 2.6 allows to reproduce the mean square velocity at thermal equilibrium  $\langle v^2 \rangle_{eq} = k_B T / m$

In classical Langevin equation, fluctuations are assumed as instantaneous. This correspond to a random force with uncorrelated variations or, in terms of coloured noises, as a white noise.

$$\langle \delta F(t) \rangle = 0, \quad \langle \delta F(t) \delta F(t') \rangle = 2B \delta(t - t') \quad (2.7)$$

with  $B$  that is a measure of the strength of the fluctuating force. At equilibrium

$$B = \zeta k_B T \quad (2.8)$$

that is the *Fluctuation-dissipation theorem*. It relates the strength of the fluctuating forces  $B$  with the magnitude of the friction  $\zeta$ .<sup>11</sup> Coupling the forces with the friction coefficient results in what is called a thermostat that maintains the energies of the system at values connected to the given temperature.

A thermostat that is based on a white noise is typical for Markovian processes and imply no memory.<sup>12</sup> However, Markovian processes are only a subset of all possible stochastic processes. In order to generate non-Markovian processes, it is possible to use coloured noises in the thermostat.

The use of coloured noises introduces the effects of degrees of freedom of the system that were excluded with white noises. In this way, Langevin equation simulates not only Brownian motions but also anomalous diffusion. Coloured noises posses memory, the history of the system influences its present value either with a correlated force that speeds up motion or with an anticorrelated force that slows down motion. In the first case the system is said superdiffusive while in the second case the system is subdiffusive. Anomalous diffusion, and so environmental noises other than white, gives information about the properties of the environment like deep traps immobilizing the diffusing particle in the case of subdiffusion.<sup>4,13</sup>

## 2.6 Applications of Coloured Noises

Generalized Langevin equation is one of the possible applications of coloured noises, but  $1/f^\beta$  noises are ubiquitous with applications, among the others, in Molecular Dynamics, Biology, Physiology and Electronics.

Here I will present a short list of examples for each one of the previous fields.

### 2.6.1 Noises in Molecular Dynamics

Molecular simulations investigate a variety of systems that exhibit stochastic fluctuations other than white.

A widely studied system is the membrane protein aquaporin. This channel protein, involved in the transport of water molecules across cell membranes, exhibits  $1/f$  fluctuations. In particular, Yamamoto *et al.*<sup>14</sup> observed that  $1/f$  fluctuations of amino acids contribute to water transportation in aquaporins. The effect of  $1/f$  noises was also observed experimentally by López *et al.* in a study on the ion channels in bacteria.<sup>15</sup>

A study on dynamic self-assembly by Lugli and Zerbetto shows how a coloured noise can improve the catalytic activity of self-assembled nanoparticles.<sup>16</sup> In particular, they observed that, randomly changing the interaction parameter for the self-assembly increases the reaction rate thanks to an higher number of interaction between reactant and catalyst. The effect of coloured noises is thought to be similar to the real behaviour in Nature where many processes are affected or driven by stochastic fluctuations.

All this studies suggest that coloured noises are far from being rare with a spreading from biomolecules to reactions.

### 2.6.2 Noises in Biology

Noises affect all biological processes from cellular level to population dynamics. Other than being a disruptive factor, noise is the driving force for all the cellular processes from gene expression to signalling to motility.<sup>17</sup> The stochastic variability introduced during the transcription processes in gene expression is thought to play a central role in stress responses and survival strategies of microorganisms.<sup>18</sup>

A remarkable example of noises as driving forces is in the antibiotic resistance: some bacteria are antibiotic-resistant thanks to the presence of a small number of slow growing cells (persisters) that are strongly resistant to antibiotic. Their resistance is not due to a different genome but to the level of expression of a particular antitoxine. Fluctuations in the level of expression of this toxin above or below an effective threshold result in the coexistence of normal and persister cells. This helps the population to survive antibiotics and spread again when antibiotics were removed.<sup>19</sup>

Fluctuations are also central in population evolution in ecosystems. Noises act on the deterministic law for population evolution as the effects of the environment. Changes in the environment results in a variation of the population equilibrium. Coloured noises are used to introduce correlations of the environment in the model. For example, in noise-delayed extinction the

environment fluctuations increase the population lifetime with respect to the one predicted in the case of constant environmental conditions.<sup>20</sup>

### 2.6.3 Noise in Physiology

Many studies on brain activity observed a  $1/f$ -like power spectrum using different techniques, including electroencephalography (EEG), local field potential (LFP) and functional magnetic resonance imaging (fMRI).

$1/f$  brain activity, called arrhythmic activity, represents the majority of the signal power in LFP, EEG and fMRI, nevertheless, it is less studied than the brain oscillations. Despite its ubiquitous presence suggests a lack of functional specificity, arrhythmic activity plays a central role in brain functionality. In particular, it has different properties depending on the different observed activity.

Among the studies on  $1/f$  brain activity, a work by He *et al.* observed that the scale-free brain activity in different arousal states possesses a characteristic scaling exponent for the different states.<sup>21</sup> Other studies have demonstrated that the power exponent can be modulated by sensory stimuli or task performance.<sup>22,23</sup>

In order to have a wider view about coloured noises in brain activity I suggest to the reader the short review by He<sup>23</sup> and the work by Toronov *et al.*<sup>24</sup> The first is a perspective on the studies on the scale free brain activity. The latter gives an overview on different models for the brain processes and makes a comparison with the results of experimental studies.

### 2.6.4 Noise in Electronics

Electric and electronic devices present a wide variety of noises.

Thermal fluctuations contribute to the current according to the Johnson-Nyquist formula

$$\langle I^2 \rangle = 4k_B T G \Delta f \quad (2.9)$$

where  $\langle I^2 \rangle$  is the mean-square noise current,  $k_B T$  is the temperature in energy units,  $G = 1/R$  is the conductance of the resistor, and  $\Delta f$  is the measurement bandwidth.<sup>25</sup> The relationship depicted in equation 2.9 is crucial for the design of all the electrical and electronic devices and it represents a baseline for the noise level in an electric circuit.

Another noise present in a circuit is the  $1/f$  noise. It is a rich source of information about the device under study that cannot be extracted using mean conductance data. Sources of  $1/f$  noises are the trapping/de-trapping processes of electrons or the scattering of electrons in metals.

The quantization of charge carrier (the charge is transported by electrons) generates the shot noise. It is due to the motion of independent electrons through a barrier. Shot noise is a Poisson white noise.<sup>26</sup>

## 2.7 Synthesis of Coloured noises

Different methods can be used to numerically generate a stochastic time series with prescribed properties, such as a certain Hurst or spectral exponent. Generation methods can be divided in two classes: methods acting in the time domain and methods acting in the frequency domain. Among those methods I present spectral processing, as an example of method in the frequency domain, and fractional calculus, as an example of method in the time domain. Both methods, starting from a white noise of given probability distribution, produces self similar time series with a predetermined  $\beta$ .

### 2.7.1 Spectral Processing

Spectral processing is a simple way to generate a coloured noise that acts in the frequency domain. With spectral processing, a white noise is converted to a coloured noise rotating its power spectrum.<sup>27,28</sup>

The basic procedure for spectral processing reads:

1. generate a pseudorandom white noise vector  $w(t)$  of the given length with the desired probability distribution;
2. Generate the Fourier transform of the white noise using the Fast Fourier Transform,  $W = FFT(w)$ ;
3. Multiply the complex spectral coefficients of the white noise by a power law of the desired coloured noise,  $W' = W f^h$  where  $h = \beta/2$ ;
4. Apply the inverse Fourier transform to obtain the coloured noise,  $y = IFFT(W')$ .

### 2.7.2 Fractional Calculus

Fractional Calculus (FC) is a time-domain method to generate coloured noises based on the use of derivative and integrals. Red noises can be generated by integration of white noises, while purple noise can be obtained by differentiation of white noises. Other colours can be generated using integrals and derivatives of non-integer orders, that is using Fractional Calculus.<sup>29</sup>

A coloured noise obtained by Fractional Calculus can be expressed as

$$D^\nu w(t) = N^\beta \quad (2.10)$$

where  $D$  is the fractional operator,  $\nu$  is the fractional order,  $N$  is the coloured noise sequence and  $\beta$  is the power exponent. For  $\nu < 0$  eq. 2.7.2 describes fractional integration, for  $\nu > 0$  describes fractional derivation. For  $\nu$  integer eq. 2.7.2 reduces to classical integration/derivation.

Different definitions of fractional derivatives and integrals exists, among them Grünwald-Letnikov(GL), Riemann-Liouville, Caputo,etc,<sup>30</sup> Here I introduce the Grünwald-Letnikov (GL) definition.<sup>31</sup>

The GL definition for fractional derivative is given by

$${}_a D_t^\nu w(t) = \lim_{h \rightarrow 0} h^{-\nu} \sum_{j=0}^{\lfloor \frac{t-a}{h} \rfloor} (-1)^j \binom{\nu}{j} w(t - jh) \quad (2.11)$$

where  $a$  is a real constant which express a limit value,  $h$  is the discretization step, and  $\lfloor \frac{t-a}{h} \rfloor$  is the floor function. The binomial coefficients are obtained through the Euler's Gamma function as

$$\binom{\nu}{j} = \frac{\nu!}{j!(\nu-j)!} = \frac{\Gamma(\nu+1)}{\Gamma(j+1)\Gamma(\nu-j+1)} \quad (2.12)$$

Coloured noises are generated starting from a white noise of the given probability distribution applying equation 2.7.2 for  $t = 0$  to  $t = N$ .

# Bibliography

- [1] P. V. McClintock, *Nature*, 1999, **401**, 23–25.
- [2] B. B. Mandelbrot and J. W. Van Ness, *SIAM Review*, 1968, **10**, 422–437.
- [3] R. Brown, *The Philosophical Magazine*, 1828, **4**, 161–173.
- [4] R. Metzler and J. Klafter, *Physics Report*, 2000, **339**, 1–77.
- [5] P. Bak, C. Tang and K. Wiesenfeld, *Phys. Rev. Lett.*, 1987, **59**, 381–384.
- [6] E. Milotti, *Phys. Rev. E*, 2005, **72**, 056701.
- [7] J. I. Yellott, *Science*, 1983, **221**, 382–385.
- [8] J. A. Hildebrand, *Marine Ecology Progress Series*, 2009, **395**, 5–20.
- [9] A. Fuliński, *Journal of Physics A: Mathematical and Theoretical*, 2017, **50**, 054002.
- [10] A. Eke, P. Herman, J. B. Bassingthwaighte, G. M. Raymond, D. B. Percival, M. Cannon, I. Balla and C. Ikrényi, *Pflugers Archiv : European journal of physiology*, 2000, **439**, 403–15.
- [11] R. Zwanzig, *Nonequilibrium statistical mechanics*, Oxford university press, New York, 2001.
- [12] M. Ceriotti, G. Bussi and M. Parrinello, *Physical Review Letters*, 2009, **102**, 020601.
- [13] R. Metzler and J. Klafter, *Journal of Physics A: Mathematical and General*, 2004, **37**, R161–R208.
- [14] E. Yamamoto, T. Akimoto, Y. Hirano, M. Yasui and K. Yasuoka, *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2014, **89**, 022718.

- [15] M. L. López, M. Queralt-Martín and A. Alcaraz, *Nanoscale*, 2016, **8**, 13422–13428.
- [16] F. Lugli and F. Zerbetto, *Journal of Physical Chemistry C*, 2019, **123**, 825–835.
- [17] L. S. Tsimring, *Reports on Progress in Physics*, 2014, **77**, 26601.
- [18] D. Fraser and M. Kærn, *Molecular Microbiology*, 2009, **71**, 1333–1340.
- [19] E. Rothem, A. Loinger, I. Ronin, I. Levin-Reisman, C. Gabay, N. Shoshitaishvili, O. Biham and N. Q. Balaban, *Proceedings of the National Academy of Sciences*, 2010, **107**, 12541–12546.
- [20] A. Fiasconaro, D. Valenti and B. Spagnolo, *Mathematical Biosciences and Engineering*, 2004, **1**, 185–211.
- [21] B. J. He, J. M. Zempel, A. Z. Snyder and M. E. Raichle, *Neuron*, 2010, **66**, 353–369.
- [22] S. El Boustani, O. Marre, S. Béhuret, P. Baudot, P. Yger, T. Bal, A. Destexhe and Y. Frégnac, *PLoS Computational Biology*, 2009, **5**, e1000519.
- [23] B. J. He, *Trends in Cognitive Sciences*, 2014, **18**, 480–487.
- [24] V. Toronov, T. Myllylä, V. Kiviniemi and V. V. Tuchin, *European Physical Journal: Special Topics*, 2013, **222**, 2607–2622.
- [25] M. T. H. Reid, A. W. Rodriguez and S. G. Johnson, *Proceedings of the IEEE*, 2013, **101**, 531–545.
- [26] Z. Balogh, G. Mezei, L. Pósa, B. Sánta, A. Magyarkuti and A. Halbritter, *Nano Futures*, 2021, **5**, 042002.
- [27] J. Timmer and M. Konig, *Astronomy and Astrophysics*, 1995, **300**, 707–710.
- [28] H. Zhivomirov, *Romanian Journal of Acoustics and Vibration*, 2018, **15**, 14–19.
- [29] A. Regadío, J. Tabero and S. Sánchez-Prieto, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2016, **811**, 25–29.
- [30] R. Gorenflo and F. Mainardi, in *Fractional Calculus*, ed. A. Carpinteri and F. Mainardi, Springer Vienna, Vienna, 1997, pp. 223–276.

- [31] R. Scherer, S. L. Kalla, Y. Tang and J. Huang, *Computers and Mathematics with Applications*, 2011, **62**, 902–917.



# Chapter 3

## Singular Spectrum Analysis

### Introduction

Singular Spectrum Analysis (SSA) is a model-free technique that decompose a time series into a sum of a small number of interpretable components like a trend, oscillatory components and noise. According to Golyandina *et al.*<sup>1</sup>, the main methodological principle of SSA arises from multivariate geometry, with the signal matrix projected into a new space of almost non-interacting components. SSA incorporates elements from classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems, and signal processing. The origin of the technique dates back in the 1980s with the seminal works of Broomhead.<sup>2,3</sup> Since then, more than three thousands articles on this topic were published mainly in Electrical and Electronic Engineering, Meteorology and Atmospheric sciences and Geosciences with an increasing interest in the last decade with less than 100 articles published in 2010 and about 300 in 2020 (Web of Science) Different monographs have been published on this technique mainly form Golyandina et al.<sup>1,4</sup> In its original formulation SSA it is widely used in engineering, climate sciences, economics and financial mathematics to decompose complex signals into a set of elementary components.<sup>5-7</sup> SSA is based on principal component analysis (PCA) in the vector space of the delayed coordinates for a time series. We will see in Chapter 4 that PCA is performed in the vector space of the multivariate data set. In the absence of a set of multiple measurements it is not possible to perform PCA in order to extract meaningful components of the signal. SSA circumvent this limit creating a vector space using the time as generator. In this way, it is possible to apply a decomposition that reduces the information contained in the signal.

Let us briefly explain the basic version of SSA. Basic SSA consist of four

steps as follow. Given a time series  $S = (s_0, s_1, \dots, s_{n-1})$  of length  $N$ , it is defined an integer window length  $L$ . Set  $K = N - L + 1$  and define the  $L$ -lagged vectors  $X_k = (s_{k-1}, \dots, s_{k+L-2})^T$ ,  $k = 1, 2, \dots, K$ , and the trajectory matrix

$$\mathbf{H} = (s_{i+j-2})_{i,j=1}^{L,K} = [X_1, \dots, X_K]$$

Matrix  $\mathbf{H}$  is a Hankel matrix, it has all the elements along the diagonal  $i + j = \text{const}$  that are equal. The second step is the singular value decomposition (SVD) of the matrix  $\mathbf{H}$  or, that is the same, the eigendecomposition of the matrix  $\mathbf{A} = \mathbf{H}\mathbf{H}^T$  of size  $L \times L$ . In this way we obtain a set of  $L$  singular values.  $\mathbf{H}$  can be represented as a sum of biorthogonal matrices  $\mathbf{H}_i$  that are related to the eigenvectors of  $\mathbf{A}$ . The third step is the grouping of the previous matrices in sets summing the matrices in the same set. The last step is the diagonal averaging of the group matrices in order to obtain a new time series from each group. This results in a decomposition of the original time series  $S$  into a sum of series. The possibility to choose the number of components in which decompose the signal makes SSA more flexible than PCA, where the number of components is defined by the dimension of the dataset.

In the next section I will explain the basic algorithm for SSA decomposition of a time series.

Section 3.2 introduces the algorithm for SSA decomposition applied to images.

Section 3.3 and 3.4 present variations of basic SSA. In the chapters of this thesis devoted to practical application I will present projects in which SSA has played a central role in obtaining results.

## 3.1 1D SSA

The simplest application of SSA is to the decomposition of a one-dimensional time series.

### 3.1.1 Method

Consider a time series  $S = (s_1, s_2, \dots, s_N)$  of length  $N > 2$ . Define a window length  $L$ , ( $1 < L < N/2$ ) and set  $K = N - L + 1$

#### Step 1: Embedding

The *embedding* maps the original time series into a trajectory matrix whose columns are lagged version of the original time series. Given  $L, K$  lagged

vectors are formed:

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

which have length  $L$ . The trajectory matrix of  $S$  then is:

$$\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} s_0 & s_1 & s_2 & \dots & s_{K-1} \\ s_1 & s_2 & s_3 & \dots & s_K \\ s_2 & s_3 & s_4 & \dots & s_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{L-1} & s_L & s_{L+1} & \dots & s_{N-1} \end{pmatrix} \quad (3.1)$$

The trajectory matrix presents equal elements on the skew diagonals  $i + j = \text{const}$  thus, the trajectory matrix is a Hankel matrix.

### Step 2: Singular Value Decomposition

In this step the singular value decomposition of the trajectory matrix is computed.

Let  $\mathbf{A} = \mathbf{X}\mathbf{X}^T$  be the correlation matrix of the trajectory matrix. It is a square matrix of size  $L \times L$ . Denote by  $\lambda_1, \dots, \lambda_L$  the eigenvalues of  $\mathbf{A}$  sorted in decreasing order of magnitude ( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L \geq 0$ ) and by  $U_1, \dots, U_L$  the eigenvectors of the matrix  $\mathbf{A}$  corresponding to these eigenvalues.

If we define  $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$ , then the SVD of matrix  $\mathbf{X}$  can be written as follows:

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_L \quad (3.2)$$

with  $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$ . The collection  $(\sqrt{\lambda_i}, U_i, V_i)$  is called the  $i$ th-eigen-triple of the SVD.

### Step 3: Grouping

With the SVD the original signal is expanded as stated in equation 3.2. In this third step the expansion is partitioned in  $m$  disjoint subset  $I_1, \dots, I_m$ ,  $I = \{i_1, \dots, i_p\}$ . The resultant matrix  $\mathbf{X}_I$  of the group  $I$  is defined as

$$\mathbf{X}_I = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_p} \quad (3.3)$$

### Step 4: Diagonal averaging or projection

The last step of Basic SSA transforms back the matrices from the grouping step into a set of time series of length  $N$ . For this step exist two different approaches: the one reported by Golyandina *et al.*<sup>1</sup>, based on the diagonal

averaging of the matrix  $\mathbf{X}$  and the one proposed by Vautard *et al.*<sup>8</sup> based on the projection of the trajectory matrix onto the eigenvectors. Here I will present both approaches.

**Diagonal Averaging (*Golyandina et al.*)** In the approach proposed by Golyandina et al. the new series are obtained from the matrices  $\mathbf{X}_I$  (3.3) by averaging on the skew diagonals. Given the matrix  $\mathbf{X}_I$  of shape  $L \times K$  and elements  $x_{ij}$ ,  $1 \leq i \leq L$ ,  $1 \leq j \leq K$  set  $L^* = \min(L, K)$ ,  $K^* = \max(L, K)$  and  $N = L + K - 1$ . Let  $x_{ij}^* = x_{ij}$ . The matrix  $\mathbf{X}$  is transferred to the series  $g_0, \dots, g_{N-1}$  by the formula:

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} x_{m,k-m+2}^* & \text{for } 0 \leq k < L^* - 1, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} x_{m,k-m+2}^* & \text{for } L^* - 1 \leq k < K^*, \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} x_{m,k-m+2}^* & \text{for } K^* \leq k < N. \end{cases} \quad (3.4)$$

The equation 3.4 corresponds to averaging of the matrix elements over the ‘diagonals’  $i + j = k + 2$ . This procedure is necessary because the grouped matrices  $\mathbf{X}_I$  are not Hankel matrices, elements on the ‘diagonals’ are not equals.

