Instituto Nacional de Matemática Pura e Aplicada

DIMENSIONALITY REDUCTION IN NEUROSCIENCE AND EPIDEMIOLOGY

Lucas Martins Stolerman

Doctoral Thesis

Rio de Janeiro
April 2017

Instituto Nacional de Matemática Pura e Aplicada

Lucas Martins Stolerman

DIMENSIONALITY REDUCTION IN NEUROSCIENCE AND EPIDEMIOLOGY

Thesis presented to the Post-graduate Program in Mathematics at Instituto de Matemática Pura e Aplicada as partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics.

Advisor: Roberto Imbuzeiro de Oliveira
Co-advisor: Nathan Kutz (University of Washington)

Rio de Janeiro
March 2017

Aos nossos filhos Maximiliano, Isabel e Maitê.

"The diversity of the phenomena of nature is so great, and the treasures hidden in the heavens so rich, precisely in order that the human mind shall never be lacking in fresh nourishment."

Johannes Kepler

AGRADECIMENTOS

Esta tese de doutorado foi desenvolvida com o apoio decisivo de algumas pessoas. Estou certo de que sem o auxílio delas eu não teria chegado até o fim. E esta certeza me faz escrever estas breves linhas com sentimentos de gratidão e paz.

Ao professor Roberto Imbuzeiro, meu orientador, agradeço pela generosidade, paciência e mansidão ao me guiar nestes quatro anos. Roberto, obrigado por me ensinar matemática e por ter dito que você não estava nervoso naquele momento antes do meu exame de probabilidade. Aquelas palavras me encorajaram. Obrigado por me animar a pensar no meu problema de tese, por permitir que eu tentasse encontrar meu caminho. Obrigado por apoiar minha viagem aos EUA, por me orientar quando voltei ao Brasil e por acompanhar de perto a redação desta tese. Obrigado por ser para mim um exemplo de cientista, pai de família, professor, orientador e por aí vai. Sem a sua ajuda eu não teria conseguido chegar até aqui.

Ao professor Nathan Kutz (que entende português), meu co-orientador, que me recebeu com grande entusiasmo na Universidade de Washington durante o período entre Fevereiro e Maio de 2016. Nathan, sem seu apoio e experiência, eu provavelmente teria sido muito menos produtivo no curto período de tempo em que passei em Seattle. Obrigado por me ensinar a valorizar meu tempo e ser mais disciplinado. Obrigado por sugerir o projeto da Dengue & Clima e por despertar em mim um grande interesse por *data science*.

Ao professor Cláudio Queiroz que nos disponibilizou os dados de crises epilépticas e tanto me ajudou ao longo destes anos. Cláudio, Obrigado pela paciência comigo e pelo entusiasmo com o nosso trabalho. Obrigado por ter nos ajudado a compreender o impacto científico da nossa metologia e por nos impulsionar na direção de uma pesquisa que pudesse realmente contribuir para o estudo da epilepsia.

Aos professores Antônio Galves, Augusto Quadros e Diego Nehab por aceitarem o convite para fazer parte da banca examinadora e também pelas correções e comentários sobre esta tese.

Ao projeto Neuromat, que proporcionou o encontro do Roberto com o Cláudio, de onde nasceu o projeto da epilepsia. Agradeço pelas oportunidades de ir a São Paulo e participar de diversas atividades organizadas sob a coordenação do professor Antônio Galves.

Ao Pedro Maia, meu amigo de graduação na UFRJ que se tornou colaborador de pesquisa. Pedro, muito obrigado pela atenção que você me deu nos EUA e também quando voltei ao Brasil. Seus conselhos me ajudaram a ser um profissional mais competente. Obrigado por me ensinar a fazer (e por me cobrar!) figuras bonitas, a ser sempre positivo com relação ao trabalho, a pensar em ciência de forma apaixonada.

À professora Stefanella Boatto (UFRJ) que me orientou nos primeiros anos de formação acadêmica e me recomendou ao programa de doutorado do IMPA. Professora, muito obrigado por também me recomendar ao Prof. Nathan Kutz e pelo entusiasmo em acompanhar nossos resultados até hoje. A senhora foi e continua sendo uma pessoa importante na minha trajetória como cientista.

Ao amigo Liev Maribondo, por toda a ajuda durante estes anos. Liev, você me ajudou a resolver o quebra-cabeças da minha vida na época da viagem para os EUA. Obrigado pelas conversas filosóficas que tivemos ao longo deste período e por me ajudar a ser uma pessoa mais cuidadosa em muitos aspectos. Obrigado por nos receber em João Pessoa no momento mais difícil de nossas vidas. Aqueles dias na Paraíba foram mais importantes para nossa família do que você pode imaginar.

Agradeço a todos os funcionários do IMPA: As pessoas da divisão de ensino, do restaurante, da xerox, biblioteca, RH, etc. Vocês fazem do IMPA um lugar especial. Quem nos dera se todas as instituições do nosso país tivesse o padrão de qualidade que vocês nos oferecem. Vocês me deram todas as condições para desenvolver este trabalho.

Agradeço aos meus familiares: À minha mãe Said, pela amizade e companheirismo de sempre. À minha irmã Agnes, que agora é uma mulher que me dá conselhos valiosos. À minha madrInhá Rita, que me ensinou a gostar de matemática. Ao meu primo Thomas, amigo de toda a vida. Aos meus avô Alberto e avós Lúcia e Olinda, por serem tão presentes em nossas vidas. À família da Carol e minha família: César, Francila, Felipe e Camila. Obrigado por nos apoiarem em tudo.

À equipe de senhores do movimento Regnum Christi e à comunidade dos Legionários de Cristo do Rio de Janeiro, que me dão o suporte necessário para que eu possa buscar ser a cada dia uma pessoa melhor em todos os aspectos. Em especial agradeço ao meu diretor espiritual Pe. Manuel Flores L.C., pela paciência em me ouvir falar sobre as dificuldades do doutorado por todos estes meses e por me ajudar a colocar todas as coisas na perspectiva correta. Padre, Obrigado por me ensinar que devo me ocupar e não me preocupar.

Agradeço aos nossos filhos: Max e Isabel que estão no céu e Maitê que tornou nossa casa uma bagunça e nossa vida uma grande alegria. Esta tese é para vocês. Eu os amo muito. Maitê, você é o motivo que faz o papai levantar todos dias e ir trabalhar. Você me faz feliz de uma maneira inexplicável.

Agradeço por fim à minha esposa Carolina. Amor, subir no altar para casar com você foi a melhor decisão que já tomei. Obrigado pela paciência durante estes quatro anos. Obrigado por me encorajar e estar sempre ao meu lado. Você me ajudou a ser um homem mais forte, o que fez crescer ainda mais minha admiração e amor por você.

Rio de Janeiro, 24 de Novembro de 2017

RESUMO

Esta tese é dedicada a novos métodos *data-driven* para análise de problemas em epilepsia e epidemiologia da Dengue. A noção de redução de dimensionalidade será importante nos dois problemas que estudamos.

Nossa primeira contribuição, em colaboração com Cláudio M. Queiroz (Instituto do Cérebro, UFRN), Nathan Kutz (Universidade de Washington) e Roberto I. Oliveira (IMPA), trata de um problema sobre detecção de crises epilépticas. Nós desenvolvemos um método baseado na Decomposição em Valores Singulares (SVD) para explorar o grau de sincronização antes, durante e após crises em um modelo animal de Epilepsia do Lobo Temporal. Com nossa metodologia criamos um algoritmo baseado em limiares de sincronização que melhora significativamente o estado da arte. Do ponto de vista neurobiológico, encontramos níveis de sincronização consideravelmente baixos durante e alta atividade síncrona após as crises, o que tem importantes consequências.

Nossa segunda contribuição, em colaboração com Pedro D. Maia (Weill Cornell Medicine) e Nathan Kutz (UW), trata da análise de séries climáticas e sua relação com epidemias de Dengue. Condições climáticas locais têm papel importante no desenvolvimento da população do mosquito *Aedes Aegypti*, responsável pela transmissão da Dengue. Nós aplicamos técnicas de redução de dimensionalidade e algoritmos de aprendizado de máquina em séries climáticas e analisamos sua conexão com a ocorrência de Dengue em sete capitais brasileiras. Especificamente, identificamos duas variáveis chaves e um período durante o ciclo anual com grande poder preditivo. Assinaturas de temperatura e chuva variam significativamente de cidade a cidade, sugerindo que a relação entre clima e Dengue é mais complexa do que se pode imaginar.

Palavras-chave: Decomposição em Valores Singulares · Máquinas de Vetores de Suporte · Validação cruzada · Aprendizado de Máquina · Dengue · Detecção de crises epilépticas

ABSTRACT

This thesis is devoted to new data-driven methods for the analysis of problems in epilepsy and Dengue epidemics. The notion of dimensionality reduction will be important throughout the thesis and in the two problems we study.

Our first contribution is a joint work with Cláudio M. Queiroz (Brain Institute, UFRN), Nathan Kutz (University of Washington) and Roberto I. Oliveira (IMPA), and it deals with a seizure detection problem. We develop a SVD-based method to explore the degree of synchronization before, during and after seizures in a Temporal Lobe Epilepsy experimental (animal) model. With our methodology we build a seizure detection algorithm based on synchronization thresholds that significantly improves the state of the art. From the neurobiological viewpoint, we have found considerably low levels of brain synchronization during seizures and higher synchronous activity after seizures, which have important consequences.

Our second contribution, joint with Pedro D. Maia (Weill Cornell Medicine) and Nathan Kutz (UW), deals with the analysis of climate time series and their relationship with Dengue epidemic outbreaks. Local climate conditions play a major role in the development of the mosquito population responsible for transmitting Dengue Fever. We apply dimensionality reduction techniques and machine-learning algorithms to climate time series data and analyze their connection to the occurrence of Dengue outbreaks for seven major cities in Brazil. Specifically, we have identified two key variables and a period during the annual cycle that are highly predictive of epidemic outbreaks. Critical temperature and precipitation signatures may vary significantly from city to city, suggesting that the interplay between climate variables and Dengue outbreaks is more complex than generally appreciated.

Keywords: Singular Value Decomposition · Support Vector Machine · Cross Validation · Machine Learning · Dengue Epidemics · Seizure Detection

List of Figures

Contents

Part I

Introduction

In this thesis we develop new data-driven methodologies based on dimensionality reduction and machine learning techniques to investigate problems in epilepsy and Dengue epidemiology. Dimensionality reduction deals with the problem of finding compact representations of high-dimensional data via a smaller number of explanatory variables [1]. Our overarching goal is to learn from data: extract important patterns, trends and ultimately understand what the data *says* [2].

Data-driven methods are important tools in a world where huge amount of information is produced everyday. The low cost of computational power and data storage in the last years has increased the volume of available data in many scientific areas. The ability to extract valuable knowledge from data, called the fourth paradigm of science [3], is the essence of Machine Learning: A "field of study that gives computers the ability to learn without being explicitly programmed" [4]. Prof. Abu-Mostafa says that the process of solving a learning problem is essentially divided into three different parts [5]:

1. **A pattern exists**: we assume that the data has an underlying structure to be explored by some computational tool.

2. **We cannot pin it down mathematically**: there is no closed mathematical formula to describe such pattern.

3. **We have data on it**: without a considerable amount data the learning problem is not feasible.

The different machine learning problems can be broadly categorized as supervised or unsupervised. On supervised problems, we have access to the outcome variables and we want to use them in the learning process. These algorithms usually starts with a training step when the classification is done according to our labelling criteria. Then new data is classified (testing step) and predictions can be made. On the unsupervised scenario, we have no information about the outcomes so it is not possible to label the data previously. So the idea is to describe how the data are organized or clustered.

Why dimensionality reduction?

Generally, a large system of measured quantities can be represented by a smaller number of explanatory variables. Dimensionality reduction methods are able to discover and extract these explanatory variables which are often not directly observed. This may not be a straightforward task, but is especially important in applications where the dimension is quite large, potentially larger than the sample sizes coming from laboratory data.

From the mathematical modeling viewpoint, machine learning techniques have been combined with nonlinear dynamical systems to discover equations from noisy measurement data. The equations governing physical or biological processes have been traditionally obtained from first principles and their predictive power has been evaluated by comparison with data. The data-science approach consists of modeling those processes by identifying nonlinear systems from data without assumptions on the form of the equations. As a worth mentioning example, Brunton, Kutz and Proctor have recently developed a novel framework to discover governing equations underlying a dynamical system simply from data measurements [6], using sparse regression and compressed sensing. When the number of measurements is very high, dimensionality reduction techniques can be used to extract coherent structures. In this work the authors were able to discover the dynamics of a fluid vortex shedding behind an obstacle, which took experts in the community nearly 30 years to resolve.

In neuroscience, network dynamics of huge populations of neurons give rise to sensory, cognitive and motor functions. Thanks to dimensionality reduction techniques, studies have begun to characterize such complex systems by seeking for simplicity at neural ensemble levels. Broome, Jayaraman and Laurent [7], studying the locust olfactory system, recorded responses of projection neurons (PN) and Kenyon cells (KC) to different odor stimulus. Using dimensionality reduction with Locally Linear Embedding techniques [1], they examined how odor representations by PNs and by KCs change as different odors coincided. Their results helped to explain the nature of odor perception or recognition. For visual attention systems, Cohen and Maunsell ([8]) found that a single trial measure of attention on a few dozen simultaneously recorded neurons can predict behavior. De-

spite the apparent complexity of single-neuron responses, the population activity showed orderly structure across different conditions. For an interesting review of dimensionality reduction for neural recording we refer the work of Cunningham and Yu [9].

In epidemiology, prediction methods have been developed with the analysis of high-dimensional data. Ju and Brasier ([10]) applied different feature selection methods to identify informative blood variables for Dengue Hemorrhagic Fever (DHF). They found that that IL-10, platelet and lymphocyte counts may be the major features for predicting Dengue DHF on the basis of blood measurements. A different mathematical approach was adopted recently by Frasca and colleagues [11], where a low-dimensional description of epidemic processes was carried with an isometric features mapping (ISOMAP) approach [12]. Using synthetic data from two epidemic models, they found an embedding dimensionality equal to three, thus revealing that more than one macroscopic variable is necessary to describe the epidemic dynamics.

Our work

Part II of this thesis is devoted to a seizure detection algorithm based on synchronization thresholds, a joint work with Cláudio M. Queiroz (Brain Institute, UFRN), Nathan Kutz (University of Washington) and Roberto I. Oliveira (IMPA). This work has been done under the auspices of the Neuromat project[1]. Data has been made available from Prof. Queiroz's Lab and consist of Local Field Potential (LFP) activity from 24 channels of mice hippocampus. LFP signals represent the electric potential in the extracellular brain tissue [13]. Our method is a SVD-based technique to analyse brain synchronous activity. We have been able to score the signal complexity by characterizing the energy tails of the singular values. We applied our method on a dataset containing 19 epileptic seizures from a Temporal Lobe Epilepsy experimental model and we were able to compare our detection times with the expert opinions. From the neurobiological viewpoint, we obtained

[1]The Research, Innovation and Dissemination Center for Neuromathematics is hosted by the University of São Paulo and funded by FAPESP (São Paulo Research Foundation), grant 2013/07699-0.

interesting results regarding desynchronization during the seizure epochs and also high synchronized activity during postictal (after seizure) phase.

In Part III we developed two methods based on machine learning algorithms to analyse climate series and their connection with Dengue outbreaks in Brazil. This is a joint work with Pedro D. Maia (Weill Cornell Medicine) and Nathan Kutz (UW). The first methodology is based on the SVD decomposition of a climate data matrix. We project temperature and precipitation time series onto a 2-mode plane and label them as epidemic or non-epidemic year. This serve as classification step and also enable us to evaluate the correlation between these critical climate signatures and Dengue epidemics. The second methodology is an application of Support Vector Machine algorithms ([14, 15]) on two key features: Mean temperature and frequency of rain events. We have also labelled those climate indicators based on an epidemic or non-epidemic criteria and we were able to find the most important seasons for Dengue epidemics for each state capital. Moreover we also show how methods could be used for Dengue forecasting.

Part II

Seizure Detection and Analysis: a SVD-based method

Chapter 1

Introduction

Epilepsy is one of the most common neurological disorders. It is well known for sudden and apparently random seizures which are mainly triggered by stroke and trauma, but also by accidents in the brain tissues, infections or inheritance [16, 17, 18]. Epileptic seizures are produced by intense electrical impulses that might spread through the entire brain (primary generalized seizures) or just through a relatively small part of it (partial seizures). Almost 50 million people worldwide have epilepsy [19, 20] and for at least for 30% of the patients, the epileptic process cannot be controlled by current medication [21]. Epilepsy surgery might be an alternative in some cases, but there is no guarantee of success and the chances of long-term seizure outcomes are considerably high [22, 23]. In a different direction, early warning systems for seizure detection prior the clinical onset have been widely studied and might be the best way to avoid potentially harmful situations [24, 25, 26, 27].

The study of the relationship between epileptic seizures and abnormal electric events started in 1912 with the seminal works of Kaufman [28] and Pravidch-Neminsky [29], using electroencephalography (EEG) data. EEG is a non-invasive electrophysiological monitoring method in which the electrodes are usually placed along the scalp [30]. Major contributions in the field EEG analysis in epilepsy were made in the subsequent decades by Berger [31, 32], Gibbs and Lennox [33, 34, 35, 36]. For a historical review we refer the work by Magiorkinis et al [37].