**Projection (*Vautard et al.*)** The method proposed by Vautard et al. is based on the projection of the trajectory matrix on the space produced by the grouped eigenvectors like what is done in PCA (see Chapter 4). The equations presented here resemble those reported in Groth and Ghil<sup>9</sup> and used in popular implementation of SSA. The first step is the projection of the trajectory matrix  $\mathbf{X}$  onto the eigenbasis  $\mathbf{U}$  producing the principal components of the signal.

$$\mathbf{P} = \mathbf{XU} \quad (3.5)$$

with  $\mathbf{P}$  containing the principal components as columns.

In order to obtain the reconstructed components of  $\mathbf{X}$ , an inverse transformation with respect to eq. 3.5 is necessary:

$$\mathbf{R} = \mathbf{XUU}' \quad (3.6)$$

or

$$\mathbf{R}_I = \mathbf{P}_I \mathbf{U}'_I \quad (3.7)$$

with  $I = \{I_1, \dots, I_m\}$   $\mathbf{R}_I$  is the matrix of the reconstructed component of  $\mathbf{X}$  related to the group  $I$ . The last step transforms back the reconstructed

components matrix to a vector. This is done by averaging the skew diagonals of  $\mathbf{R}_I$ :

$$g_I(n) = \frac{1}{L_d} \sum_{m=1}^n \mathbf{R}_I(m, n - m + 1) \quad (3.8)$$

series  $\mathbf{g}_I$  is the  $i$ -th component of the original time series.

### 3.1.2 Window length and decomposition

The window size  $L$  is the only parameter in the SSA decomposition and its choice determines the performances of the entire process. As introduced in section 3.1.1, the recommended range  $1 < L < N/2$  is due to the fact that, for  $L > N/2$ ,  $L$  and  $K$  change order of magnitude and the correlation matrix becomes of size  $K \times K$  with  $K < N/2$ . Except for the previous condition, no further constraint exist on  $L$ . The choice of the window size depends on the time series properties. In the presence of a periodic component a good choice is to take  $L$  proportional to the period. A large value of  $L$  corresponds to a detailed decomposition of the signal, but with the drawback of a trend that is spread in a large number of components and contaminated by other components. A small value of  $L$  is good for the extraction of the trend from the signal, but it is limited in the ability of removing the noise from the signal. A general rule for the choice of the window size does not exist and the choice is made based on the purposes of the decomposition. In the book by Golyandina *et al.*<sup>1</sup> and in<sup>10</sup> an entire section is devoted to the discussion of the choice of SSA parameters. Sometimes the basic SSA is unable to produce good results. In section 3.4 I present the reasons for the fail of basic SSA and some methods that try to overcome the problem.

### 3.1.3 Separability

The main purpose of SSA is the decomposition of a given series into a sum of identifiable series such as a trend, periodic components and noise. SSA succeeds at its purpose if the additive components are in different groups: the eigenvectors in one group are orthogonal to the eigenvectors of the others groups. In this case the series are *weakly separable*. If the eigendecomposition of the main series can be grouped to obtain the trajectory matrices of the sub-series then the series are *strongly separable*. Separability depends on the size of the embedding window  $L$ . Adjusting  $L$  increases separability as presented in Golyandina<sup>11</sup>.

### W-correlation Matrix

A measure of separability is given by the so called *weighted correlation matrix*, **w**-correlation matrix. This matrix consists of weighted cosines of angles between the reconstructed time series components and can be easily generated.

Defined  $L^* = \min(L, K)$  and  $K^* = \max(L, K)$  it is possible to define the weights as

$$w_i = \begin{cases} i + 1 & \text{for } 0 \leq i \leq L^* - 1 \\ L^* & \text{for } L^* \leq i \leq K^* \\ N - i & \text{for } K^* \leq i \leq N - 1 \end{cases} \quad (3.9)$$

and the inner product of two series  $S^{(1)}, S^{(2)}$  as

$$(S^{(1)}, S^{(2)})_w = \sum_{i=0}^{N-1} w_i s_i^{(1)} s_i^{(2)} \quad (3.10)$$

If  $(S^{(1)}, S^{(2)})_w = 0$ ,  $S^{(1)}$  and  $S^{(2)}$  are called ***w**-orthogonal*. Separability implies **w**-orthogonality. **w**-orthogonality can be expressed in the **w**-correlation matrix where the inner products are normalized producing the weighted correlation

$$\rho_{12}^{(w)} = \frac{(S^{(1)}, S^{(2)})_w}{\|S^{(1)}\|_w \|S^{(2)}\|_w} \quad (3.11)$$

where  $\|S^{(i)}\|_w = \sqrt{(S^{(i)}, S^{(i)})_w}$ ,  $i = 1, 2$ .

If the **w**-correlation is small in absolute value, then the two series can be considered **w**-orthogonal, otherwise, if it is large, the series are not orthogonal and therefore badly separable.

Looking at the **w**-correlation matrix one can find groups of correlated series components and use them for the grouping. If two components are correlated, usually they belong to the same signal. If this is not true, the two series are not separable using the given embedding window.

## 3.2 2D SSA

A first example of bidimensional implementation of SSA is in Golyandina *et al.*<sup>12</sup> where they apply 2D-SSA to a digital terrain model.<sup>12</sup> Other applications of 2D-SSA is in the analysis of hydrological data and in processing of hyperspectral images. As its monodimensional version, bidimensional SSA is based on the Singular Value Decomposition of a trajectory matrix. In bidimensional case this matrix is in the form of an Hankel-block-Hankel matrix

and it is generated from the 2D array corresponding to the input signal. Here we report a brief explanation of the method. A detailed implementation of 2D-SSA is presented in Golyandina and Usevich<sup>13</sup>.

### 3.2.1 Method

As for basic SSA, also 2D-SSA can be divided in four steps. Only step 1 and 4 differ from the 1D version. Let consider the input greyscale image as a 2D-array of  $N_r$  rows and  $N_c$  columns

$$P = \begin{pmatrix} p(1,1) & p(1,2) & \cdots & p(1,N_c) \\ p(2,1) & p(2,2) & \cdots & p(2,N_c) \\ \vdots & \vdots & \ddots & \vdots \\ p(N_r,1) & p(N_r,2) & \cdots & p(N_r,N_c) \end{pmatrix} \quad (3.12)$$

Define the window sizes  $(L_r, L_c)$  with  $1 \leq L_r \leq N_r$ ;  $1 \leq L_c \leq N_c$ ;  $1 < L_r L_c < N_r N_c$ . Set  $K_r = N_r - L_r + 1$ ;  $K_c = N_c - L_c + 1$  and  $K = K_r K_c$

#### Step 1: Embedding

At this step, the input 2D-array is arranged into a Hankel-block-Hankel (HbH) matrix corresponding to the trajectory matrix. To build the HbH matrix the columns are arranged in a set of 2D matrices

$$\mathbf{H}_i = \begin{pmatrix} p(1,i) & p(2,i) & \cdots & p(K_r,i) \\ p(2,i) & p(3,i) & \cdots & p(K_r+1,i) \\ \vdots & \vdots & \ddots & \vdots \\ p(L_r,i) & p(L_r+1,i) & \cdots & p(N_r,i) \end{pmatrix} \quad (3.13)$$

$1 \leq i \leq N_r$ . Then the matrices are arranged in a HbH matrix:

$$\mathbf{HbH} = \mathbf{W} = \begin{pmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \cdots & \mathbf{H}_{K_c} \\ \mathbf{H}_2 & \mathbf{H}_3 & \cdots & \mathbf{H}_{K_c+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{L_c} & \mathbf{H}_{L_c+1} & \cdots & \mathbf{H}_{N_c} \end{pmatrix} \quad (3.14)$$

#### Step 2: Singular Value Decomposition

The SVD is applied to the HbH matrix obtaining the eigenvalues and the eigenvectors of the correlation matrix of  $\mathbf{W}$ ,  $\mathbf{W}\mathbf{W}^T$ .

$$\mathbf{W} = \sum_{i=1}^d \mathbf{W}_i = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T \quad (3.15)$$

with  $\lambda_i, \dots, \lambda_d$  eigenvalues in decreasing order of magnitude,  $U_i, \dots, U_d$  eigenvectors and  $V_i, \dots, V_d$  factor vectors.  $V_i = \mathbf{W}^T U_i / \sqrt{\lambda_i}$ . The set  $(\lambda_i, U_i, V_i)$  is called the  $i$ -th eigentriple.

### Step 3: Grouping

Grouping is the most important step of SSA and determines the properties of the reconstructed components. In this step the eigentriples from the previous step are grouped in  $M$  disjoint groups based on their properties,  $M \leq L_r L_c$ .

$$\mathbf{W} = \sum_{k=1}^M \mathbf{W}_{\mathbf{I}_k} \quad (3.16)$$

with  $\mathbf{W}_{\mathbf{I}} = \sum_{i \in \mathbf{I}} \sqrt{\lambda_i} U_i V_i^T$  and  $I_1 \cup I_2 \cup \dots \cup I_m = \{1, \dots, d\}$ .

### Step 4: Projection

In order to obtain a decomposition of the input array a projection step is necessary moving back from a set of 4D objects to a set of 2D-arrays corresponding to the reconstructed images.

Let us explain in details the projection algorithm. Take one of the matrices from the previous step:

$$\mathbf{W}_{\mathbf{I}} = \begin{pmatrix} W_{1,1} & W_{1,2} & \cdots & W_{1,K_c} \\ W_{2,1} & W_{2,2} & \cdots & W_{2,K_c} \\ \vdots & \vdots & \ddots & \vdots \\ W_{L_c,1} & W_{L_c,2} & \cdots & W_{L_c,K_c} \end{pmatrix}$$

where  $W_{i,j}$  are  $L_r \times K_r$  blocks.

For the signal reconstruction,  $\mathbf{W}_{\mathbf{I}}$  and  $W_{i,j}$  have to be Hankel matrices. To make the Hankel matrices two sequential averaging procedures are necessary:

1. ‘*Within block*’ hankelization. Averaging  $(k,l)$ -entries of  $W_{i,j}$  with  $k+l = s$ ,  $1 \leq s \leq N_r$  to make a matrix with Hankel blocks  $\mathbf{W}'_{i,j}$ :

$$\mathbf{W}'_{\mathbf{I}} = \begin{pmatrix} W'_{1,1} & W'_{1,2} & \cdots & W'_{1,K_c} \\ W'_{2,1} & W'_{2,2} & \cdots & W'_{2,K_c} \\ \vdots & \vdots & \ddots & \vdots \\ W'_{L_c,1} & W'_{L_c,2} & \cdots & W'_{L_c,K_c} \end{pmatrix} \quad (3.17)$$

2. ‘Between block’ hankelization. Averaging the blocks  $\mathbf{W}'_{i,j}$  with  $i+j = s$ ,  $1 \leq s \leq N_c$  to make Hankel the blocks obtaining an Hankel-bolck-Hankel matrix

$$\widetilde{\mathbf{W}}_{\mathbf{I}} = \begin{pmatrix} \tilde{\mathbf{H}}_1 & \tilde{\mathbf{H}}_2 & \cdots & \tilde{\mathbf{H}}_{K_c} \\ \tilde{\mathbf{H}}_2 & \tilde{\mathbf{H}}_3 & \cdots & \tilde{\mathbf{H}}_{K_c+1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{H}}_{L_c} & \tilde{\mathbf{H}}_{L_c+1} & \cdots & \tilde{\mathbf{H}}_{N_c} \end{pmatrix} \quad (3.18)$$

reverting step 1 the reconstructed images are obtained.

For the practical implementation of the algorithm, the grouping step has been performed after step 4. Each component will produce an output that is subsequently summed with other components in order to obtain the proper grouping.

### 3.3 Circulant SSA

Here I want to introduce a method that differs from basic SSA: Circulant SSA. It has been introduced by Bógalo in 2017<sup>14</sup> and developed by the same group in two articles Bógalo *et al.* 2018,<sup>15</sup> 2021.<sup>16</sup>

This method is not based on the SVD of the trajectory matrix but on the use of a circulant matrix (the Fourier matrix) for the decomposition. The advantage of using the Fourier matrix is that the signal is decomposed in band of frequencies with the reconstructed components ordered in term of frequency and not of magnitude of the associated eigenvalue. Sorting the components in terms of frequency enable an automatic extraction of the signal.

The method proposed differs from basic SSA in step 2 and 3. Let discuss Circulant SSA in details.

#### Step 2: Decomposition

In basic SSA the decomposition step is based on the SVD of the trajectory matrix or on the eigendecomposition of the covariance matrix. In Circulant SSA the eigendecomposition is applied to a circulant matrix (in a circulant matrix every row is a right cyclic shift of the row above) whose elements are

a scaled form of the sample second moment.

$$\mathbf{S}_C = \begin{pmatrix} \hat{c}_0 & \hat{c}_1 & \hat{c}_2 & \cdots & \hat{c}_{L-1} \\ \hat{c}_{L-1} & \hat{c}_0 & \hat{c}_1 & \cdots & \hat{c}_{L-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{c}_1 & \hat{c}_2 & \hat{c}_3 & \cdots & \hat{c}_0 \end{pmatrix} \quad (3.19)$$

$$\hat{c}_m = \frac{L-m}{L} \hat{\gamma}_m + \frac{m}{L} \hat{\gamma}_{L-m} \quad m = 0, 1, \dots, L-1 \quad (3.20)$$

$$\hat{\gamma}_m = \frac{1}{T-m} \sum_{t=1}^{T-m} x_t x_{t+m} \quad (3.21)$$

In equation 3.20  $\{\hat{\gamma}_m\}_{m=0}^{L-1}$  are the sample second moments defined as in equation 3.21. The diagonalization of  $\mathbf{S}_C$  produces the eigenvalues  $\hat{\lambda}_k$  that are associated to the  $k$ -th frequency  $w_k = \frac{k-1}{L}$ ,  $k = 1, \dots, L$ . They correspond to the power spectral density evaluated at frequency  $w_k$ . The analytical form for the eigenvalues is:

$$\hat{\lambda}_{L,k} = \sum_{m=0}^{L-1} \hat{c}_m \exp\left(i2\pi m \frac{k-1}{L}\right) \quad (3.22)$$

and for the associated eigenvectors is:

$$\mathbf{u}_k = L^{-1/2} (u_{k,1}, \dots, u_{k,L})' \quad (3.23)$$

with  $u_{k,j} = \exp(-i2\pi(j-1)\frac{k-L}{L})$  for  $k = 1, \dots, L$ .

### Step 3: Grouping

The power spectral density is symmetrical so  $\hat{\lambda}_k = \hat{\lambda}_{L+2-k}$ . Their corresponding eigenvectors given by 3.23 are complex, therefore, they are conjugated complex by pairs,  $\mathbf{u}_k = \mathbf{u}_{L+2-k}^*$  with  $\mathbf{u}^*$  complex conjugate of  $\mathbf{u}$ . In order to reconstruct the signal, the reconstructed matrices must be real. This is done using the conjugate transpose of the eigenvectors instead of the transposed as in equation 3.6.

$$\mathbf{X}_k = \mathbf{u}_k \mathbf{u}_k^\dagger \quad (3.24)$$

Due to the pairs of complex conjugate vectors  $\mathbf{X}_k$  and  $\mathbf{X}_{L+2-k}$  can be grouped producing a new elementary matrix that contains all the contribution at frequency band  $k$  for  $k = 2, \dots, M$  with  $M = \lfloor \frac{L+1}{2} \rfloor$

$$\begin{aligned} \mathbf{X}_{B_k} &= \mathbf{X}_k + \mathbf{X}_{L+2-k} \\ &= \mathbf{u}_k \mathbf{u}_k^\dagger \mathbf{X} + \mathbf{u}_{L+2-k} \mathbf{u}_{L+2-k}^\dagger \mathbf{X} \\ &= (\mathbf{u}_k \mathbf{u}_k^\dagger + \mathbf{u}_k^* \mathbf{u}_k') \mathbf{X} \\ &= 2(R_{\mathbf{u}_k} R_{\mathbf{u}_k}' + I_{\mathbf{u}_k} I_{\mathbf{u}_k}') \mathbf{X} \end{aligned} \quad (3.25)$$

where  $R_{\mathbf{u}_k}$  denotes the real part of  $\mathbf{u}_k$ ,  $I_{\mathbf{u}_k}$  its imaginary part and  $\mathbf{u}^\dagger$  the conjugate transpose of  $\mathbf{u}$ .  $B_k$  denotes the grouping  $B_k = \{k, L + 2 - k\}$  for  $k = 2, \dots, M$ ,  $B_1 = \{1\}$  and  $B_{\frac{L}{2}+1} = \{\frac{L}{2} + 1\}$  if  $L$  is even. After this procedure  $\mathbf{X}_{B_k}$  are real matrices that can be used in the last step of SSA.

### 3.4 New SSA methods

Basic SSA relies only on the parameter  $L$ . This can be a limit. Very long noisy time series or time series with a variable content are poorly extracted and decomposed if processed at once. In fact, using a big embedding window spreads the trend over a large number of components and is computationally expensive; on the other side, a small window has good performances for region with low noise intensity but is unable to remove it in region with high noise intensity.

If the trend has a complex shape, a large number of components is needed for reconstruction, this can be a problem if the last components of the trend contain not only the trend but also other components.

In this section I introduce a segmentation approach and an iterative procedure that can improve the performances of SSA solving that issues.

#### 3.4.1 Windowed (Overlap) SSA

The easiest way to deal with a very long series is to chop it and work with the shorter segments produced. In this way not only SSA computational cost is reduced but also the time-frequency characterization of the signal is improved.<sup>17</sup>

A pure segmentation approach has been proposed by Rekapalli and Tiwari<sup>18</sup>. In their work, a long time series containing data on the surface air temperature has been divided into seven window before applying SSA decomposition. The processed segments are then joint in order to reconstruct the entire length of the original series. As pointed out in Leles *et al.*<sup>19</sup>, in the approach proposed by Rekapalli and Tiwari<sup>18</sup> there is not treatment for the boundary effects arising at the union of consecutive segments.<sup>19</sup> The solution proposed by Leles is to use not a set disjointed segments, as in Rekapalli, but a set of overlapping segments.<sup>17,19</sup> With the overlap, the overlapped region are computed twice: once in the right segment and once in the left segment. In this way the points at the junctions are inner points and do not suffer for border effects. The reduction of border effects is due to the fact that the surroundings of the junction point are almost the same in the left and right segment.

If the properties of the signal strongly changes from one segment to the following, the border effect are small but still noticeable. In order to reduce the gap between one segment and the other I propose a new way to join segments. In particular, I have changed the kind of junction: from one point to a region. In this way, the transition between the two segments becomes gradual. Let define  $x_T$  the transition point,  $S_L$  the segment on the left of  $x_T$  and  $S_R$  the segment on the right. In the approaches by Rekapalli and Leles the jointed signal  $S$  is:

$$S(x) = \begin{cases} S_L(l_x) & \text{for } x < x_T \\ S_R(r_x) & \text{for } x \geq x_T \end{cases} \quad (3.26)$$

with a discontinuity at  $x = x_T$ .

In my approach with a transition region, the problem of a discontinuity is removed. In fact, there is a gentle transition between one segment to the other through the transition region. If  $N_T$  is the number of points involved in the transition region:

$$S(x) = \begin{cases} S_L(l_x) & \text{for } x \leq x_T - \frac{N_T}{2} \\ (1 - w_{i_x})S_L(l_x) + w_{i_x}S_R(r_x) & \text{for } x_T - \frac{N_T}{2} < x < x_T + \frac{N_T}{2} \\ S_R(r_x) & \text{for } x \geq x_T + \frac{N_T}{2} \end{cases} \quad (3.27)$$

with  $i_x = \{1, \dots, N_T - 1\}$  and  $w = \frac{1}{N_T}$ .

$l_x, r_x$  indices in the left (right) segment that correspond to the index  $x$  in the entire signal.

### 3.4.2 Iterative Trend Extraction SSA (ITE-SSA)

For a signal with a sharp trend, basic SSA acts as a smoothing. This can be a problem if we want to preserve the shape of the signal. The use of a large window increases the computational cost and spread the trend in many components producing a suboptimal decomposition. A small  $L$  easily extracts the gentle trend from the signal but completely fails to extract the other features of the signal.