It has long been speculated that epileptic seizures are related to synchronized electrical activity in brain neurons, heretofore called synchronization. The first studies were done by Penfield and Jasper in 1954, where seizures were characterized as *hypersynchronous* neural activity caused by decreased inhibition and enhanced excitation [38]. In subsequent years many authors tried to unveil the intrinsic nature of seizures without defining the term "synchronization" rigorously. More recently, new mathematical tools have been developed to better quantify synchronized neural activity such as cross-correlation, mutual information, spectrum-based coherence, nonlinear interdependence, phase synchronization and random matrix theory [39, 40, 41, 42, 43, 44, 45, 46].

As it turns out, the same techniques that were designed to quantify synchronization have been used to suggest that desynchronization is also present in the epileptic process [47, 48, 49, 50, 51, 52, 53, 54]. In 2002, Netoff and Schiff found desynchronization during seizures by comparing various linear and non-linear detection methods for an experimental *in vitro* model with CA1 pyramidal neurons [48]. Their findings were the first experimental evidence that strongly coupled neurons may remain desynchronized on faster time scales. One year later, Mormann et al. applied an automated technique on EEG intra-cranial recordings from a group of patients with focal epilepsy [49]. They found a characteristic decrease in synchronization ranging from several minutes up to a few hours prior to the seizure onset. They have also investigated whether such drop on the synchronization degree where indeed a good criterion for the definition of a pre-seizure state. Many other authors have found similar results on desynchronization during seizures. Such discoveries on the mechanisms of neural synchrony during seizures have opened new avenues for understanding the spatiotemporal dynamics of epileptogenesis. For a recent review about we refer reader to the work of Jiruska et al [55].

The capability to perform seizure detection methods is one of the biggest challenges in epilepsy research. From a historical perspective, automated analysis of EEG recordings as warning systems for the occurrence of epileptic seizures started in the 1970s. Early methods were based on relative EEG amplitude thresholds [56, 57, 58] and were aimed to facilitate the recording process by making sure no seizure was missed. Different approaches for seizure detection have been developed ever since. Review studies can be found in [59]

and [25]. Most seizure detection methods consist of finding discriminative features in the EEG signals during the seizure period and then performing a classification step. The techniques for feature extraction are usually based on Wavelet and Fast Fourier transforms [60, 61, 62], Lyapunov exponents [63], cross correlation functions [64], entropy [65] and principal component analysis (PCA) [66]. The classification methods are important for deciding whether a piece of EEG signal comes from a seizure or not. Most classifiers are based on linear classifiers [67], support vector machines (SVM) [68], artificial neural networks (ANN) [69], k-nearest neighbour [70], decision trees [71] and Gaussian mixture models [72].

Our Work

In this work we develop a seizure detection algorithm based on a new method for scanning synchronized activity of a multichannel neural recording system. We analyse data coming from Local Field Potential (LFP) signals recorded at the hippocampus (HPC) for a Temporal Lobe Epilepsy (TLE) animal model. Here we give a brief explanation about these terms:

- *Temporal Lobe Epilepsy* is the most common form of partial or localization related epilepsy [76]. It was defined in 1985 by the International League Against Epilepsy (ILAE) as "a condition characterized by recurrent, unprovoked seizures originating from the medial or temporal lobe" [77].

- A *Local Field Potential* ([13]) is an electro-physiological signal recorded with micro-electrodes inside the neural tissue. It is generated by an electric current from multiple neuron Voltage and is produced on the extracellular space by a large number of action potentials in a small volume of the brain.

- The *Hippocampus* is a small region of the brain that is primarly associated with memory and spatial navigation [73, 74]. It this the region where TLE seizures commonly begins [75, 76], so it may be important to keep the LFP recordings in this brain structure.

Data has been made available from prof. Cláudio Queiroz's lab at the Brain Institute of the Federal University of Rio Grande do Norte (UFRN). In total, we analyzed a dataset of 19 seizures from 4 mice, which were submitted to a pilocarpine animal model of TLE [80, 81, 82, 83]. Our method is based on a synchrony indicator for the activity of an arbitrary number of channels which is obtained from the Singular Value Decomposition (SVD) of a data matrix [139]. Two consequences of this method are described:

1. We contribute to the debate on synchronization *vs* desynchronization by showing that ictal activity (i.e brain activity during the seizure) is desynchronized for most seizures in our dataset. For seizure detection, we apply a training and testing algorithm based on threshold values of the synchrony indicator.

2. We characterize both short and long term effects of the seizure in the HPC, by observing an increase of the synchronization level during the postictal (i.e post-seizure) depression for most seizures.

This part of the thesis is outlined as follows. In chapter 2, we describe the data acquisition procedure and give some information about each seizure in our dataset (section 2.1). We also explain our data analysis method (the α - series), which consists on defining synchronization scores on sliding time windows (section 2.2). We then check the sanity of our method with a stochastic Kuramoto model (section 2.3). In chapter 3 we present our main results. First we show the neurobiological implications of our scanning method (section 3.1) and then we describe the seizure detection process (section 3.2). The optimal thresholds are obtained for each seizure and the training and testing algorithm is explained. The results for each seizure and different sliding time windows are shown in subsections 3.2.1 and 3.2.2 respectively. In chapter 4 we discuss the neurobiological implications of our method, we compare our results with previous studies and discuss limitations and future perspectives on this work from the practical implementation viewpoint. We conclude this work with an outlook of the developed methods.

Chapter 2

Materials and Methods

In this chapter we describe the TLE Experimental model and our data-driven methods for analyzing synchronization levels across the Local Field Potential (LFP) time series. Our mathematical tool is a synchrony indicator based on the energy distribution of the singular values from data matrices. There are three major phases of seizures: preictal (immediately before seizures), ictal (during seizures) and postictal (after seizures). All such phases are important for the understanding of the epileptic process. Control strategies should target preictal and ictal periods. On the other hand, it is important to analyze the postictal stage in order to discover the impact of the seizure in the neural systems. We have been able to explore all these stages with our methodology.

2.1 Data acquisition

In this work we have analyzed hippocampal Local Field Potentials (LFP) of 19 seizures from 4 different mice (named Hx, A3, B10 and F10). We label them as S_i for $i = 1, 2, \ldots, 19$. Data was collected at the Neural Networks and Epilepsy Laboratory of the Brain Institute (UFRN)[1]. The epileptic condition was produced by the systemic administration of Pilocarpine in high doses (280mg/Kg i.p.)[2], which produced a long term status

[1] http://www.neuro.ufrn.br/research/groups/4

[2] i.p. stands for *intraperitonial injection.*

epilepticus [80, 81, 82, 83]. These brain insults produce neuronal death, synaptic reorganization and spontaneous limbic seizures.

The LFP traces have been acquired at 1000 Hz for S_1 and S_2 and 2000 Hz for the other series. A total of 24 channels were equally distributed on both sides of the HPC (except for mouse F10 with 23 channels with 11 on HPC right and 12 on HPC left). Table 2.1 summarizes information about each seizure. The onset has been defined as the first spike before the flattening of the LFP that precedes the paroxysmal activity. The end of the paroxysmal activity was characterized by its disappearance, by depression of the EEG amplitude and occasionally by low frequency postictal bursts. Seizure duration was defined as the the difference between the end and onset times ([79]).

2.2 The α - series.

The main novel component of our methodology is what we call the α - series. We begin our methods chapter by describing how we have produced this series for analysing the levels of synchronized activity across the filtered LFP data (data filtering is described in the Supporting Information chapter). We have introduced two temporal parameters: The time window size W_s and the overlap gap W_g, both in seconds. For fixed values of these parameters we divide the total recording time of $\mathcal{T}$ seconds into overlapping intervals of the form $\big[(k-1)W_g, t_k\big)$ with $t_0 = 0$ and

$$t_k = W_s + (k-1)W_g \quad k = 1, 2, \ldots k_{end},$$

where the last window index

$$k_{end} = \Big\lfloor 1 + \frac{\mathcal{T} - Ws}{W_g} \Big\rfloor$$

can be obtained after some algebraic manipulation. Fig. 2.1 illustrates our temporal parameters and the first two consecutive sliding windows.

For a given seizure S and for each t_k we define the data matrices $\mathcal{M} = \mathcal{M}(t_k, S)$ that contains the filtered LFP signals from the N channels during the intervals $[t_{k-1}, t_k)$. For a sampling rate of f_r Hz, each $\mathcal{M}$ is a $N \times (W_s f_r + 1)$ data matrix. Fig. 2.3a) illustrates this step: Each row of $\mathcal{M}$ is filled with data from a single channel during such time window.

Table 2.1: The 19 seizures from our dataset.