In this section I propose a new iterative procedure that improve the extraction of the trend from the signal. In this procedure, after each SSA decomposition, the first reconstructed component is stored while the other components are grouped and undergo a new cycle of SSA. The reconstructed signal then is the sum of all the first components. In this way at each iteration the signal is concentrated in the first component while the noise remains in the other components. A pseudo code for the procedure is given in Algorithm 1

---

**Algorithm 1** Iterative trend extraction SSA

---

**Require:** Series  $\mathbf{X}$ , window length  $L$  and iterations number  $C$ **Ensure:** Complete extraction of the low frequency signal from the raw signal

- 1: Decompose the signal  $\mathbf{X}$  using basic SSA (Section 3.1.1)
  - 2:  $\mathbf{S} \leftarrow \mathbf{g}_1$
  - 3:  $\mathbf{Y} \leftarrow \sum_{i=2}^L \mathbf{g}_i$
  - 4:  $c \leftarrow 2$
  - 5: **while**  $c \leq C$  **do**
  - 6: Decompose  $\mathbf{Y}$  using basic SSA
  - 7:  $\mathbf{S} \leftarrow \mathbf{S} + \mathbf{g}_1^Y$
  - 8:  $\mathbf{Y} \leftarrow \sum_{i=2}^L \mathbf{g}_i$
  - 9:  $c \leftarrow c + 1$
  - 10: **end while**
  - 11: **Return**  $\mathbf{S}$
- 

The number of iterations  $C$  has to be selected empirically in order to make  $\mathbf{S}$  matching the expected properties.

**Improvement**

If the signal is strongly contaminated by noise, basic SSA should be unable to completely recover the signal. In basic SSA components are sorted according to their prevalence in the mixture, this does not assure that all signal components appear before noise ones. If some noise components precede signal ones, then part of the signal is lost or, in order to recover all the signal, part of the noise is included.

To overcome this shortcoming, I propose to apply after the ITE-SSA, a small number of refinement cycles of ITE-SSA with circulant SSA instead of basic SSA. In CiSSA components are sorted according to their frequency and not by their prevalence, this assure a univocal sorting of the components. Introduced in the iterative trend extraction, CiSSA extracts from residuals the components of the signal that weight less than the noise. In this way all the signal is reconstructed.



# Bibliography

- [1] N. Golyandina, V. Nekrutkin and A. Zhigljavsky, *Analysis of Time Series Structure SSA and related Techniques*, CHAPMAN & HALL/CRC, New York, 1st edn., 2001, p. 320.
- [2] D. S. Broomhead and G. P. King, *Physica D: Nonlinear Phenomena*, 1986, **20**, 217–236.
- [3] D. S. Broomhead and G. P. King, *Nonlinear phenomena and chaos*, 1986, 113–144.
- [4] N. Golyandina, A. Zhigljavsky and A. Korobeynikov, *Singular Spectrum Analysis with R*, Springer, 2018.
- [5] K. Lakshmi, A. R. M. Rao and N. Gopalakrishnan, *Structural Control and Health Monitoring*, 2017, **24**, e1960.
- [6] E. Bozzo, R. Carniel and D. Fasino, *Computers and Mathematics with Applications*, 2010, **60**, 812–820.
- [7] S. M. Shaharudin, N. Ahmad and F. Yusof, *AIP Conference Proceedings*, 2015, **1643**, 321–326.
- [8] R. Vautard, P. Yiou and M. Ghil, *Physica D: Nonlinear Phenomena*, 1992, **58**, 95–126.
- [9] A. Groth and M. Ghil, *Journal of Climate*, 2015, **28**, 7873–7893.
- [10] N. Golyandina and A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [11] N. Golyandina, *Statistics and Its Interface*, 2010, **3**, 259–279.
- [12] N. Golyandina, K. D. Usevich and I. V. Florinsky, *International Journal of Ecology & Development*, 2005, **8**, 81–94.

- [13] N. Golyandina and K. D. Usevich, *Matrix Methods: Theory, Algorithms and Applications: Dedicated to the Memory of Gene Golub*, 2010, 449–473.
- [14] J. Bógalo, P. Poncela and E. Senra, *MPRA Munich Personal RePEc Archive*, 2017, 76023.
- [15] J. Bógalo, P. Poncela and E. Senra, *Nonparametric Statistics*, Cham, 2018, pp. 295–309.
- [16] J. Bógalo, P. Poncela and E. Senra, *Signal Processing*, 2021, **179**, 107824.
- [17] M. C. R. Leles, J. P. H. Sansão, L. A. Mozelli and H. N. Guimarães, *Digital Signal Processing*, 2018, **77**, 63–76.
- [18] R. Rekapalli and R. K. Tiwari, *Journal of Geological Resource and Engineering*, 2014, **2**, 167–173.
- [19] M. C. R. Leles, J. P. H. Sansão, L. A. Mozelli and H. N. Guimarães, *SoftwareX*, 2018, **8**, 26–32.

# Chapter 4

## Principal Component Analysis

### Introduction

Principal Component Analysis (PCA) is crucial technique in multivariate data analysis. PCA has the objective to make a dataset more readable. This can be done searching for the most meaningful basis to re-express the dataset. The change of base allows also for a reduction of the system dimensionality. The axes of the new bases that do not contain information about the process are dropped or, in other word, only the directions containing the majority of the information are used to re-express the data.

The first idea of PCA appeared in 1901 in the work of Pearson<sup>1</sup> in which he introduced a method of lines and planes to rationalize a dataset. Further developments of the method were done by Hotelling in 1933.<sup>2,3</sup> It is after Hotelling that PCA received its name. Since then, PCA has gained interest in many diverse scientific fields, often with a different name. Among them, Singular Value Decomposition (SVD) and Karhunen-Loève expansion are often used to refer to PCA. The main fields of application of PCA are Engineering, Computer Sciences, Biological Sciences and Analytical chemistry with more than 10000 articles per year published in the last ten years (Scopus).

In this chapter I present PCA starting from its history then I introduce the basic method for PCA and some properties.

## 4.1 Basic PCA

### 4.1.1 History

#### Pearson

The first idea of PCA proposed by Pearson in 1901 was a method of “best fitting” that can be used to represent a set of points in plane or space using the “best fitting” line or plane.<sup>1</sup> In his seminal work he used the Least squares method to define the best line or plane that fits a dataset. He proposed also a geometrical interpretation using what he calls “ellipsoid of residuals” . He defined the ellipsoid of residuals and the correlation ellipsoid, then he express the properties of the best-fitting plane and line in terms of the correlation ellipsoid. Using the relation between the two ellipsoids he observed that the directions of independent variation and the standard deviation of the independent variables may be found from the ellipsoid of residuals. The use of the ellipsoid of residuals involves a much simpler arithmetic with respect to the use of the correlation ellipsoid.<sup>1</sup>

#### Hotelling

It is in the works of Hotelling that the method of principal component analysis starts developing.<sup>2-4</sup> In his 1933 work he shows the first examples of the use of PCA in order to reduce the dimension of a dataset.<sup>2,3</sup> In the same work he proposed a recursive method to compute the principal components sorted in terms of their weights in the dataset. With that method one can compute PCs until the desired amount of information is extracted from the dataset.<sup>3</sup>

### 4.1.2 Method

The main context for the application of PCA is in the reduction of dimensionality of a dataset with observation on many different variables. The method proposed in this section is inspired by Jolliffe and Cadima<sup>5</sup>.

If we define  $m$  the number of variables and  $n$  the number of observations, we have a set of  $m$  vectors of length  $n$   $\mathbf{x}_1, \dots, \mathbf{x}_m$  that can be seen as a matrix  $\mathbf{X}$  of shape  $n \times m$  whose columns are the observation of the variables. The first way we can use PCA is to search for a new basis of the dataset that is orthogonal. PCA searches for a linear combination of the original basis set that is orthogonal.

$$\mathbf{Y} = \mathbf{P}\mathbf{X} \tag{4.1}$$

with  $\mathbf{P}$  is the combination matrix. In order to obtain advantages from the change of bases, PCA searches for a linear combination that maximize vari-

ance. In this way the information content is preserved. In Pearson's definition of PCA this is done finding the line or plane that best fit the data.<sup>1</sup> It is possible to extract all the  $m$  orthogonal components of the dataset in a recursive way:

1. Select a normalized direction in  $n$ -dimensional space that maximize the variance of  $\mathbf{X}$ . Save the vector as  $\mathbf{p}_1$ .
2. Find another direction, orthogonal to the previous direction, along with the variance of residuals is maximized. Save the new vector as  $\mathbf{p}_i$
3. Repeat until  $m$  vectors are selected.

The resulting set  $\mathbf{P} = \{\mathbf{p}_1, \dots, \mathbf{p}_m\}$  contains the principal components of the dataset.

Maximizing the variance of a vector means that the autocovariance is maximum and the covariance with the other vectors is zero. A matrix of this form is diagonal. Starting from this property we can use a linear algebra approach that considers the linear combination that diagonalize the covariance matrix of the dataset. A diagonal covariance matrix,  $\mathbf{C}_Y$ , assure that the basis vectors are orthogonal and the variance is maximized

$$\mathbf{C}_Y = \frac{1}{n-1} \mathbf{Y}\mathbf{Y}^T \quad (4.2)$$

where  $\mathbf{Y}$  is an  $(n \times m)$  matrix with  $Y_{(i,j)} = x_{ij} - \bar{x}_j$ ,  $\bar{x}_j$  mean for the variable  $j$ .

If we express  $\mathbf{C}_Y$  in terms of  $\mathbf{P}$  we have:

$$\begin{aligned} \mathbf{C}_Y &= \frac{1}{n-1} \mathbf{Y}\mathbf{Y}^T \\ &= \frac{1}{n-1} (\mathbf{P}\mathbf{X})(\mathbf{P}\mathbf{X})^T \\ &= \frac{1}{n-1} \mathbf{P}\mathbf{X}\mathbf{X}^T\mathbf{P}^T \end{aligned} \quad (4.3)$$

$$\begin{aligned} &= \frac{1}{n-1} \mathbf{P}(\mathbf{X}\mathbf{X}^T)\mathbf{P}^T \\ \mathbf{C}_Y &= \frac{1}{n-1} \mathbf{P}\mathbf{A}\mathbf{P}^T \end{aligned} \quad (4.4)$$

with  $\mathbf{A}$  that is symmetric. We can then diagonalize  $\mathbf{A}$  as follows

$$\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^T \quad (4.5)$$

where  $\mathbf{D}$  is the diagonal matrix of eigenvalues and  $\mathbf{E}$  is the matrix of eigenvectors. If we take  $\mathbf{P} = \mathbf{E}$ , then we can re-express  $\mathbf{C}_Y$ :

$$\begin{aligned}\mathbf{C}_Y &= \frac{1}{n-1} \mathbf{P}(\mathbf{P}^T \mathbf{D} \mathbf{P}) \mathbf{P}^T \\ &= \frac{1}{n-1} (\mathbf{P} \mathbf{P}^T) \mathbf{D} (\mathbf{P} \mathbf{P}^T)\end{aligned}\tag{4.6}$$

$$\begin{aligned}&= \frac{1}{n-1} (\mathbf{P} \mathbf{P}^{-1}) \mathbf{D} (\mathbf{P} \mathbf{P}^{-1}) \\ \mathbf{C}_Y &= \frac{1}{n-1} \mathbf{D}\end{aligned}\tag{4.7}$$

using the property  $\mathbf{P}^T = \mathbf{P}^{-1}$ . In this way we have proved that the normalized eigenvalues of  $\mathbf{X} \mathbf{X}^T$  are the variance of  $\mathbf{X}$ . The products of the eigenvectors  $\mathbf{X} \mathbf{p}_k$  are the principal components of  $\mathbf{X}$ .

In standard PCA terminology, the elements of the eigenvectors  $\mathbf{p}_k$  are called PC *loadings*, whereas the elements of the linear combinations  $\mathbf{X} \mathbf{p}_k$  are called PC *scores*, as they are the values that each individual would score on a given PC.

## SVD

Another way to express PCA is through the singular value decomposition (SVD) of  $\mathbf{X}$ .

$$\mathbf{X} = \mathbf{U} \mathbf{S} \mathbf{V}^T\tag{4.8}$$

where  $\mathbf{X}$  is the  $n \times m$  matrix of the measurements,  $\mathbf{U}$  is an  $(n \times m)$  orthogonal matrix of the left singular vectors,  $\mathbf{S}$  is the diagonal matrix of the singular values and  $\mathbf{V}$  is an  $(m \times n)$  orthogonal matrix of the right singular vectors. The right singular vectors correspond to the eigenvectors of the square matrix covariance matrix.

**proof** : Consider  $\mathbf{X}^T \mathbf{X}$

$$\mathbf{X}^T \mathbf{X} = \mathbf{V} \mathbf{S}^T \mathbf{U}^T \mathbf{U} \mathbf{S} \mathbf{V}^T = \mathbf{V} \mathbf{S}^2 \mathbf{V}^T\tag{4.9}$$

In this way the eigenvalues of the covariance matrix are equals to the square of  $\mathbf{S}$

## Assumptions and Limits

The PCA method discussed above starts from strong assumption and has some limitations. The first assumption is *linearity*. It considers the dataset

as a linear combination of independent processes.

The second assumption is that the probability distribution of the data is considered fully described by mean and variance. This is not always true. In this case the decomposition leads to unsatisfactory results.

The sorting of PCs is based on the associated variance, this is possible if the assumption that large variance correspond to high information content is true.

### 4.1.3 PCA Reconstruction

Using the principal components it is possible to reconstruct the dataset using only a selected set of PCs in order to remove the unwanted information. The reconstruction is done projecting the dataset onto the PCs

$$\hat{\mathbf{X}} = \mathbf{P}\tilde{\mathbf{X}} + \mu = \tilde{\mathbf{X}}\mathbf{W}\mathbf{W}^T + \mu \quad (4.10)$$

with  $\tilde{\mathbf{X}}$  matrix of the centered and normalized signal,  $\mu$  vector of the means,  $\mathbf{P}$  matrix containing the selected PCs and  $\mathbf{W}$  matrix of the corresponding eigenvectors organized as columns.

$\hat{\mathbf{X}}$  differs from  $\mathbf{X}$  due to the different information content. The difference  $\mathbf{X} - \hat{\mathbf{X}}$  is called reconstruction error, or residuals, and accounts for the information content that is not included in the reconstructed signal.

### 4.1.4 Implementation

In the sections above I have presented the main properties of PCA. In this section I make a summary on how to apply PCA on a real dataset.

#### Step 1: Preprocessing

For an optimal performance, PCA needs that the variables in the dataset contribute equally to the analysis. To assure this, each variable is centered and normalized.

$$\tilde{\mathbf{X}}' = \mathbf{X} - \mu_v \quad (4.11)$$

$$\tilde{\mathbf{X}} = \tilde{\mathbf{X}}'/\sigma_v \quad (4.12)$$

#### Step 2: Covariance matrix computation

The standardized dataset is used to compute the covariance matrix between the different variables. In this way we obtain a  $(m \times m)$  square matrix. Covariance matrix is symmetric.

**Step 3: Eigendecomposition**

The covariance matrix is diagonalized in order to maximize the variance and minimize the correlation between the variables. This step produces a set of  $m$  eigenvalues that are sorted in decreasing order of magnitude. In this way we have a unique way to express the decomposition with the eigenvectors ordered by their significance.

**Step 4: Selection of the PCs**

From the set of eigenvectors, only those associated with the desired level of variance are selected to represent the dataset.

**Step 5: Reconstruction**

The set of selected PCs is used to build the new dataset using the equation proposed in subsection 4.1.3. In this way a new dataset containing only the information enclosed in the selected PCs is obtained.

# Bibliography

- [1] K. Pearson, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1901, **2**, 559–572.
- [2] H. Hotelling, *Journal of Educational Psychology*, 1933, **24**, 417–441.
- [3] H. Hotelling, *Journal of Educational Psychology*, 1933, **24**, 498–520.
- [4] H. Hotelling, *Biometrika*, 1936, **28**, 321–377.
- [5] I. T. Jolliffe and J. Cadima, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2016, **374**, 20150202.



# Part II

## Applications



# Chapter 5

## SSA and XRD

### Introduction

Crystals play a key role in materials science with new properties arising from their structure. Knowing or predicting the crystal structure of a solid is of great importance in order to rationalize its properties and design new materials.

Among the numerous techniques available for solid state analysis, X-ray diffraction (XRD) is the method of choice for crystal structure determination, in particular, single crystal diffraction. Unfortunately, single crystals are not available for many materials and the structure determination relies on powder diffraction methods.

In a collaboration with professor Giuseppe Falini of Biocrystallography and Biomineralization group in our department and the group of professor Boaz Pokroy from the Department of Materials Science & Engineering, Technion-Israel Institute of Technology (Haifa, Israel), we developed a pre-processing technique to improve the quality of diffractograms for structure resolution. In particular, a novel recursive application of Singular Spectrum Analysis has been used in order to remove the background from our samples retrieving the majority of diffraction information.

### 5.1 Powder X-ray diffraction and Structure Determination

#### 5.1.1 Diffraction

In crystals, atoms or molecules are arranged in a three-dimensional periodic structure. It is this lattice that plays a central role in diffraction. When

X-rays, whose wavelength similar to the lattice distances, are incident on a crystal, interact in different ways: they can be scattered with exactly the same wavelength (elastic scattering), or with a small loss of energy (Compton scattering); other photons can be absorbed causing photoelectric effects.<sup>1</sup> Looking at the total intensity of the scattered X-ray beam,  $I_T$ , it is composed of several parts

$$I_T = I_E + I_C + I_{MC} + I_{BG}$$

where  $I_E$  is the elastic scattering,  $I_C$  Compton,  $I_{MC}$  the multiple scattering and  $I_{BG}$  the background intensity.<sup>2</sup> The elastic scattering for a crystal is not diffuse in all directions but it is arranged in a regular pattern that is related to the crystal lattice. Lattice in fact acts as interference grid with the position of diffraction spots that depends on the crystallographic distances according to the Bragg's Law<sup>3</sup>

$$\lambda = 2d_{hkl} \sin \Theta \quad (5.1)$$

where  $\lambda$  is the wavelength,  $\Theta$  is the diffraction angle and  $d_{hkl}$  is a normalized distance between crystallographic planes.

For a single crystal Bragg's law results in a grid of spots, called reflections that generate the reciprocal lattice. It is the Fourier transform of the crystal lattice. Given a set of randomly oriented crystals, reciprocal lattices are also oriented randomly resulting in a sequence of concentric bright circles, remember the dependence on  $\Theta$  (eq. 5.1). In powder X-ray diffraction, X-ray intensity is measured along a line producing a diagram Intensity *vs* diffraction angle  $\Theta$ .<sup>4</sup>

### 5.1.2 X-ray Sources

The typical X-ray source is the Sealed-tube, a glass tube under vacuum with two electrodes. When a very high voltage is applied through the electrodes a flux of electrons is generated. The electrons hit the anode and lose energy producing a continuous spectrum of X-rays known as white radiation. If the energy of electrons is higher than a threshold (that depends on the metal anode), a second spectrum, named characteristic radiation, is obtained. It is caused by electrons being knocked out of the K shell of an atom and then the electrons from the L shell cascading down into the vacancies in this K shell. The energy emitted in this process corresponds to the so-called K alpha and K beta lines.<sup>4,5</sup> A system of slits then selects the line and focus the beam.

The wavelength and intensity of this radiation are limited by the metal used for the anode and the electric potential applied to the tube. Another defect of conventional X-ray generators is that radiation is not focused and coherent. In order to solve this issue, in the 1980s, scientist started to use

synchrotrons as generators of bright X-rays. In synchrotrons, X-rays are produced using the property of accelerated electrons: when they change direction they emit energy. In this way, a strong pulse of highly coherent and monochromatic X-rays is produced.<sup>46</sup>

The use of synchrotron X-rays opened up the possibility to study new specimens with particular interest in unstable or small crystals that are impossible to study with traditional sources.

### 5.1.3 Rietveld method

The Rietveld method generates an analytical representation of a powder diffractogram.<sup>7</sup> This is extremely useful for quantitative phase analysis where we are interested in the amount of each phase in the mixture. The idea behind the Rietveld method is simple: Instead of analyzing the integrated peak intensities from a powder pattern in a single crystal-like fashion, the entire information content of a powder pattern is fitted with a model whose parameters are refined using a least squares procedure to optimize the fit. In this way it is possible to account for the accidental or systematic overlap of some reflections typical of powder diffractograms. Given a powder with  $p$  phases, the expression for the intensity ( $Y_c$ ), used in the Rietveld method is

$$Y_c(s) = \text{BKG} + \sum_p S_p \left( \sum_{hkl} m |F_T|^2 I C \right) \quad (5.2)$$

where  $s = 2 \sin \Theta / \lambda$ ,  $\lambda$  wavelength; BKG is a smooth function reproducing the background;  $S_p$  is a scale factor typical for each phase;  $m$  multiplicity;  $|F_T|^2$  square of the structure factor;  $C$  correcting factors;  $I$  line profile function.<sup>8</sup> The value of the difference  $Y_o(s) - Y_c(s)$  is then minimized.