Seizure	**Mouse name**	**onset (sec)**	**offset (sec)**	**duration (sec)**
1	Hx	1800	1835	35
2	Hx	1813	1864	51
3	Hx	1320	1356	36
4	A3	1800	1862	62
5	B10	314.4	343.2	28.8
6	B10	1791	1870	79
7	B10	1402	1444	42
8	B10	1041	1097	56
9	F10	1105	1156	51
10	F10	977.1	1015	37.9
11	F10	1803	1842	39
12	F10	1709	1753	44
13	F10	1805	1835	30
14	F10	1802	1840	38
15	F10	1387	1440	53
16	F10	540.9	590	49.1
17	F10	1788	1850	62
18	F10	1800	1836	36
19	F10	1210	1242	32

We then compute the SVD of $\mathcal{M}(t_k, S)$, which can be written as

$$\mathcal{M} = U\Sigma V^{\dagger},$$

where † is the symbol for the transpose of V. The matrices $U = U(t_k, S)$, $\Sigma = \Sigma(t_k, S)$ (both with dimension $N \times N$) and $V = V(t_k, S)$ (dimension $(W_s f_r + 1) \times N$) provide a low-dimensional representation of the internal structure of the data from its most informative

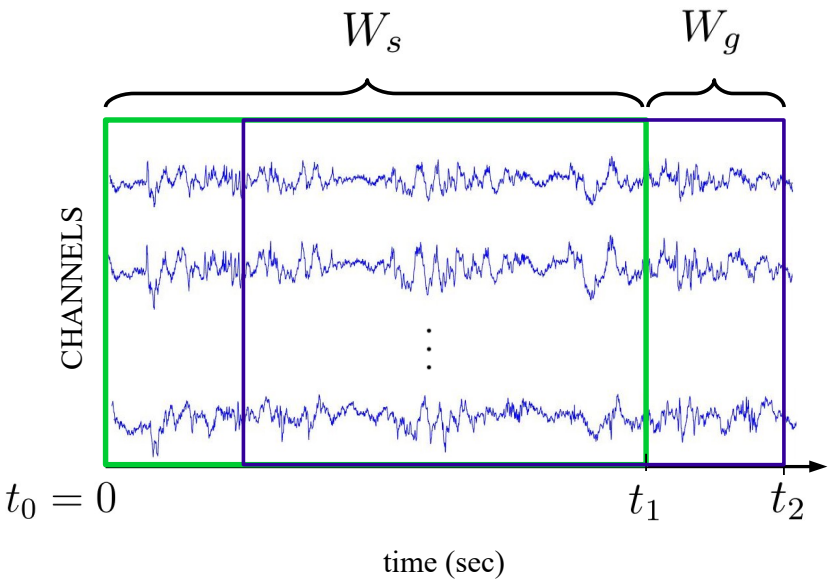

Figure 2.1: **Sliding window parameters.** We illustrate the first two sliding windows. Both time window size W_s and the overlap gap W_g are measured in seconds.

(correlated) viewpoint. In fact, by denoting u_i and v_i the column vectors of U and V respectively, we know from theory that $\{u_i\}_{i=1}^N$ forms a orthonormal basis for the span of the columns vectors of $\mathcal{M}$. The matrix Σ is diagonal and contains the singular values $\{\sigma_i\}_{i=1}^N$ in a decreasing order. Thus the best *rank - r* linear approximation of $\mathcal{M}$ is given by

$$\mathcal{M}^{(r)} = \sum_{j=1}^{r} \sigma_j u_j v_j^\dagger$$

For each singular value σ_i, we compute the associated energy $E(\sigma_i)$ defined by

$$E(\sigma_i) = \frac{\sigma_i}{\sum_{j=1}^N \sigma_j}$$

and the tail of the energy distribution is a measure for data complexity (Fig. 2.3b). Fast decay indicates low-dimensional dynamics: The first SVD-modes play a significant role in data representation. From the neurological point of view this indicates highly correlated LFP signals across channels, which may be seen as a synchronized activity. On the other hand, slow decay tails means that data has more complex structure which implies, low signal correlation, thus indicating desynchronized activity in the brain network.

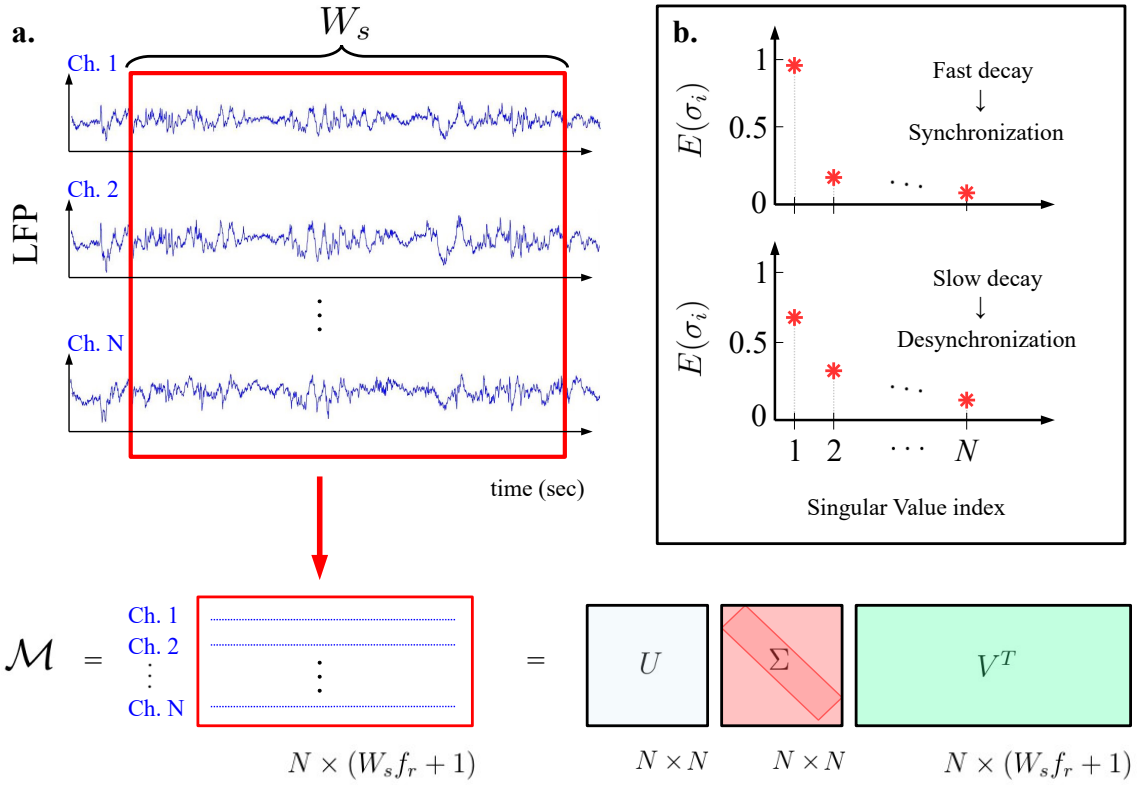

Figure 2.2: **The Singular Value Decomposition of the data matrix.** For each time interval $\big[(k-1)W_g, t_k\big)$ and given the sampling rate f_r (in Hz), we proceed as folows: **a.** We build a data matrix $\mathcal{M} = \mathcal{M}(t_k, S)$ with N rows and $W_s f_r + 1$ columns with the LFP measured values in each of the N channels. We compute the SVD of $\mathcal{M}_k(W_s, f_r)$ and also the Energy Distribution ($E(\sigma_i)$ for $i \in \{1, 2, \ldots N\}$) for the singular values of this decomposition. **b.** The decay of the tail of such distribution indicates redundancy in the data and is associated with synchronization in the LFP channels network.

Once the Energy from the singular values of $\Sigma(W_s, f_r)$ is obtained, we give a score for its distribution tail by fitting a Pareto density function ([84])

$$\rho(x, \alpha) = \frac{\alpha}{x^{\alpha+1}}, \quad x \geq 1$$

via least-squares optimization ([85]). Fig. 2.3a illustrates this step in our methodology. The Pareto probability density has the following property: For high values of the α parameter, the distribution tail exhibit a fast decay behaviour. On the contrary, low α's

represents slow decay tails. In the context of this work, high and low α's may be associated to synchronized and desynchronized activity across different channels, respectively. For given values of Ws and W_g, each matrix $\mathcal{M}(t_k, S)$ we associate a positive number $\alpha(t_k)$ that gives a synchronization score for the data on a W_s - seconds window. Fig. 2.3b illustrates this step for the first two sliding windows.