In order to obtain a set of parameters with a physical meaning it is necessary to set the initial guess according to an hypothetical crystal structure. Many different method exist for structure determination from powder diffraction working in reciprocal (diffraction) space and in direct (crystal) space. In David *et al.*<sup>9</sup> a comprehensive review on the different methods for structure determination. All this method should be applied before Rietveld in order to define the initial guess.<sup>10</sup>

## 5.2 Method

Looking at a powder X-ray diffractogram (Figure 5.1a), it presents a sequence of sharp peaks of different height and width emerging from a fluctuating

background that increases its intensity at increasing diffraction angles. For this reason, small diffraction peaks may be covered by noise, especially at high angles, making difficult to solve the crystal structure. The peak position is one of the most important parameter for the structure determination and it is determined by the crystal lattice. Using also peaks of higher order increases the quality of the final structure.

In order to solve a structure using a powder diffractogram it is necessary to reduce the amount of noise. This is usually done averaging over a set of different measurement on the same sample. Another way should be the direct subtraction of the noise, but it needs a good model for the noise that it is not available.

Singular Spectrum Analysis (SSA) is a non-parametric technique that decomposes a signal into a set of interpretable components, namely a trend, periodic components and noise. In the case of powder x-ray diffractograms, the signal has a complex form with sharp peaks that are difficult to reconstruct.

In order to apply SSA to XDR it is necessary to consider that:

1. the signal has a complex form;
2. usually the diffractogram is a long series  $N \simeq 10^4$ ;
3. the content of the signal changes for the different region with the noise that increases in amplitude at the increasing of the diffraction angle;
4. some noise components can lead the components of the diffractogram.

All these considerations may prevent the application of SSA in an automatic way to XRD. We developed a new procedure that introduces the use of SSA in denoising of XRD.

Our procedure starts slicing the original series in a set of overlapping windows. Each slice is then processed individually using Iterative Trend Extraction SSA (ITE-SSA). In this method, the signal is reconstructed applying iteratively basic SSA to the residuals after storing the first component. In section 3.4.2 ITE-SSA is presented in details.

For SSA decomposition we used a fixed embedding window and changed the number of iteration for each slice. The number of iteration was set empirically as the maximum number of iteration before the introduction of noise in the reconstructed signal. If, at the end of iterations, residuals still contain signal, a small number of iteration using CiSSA should be applied. The reason is that CiSSA sorts components according to their frequency and not according to their weight. In this way, signal components, usually

containing low frequencies, precede the noise ones and can be easily extracted in an automatic procedure.

The reconstruction of the entire diffractogram is obtained joining the different slices with a smooth transition between two adjacent slices. The smooth transition is done using a weighted average for the 100 points around the transition point with the weight of the left slice decreasing and the weight of the right slice increasing

$$XRD_i = \begin{cases} XRD_{L_i} & \text{for } i \leq t - 50 \\ XRD_{L_i} * (1 - j) + XRD_{R_i} * j & \text{for } t - 50 < i < t + 50 \\ XRD_{R_i} & \text{for } i \geq t + 50 \end{cases} \quad (5.3)$$

with  $j = (i - (t - 50)) / 100$ ,  $t$  index of the transition point,  $XRD_L$  left slice,  $XRD_R$  right slice. In this way discontinuities at the end of each section are smoothed out and the reconstructed signal looks continuous. The overlap between adjacent windows is necessary to avoid border effects that distort the reconstructed signal.<sup>11</sup>

## 5.3 Results and Discussion

We tested our method on different sets of powder diffractograms recorded at ESRF in Grenoble.

### 5.3.1 First Set

The first set consists of a diffractogram of hydroxyapatite sample. It was recorded at room temperature,  $v_{scan} = 5 \text{ deg / min}$ , wavelength:  $0.4 \text{ \AA}$ . with an angular resolution  $\Delta\Theta = 0.001^\circ$ . The total number of points,  $N$ , is  $N = 32913$  The diffractogram is an average over 3 runs recorded with the same settings.

In figure 5.1a it is shown the given diffractogram. We can see how noise fluctuations increase for increasing values of  $2\Theta$ . Figure 5.1b presents the first attempt to denoise the diffractogram. We applied SSA on the entire dataset using an embedding window  $L = 329$ , a tenth of the diffractogram length. For the reconstruction only the first 10 components were used. Applying SSA on the entire diffractogram has the disadvantage that the content of each component changes changing the diffraction angle.

In order to avoid big differences in the content of SSA components, we divided the diffractogram in a set of overlapping windows and then we applied SSA on each of them. In this way, it is possible to tune both  $L$  and the number of components used for reconstruction.

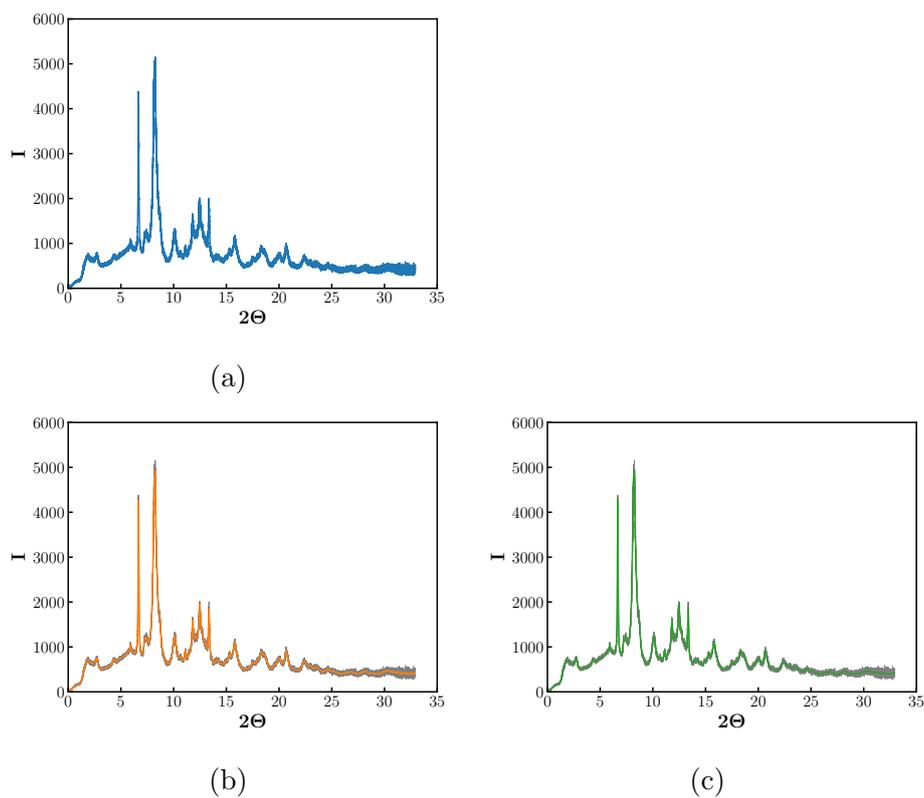


Figure 5.1: HRPXRD for Hydroxiapatite: (a) raw diffractogram for Hydroxiapatite; (b) denoised diffractogram using SSA on the entire series ( $L = 329$ , the first 10 components are used for reconstruction); (c) denoised diffractogram applying SSA on windowed signal. The original signal has been divided into 4 overlapping segments and reconstructed using a different number of components for each segment.

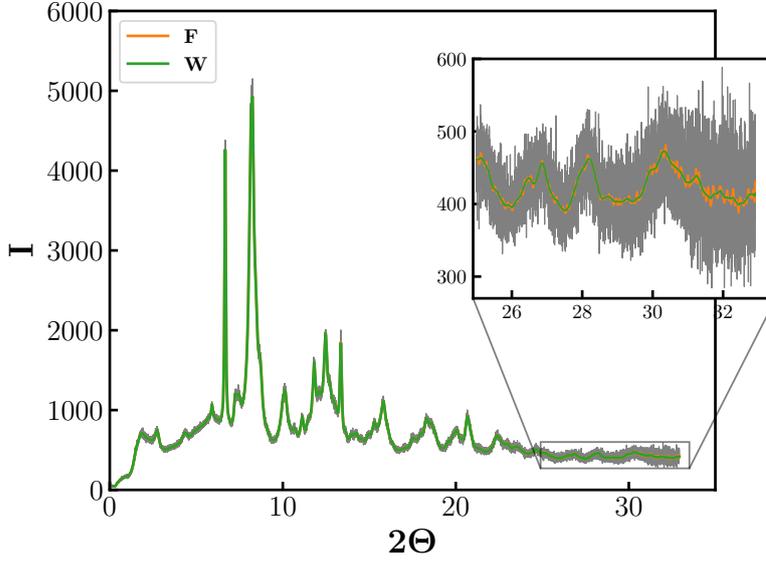


Figure 5.2: Comparison between the denoised diffractogram applying SSA on the entire signal (orange line) and on overlapping segments (green lines). We can notice that the latter approach removes also the fluctuations in the last part of the signal while SSA applied on the entire signal does not.

In figure 5.2 we show a comparison between our method and SSA applied on the entire SSA. In the inset it is evident how our method improves the reconstruction of the diffractogram also for high diffraction angles.

### 5.3.2 Second Set

The second set contains diffractograms for two Vaterite samples obtained at different scan rate. The measurement were carried at room temperature, wavelength:  $0.4 \text{ \AA}$ . with an angular resolution  $\Delta\Theta = 0.005^\circ$ . The scanning rate is changed from  $v_{scan} = 5 \text{ deg/min}$  to  $v_{scan} = 25 \text{ deg/min}$  with step of  $5 \text{ deg/min}$

Due to the low angular resolution, the embedding window has to be small in order to reduce smoothing effects. Using a small window has the disadvantage that signal and noise are poorly divided. Our iterative method improves the reconstruction extracting more information than basic SSA.

In Figures 5.4 and 5.5 we show the results of this method of reconstruction. Looking in particular at Figure 5.5c,d we notice that reconstruction is not complete, with a big amount of information left in residuals.

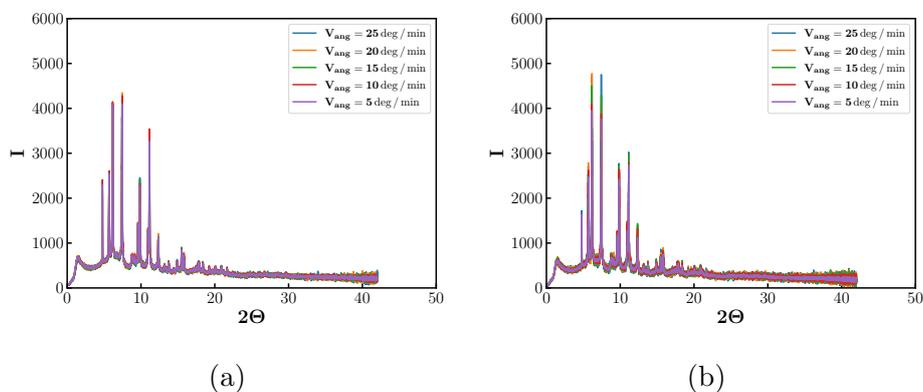


Figure 5.3: HRPXRD for Vaterite: (a) Vaterite sample 1; (b) Vaterite sample 2. The plot is an average over 5 runs with the same experimental setup for each angular speed. At the increasing of the angular speed, the amplitude of noise increases.

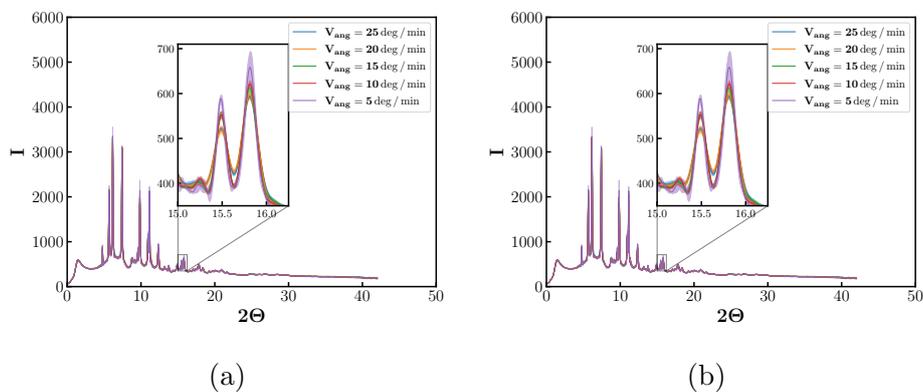


Figure 5.4: Iterative Trend Extraction SSA applied to HRPXRD for Vaterite: (a) Vaterite sample 1; (b) Vaterite sample 2. For each angular speed the average reconstructed signal is plotted with the standard deviation area. The reconstructed signal is similar for the different scan speeds. The inset zoom shows that, increasing the scan speed, the signal extracted decreases.

From this result we conclude that iterative application of SSA, as ITE-SSA, are not enough for a good reconstruction in an automated way.

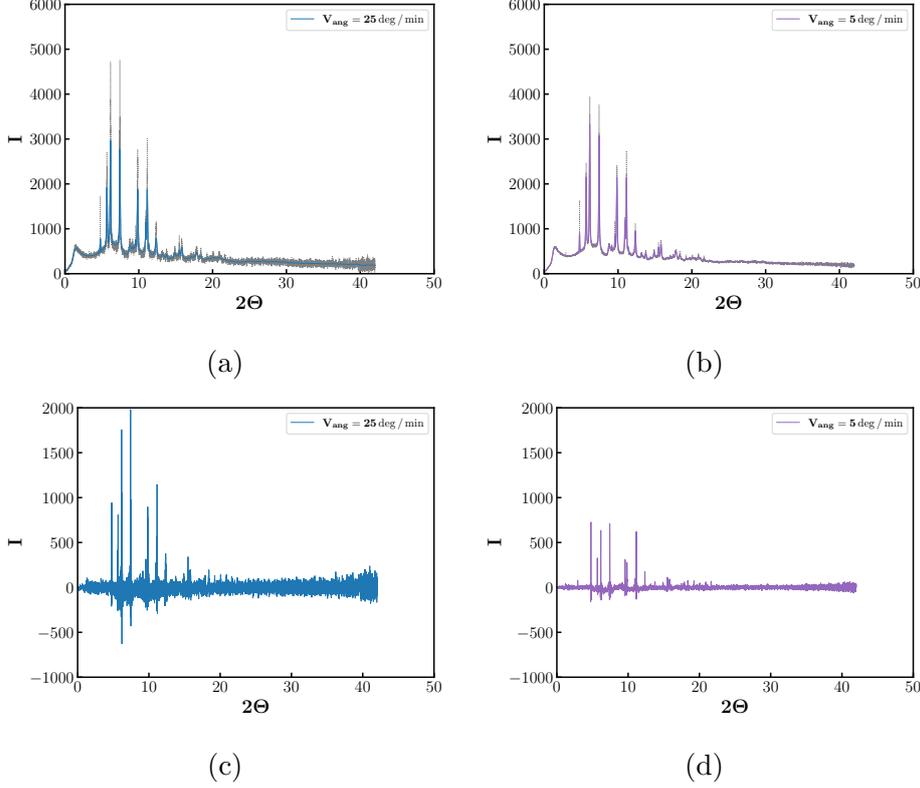


Figure 5.5: Vaterite HRPXRD denoised using ITE-SSA,  $L = 64$ . The number of cycles depends on the section of the signal. (a) Vaterite sample 2 recorded at  $V_{ang} = 25 \text{ deg/min}$ ; (b) Vaterite sample 2 recorded at  $V_{ang} = 5 \text{ deg/min}$ . Using ITE-SSA the signal is reconstructed, but a large amount of signal remains in the residuals as shown in (c) for  $V_{ang} = 25 \text{ deg/min}$  and in (d) for  $V_{ang} = 5 \text{ deg/min}$ . The higher the scanning speed the greater the amount of signal left.

The bigger problem for an automated reconstruction process is that periodic components come before components containing information of the peaks in the decomposition sequence making the selection of components not automatable.

Circulant SSA (CiSSA) decomposes a given signal according to the Fourier matrix generated by the embedding window. In this way, components are sorted by frequency and not according to the magnitude of their eigenvalue making automatic reconstruction easier.

CiSSA alone is unable to perform well in the first cycle of decomposition and extraction. In order to improve the automatic reconstruction, CiSSA is applied after the usual recursive reconstruction. Doing so allow the extraction from residuals of diffractogram components that come after periodic components of the noise. Carefully tuning the number of recursion for classical SSA and for CiSSA allow for an almost complete denoising.

For our dataset the number of iteration for ITE-SSA was manually selected depending on the given segment: we stop iteration one cycle before the appearance of periodic oscillation. After SSA cycles a sequence of CiSSA iteration is performed on residuals, the number of cycles is fixed for all the segments. The signal from CiSSA is the sum of the first component for each iteration plus the second component of the last iteration. Summing the signal from ITE-SSA and from CiSSA refinement results in the reconstructed signal.

As shown in Figure 5.6 and Figure 5.7, the application of CiSSA on the residuals improves the reconstruction of the signal, in particular, for the part with small peaks whose contribution to the signal is smaller than the one of the noise.

Usually, the reduction of noise is done accumulating many diffractograms. In this way noise cancels out and the signal is less noisy. With our denoising method, less runs are necessary for a good signal. This can speed up measurements.

In figure 5.8 we present the results for the average over five runs and two runs. In order to increase reproducibility, the accumulation of two runs has been computed for three independent couples.

In order to test the goodness of our method we perform the structure resolution using The Rietveld method comparing the results for the raw data and the denoised data. In table 5.1 and table 5.3, the results for the raw data are presented. At the increasing of the scan speed the quality of the structure reduces. In table 5.2 and table 5.4, the results for the processed data are presented. The structure resolution is more stable for the processed data. We can see a generalized increments in the values of the cell parameters for the processed XRD.

Looking at table 5.5 it is possible to see that our denoising method improves the quality of the structure resolution at all the scan speeds with a slightly bigger improvements for the high scan speeds. Remember that the high speeds correspond to a higher noise level in the raw data.