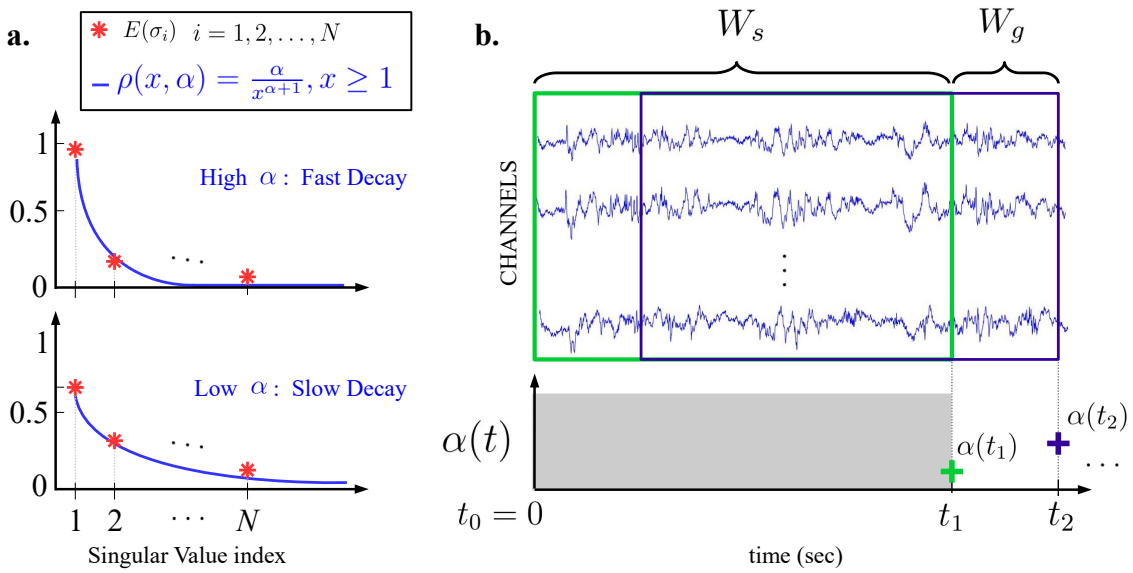

Figure 2.3: **Building the α - series.** **a.** The energy distribution $E(\sigma_i)$ of the singular values (red stars) is fitted with a Pareto density function $\rho(x, \alpha)$ (blue curve). From the fitting process we keep the α parameter as a score of the distribution. High/low α are indicators of fast/slow decay tails and thus synchronized/desynchronized activity across channels. **b.** Two temporal parameters: W_s is the size of the time window and W_g is the time gap between two overlapped windows. Both assume a constant value across the N different channels. We show the first two consecutive time windows at $t_1 = W_s$ and $t_2 = W_s + W_g$ with their respective α -score. This process spans the entire time interval and the resulting α-series can be used as tools for analysing the evolution of synchronized activity in the network.

2.3 Sanity Check: Kuramoto model

We apply our data-driven method on a theoretical model of in order to validate our assumption that high and low values of the parameter α from our Pareto density function are correlated with synchronized / desynchronized activity. We have chosen the well known non-linear mathematical model of synchronization - the Stochastic Kuramoto Model - where the phase $\left(\theta_t^i \in [0, 2\pi]\right)$ of N oscillators are coupled under the following dynamics

$$d\theta_t^i = \left(\omega_i + \frac{K(t)}{N} \sum_{j=1}^{N} \sin(\theta_t^j - \theta_t^i) \right) dt + \varepsilon d\xi_t^i \quad i = 1, 2, \ldots, N$$

Here $K(t)$ is a real function representing the strength between any two oscillators. Each element i has a natural frequency ω_i and ξ_t^i are independent standard Brownian motions describing the effects of noise in the system. The electric oscillations in the signal are represented by $X_t^i = \cos(\theta_t^i)$ for $i = 1, 2, \ldots, N$. Fig. 2.4 shows an example that confirms our hypothesis. In this simulation we have used the following parameters:

- Number of vertices $N = 24$.
- Noise intensity $\epsilon = 0.05$.
- Natural frequencies $\{\omega_i\}_{i=1}^N$ and initial conditions $\{\theta_0^i\}_{i=1}^N$ independent and uniformly distributed on $[0, 2\pi]$.
- Total of $T = 90$ seconds sampled at 2000 Hz.

We numerically solve the Kuramoto model with the Euler-Maruyama ([86]) scheme for a coupling strength function given by

$$K(t) = \begin{cases} 0 & \text{if } t < 30 \\ 10 & \text{if } 30 \leq t < 60 \\ 0 & \text{if } t \geq 60. \end{cases}$$

The α - series have been produced with a time window size $W_s = 1$ seconds and time overlap $W_g = 0.25$ seconds. As we expected, higher values of $\alpha(t)$ (around 0.8 on Fig. 2.4

) follows synchronized activity when $t \in [30, 60)$ and $K(t) = 10$. For a system without coupling ($K = 0$), the natural frequencies dominate the dynamics and a desynchronized activity with lower $\alpha(t)$ (around 0.4) is observed. This theoretical example illustrates how the α - series gives rich information about the evolution of redundancy and therefore synchrony on artificial networks of oscillatory elements.

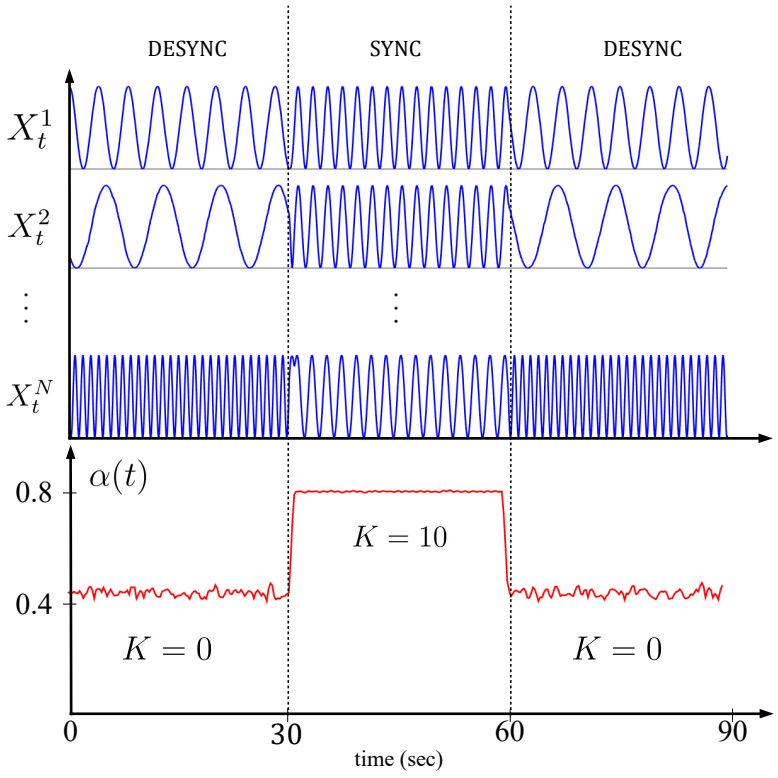

Figure 2.4: **A sanity check example.** We simulate an artificial network using the stochastic Kuramoto Model with coupling strength given by a step function $K(t)$. For $K(t) = 0$ the system is decoupled and a desynchronized activity is followed by lower values of $\alpha(t)$. When synchronization takes place (strong coupling with $K(t) = 10$ for $t \in [30, 60)$), we observe an instantaneous increasing for $\alpha(t)$, which remains at higher values until $t = 60$, when $K(t)$ turns back to 0 again. High values of the α parameter are correlated with strong levels of synchronization in the network.

Chapter 3

Results

In this chapter we present the results of our methods into two parts. In the first part we highlight the contributions of the α - series method for a deeper understanding of the epileptic process in the brain. In the second part we introduce a seizure detection algorithm based on the same methodology. By defining new variables and notation for explaining our results, we show the results with respect to different seizures and window parameters.

3.1 Neurobiological implications of the α - series method

3.1.1 Desynchronized activity during Seizures epoch: The α - drop seizures

We have applied the α - series methodology on our data. By scanning the whole LFP time series, it has been possible to characterize, from the synchronization point of view, the different regimes across the recording periods. The first interesting result is the sudden changes in the signal at the seizure epoch. Fig. 3.1 shows an example of this result for S_7, $(W_s, Wg) = (1.5, 0.25)$, both time windows measured in seconds. The whole α - series is shown in Fig. 3.1a. By zooming a time window around the seizure, a clear drop on the α parameter appears during seizures onset, followed by a rapid increase during the offset (Fig. 3.1b). This pattern has been found for 15 seizures, which represents 79% of our

dataset. We have called them α - *drop seizures*. Figs. 5.3, 5.4 and 5.5 exhibits the LFP trace (channel 1) for such seizures with their respective α - series. This results highlights the interplay between synchronized and desynchronized activity during the seizure. It shows that desynchronization may play a major role in the epileptic process and that seizure termination may be promoted by the increase of the synchronization level.