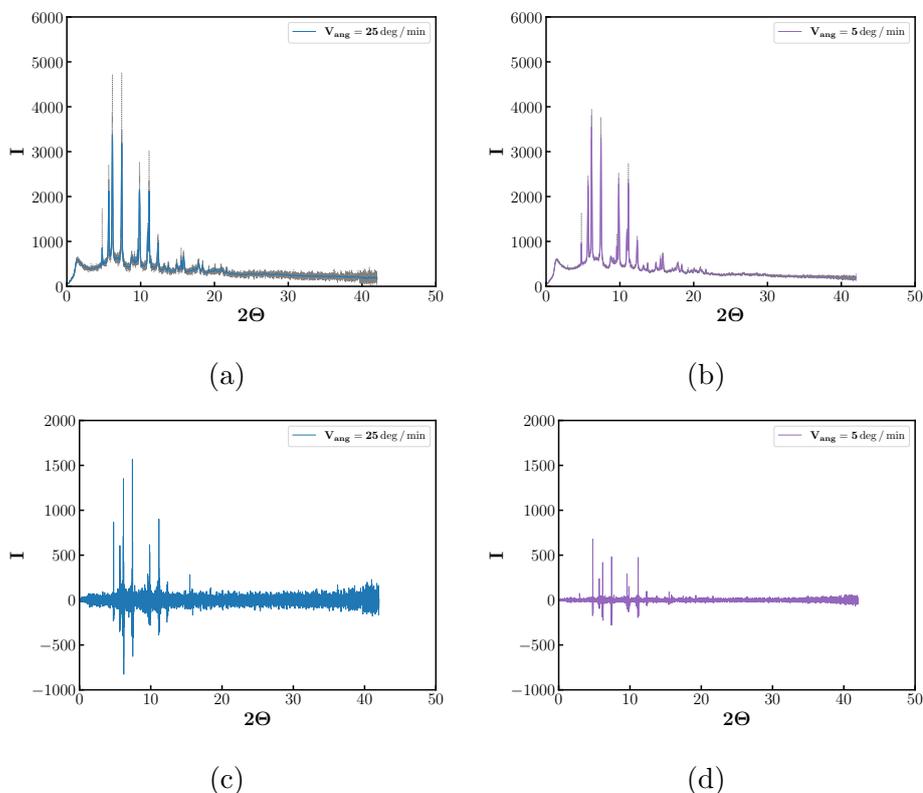


Figure 5.6: Vaterite HRPXRD denoised using ITE-SSA and subsequently iterative CiSSA refinement,  $L = 64$  for SSA and CiSSA. The number of cycles for SSA depends on the section, for CiSSA the number of iteration is fixed for all section. (a) Vaterite sample 2 recorded at  $V_{ang} = 25$  deg / min; (b) Vaterite sample 2 recorded at  $V_{ang} = 5$  deg / min. (c) Residuals for  $V_{ang} = 25$  deg / min; (d) Residuals for  $V_{ang} = 5$  deg / min. Applying CiSSA after ITE-SSA the residuals decrease

Table 5.1: Crystallographic parameters obtained with Rietveld method for the Vaterite sample 1 before SSA denoising

scan speed [deg/min]	$a$ [Å]	$da$ [Å]	$da/a$ [%]	$c$ [Å]	$dc$ [Å]	$dc/c$ [%]	$wR$ [%]	$GoF$
5	4.1255	0.0002	0.00485	8.4787	0.0005	0.0059	11.71	2.96
10	4.1256	0.0002	0.00485	8.4788	0.0005	0.0059	12.38	2.21
15	4.1256	0.0003	0.00727	8.4780	0.0005	0.0059	13.05	1.85
20	4.1257	0.0003	0.00727	8.4777	0.0005	0.0059	13.42	1.65
25	4.1261	0.0003	0.00727	8.4784	0.0005	0.0059	14.02	1.54

Table 5.2: Crystallographic parameters obtained with Rietveld method for the Vaterite sample 1 after SSA denoising

scan speed [deg/min]	$a$ [Å]	$da$ [Å]	$da/a$ [%]	$c$ [Å]	$dc$ [Å]	$dc/c$ [%]	$wR$ [%]	$GoF$
5	4.1268	0.0007	0.016 35	8.480	0.001	0.016 52	9.14	0.09
10	4.1267	0.0007	0.016 74	8.480	0.001	0.017 49	9.20	0.08
15	4.1265	0.0007	0.016 91	8.479	0.001	0.017 43	9.26	0.09
20	4.1268	0.0007	0.016 91	8.479	0.001	0.017 14	9.44	0.09
25	4.1270	0.0007	0.018 10	8.481	0.001	0.018 52	9.59	0.1

Table 5.3: Crystallographic parameters obtained with Rietveld method for the Vaterite sample 2 before SSA denoising

scan speed [deg/min]	$a$ [Å]	$da$ [Å]	$da/a$ [%]	$c$ [Å]	$dc$ [Å]	$dc/c$ [%]	$wR$ [%]	$GoF$
5	4.1257	0.0003	0.007 27	8.4824	0.0005	0.005 89	14.00	3.35
10	4.1256	0.0003	0.007 27	8.4825	0.0005	0.005 89	15.00	2.56
15	4.1257	0.0003	0.007 27	8.4827	0.0006	0.007 07	15.65	2.18
20	4.1261	0.0003	0.007 27	8.4824	0.0006	0.007 07	15.78	1.91
25	4.1264	0.0003	0.007 27	8.4818	0.0006	0.007 07	16.47	1.78

Table 5.4: Crystallographic parameters obtained with Rietveld method for the Vaterite sample 2 after SSA denoising

scan speed [deg/min]	$a$ [Å]	$da$ [Å]	$da/a$ [%]	$c$ [Å]	$dc$ [Å]	$dc/c$ [%]	$wR$ [%]	$GoF$
5	4.1277	0.0007	0.016 69	8.485	0.001	0.015 12	9.46	0.1
10	4.1273	0.0007	0.016 21	8.484	0.001	0.014 90	8.91	0.09
15	4.1272	0.0007	0.015 87	8.483	0.001	0.014 71	8.78	0.09
20	4.1270	0.0007	0.016 62	8.484	0.001	0.015 21	9.01	0.09
25	4.1281	0.0007	0.017 64	8.483	0.001	0.016 02	9.36	0.09

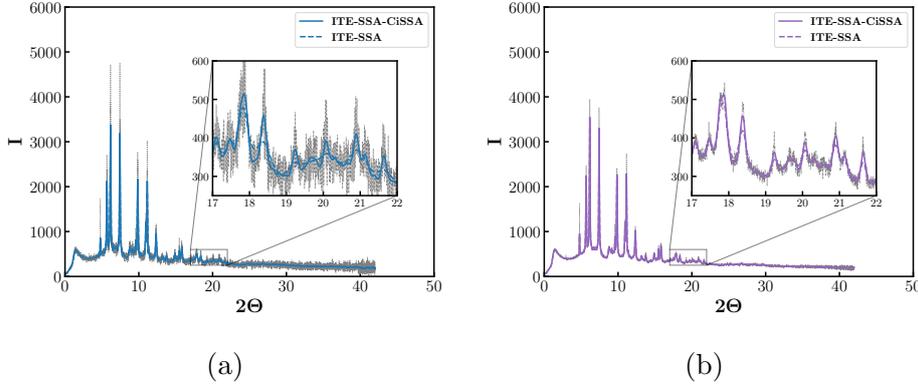


Figure 5.7: Comparison between ITE-SSA and ITE-SSA with CiSSA refinement. (a) comparison for  $V_{ang} = 25$  deg/min; (b) comparison for  $V_{ang} = 5$  deg/min. From the insets it is possible to see that CiSSA improves the reconstruction with new peaks emerging in the reconstructed signal

### 5.3.3 Third Set

The third set contains three diffractograms of a Vaterite sample. It was recorded at room temperature with  $v_{scan} = 5$  deg/min and an angular resolution  $\Delta\Theta = 0.002^\circ$ , wavelength:  $0.4 \text{ \AA}$ . The total number of points,  $N$ , is  $N = 14452$ . In this set diffractograms are the accumulation of many runs. With accumulation the amount of noise is reduced. Applying our denoising method produces a cleaner signal. The higher angular resolution, and so the greater number of points, improves the performances of the analysis: a larger embedding window can be used increasing the separability between signal and noise.

From the results emerges that, when the noise level is low, ITE-SSA is almost enough for the reconstruction with a very small amount of information left in the residuals. Only for high diffraction angles CiSSA refinement increases the amount of signal extracted.

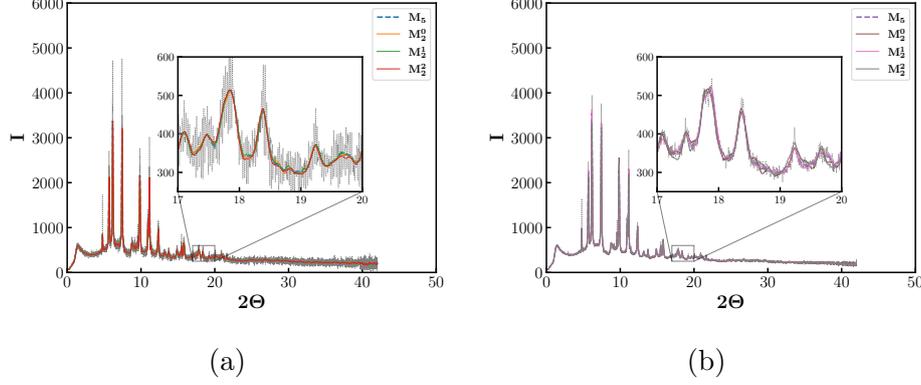
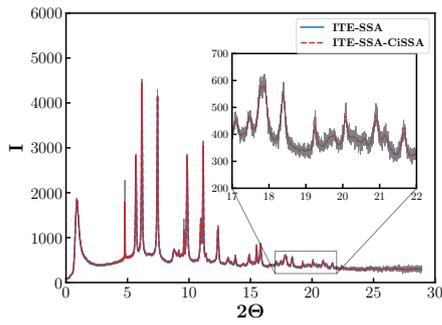


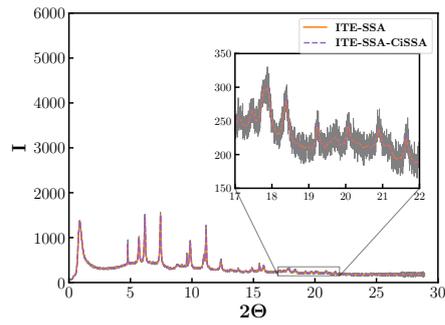
Figure 5.8: Analysis on the effect of the number of runs accumulated for the reconstruction of the signal. (a) comparison for  $V_{ang} = 25$  deg / min; (b) comparison for  $V_{ang} = 5$  deg / min.  $\mathbf{M}_5$  accumulation of five runs;  $\mathbf{M}_2^0$ ,  $\mathbf{M}_2^1$ ,  $\mathbf{M}_2^2$ , accumulation of two runs. For accumulation of two runs, three independent couples of runs are used. The variability of the independent accumulation is smaller for the slowest scanning speed. From the inset it is possible to see that, with CiSSA, the reconstruction accumulating two runs is similar to the one produced with accumulation of five runs. This can reduce the time of acquisition.

Table 5.5: Improvement for the Rietveld parameters after the SSA denoising

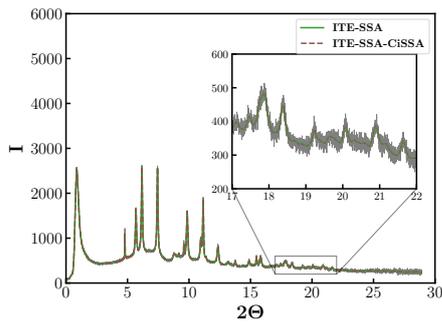
sample	scan speed [deg/min]	$wR$ [%]	$GoF$ [%]
V1	5	21.95	96.96
	10	25.69	96.38
	15	29.04	95.14
	20	29.66	94.55
	25	31.60	93.51
V2	5	32.43	97.01
	10	40.60	96.48
	15	43.83	95.87
	20	42.90	95.29
	25	43.17	94.94



(a)



(b)



(c)

Figure 5.9: Denoising for three different samples of Vaterite.  $L = 160$  for ITE-SSA,  $L = 80$  for CiSSA refinement. The low noise level makes CiSSA almost unnecessary. Improvement due to CiSSA are visible at high diffraction angles as shown in the insets. (a) Vaterite sample 1; (b) Vaterite sample 2; (c) Vaterite sample 3.

## 5.4 Conclusions

High Resolution Powder X-ray Diffraction is a powerful technique used to determine the crystal structure for a microcrystalline solid. The scan speed strongly affects the quality of the diffractogram and, consequently, the quality of the crystal parameters.

Our denoising method based on SSA can be used to improve the quality of diffractograms as a preprocessing for the Rietveld method. Improving the quality of the data after the acquisition can change the acquisition protocols. A good data quality can be obtained also for faster scan speed or a smaller number of scans for each sample. In this way, the number of analysed samples is increased or a sample that undergoes degradation is preserved due to the shorter exposure to the X-rays.

# Bibliography

- [1] C. Giacovazzo, H. L. Monaco, G. Artioli, D. Viterbo, M. Milanesio, G. Gilli, P. Gilli, G. Zanotti, G. Ferraris and M. Catti, *Fundamentals of Crystallography*, Oxford University Press, 2012.
- [2] *Underneath the Bragg Peaks*, ed. T. Egami and S. J. Billinge, Pergamon, 2012, vol. 16 of Pergamon Materials Series, pp. 2–480.
- [3] W. L. Bragg, *Proceedings of the Cambridge Philosophical Society*, 1913, **17**, 43–57.
- [4] *Powder Diffraction*, ed. R. E. Dinnebier and S. J. L. Billinge, The Royal Society of Chemistry, 2008, pp. p001–p582.
- [5] M. Laing, *An Introduction to the Scope , Potential and Applications of X-ray Analysis*, 1981, [http://www.iucr.org/\\_\\_data/assets/pdf\\_file/0008/13013/2.pdf](http://www.iucr.org/__data/assets/pdf_file/0008/13013/2.pdf).
- [6] P. Sedigh Rahimabadi, M. Khodaei and K. R. Koswattage, *X-Ray Spectrometry*, 2020, **49**, 348–373.
- [7] H. M. Rietveld, *Journal of Applied Crystallography*, 1969, **2**, 65–71.
- [8] P. Scardi, *Crystal Growth & Design*, 2020, **20**, 6903–6916.
- [9] *Structure Determination from Powder Diffraction Data*, ed. W. I. F. David, K. Shankland, L. B. McCusker and C. Bärlocher, Oxford University Press, Oxford, 2006.
- [10] R. E. Dinnebier, A. Leineweber and J. S. Evans, *Rietveld Refinement: Practical Powder Diffraction Pattern Analysis using TOPAS*, De Gruyter, 2018.
- [11] M. C. R. Leles, J. P. H. Sansão, L. A. Mozelli and H. N. Guimarães, *Digital Signal Processing*, 2018, **77**, 63–76.



# Chapter 6

## SSA and ECL Imaging

### Introduction

Electrochemiluminescence (ECL) is the emission of light induced by an initial electrochemical reaction at the electrode surface. It has been a successful transduction method in many analytical application thanks to its outstanding properties. The combination between electrochemical and spectroscopic methods allows for a better temporal and spatial control on light emission, a very low background and high sensitivity compared to chemiluminescence and photoluminescence.<sup>1</sup> All these properties make ECL a powerful analytical technique widely applied also in commercial applications.

Recently ECL was coupled with microscopy paving the way for new analytical application with imaging at the nanoscale.<sup>1</sup> As examples, with ECL were successfully observed microbeads<sup>2</sup> and cells.<sup>3</sup>

Even if ECL has intrinsically a good signal to noise ratio, the low emission intensity can be a limit. In order to capture even low intensities, the sensitivity of the detector and the time of acquisition have to be increased. For static measurement it is possible to increase the integration time in order to collect a greater amount of light. However, if we are interested in dynamical properties of the system, the integration time has to be short so it is the sensitivity of the detector that must be increased. One way to increase the sensitivity of detector is the use of CCD cameras with electron multiplier (EM-CCD). In EM-CCD a signal amplifier is placed immediately after the CCD sensor increasing the current generated by the photons.

Even if EM-CCD cameras are optimized for very low light emission, in ECL imaging we observed a very strong noise that corrupt the output images. Given the low intrinsic noise of ECL, it is crucial to find a method that is able to remove the instrumental noise. Another issue in ECL is the fading of

emission in time. The reason is not completely understood but it is supposed to be related to the degradation of the electrode.

In this work we propose 2D-SSA as a valid tool to remove the strong noise corrupting ECL and to allow the observation for short integration time. Decomposing the signal into a set of almost orthogonal components, SSA can separate the ECL emission from noise in a precise way.

We tested the performances of our procedure using images of microbeads acquired for different beads diameters and integration times. We used spherical beads because a well established acquisition protocol exists.

For the images acquisition we collaborated with Sara Rebecani and Giovanni Valenti from the EMFM (Electrochemistry of Molecular and Functional Materials) group of our department.

## 6.1 Electrochemiluminescence

Electrochemically induced chemiluminescence uses an electrochemical stimulus to activate a light emission. It is the electric potential between the electrodes in the electrochemical cell that promotes the formation of the active species responsible for the chemiluminescence reaction.

The first time ECL has been observed was in the 1960s through the use of rubrene, 9,10-diphenylanthracene (DPA), and similar compounds.<sup>4,5</sup> These compounds react through the annihilation pathway with oxidized and reduced form of the luminophore produced at the electrode during two different potential steps. This pathway has many limitation, in particular, the potential window needed for this reaction may exceed the potential window of water making the use of deoxygenated organic solvent necessary.<sup>6</sup>

Another ECL system is based on the coreactant pathway. It is not only the luminophore that reacts, but also another reactive species is reduced or oxidized at the electrode. In this way, the active species are generated at the same potential step. A key advantage of using the coreactant pathway is that it is possible to work in aqueous solution. In particular, tris(2,2'-bipyridine)ruthenium(II),  $[\text{Ru}(\text{bpy})_3]^{2+}$ , as the light emitting species, and tri-*n*-propylamine, TPrA, as coreactant, is the system of choice in many applications.<sup>1,7-9</sup> Many other ECL system are used on different applications, a short collection can be found in Forster *et al.*<sup>6</sup> and Richter<sup>10</sup>.

ECL emission can happen both in solution, with luminophore and coreactant freely diffusing in the electrochemical cell or in heterogeneous setup with the dye immobilized on the electrode or on objects adhering on the electrode.

The reaction mechanism of  $[\text{Ru}(\text{bpy})_3]^{2+}/\text{TPrA}$  has been widely explored

for liquid<sup>10,11</sup> and heterogeneous system.<sup>11</sup> In Sentic *et al.* they explored also the ECL emission in space studying the vertical distribution of emission. From the experiments, they observed that emission is strictly related to the stability of radicals generated at the electrode: the more stable they are the further they diffuse. This limits the emission to a thin layer of  $3/4\ \mu\text{m}$  of height.<sup>9</sup>

## 6.2 Method

### 6.2.1 ECL Images

The ECL emission of microbeads functionalised with  $[\text{Ru}(\text{bpy})_3]^{2+}$  has been studied for different acquisition time and beads size using an optical microscope equipped with a CCD camera with electron multiplier. All the measurements have been performed using the protocol defined in Zanut *et al.*<sup>12,13</sup> and Rebecani *et al.*<sup>14</sup>. According to the protocol, the electrochemical cell was equipped with a platinum working electrode and a Pt wire as counter electrode. A constant potential of 1.4 V (vs. Ag/AgCl) was applied for the established amount of time and the ECL emission was collected using an EM-CCD camera.

Two different types of measurements have been performed:

- *Long integration time*: a constant potential of 1.4 V (vs. Ag/AgCl) is applied for 4 s in 180 mM TPrA and 0.2 M phosphate buffer (PB). Integration time: 8 s; magnification:  $\times 100$ .
- *Transient ECL emission*: a constant potential of 1.4 V (vs. Ag/AgCl) is applied for 15 s. Pictures have been acquired sequentially with exposure of 200 ms; magnification:  $\times 100$ .

### 6.2.2 Images Elaboration

The acquired images have been processed using bi-dimensional Singular Spectrum Analysis (2D-SSA). 2D-SSA decomposes the raw signal into a set of components dividing the noise from the signal. In this way the ECL emission is cleaned from instrumental noise. Look at chapter 3 for details on SSA and 2D-SSA.

SSA decomposition has been performed using a FORTRAN subroutine coupled with a Python interface in order to exploit the capabilities of both programming languages. In particular, FORTRAN has been used for the

computationally intensive part of the analysis with capabilities further improved using GPGPU paradigms *OpenACC*<sup>15</sup> while Python has been used for the input-output interface.

The  $512 \times 512$  pixel images have been decomposed using a square embedding window of size  $L_x = L_y = 32$ . In order to avoid border effects, the images have been expanded adding on each side a border of size  $L_x(L_y)$  pixels containing the values of the last pixels on the edge. In this way picture is preserved with border effects acting only on dummy pixels.

ECL emission has been reconstructed looking at the W-correlation matrix in order to select the SSA components to be used in the process. The first  $n$  leading components in the W-correlation matrix have been selected in such a way that the components used for the reconstruction correlated among them and are uncorrelated to the others. Remember from section 3.1.3 that the W-correlation matrix gives information on the orthogonality between the SSA components and, as a consequence, on the intrinsic groups of the signal.

In order to evaluate the reconstruction, the signal to noise ratio (S/N) was computed as the ratio between the mean and the standard deviation of the pixels containing a bead.

$$\text{S/N} = \frac{\mu_{sig}}{\sigma_{sig}} \quad (6.1)$$

## 6.3 Results and Discussion

The different set of images have been processed with 2D-SSA. For images acquired with long acquisition time, the signal to noise ratio is good also for raw images ( $\text{S/N} > 1$ ). In the case of short acquisition time, the S/N ratio decreases in time, with ECL emission that decreases due to degradation of the electrode surface.

### 6.3.1 Long integration time

With long integration times the ECL emission of the beads clearly emerges from the background as presented in figure 6.1 and figure 6.3.

The signal to noise ratio, reported in table 6.1 and table 6.2 is good also for the raw image with values always larger than one.

Keeping fixed the embedding window for the SSA decomposition, the number of components containing the ECL emission changes changing the bead size as reported in figure 6.2 and figure 6.4.