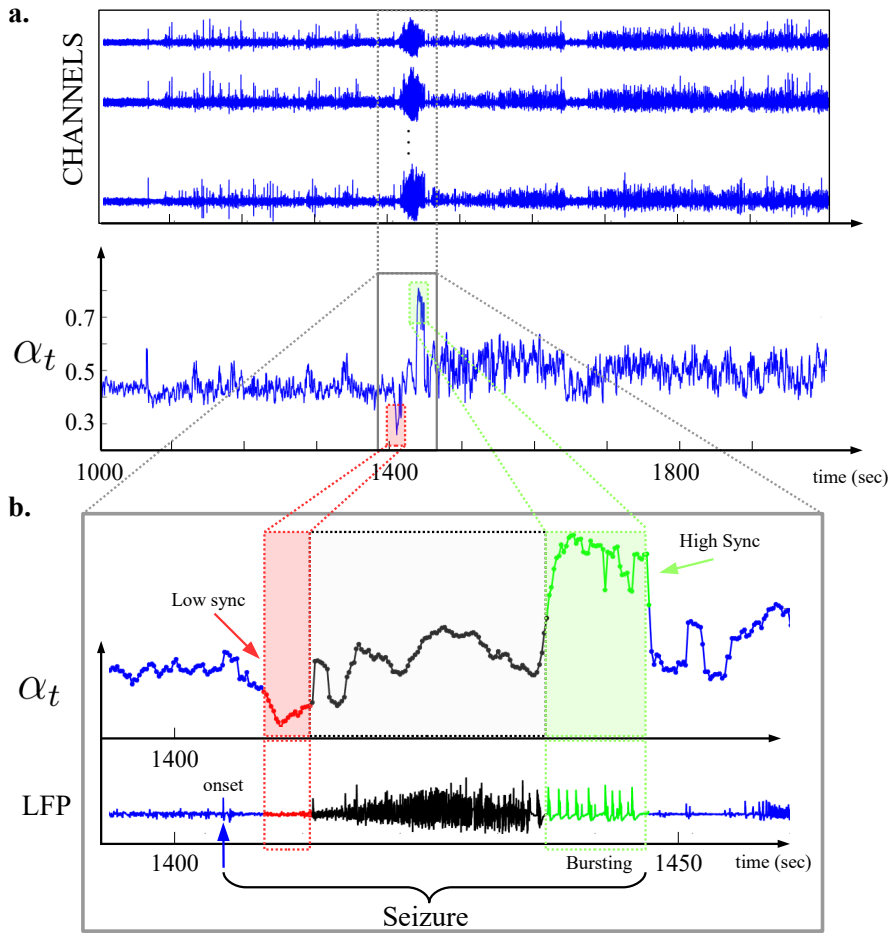

Figure 3.1: **Critical changes on synchronization during epileptiform acitivity.** For most of the seizures, we have found a significant drop of the α-series near the seizures onset, thus indicating desynchronized activity. On the other hand, the offset is characterized by high synchronization during the burst activity. This result allow us to recognize the complexity of the seizure as a dynamic process. In this example we have used the LFP signal of Seizure S_7, with $(W_s, Wg) = (1.5, 0.25)$.

3.1.2 Postictal depression characterized by high synchronized activity

Our method have allowed us to investigate both short and long-term effects of the seizures in the HPC. Once the α - series have been obtained, we have analysed the different synchronization levels during both pre and post seizure epochs. By choosing time intervals of length $T = 1$ or 10 minutes, we have built histograms with α values before and after the epileptiform activity (from now on called α - histograms). Fig. 3.2a highlights this procedure and Fig. 3.2b shows the results for S_7. The mean of α values during the pre and post seizure epochs are indicated by vertical red and green lines respectively. In this example we have chosen $(W_s, Wg) = (1.5, 0.25)$. Table 3.1 shows the average difference between such mean of α values for each α - drop seizure, for all $(W_s, W_g) \in \Omega_s \times \Omega_g$ where $\Omega_s = [1.5 : 0.5 : 4]$ and $\Omega_g = [0.25 : 0.25 : 1.5]$. For $T = 1$ (short term), the average difference was positive for all seizures except S_5, S_{17} and S_{18}. On the other hand for $T = 10$ (long term) only seizure 5 has shown a negative difference. Hence for almost all seizures we have found higher α - values during the post seizure epoch, thus indicating that the postictal depression is characterized by high synchronization levels.

3.2 The Seizure Detection algorithm

The optimal threshold. Training and testing procedure.

We have built normalized versions of the α - series for the seizure detection process. For each positive integer $k \geq 2$ and $t_k = t_k(W_s, W_g)$ as defined in section 2.2, we have introduced the β_{t_k} parameter of a given seizure as

$$\beta_{t_k} = \frac{\alpha_{t_k} - \langle \alpha_{t_1}, \alpha_{t_2}, \ldots \alpha_{t_k} \rangle}{\sigma(\alpha_{t_1}, \alpha_{t_2}, \ldots \alpha_{t_k})}$$

where the symbols $\langle . \rangle$ and $\sigma(.)$ represents mean and standard deviation respectively. The real numbers β_{t_k} represent a measure of the discrepancy between α_{t_k} and the whole α - series until time t_k. For notational purposes, we may write $\beta_{t_k} = \beta_{t_k}(S)$ if we want to stress the dependence on a seizure S.

In order to quantify the seizure detection process, for a given real number $\bar{\beta}$, we define the $\bar{\beta}$ -*detection time* for seizure S as

Table 3.1: The average difference of mean α values for pre and post seizure epochs across $(W_s, W_g) \in \Omega_s \times \Omega_g$.

Seizure	$T = 1$ **min**	$T = 10$ **min**
3	0.0417	-0.0124
4	0.2250	0.0814
5	0	-0.0001^*
6	0.0454	0.0326^*
7	0.0942	0.0650
8	0.0364	0.0157
9	0.0684	0.0322
10	0.0533	0.0486
11	0.0612	0.0477^*
12	0.0661	0.0309
13	0.0518	0.0226
15	0.0229	0.0194
16	0.0304	0.0235^*
17	-0.0002	0.0090
18	-0.0092	0.0105

*For seizures S_5, S_6, S_{11} and S_{16} we didn't have access to an interval of $T = 10$ minutes before or after the seizure epoch. Therefore the maximum interval was chosen for both pre and post seizure: $T = 5.2$ for S_5, $T = 3.7$ for S_6, $T = 5.02$ for S_{11} and $T = 8.9$ for S_{16}.

$$\tau_d(\bar{\beta}, S) = \min\{t_k : \beta_{t_k}(S) \leq \bar{\beta}\}.$$

In words, $\tau_d(\bar{\beta}, S)$ is the first time where the β - series for seizure S reaches threshold $\bar{\beta}$.

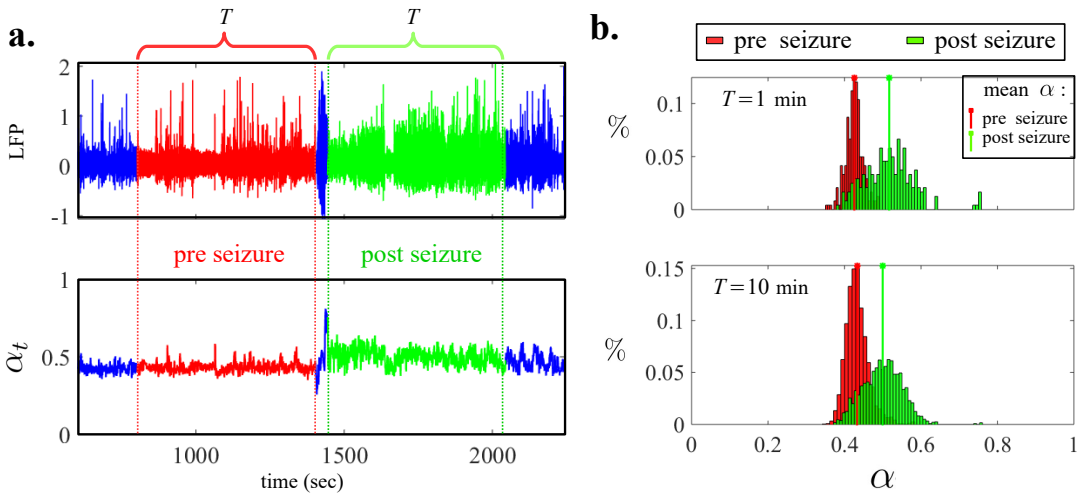

Figure 3.2: **Postictal depression is characterized by short and long-term synchronized activity**. **a.** In the top we show the LFP near the S_7 epoch. We set a time window with size T minutes before and after the seizure. In this work we have chosen $T = 1$ or 10 minutes. The bottom plot shows the same time windows for the α - series. **b.** The distribution of the α parameter for pre and post seizure epochs is evaluated. This example shows higher values for the post seizure epoch, thus indicating strong correlation between the occurrence of the seizure and the increased synchronization level in the HPC. The means of the α parameter are represented by solid vertical lines, indicating the shift between the distributions. For this simulations we have used $(W_s, Wg) = (1.5, 0.25)$ seconds.