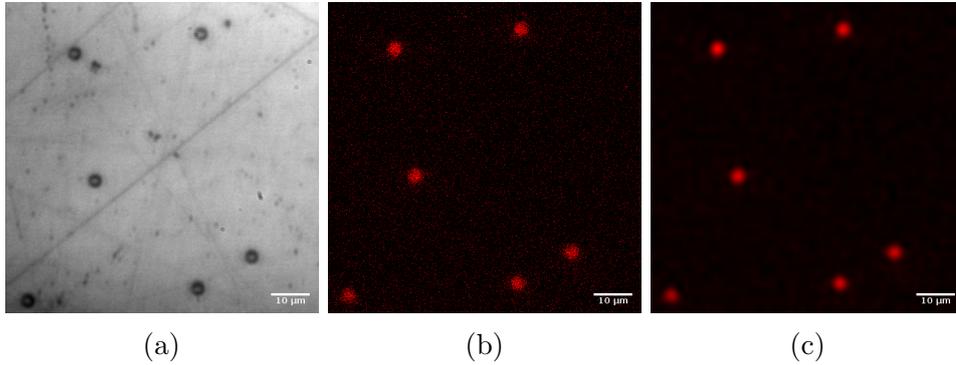


Figure 6.1: ECL emission for 2.8 $\mu\text{m}$  microbeads. Magnification:  $\times 100$ ; scale bar 10 $\mu\text{m}$ ; integration time: 8s. (a) optical; (b) raw ECL emission; (c) denoised ECL emission; SSA embedding windows,  $L_x = L_y = 32$ , reconstruction using the first 10 components.

Table 6.1: Signal to noise ratio for the 2.8 $\mu\text{m}$  microbeads depicted in figure 6.1. The intensities have been averaged over a region of  $11 \times 11$  pixels centred on the maximum of ECL emission

Bead	Raw	SSA processed
1	1.797	11.719
2	1.726	12.261
3	1.581	12.539
4	1.424	11.015
5	1.621	12.631
6	1.409	11.599
Mean	1.593	11.961

Table 6.2: Signal to noise ratio for the 8.9 $\mu\text{m}$  microbeads depicted in figure 6.3. The intensities have been averaged over a region of  $25 \times 25$  pixels centred on the maximum of ECL emission

Bead	Raw	SSA processed
1	2.428	10.062
2	2.467	11.691
3	2.217	9.012
Mean	2.371	10.255

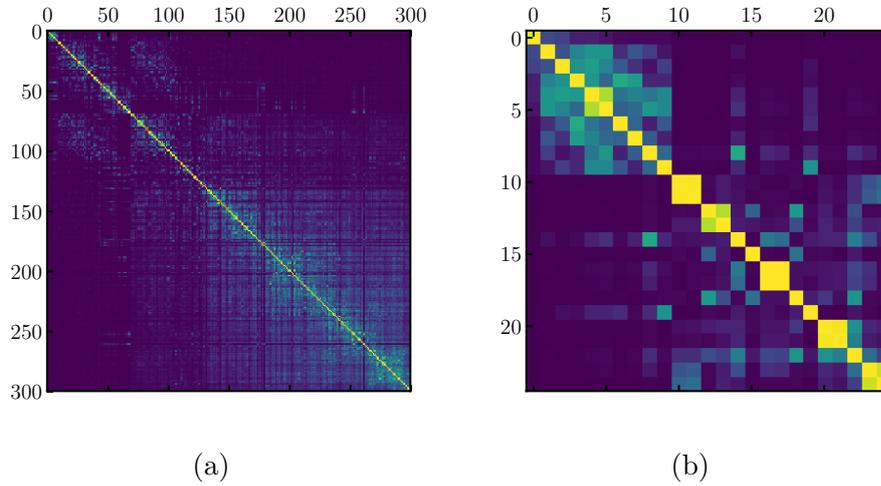


Figure 6.2: W-correlation matrix for the SSA decomposition for  $2.8\mu\text{m}$  microbeads in figure 6.1; SSA embedding windows  $L_x = L_y = 32$ . (a) correlation for the first 300 components; (b) correlation for the first 25 components. The first 10 components contain the ECL emission.

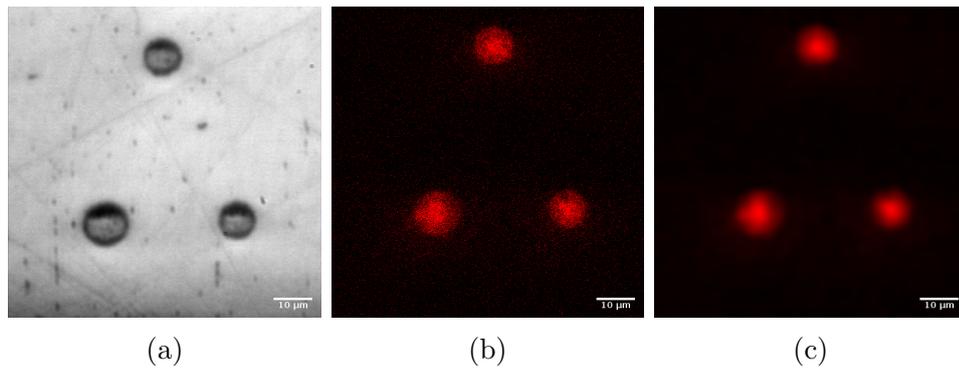


Figure 6.3: ECL emission for  $8.9\mu\text{m}$  microbeads. Magnification:  $\times 100$ ; scale bar  $10\mu\text{m}$ ; integration time: 8s. (a) optical; (b) raw ECL emission; (c) denoised ECL emission; SSA embedding windows,  $L_x = L_y = 32$ , reconstruction using the first 6 components.

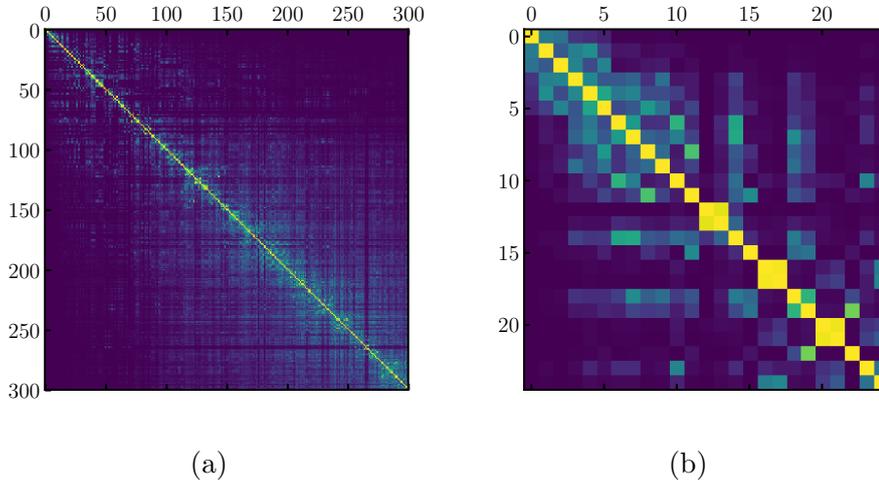


Figure 6.4: W-correlation matrix for the SSA decomposition for 8.9 $\mu\text{m}$  microbeads in figure 6.3; SSA embedding window  $L_x = L_y = 32$ . (a) correlation for the first 300 components; (b) correlation for the first 25 components. The first 6 components contain the ECL emission.

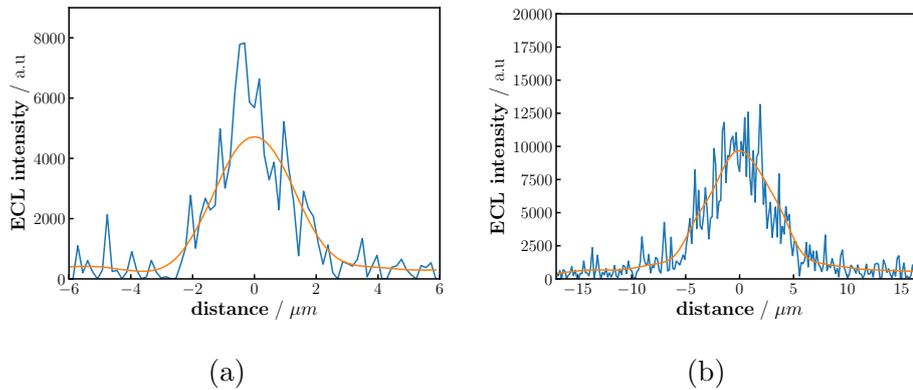


Figure 6.5: Examples of ECL emission profiles for microbeads, blue line raw data, orange line denoised data. Magnification:  $\times 100$ ; integration time: 8s. The reported profiles are an average over a line three pixels deep. (a) 2.8 $\mu\text{m}$  microbeads; (b) 8.9 $\mu\text{m}$  microbeads. SSA strongly reduces the noise highlighting the bell-shape of the intensity distribution.

From the analysis of the raw Intensities it is not clear if the emission happens through flashes of the individual luminophores or if it is constant and the flashes are due to the noise. Looking at the intensity distribution after SSA denoising it appears that emission is uniform on the surface of the beads with a bell-shaped distribution due to the spherical shape of the beads. For the 2.8 $\mu\text{m}$  microbeads the ECL emission show a regular distribution along the beads diameters, this should indicate that the diffusion of the coreactant is not a limiting factor for the reaction. In figure 6.5a the emission profiles, along the diameter, for a sample bead.

### 6.3.2 Transient ECL emission

With short integration times it is possible to record the variation of the ECL emission in time. For raw data the ECL emission is masked by the noise as shown in pictures 6.6b,c and pictures 6.8b,c. Applying SSA to the data has a huge effect in the S/N ratio moving from a prevalence of noise to a prevalence of signal as reported in table 6.3 and table 6.4. In pictures 6.6d,e and pictures 6.8d,e the denoised ECL emission are reported.

Looking at the ECL distribution for the 8.9 $\mu\text{m}$  microbeads (figure 6.8) emerges that it is not regularly distributed around the centre of the beads indicating an effect of the size on the ECL emission. This is due to the fact that the active coreactant is unable to reach the top of the beads.

The study of beads of different diameters shows that emission efficiency decreases at the increasing of bead diameter reaching a plateau. If we consider the emission efficiency of ECL, it is possible to define a Turnover frequency (TOF) as a function of bead size, expressed as the number of photons generated by a mole of luminophore per time unit

$$\text{TOF} = \frac{(\text{ECL}_{Ru@Bead} - \text{ECL}_{Bead})}{n^\circ \text{ of } [\text{Ru}(\text{bpy})_3]^{2+} \times t} \quad (6.2)$$

where  $\text{ECL}_{Ru@Bead}$  is the integrated ECL signal of a single bead,  $\text{ECL}_{Bead}$  is the background and  $[\text{Ru}(\text{bpy})_3]^{2+}$  is the amount of Ru luminophores.<sup>13</sup>

Another effect that is observable in the transient analysis is the fading of the ECL intensity in time. Figure 6.10 shows the emission profiles for two beads, of the two diameters, at two different times: (6.10a) and (6.10c) immediately after the potential step; (6.10b) and (6.10d) 1s after the potential step; the ECL intensity is clearly diminished.

In figure 6.11 the ECL intensity for the entire transient is reported. The decay is similar to a second order kinetics

$$\frac{1}{[ECL]} = \frac{1}{[E\tilde{C}L]_0} + kt^\alpha \quad (6.3)$$

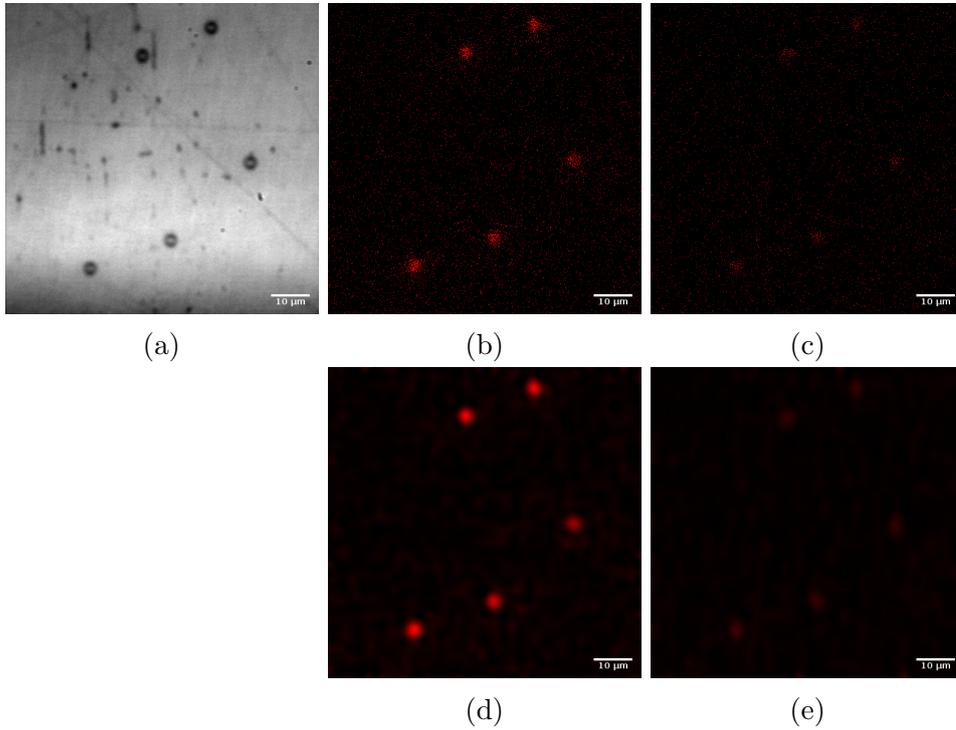


Figure 6.6: ECL emission for 2.8 $\mu$ m microbeads. Magnification:  $\times 100$ ; scale bar 10 $\mu$ m; integration time: 0.2s; SSA embedding windows,  $L_x = L_y = 32$ . (a) optical; (b) raw ECL emission frame 11, the first frame after the potential step; (c) raw ECL emission frame 16, 1s after the potential step; (d) denoised ECL emission frame 11, reconstruction using the first 4 components; (e) denoised ECL emission frame 16, reconstruction using the first 4 components;

Table 6.3: Signal to noise ratio for the 2.8 $\mu$ m microbeads depicted in figure 6.6. The intensities have been averaged over a region of  $11 \times 11$  pixels centred on the maximum of ECL emission. F11, frame 11; F16 frame 16.

Bead	Raw F11	SSA processed F11	Raw F16	SSA processed F16
1	0.723	14.156	0.386	15.624
2	0.725	12.859	0.345	20.152
3	0.668	11.934	0.414	13.010
4	0.704	13.256	0.428	17.107
5	0.515	14.101	0.324	15.546
Mean	0.667	13.261	0.380	16.288

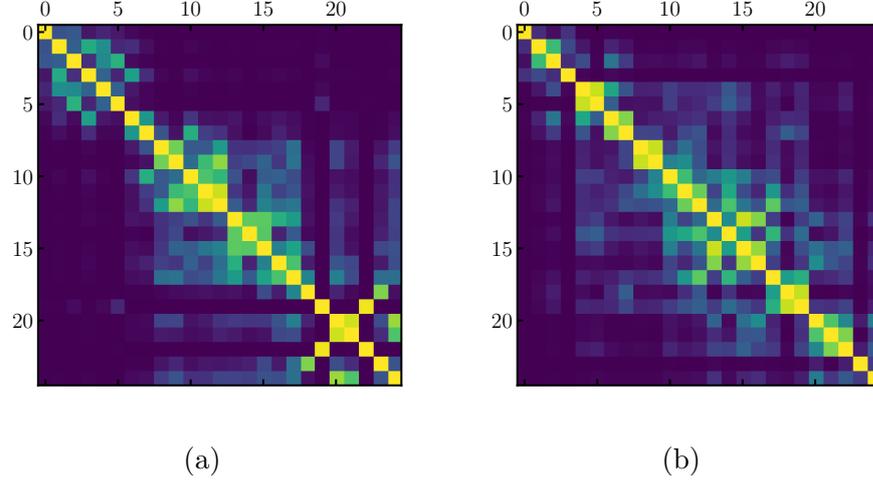


Figure 6.7: W-correlation matrix for the SSA decomposition for 2.8 $\mu$ m microbeads in figure 6.6, SSA embedding windows  $L_x = L_y = 32$ . (a) correlation for the first 25 components frame 11; (b) correlation for the first 25 components frame 16

Table 6.4: Signal to noise ratio for the 8.9 $\mu$ m microbeads depicted in figure 6.8. The intensities have been averaged over a region of  $25 \times 25$  pixels centred on the maximum of ECL emission. F11, frame 11; F16 frame 16.

Bead	Raw F11	SSA processed F11	Raw F16	SSA processed F16
1	0.562	6.024	0.384	8.997
2	0.572	8.760	0.399	12.019
3	0.512	5.343	0.402	9.589
Mean	0.549	6.709	0.395	10.202

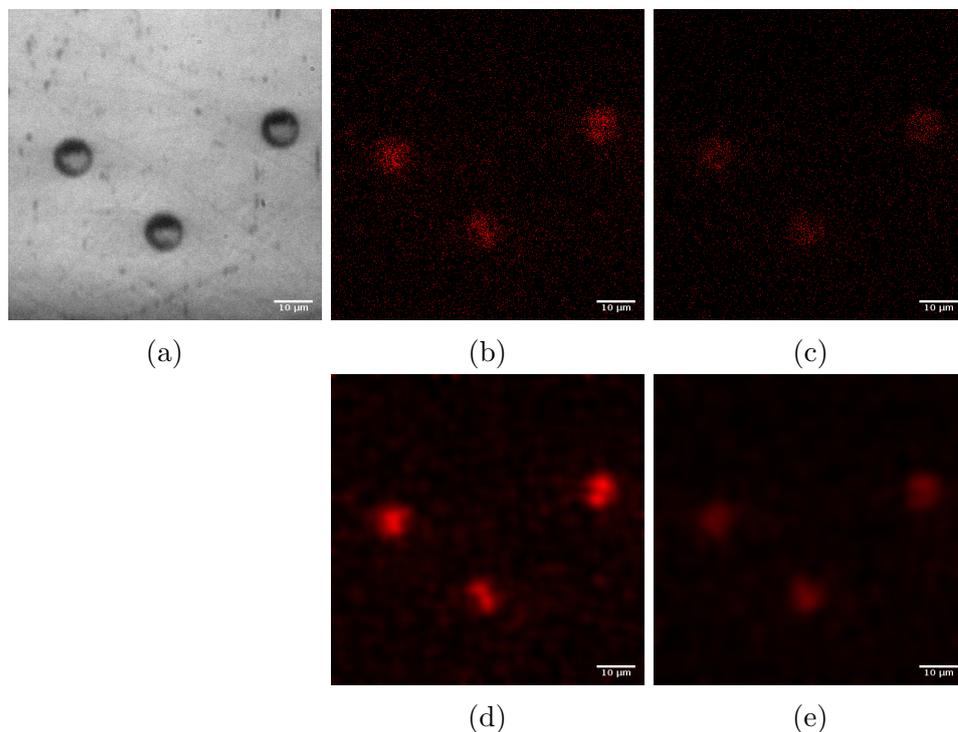


Figure 6.8: ECL emission for 8.9 $\mu\text{m}$  microbeads. Magnification:  $\times 100$ ; scale bar 10 $\mu\text{m}$ ; integration time: 0.2s; SSA embedding windows,  $L_x = L_y = 32$ . (a) optical; (b) raw ECL emission frame 11, the first frame after the potential step; (c) raw ECL emission frame 16, 1s after the potential step; (d) denoised ECL emission frame 11, reconstruction using the first 6 components; (e) denoised ECL emission frame 16, reconstruction using the first 3 components.

with  $[E\tilde{C}L]_0$  ECL intensity at  $t = 0$ ,  $k$  decay constant and  $\alpha$  time correction.

Superimposed to the decay the intensity presents a sequence of fluctuations. The origin of those fluctuations is not clear. They should be due to the fluctuation of the local concentration of the active coreactant that is consumed during the emission process, or they should be due to instrumental noise that SSA has not removed.

We observed that fluctuations are almost white with anticorrelated increments and the PDFs of the increments are not Gaussian with left tails.