Moreover, given the onset time $t_b(S)$, the $\bar{\beta}$ - *detection delay*

$$\Delta(\bar{\beta}, S) = \tau_d(\bar{\beta}, S) - t_b(S)$$

is the time the distance between our detection time and the beginning of the seizure as defined by experienced investigators ([79]). Fig. 3.3 illustrates both α and β - time series and the detection parameters $\tau_d(\bar{\beta}, S)$ and $\Delta(\bar{\beta}, S)$ for $S = S_7$, $\bar{\beta} = -5$ and $(Ws, Wg) = (1.5, 0.25)$. The drops of the α and β - series on the seizure epoch are aligned and represent abnormal desynchronized activity.

The first part of the detection process consists on finding an optimal β - threshold for

each α - drop seizure. Thus we define $\beta_{opt} = \beta_{opt}(S)$ by

$$\beta_{opt} = \big\langle \operatorname*{argmin}_{\beta} |\Delta(\beta, S)| \big\rangle.$$

In words, the optimization is done by evaluating the mean of all β such that the minimum of $|\Delta(\beta, S)|$ is achieved. For me optimization process, β ranges on the discrete interval $[\min_k(\beta_{t_k}) : dt : 1]$ for $dt = 10^{-3}$. Table 3.2 shows the list of $\beta_{opt}(S)$ and $\Delta(\beta_{opt}, S)$ for each seizure S and $(W_s, W_g) = (1.5, 0.25)$.

Table 3.2: Optimal $\beta_{opt}(S)$ and $\Delta(\beta_{opt}, S)$ for each α - drop seizure S. We have chosen $(W_s, Wg) = (1.5, 0.25)$

Seizure	β_{opt}	$\Delta(\beta_{opt}, \cdot)$ **(sec)**
3	4.7	5.5
4	2.6	4.8
5	5.3	10.0
6	4.1	6.5
7	7.8	5.0
8	5.9	14.0
9	2.9	4.5
10	3.7	6.2
11	5.7	23.0
12	5.1	4.7
13	3.5	14.0
15	3.5	13.0
16	4.9	19.0
17	4.7	9.2
18	4.4	9.8

For fixed W_s and W_g and given the optimal $\beta_{opt}(S)$ values for all α - drop seizures

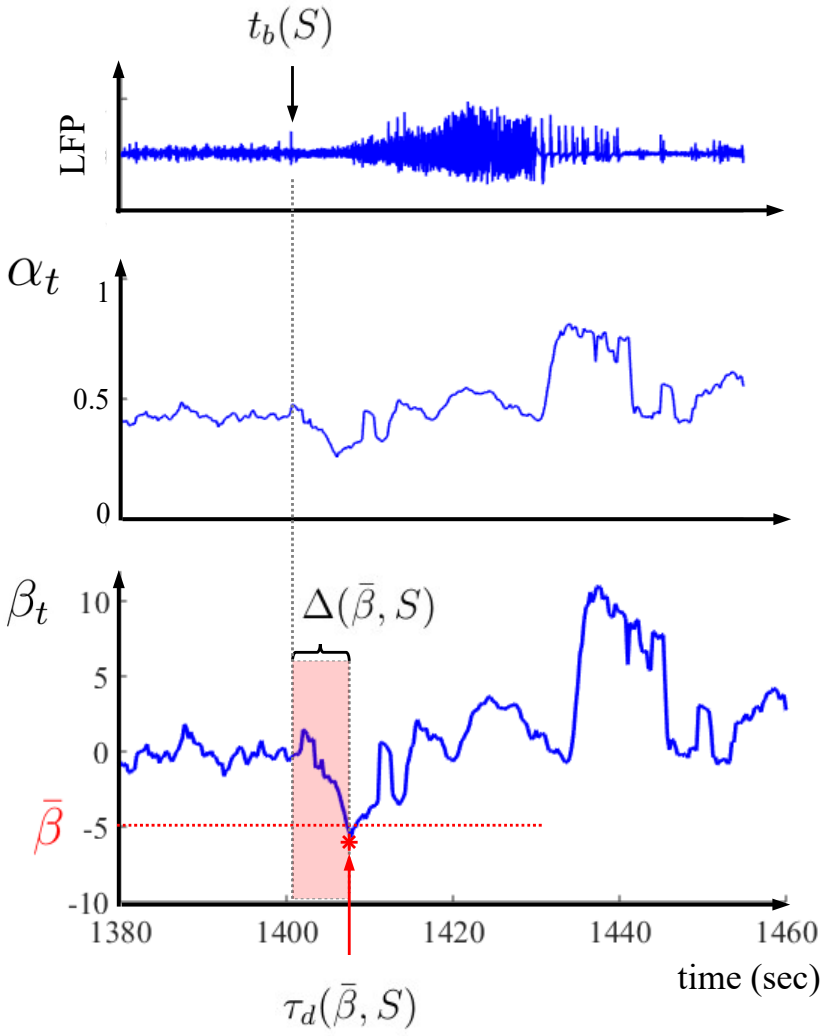

Figure 3.3: **Example of β - series and the seizure detection parameters for S_7.** The spike criterion gives the onset time $t_b(S_7)$. For $Ws = 1.5$ and $Wg = 0.25$, the β - series are calculated by normalizing the α - series. The detection time $\tau_d(\bar{\beta}, S_7)$ is defined as the first time at which the β parameter crosses down the threshold $\bar{\beta}$ (equals -5 in the example). The detection delay $\Delta(\bar{\beta}, S_7)$ is time distance between the detection time and the seizures onset.

S, we have been able to perform a seizure detection process by a training and testing procedure with three steps:

1. Pick a random subsample of d seizures and label them as r_1, r_2, ... r_d.

2. Define the training threshold $\beta^* = \langle \beta_{opt}(r_1), \beta_{opt}(r_2), \ldots, \beta_{opt}(r_d) \rangle$

3. Evaluate the detection delay $\Delta(\beta^*, \cdot)$ for each one of the other $15 - d$ seizures.

By repeating these steps we have been able to measure the mean detection delay for each seizure and for different values of W_s and W_g.

3.2.1 Results for each α - drop seizure.

The training/testing procedure has been performed for W_s and W_g taking values at the discrete intervals $\Omega_s = [1.5 : 0.5 : 4]$ and $\Omega_g = [0.25 : 0.25 : 1.5]$ respectively. For each pair $(W_s, W_g) \in \Omega_s \times \Omega_g$, the algorithm has been repeated 100,000 times. By denoting $n(S)$ as the number of iterations in which seizure S has been chosen for test and for a training threshold β^* (as defined in the previous subsection), we say that a seizure is detected at some iteration if

$$|\Delta(\beta^*, S)| \leq 30.$$

We say that a false positive occurs if at some iteration the absolute time distance between the detection time and the onset time is greater than 30 seconds. We then define

$$g(S) = \frac{\{\text{num. of iterations where} \quad |\Delta(\beta^*, S)| \leq 30\}}{n(S)}$$

as the *detection rate* for the seizure S. Hence $1 - g(S)$ is the detection rate of false positives before S. The average results across $(W_s, W_g) \in \Omega_s \times \Omega_g$ are shown in table 3.3. Seizures S_3, S_6, S_7, S_{10} and S_{16} have been perfectly detected, whereas seizures S_8,S_{17} and S_{18} were almost never detected. The other seizures were partially detected at different levels of $g(S)$. The best detection delay was $\Delta(\beta^*, S_{10}) = -8.1$ seconds, which means that our method was able to identify S_{10} before its onset time. For the perfectly detected seizures, the worst detection delay was 14 seconds for S_6. The average $g(S)$ and mean $\Delta(\beta^*, S)$ across the α - drop seizures were 0.69 and 9.53 respectively.