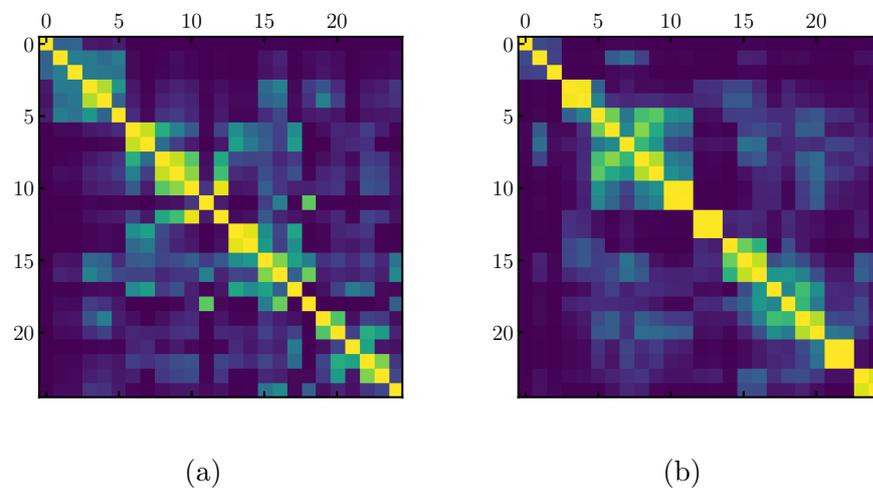


Figure 6.9: W-correlation matrix for the SSA decomposition for 2.8 $\mu$ m microbeads in figure 6.8; SSA embedding windows  $L_x = L_y = 32$ . (a) correlation for the first 25 components frame 11; (b) correlation for the first 25 components frame 16

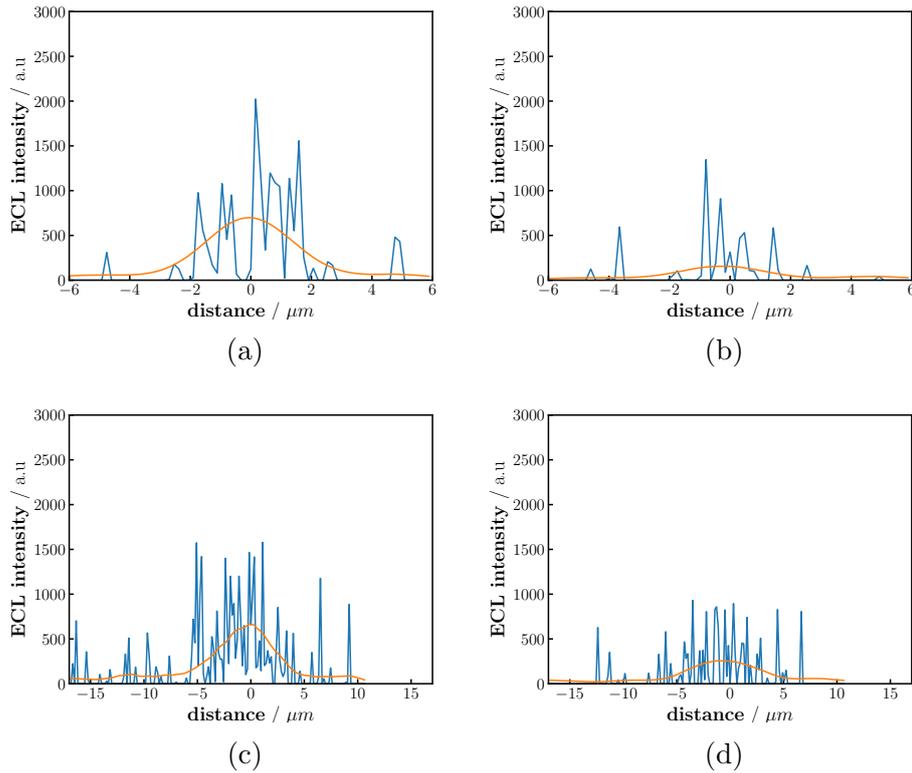


Figure 6.10: Examples of ECL emission profiles for microbeads, blue line raw data, orange line denoised data. Magnification:  $\times 100$ ; integration time: 8s. The reported profiles are an average over a line three pixels deep. (a)  $2.8\mu\text{m}$  microbeads at frame 11; (b)  $2.8\mu\text{m}$  microbeads at frame 16; (c)  $8.9\mu\text{m}$  microbeads at frame 11; (d)  $8.9\mu\text{m}$  microbeads at frame 16; SSA strongly reduces the noise highlighting the bell-shape of the intensity distribution. the ECL intensity strongly decreases int time.

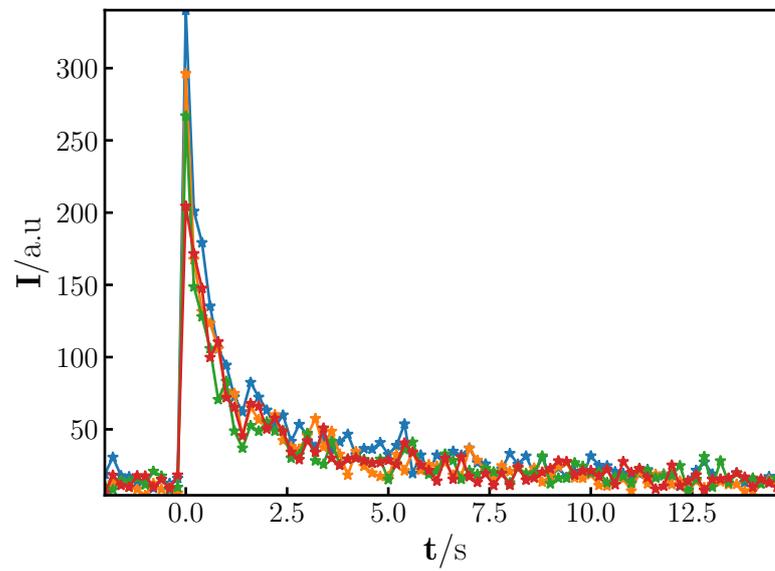


Figure 6.11: ECL emission intensity of the 1<sup>st</sup> SSA component for 2.8 $\mu$ m microbeads. Integration time: 0.2s; SSA denoising, with embedding windows  $L_x = L_y = 20$ , performed on each individual bead processing an area of  $100 \times 100$  centred around the maximum of emission at frame 11. Beads have been selected processing with SSA the entire image as in figure 6.6d.

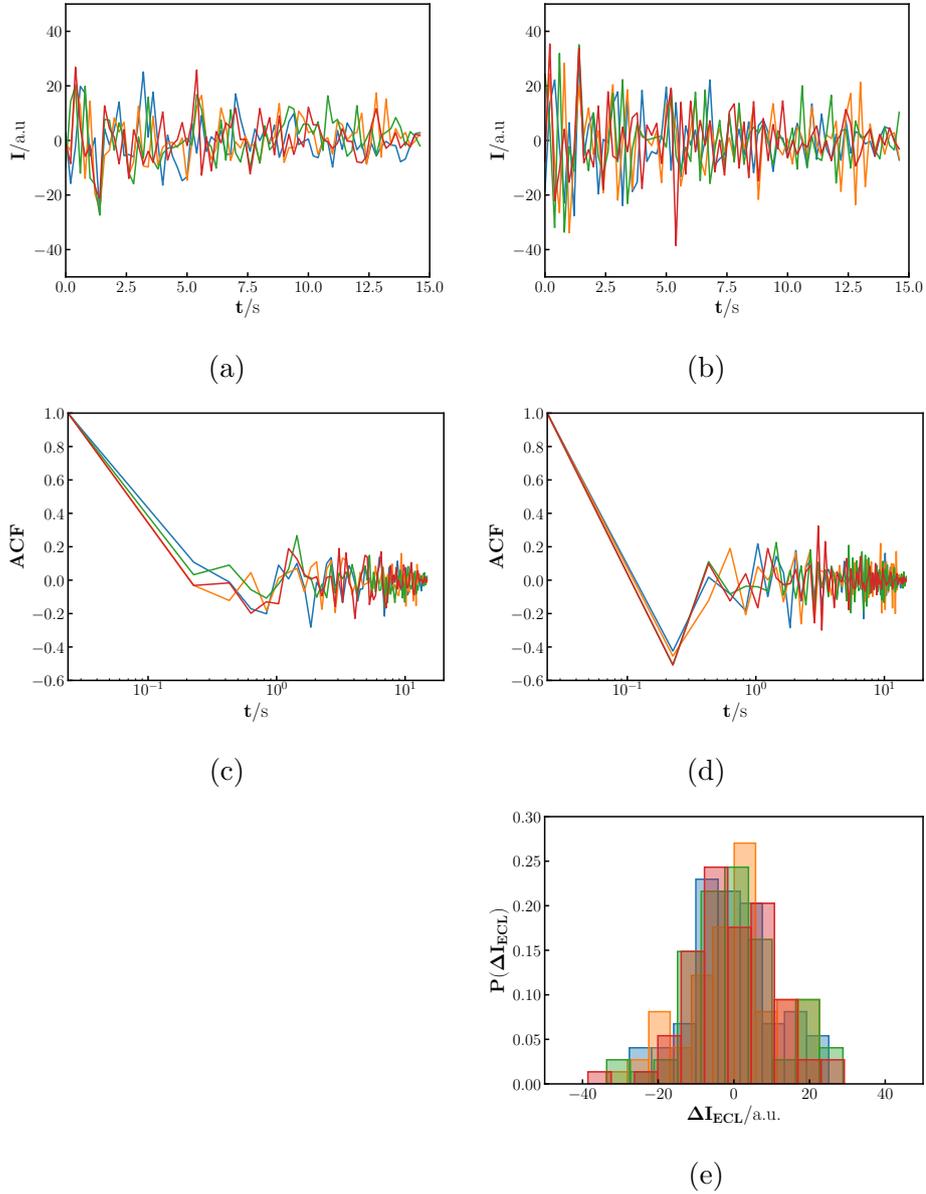


Figure 6.12: Analysis of the fluctuations of ECL intensity. (a) fluctuations; (b) increments; (c) ACF for the fluctuations; (d) ACF for the increments; (e) PDF for the increments. Fluctuations are almost white with anticorrelated increments. the PDFs of the increments are not Gaussian with left tails.

## 6.4 Conclusions

ECL imaging is a powerful technique that can be used for analytical purposes, but also for the study of the ECL mechanisms. Using Singular Spectrum Analysis we improved the quality of the ECL images reducing the S/N ratio. In this way we shed new light on the properties of ECL. The main result is on the distribution of ECL intensity on the bead, from the noisy images it was not clear if the emission was uniformly distributed on the entire surface or if the individual luminophores were emitting with flashes. The noise removal reveals that emission is uniform on the bead surface.

The study of the transient emission revealed a non-constant decay rate with fluctuations superimposed to a second order decay. Further studies are necessary to determine the nature of that fluctuations. Among the different hypothesis they should be due to noise components that remain after SSA decomposition or to a temporary reduction of the local concentration of active coreactant due to the reaction at the bead surface and the diffusion processes from and to the reaction sites.

# Bibliography

- [1] A. Zanut, A. Fiorani, S. Rebecani, S. Kesarkar and G. Valenti, *Analytical and Bioanalytical Chemistry*, 2019, **411**, 4375–4382.
- [2] G. Valenti, M. Zangheri, S. E. Sansaloni, M. Mirasoli, A. Penicaud, A. Roda and F. Paolucci, *Chemistry – A European Journal*, 2015, **21**, 12640–12645.
- [3] J. Xu, P. Huang, Y. Qin, D. Jiang and H.-y. Chen, *Analytical Chemistry*, 2016, **88**, 4609–4612.
- [4] R. E. Visco and E. A. Chandross, *Journal of the American Chemical Society*, 1964, **86**, 5350–5351.
- [5] K. S. Santhanam and A. J. Bard, *Journal of the American Chemical Society*, 1965, **87**, 139–140.
- [6] R. J. Forster, P. Bertoncello and T. E. Keyes, *Annual Review of Analytical Chemistry*, 2009, **2**, 359–385.
- [7] A. Fiorani, D. Han, D. Jiang, D. Fang, F. Paolucci, N. Sojic and G. Valenti, *Chemical Science*, 2020, **11**, 10496–10500.
- [8] S. Voci, B. Goudeau, G. Valenti, A. Lesch, M. Jović, S. Rapino, F. Paolucci, S. Arbault and N. Sojic, *Journal of the American Chemical Society*, 2018, **140**, 14753–14760.
- [9] M. Sentic, M. Milutinovic, F. Kanoufi, D. Manojlovic, S. Arbault and N. Sojic, *Chemical Science*, 2014, **5**, 2568.
- [10] M. M. Richter, *Chemical Reviews*, 2004, **104**, 3003–3036.
- [11] W. Miao, J.-P. Choi and A. J. Bard, *Journal of the American Chemical Society*, 2002, **124**, 14478–14485.

- [12] A. Zanut, F. Palomba, M. Rossi Scotto, S. Rebecani, M. Marcaccio, D. Genovese, E. Rampazzo, G. Valenti, F. Paolucci and L. Prodi, *Angewandte Chemie - International Edition*, 2020, **59**, 21858–21863.
- [13] A. Zanut, A. Fiorani, S. Canola, T. Saito, N. Ziebart, S. Rapino, S. Rebecani, A. Barbon, T. Irie, H. P. Josel, F. Negri, M. Marcaccio, M. Windfuhr, K. Imai, G. Valenti and F. Paolucci, *Nature Communications*, 2020, **11**, 1–9.
- [14] S. Rebecani, C. Wetzl, V. A. Zamolo, A. Criado, G. Valenti, F. Paolucci and M. Prato, *Chemical Communications*, 2021, **57**, 9672–9675.
- [15] *OpenACC*, <https://www.openacc.org/>.

# Chapter 7

## Electrocorticogram Analysis

### Introduction

The study of the brain involves many scientific areas from physiology and psychology to chemistry, physics and statistics. All these sciences strive to understand the intimate mechanisms that result in the behaviours of animals and humans, in the development of brain diseases, and in the effects of the environment on the brain.

In order to study the brain activity, different techniques are used such as functional magnetic resonance imaging (fMRI), magneto-encephalography (MEG), Electroencephalography (EEG) and electrocorticography (ECoG). The first three techniques non-invasively investigate the brain from the outside while ECoG is an invasive technique with electrodes placed directly on the brain in the skull.

In this project we want to obtain new information on the brain processes involved in naming task using ECoG for measurements. Using time series analysis we investigated the brain activity searching for a correlation with the speech in order to find a precursor signal for the speech. This will be useful for Brain Computer Interfaces.

### 7.1 State of the art

Electrocorticography is an electrophysiological invasive technique that allows the study of the brain activity with high spatial and temporal resolution. In this technique, a grid of electrodes is placed directly on the cortical surface. A craniotomy is necessary in order to expose the cortex thus, ECoG is performed on humans only in clinical applications for which surgery is already

planned. Among the reasons for a craniotomy the main ones are pharmacoresistant epilepsy and brain tumours.

Human ECoG data have been used to address neuroscience issues, such as map cortical function and investigate functional mechanism of brain oscillations. A limited use of ECoG data has been done in cognitive studies for the lack of data on the areas involved in cognitive processes; ECoG recording are performed only alongside surgical intervention and the grid placement is led by clinical requirements and not for neuroscientific purposes (Todaro *et al.*<sup>1</sup> and references therein).

It is known that the brain is a dynamic system working out of equilibrium with oscillating electric potential. The oscillations can be divided into two components, namely, rhythmic and arrhythmic. The rhythmic component is the oscillatory periodic component in the oscillations of the brain activity. The rhythm can be divided in frequency bands: delta (1-3 Hz), theta (4-8 Hz), alpha (9-12 Hz), beta (12-30 Hz) and gamma (> 30 Hz). These oscillatory rhythms characterise specific behavioural states e.g. delta oscillations during sleep, alpha oscillations during eye-closure. The arrhythmic oscillations are those with no periodicity, they show a power law in the form  $1/f^\beta$  that is typical of scale free (fractal) processes and is ubiquitous in nature. Regardless of the fact that arrhythmic oscillations constitute a significant part of the brain activities and strongly contribute to the EEG and ECoG records, classical analysis methods remove them because of their ubiquity.<sup>2</sup>

In the last two decades an increasing number of studies has focused on the scale-free component of the brain activity.<sup>2-5</sup> From these studies emerged that the noise (another name for the aperiodic oscillations) plays a central role in neuronal activity and in the development and maintenance of life in general. As pointed out in a review by McDonnell and Abbott,<sup>6</sup> noise acts with the so-called Stochastic Resonance (SR) that increase the effectiveness of the neuron signals.

All the studies that analyse the arrhythmic component of the brain activity focused on its power exponent or, in addition, on the Hurst exponent. Those indicators are related to the scaling and self-similarity of the neuronal signal. With these studies it is possible to distinguish between a resting state and a task or between normal activity and an epileptic seizure.

Even if there is an increasing effort in the study of the properties of the arrhythmic brain activity, studies do not exist that extract the stochastic mechanisms that underlie these processes. In particular, the stochastic analysis of ECoG signals is scarce. With this work we will contribute to the knowledge of the stochastic properties of ECoG signals.

### 7.1.1 Speech Brain Computer Interface

Recent works on speech Brain Computer Interfaces (sBCI) focus on the identification of speech-preparatory neural signals to detect the speech onset.

It is known that the most reliable speech onset/offset neuronal signals are found in the temporal cortex.<sup>7,8</sup> Those signals unfortunately are related to the auditory feedback of the subject's own voice, that is not available for real life sBCI deployment that will be used for patients that can no longer speak.

A good candidate should be a signal that precede the speech onset, in order to give enough time to trigger the decoder, and is highly correlated with the speech preparation process, being available also in the absence of speech emission.

One suitable region of the brain where to find a physiological signal related speech preparation is the speech arrest in Broca's area. In this region, a direct electrical stimulation during speech production induces the so called speech arrest phenomenon, the complete interruption of ongoing speech.<sup>9-12</sup>

Broca's area is known to be active prior to articulation with an increase of the high-gamma activity immediately before the speech onset.<sup>13,14</sup>

## 7.2 Project description

Human ECoG data are rare because the technique is invasive and data recorded during a task are even rarer. Because of the scarce availability of ECoG data, it is important to perform an exhaustive analysis of the associated time series. In order to achieve good results it is necessary to use robust and trusted analysis methods and have a good knowledge of them.

The data recording is part of a wider project that involves the "Center for Translational Neurophysiology" at Istituto Italiano di Tecnologia (Ferrara) and the Section of Physiology of University of Ferrara. Data were collected from a patients undergoing awake neurosurgery for tumour resection (low-grade glioma). The patient gave his informed consent, and the protocol was approved by the Ethics Committee of Azienda Ospedaliera Universitaria Santa Maria della Misericordia (Udine, Italy) after verification of the Italian Ministry of Health.

The data set we analysed is from a male Italian native speaker (hereinafter subject). the data are collected during a session of naming task. The task consisted in naming 10 different images shown on a screen and representing Italian nouns. The session consisted of three repetition of the 10 pictures for a total of 30 trials. The neuronal data were collected using the Epi array of

Table 7.1: List of the speech production tasks. In the naming task, the subject has to name the object presented on the screen as listed in this table

Italian	Translation
Antenna	Antenna
Ananas	Pineapple
Elefante	Elephant
Pecora	Sheep
Stivale	Boot
Coltello	Knife
Bottiglia	Bottle
Fungo	Mushroom
Elicottero	Helicopter
Pavone	Peacock

electrodes\*. Using the same data acquisition equipment, the voice and the neural signals have been recorded. The sampling frequency for the neural signal was 3051.8Hz while for audio was 24kHz. Before the analysis audio was downsampled to 3051.8Hz.<sup>15</sup>

### 7.3 Results

Our analysis started from the routine methods for the analysis of time series with a stochastic component. In particular: check for stationarity, autocorrelation function, power spectral density, first and second moments,... All this techniques show no differences between the time regions when the task was performed and the resting periods.

In figure 7.1 an example of our dataset is presented. Looking at the second moment, presented here as an example of the time series analysis performed, it is possible to see that signal differs from one task to the other making the analysis almost useless for our purposes.

In order to improve our results we compared the signal with the recorded audio using cross-correlation.

$$\hat{R}_{xy,k} = \frac{1}{N - k - 1} \sum_{i=k+1}^N (x_i - \mu_x)(y_{i-k} - \mu_y) \quad (7.1)$$

---

\***Epi array**: Total working area:  $4.3 \times 4.3\text{mm}$ ; Electrodes diameter:  $140\mu\text{m}$ ; Electrodes pitch:  $600\mu\text{m}$ ; electrodes number: 64,  $8 \times 8$  grid<sup>15</sup>

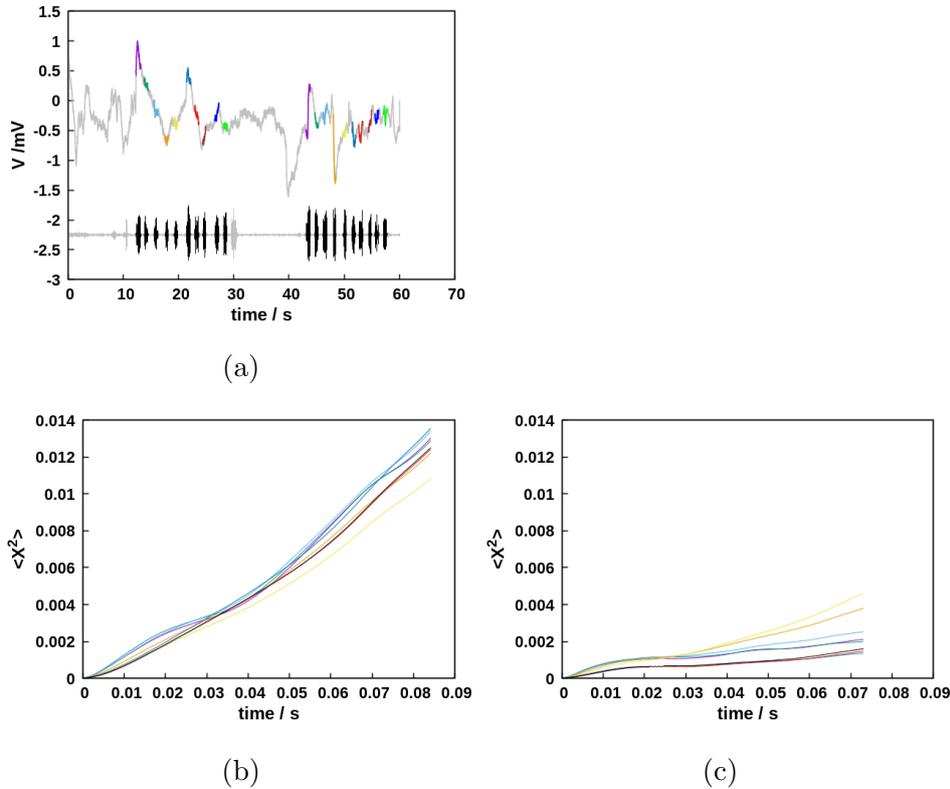


Figure 7.1: Example of the time series analysis. **a)** presents the ECoG for one electrode and the audio track. Colours highlight the portion containing a speech emission; the same color refers to the same task. **b)**, **c)** show the second moment for two different tasks. It is possible to see that the two tasks differs a lot making a direct comparison between the tasks almost impossible.

where  $\mu$  is the mean and  $k$  is the lag.

We cross-correlated, for each ECoG channel, the region containing the entire sequence of ten recognitions with the track for the voice sampled at the same frequency of the ECoG. We analysed the raw signal and different filtered versions of the signal. Filtering has been done using a rectangular low pass Fourier filter, placing to zero all the unwanted frequencies in the Fourier transform of the signal and making the inverse transform of the obtained sequence. The upper boundary of the filter was set from 1Hz to 15Hz with step 1Hz.

From the results we observed that cross-correlation between voice and signal is at maximum few hundreds milliseconds before speech. In a work on the same dataset,<sup>15</sup> the group of the IIT that collected the data, observed that, around 400 ms before speech there is a spike in the high gamma activity.

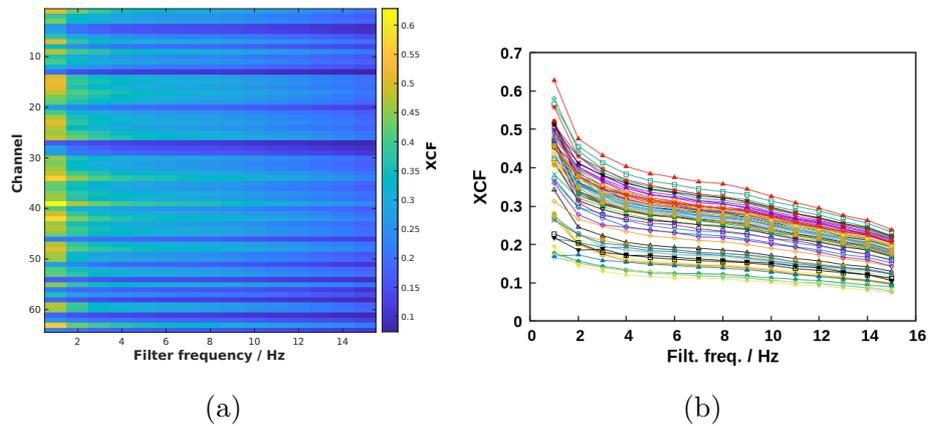


Figure 7.2: Maximum of x-corr for the different filter frequencies. From colormap (a) emerges that the XCF with the voice decreases with the increasing of the filter frequency. (b) shows a plateau for frequencies between 5Hz and 8Hz.

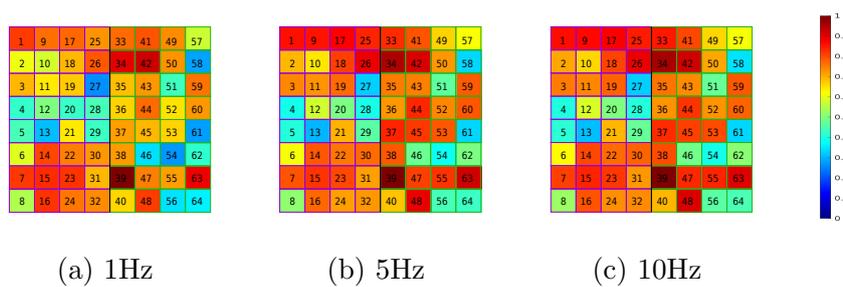


Figure 7.3: Maximum cross-correlation.  $X_{\text{corr}}$  normalized with respect to the maximum value. At 1Hz filtering frequency the variability of the channels activity is higher than for 10Hz filtering frequency

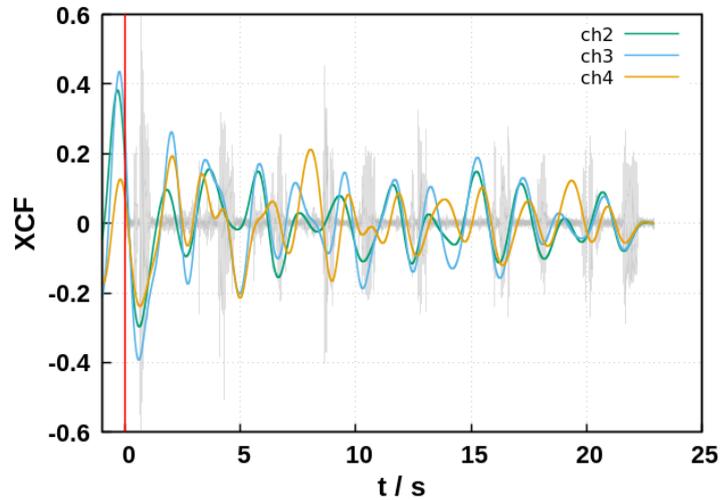


Figure 7.4: Comparison between the audio track and the XCF for three channels with filter at 1Hz. Voice and ECoG signal correlates in the pauses between words with a peak before the words

Our work shows that, also the low frequencies can provide information on the speech processes.

In order to improve our results we used the Principal Component Analysis (PCA) with the aim of finding a combination of channels that are related to the voice emission. Firstly we computed PCA including the audio track in the input matrix. In this way it is possible to find how the voice mixes with the ECoG channels. We can see, from figure 7.5, that voice does not mix with ECoG when unfiltered. After filtering voice mixes with ECoG signals contributing to different sets of PCs at the different filter frequencies.

We used the geometric distance ( $\sqrt{x^2 + y^2}$ ) between the eigenvectors in PCA and the maximum XCF for the different filter frequencies to find out which electrodes correlate the most with the voice.

Looking at figure 7.6 we can see how the different electrodes contribute to the eigenvectors containing also the voice contribution. For 1Hz filter many electrodes contribute in the mixture while, for 5Hz and 10Hz filters, a small number of electrodes is involved with the contributing electrodes that are the same for both frequencies.

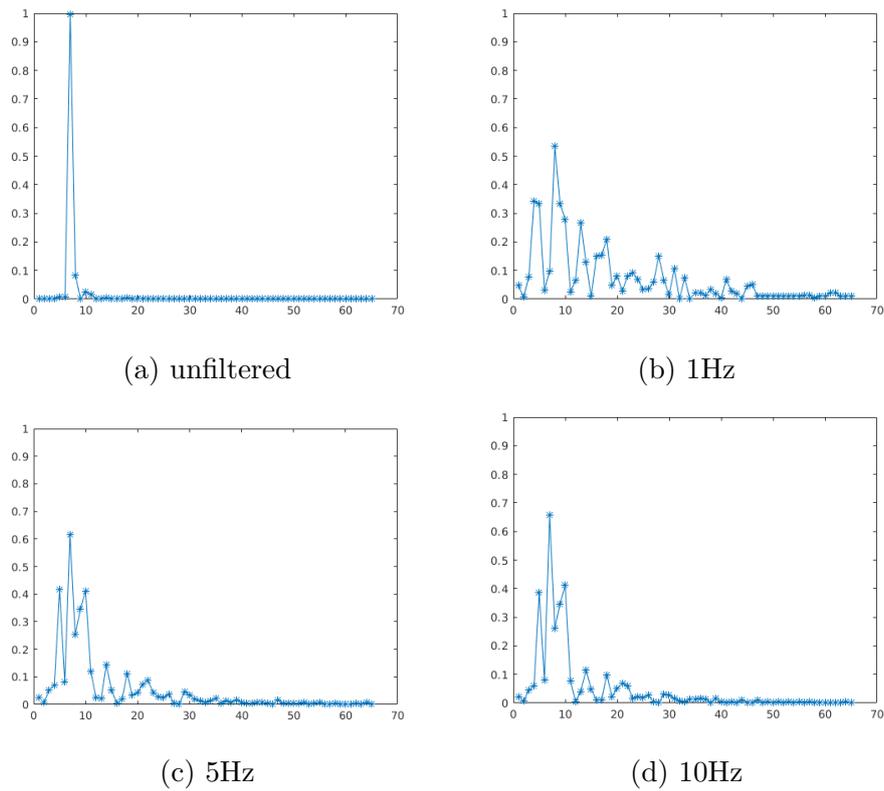


Figure 7.5: Contribution of the voice to the different PCs at the different filter frequencies. In the abscissae the number of the PCs and in ordinate the contribution of the voice. With the signal unfiltered the voice is in an eigenvector on its own. Changing the filter frequency changes the eigenvectors containing contribution from the voice.

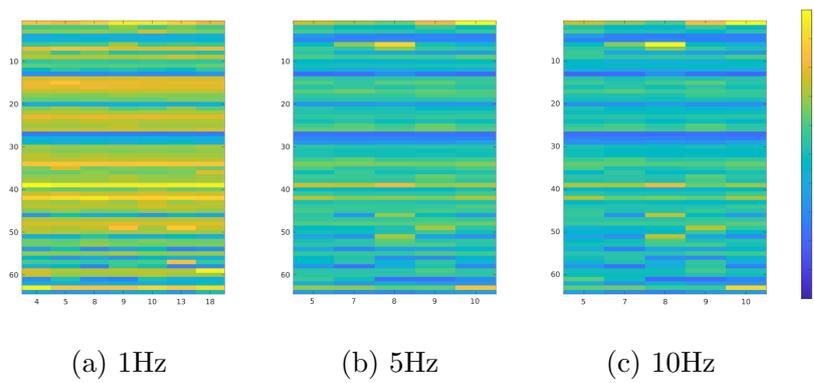


Figure 7.6: Contribution of the different electrodes to the PCs containing also a contribution from the voice. In order to obtain a compact representation we used the geometric distance between the PC and the XCF. In the colormap, the different PCs are arranged in columns with the electrodes in the rows.

## 7.4 Conclusions

Electrophysiology data are a valuable source of information on the mechanisms in the human brain. The information content is so rich that a unique method for its extraction does not exist. This poses a challenge in the design of the best analysis procedure for the given task or to extract unnoticed informations.

Too naive approaches, like the direct application of stochastic time series analysis methods, fail to obtain insight on such complex phenomena. On the other side, methods based on multivariate analysis are good in finding time or spatial regions where something different from the surroundings happens, but are unable to fully describe the involved mechanisms.

In this project, at its early stages, we presented a new exploratory method, based on frequency filters and PCA, that can be used in assessing the time and spatial region where a brain process related to a given output (speech in this project) happens.

The results of the exploratory methods tell little or nothing about the properties of the signal that characterize the speech emission process, but indicate where to look. In particular, our analysis found out that brain activity changes few hundreds milliseconds before speech emission also for low frequencies and not only for high frequencies as previously observed. Regarding the spatial region for the activity changes, our method measures the different correlation between brain activity and speech, corresponding to different levels of activity. According to the needs one can group electrodes on the basis of the correlation and then process the different groups.

In future analysis we will try to assess the stochastic properties that characterize the speech emission processes applying classical techniques to the regions indicated by our preliminary analysis. The idea is to find if the brain activity changes, not only in the rhythmic part but also in the scale free.

# Bibliography

- [1] C. Todaro, L. Marzetti, P. A. Valdés Sosa, P. A. Valdés-Hernandez and V. Pizzella, *Brain Topography*, 2019, **32**, 583–598.
- [2] B. J. He, J. M. Zempel, A. Z. Snyder and M. E. Raichle, *Neuron*, 2010, **66**, 353–369.
- [3] B. J. He, *Trends in Cognitive Sciences*, 2014, **18**, 480–487.
- [4] M. Kelsey, D. Politte, R. Verner, J. M. Zempel, T. Nolan, A. Babajani-Feremi, F. Prior and L. J. Larson-Prior, *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2012, 4336–4340.
- [5] S. Janjarasjitt and K. A. Loparo, *Journal of Medical and Biological Engineering*, 2014, **34**, 535–541.
- [6] M. D. McDonnell and D. Abbott, *PLoS Computational Biology*, 2009, **5**, e1000348.
- [7] V. G. Kanas, I. Mporas, H. L. Benz, K. N. Sgarbas, A. Bezerianos and N. E. Crone, *IEEE Transactions on Biomedical Engineering*, 2014, **61**, 1241–1250.
- [8] V. G. Kanas, I. Mporas, H. L. Benz, K. N. Sgarbas, A. Bezerianos and N. E. Crone, 2014 19th International Conference on Digital Signal Processing, 2014, pp. 862–865.
- [9] E. F. Chang, J. D. Breshears, K. P. Raygor, D. Lau, A. M. Molinaro and M. S. Berger, *Journal of Neurosurgery JNS*, 2017, **126**, 114 – 121.
- [10] E. Mandonnet, S. Sarubbo and H. Duffau, *Neurosurgical Review*, 2017, **40**, 29–35.
- [11] M. C. Tate, G. Herbet, S. Moritz-Gasser, J. E. Tate and H. Duffau, *Brain*, 2014, **137**, 2773–2782.

- [12] V. Ferpozzi, L. Forna, M. Montagna, C. Siodambro, A. Castellano, P. Borroni, M. Riva, M. Rossi, F. Pessina, L. Bello and G. Cerri, *Frontiers in Human Neuroscience*, 2018, **12**, 64.
- [13] A. Flinker, A. Korzeniewska, A. Y. Shestyuk, P. J. Franaszczuk, N. F. Dronkers, R. T. Knight and N. E. Crone, *Proceedings of the National Academy of Sciences of the United States of America*, 2015, **112**, 2871–2875.
- [14] X. Pei, E. C. Leuthardt, C. M. Gaona, P. Brunner, J. R. Wolpaw and G. Schalk, *NeuroImage*, 2011, **54**, 2960–2972.
- [15] E. Delfino, A. Pastore, E. Zucchini, M. F. P. Cruz, T. Ius, M. Vomero, A. D’ausilio, A. Casile, M. Skrap, T. Stieglitz and L. Fadiga, *International Journal of Neural Systems*, 2021, **31**, 2150025.

# General Concluding Remarks

The subject of this PhD research thesis has been the development of analysis procedures to investigate noisy data along with the related computer software. Noise has the Janus-like nature of being a nuisance and a resource. It is an unavoidable component of all measurements, where it interferes with the main signal originated by the system of interest masking the desired information. Noise is also intrinsically connected to many physical, chemical and biological processes, playing an important role in vital physiological processes.

Many mathematical tools are available for the analysis of stochastic processes (noises). They can be exploited to provide a deeper insight into the underlying system. A variety of tools exist for denoising a signal, although most of them are unsuitable for retrieving the noise for further analysis.

Singular Spectrum Analysis (SSA) has proved to be a powerful technique for data decomposition and denoising. It is non-parametric and data-driven and it satisfies very well the double purpose of decomposing deterministic and stochastic components of a signal.

We proved the ability of SSA on mono-dimensional and bi-dimensional set of data. In order to improve the capability of SSA in the decomposition of data with complicated trends, we developed a procedure that iteratively extracts the signal from the dataset. Using a combination of Basic SSA and CiSSA our method is able to completely separate noise from signal.

The method has been used to improve the quality of diffractograms as a preprocessing tool for the Rietveld method. With preprocessing, a good data quality can be obtained also for faster scan speeds or for a smaller number of scans for each sample. This can change the acquisition protocols.

The performances of 2D-SSA have been tested on images obtained by ECL imaging. The high dynamical range and the low S/N ratio make denoising challenging. Using 2D-SSA we improved the quality of the ECL images reducing the S/N ratio. From the study of transient emission emerges a non-constant decay rate with fluctuations superimposed to a second order decay. Further studies will be necessary to determine the nature of the fluctuations.

Combining PCA and cross-correlation in a new way resulted in a good exploratory method for the analysis of electrophysiology data. With our method, we observed that brain activity changes before speech also for low frequencies and not only for high frequencies as previously observed using only Fourier analysis.

We conclude that, in order to extract all the information contained into a dataset, it is useful to combine decomposition techniques and tools for fractal analysis. In this way, it is possible to divide a given signal into a set of different components that can be analysed individually obtaining insight into all the processes that generate the signal.

# List of Figures

5.1	HRPXR for Hydroxiapatite . . . . .	76
5.2	Comparison Sliding SSA . . . . .	77
5.3	Raw Vaterite . . . . .	78
5.4	ITE-SSA Vaterite . . . . .	78
5.5	Vaterite HRPXR denoised using ITE-SSA . . . . .	79
5.6	Vaterite HRPXR denoised using ITE-SSA and iterative CiSSA . . . . .	81
5.7	Comparison ITE-SSA and ITE-SSA with CiSSA refinement . . . . .	83
5.8	Effects of the averagin on different number of runs . . . . .	84
5.9	SSA denoising Vaterite dataset 2 . . . . .	85
6.1	ECL 2.8 $\mu$ m microbeads 8s . . . . .	93
6.2	W-corr matrix 2.8 $\mu$ m microbeads 8s . . . . .	94
6.3	ECL 8.9 $\mu$ m microbeads 8s . . . . .	94
6.4	W-corr matrix 8.9 $\mu$ m microbeads 8s . . . . .	95
6.5	ECL intensity profiles 8s . . . . .	95
6.6	ECL 2.8 $\mu$ m microbeads 0.2s . . . . .	97
6.7	W-corr matrix 2.8 $\mu$ m microbeads 0.2s . . . . .	98
6.8	ECL 8.9 $\mu$ m microbeads 0.2s . . . . .	99
6.9	W-corr matrix 8.9 $\mu$ m microbeads 0.2s . . . . .	100
6.10	ECL intensity profiles 0.2s . . . . .	101
6.11	ECL intensity decay for 2.8 $\mu$ m microbeads . . . . .	102
6.12	Analysis of the fluctuations of ECL intensity in time . . . . .	103
7.1	Example of the ECoG time series analysis . . . . .	111
7.2	Maximum of x-corr for the different filter frequencies . . . . .	112
7.3	Maximum cross-correlation . . . . .	112
7.4	Comparison between audio and XCF . . . . .	113
7.5	Contribution of the voice to the different PCs . . . . .	114
7.6	Contribution of the electrodes to the PCs with voice contribution . . . . .	115



# List of Tables

5.1	Rietveld method for Vaterite 1 before SSA . . . . .	81
5.2	Rietveld method for Vaterite 1 after SSA . . . . .	82
5.3	Rietveld method for Vaterite 2 before SSA . . . . .	82
5.4	Rietveld method for Vaterite 2 after SSA . . . . .	82
5.5	Improvement of Rietveld after SSA . . . . .	84
6.1	SNR 2.8 $\mu$ m microbeads 8s . . . . .	93
6.2	SNR 8.9 $\mu$ m microbeads 8s . . . . .	93
6.3	SNR 2.8 $\mu$ m microbeads 0.2s . . . . .	97
6.4	SNR 2.8 $\mu$ m microbeads 0.2s . . . . .	98
7.1	List of the speech production tasks . . . . .	110