Table 3.3: Results: average for $W_s \in [1.5 : 0.5 : 4]$ and $W_g \in [0.25 : 0.25 : 1.5]$

Seizure	$g(\cdot)$	**mean** $\Delta(\beta^*, \cdot)$
3	1.0	7.6
4	0.84	9.3
5	0.13	6.9
6	1.0	14.0
7	1.0	5.4
8	0	5.0
9	0.71	9.9
10	1.0	-8.1
11	0.82	8.2
12	0.97	10.0
13	0.55	8.7
15	0.35	19.0
16	1.0	18.0
17	0	15.0
18	0	14.0
average	0.69	9.53

3.2.2 Best window parameters (W_s, W_g)

The detection results can also be analysed with respect to the window parameters (W_s, W_g). Tables 3.4 and 3.5 shows, for each $(W_s, W_g) \in \Omega_g \times \Omega_s$, the results of the average detection rates $g(S)$ and detection delay $\Delta(\beta^*, S)$ across the 15 α - drop seizures. The highest/lowest average $g(S)$ were found at $(W_s, W_g) = (4.0, 1.5)$ and $(W_s, W_g) = (1.5, 0.25)$ respectively. The best/worst average detection delay (min/max $\Delta(\beta^*, S)$) were found at $(W_s, W_g) = (4.0, 0.25)$ and $(W_s, W_g) = (1.5, 1.5)$ respectively. In order to obtain the best window parameters (W_s, W_g), we compute the quotient between the average $g(S)$ and $\Delta(\beta^*, S)$

as a joint detection score. Table 3.6 shows this parameter for each $(W_s, W_g) \in \Omega_g \times \Omega_s$. We have found the best and worst detection parameters at $(W_s, W_g) = (4.0, 0.25)$ and $(W_s, W_g) = (1.5, 1.5)$ respectively. It is worth noting that the maximum quotient has been found with the largest W_s and smallest W_g among the chosen values. This indicates that bigger windows with a strong overlap should increase the accuracy of the seizure detection process. Table 3.7 shows $g(S)$ and mean $\Delta(\beta^*, S)$ across the iterations at which seizure S was detected, for $W_s = 4$ and $W_g = 0.25$ seconds. The mean parameters across seizures have also been computed: the mean detection rate and mean detection delay were given by 0.6 and 7.08 seconds respectively.

Table 3.4: Average $g(S)$ (detection rate) across seizures for each pair $W_s \in \Omega_s = [1.5 : 0.5 : 4.0]$ and $W_g \in \Omega_g = [0.25 : 0.25 : 1.5]$

W_g \ W_s	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**
0.25	0.54	0.57	0.59	0.59	0.59	0.6
0.5	0.6	0.62	0.59	0.6	0.61	0.61
0.75	0.59	0.66	0.6	0.62	0.65	0.64
1.0	0.64	0.67	0.65	0.62	0.66	0.65
1.25	0.66	0.63	0.62	0.65	0.65	0.64
1.5	0.58	0.65	0.68	0.64	0.66	0.7

The minimum detection score $g(S)$ (blue color) was found for $(W_s, W_g) = (1.25, 0.25)$ and the maximum (red) for $(W_s, W_g) = (4.0, 1.5)$.

Table 3.5: Average mean $\Delta(\beta^*, S)$ (detection delay) across seizures for each pair $W_s \in \Omega_s = [1.5 : 0.5 : 4.0]$ and $W_g \in \Omega_g = [0.25 : 0.25 : 1.5]$

W_g \ W_s	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**
0.25	9.93	9.28	8.81	8.2	8.12	7.08
0.5	9.95	9.39	8.76	8.19	8.2	7.32
0.75	10.3	9.77	8.22	8.44	8.71	7.89
1.0	10.4	9.65	8.95	7.99	8.96	8.38
1.25	9.98	9.71	9.61	9.57	9.37	8.5
1.5	11.1	9.8	9.65	9.63	9.47	9.3

The maximum $\Delta(\beta^*, S)$ (blue color) was found for $(W_s, W_g) = (1.5, 1.5)$ and the minimum (red) for $(W_s, W_g) = (4.0, 0.25)$.

Table 3.6: Average $g(S)$ / Average $\Delta(\beta^*, S)$ across seizures for each pair $W_s \in \Omega_s = [1.5 : 0.5 : 4.0]$ and $W_g \in \Omega_g = [0.25 : 0.25 : 1.5]$.

W_g \ W_s	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**
0.25	5.47*	6.18	6.68	7.16	7.29	8.5
0.5	6.08	6.57	6.71	7.29	7.46	8.32
0.75	5.75	6.79	7.25	7.32	7.43	8.14
1.0	6.1	6.96	7.25	7.72	7.38	7.74
1.25	6.59	6.5	6.43	6.83	6.97	7.58
1.5	5.26	6.59	7.04	6.62	6.94	7.55

* All results of this table are multiplied by 100.

Table 3.7: optimal results: $(W_s, Wg) = (4, 0.25)$

Seizure	$g(\cdot)$	**mean** $\Delta(\beta^*, \cdot)$
3	1.0	2.4
4	0.94	8.7
5	0.081	6.6
6	1.0	14.0
7	1.0	4.9
8	0	NaN
9	0.52	9.8
10	1.0	-8.9
11	0.56	8.1
12	1.0	10.0
13	0.92	7.3
15	0	NaN
16	1.0	15.0
17	0	NaN
18	0	NaN
Average	0.6	7.08

Chapter 4

Discussion

In this work we developed a seizure detection algorithm based on a data-driven method for scanning synchronization levels of a neural network, based on the low-dimensional structure of sliding data matrices. The energy distribution of the singular values were fitted with a Pareto density function and its α parameter was used as a score for the tail decay for each time window. A threshold based on the drop of the synchronization level during the seizure epoch was used as the detection parameter. Hence our methodology has two main features: (i) the capability to provide useful information about the neural mechanisms of a seizure based on the α - series and (ii) a detection algorithm that could be utilized as the basis of a seizure warning system for epileptic patients.

Desynchronization at seizures onset.

The α - series method have shown a drop on the synchronization level during the seizure epoch for 15 of the 19 (79%) seizures in our dataset. Similar results about desynchronized activity were found recently in other contexts and with other methods. Specifically for the pilocarpine model of TLE, our findings showed substantial agreement with the recent work of Wang et al ([54]), where connection probabilities where calculated in order to measure the change in cross-correlogram (CCG) peaks. This desynchronization phenomenon may be explained by the fact that seizure initiation could be caused by multiple distributed cortical domains, as suggested by Jiruska et al [55]. Another possible interpretation of this

phenomenon was suggested by Le van Quyen et al ([87]): Early ictal desynchronization could be the result of localized activity in the region that is generating the seizure, while the other regions are still unaffected. Further studies could be done, for instance, by removing channels and evaluating the α - series of the remaining dataset for locating the seizure focus area. It is also worth noting that our result about desynchronization during the ictal period should not be extended as a general principle for all kinds of seizure: Each experimental model and epilepsy syndrome has its own particularities and synchronization mechanisms. A distinguishing feature of our α - drop biomarker is that it is actionable, in so far as it is the basis for our seizure detection algorithm.

Seizure termination and postictal depression

The mechanisms of spontaneous seizure termination are still poorly understood. The α - series method indicates a possible explanation based on the sudden increase of the synchronization level just after the α - drop. This pattern was observed in most of the seizures in our dataset and indicates that increasing synchronization may be casually related with seizure termination. Similar results were found by Schindler et al ([88]) for human intra-cranial and surface EEG recordings. For the postictal state, our methodology was capable of characterizing such periods from the synchronization viewpoint. The electrophysiological effects of the postical depression are well known for TLE experimental models ([89], [90], [91],[92]). In this work the post ictal state has been characterized by high synchronization levels for most seizures if compared to the pre seizure epoch, even for a 10 minutes time window. We have measured this feature with histograms for the α values for both epochs. The post seizure mean α's were higher than the pre seizure ones in all cases, except for S_5. This indicates that the seizure produces a reverberant activity in the hipocamppus. We hypothesize that such phenomenon for TLE is also an indicative of a plasticity process in the epileptic brain ([93],[94], [95],[96],[97], [98]).

The detection process

The detection times were compared with the spike criterion which embodies the opinion of a experienced investigators and was defined by Queiroz et al previously ([79]). The results for each seizure (averaged across different sliding windows) yielded an average (across seizures) detection rate 0.69 and detection delay 9.53 seconds. By defining the quotient between detection rates and delays as a joint score, the optimal sliding window values were given by $W_s = 4$ seconds and $W_g = 0.25$. For these time parameters we have found average (across seizures) detection rate 0.6 and detection delay 7.08 seconds.

We provide a short list with both classical and recent interesting papers which have also calculated average detecion delays, with their respective summary of methods. For the optimal window parameters, our delay results were better than Osorio et al. [99], Qu et al. [100], Saab et al. [61] and Aarabi et al. [101]. On the other hand, better results are found in Gardner et al. [102], Kharbouch et al. [103] and Ahammad et al. [104]. It is worth noting that Kharbouch et al. and Ahammmad et al. had 67 and 184 seizures to analyse, respectively, therefore being larger scale studies. Gardner et al had 29 seizure, 10 more than our dataset. Finally, our method is considerably simpler from the mathematical point of view, but even though it has shown good results if compared with previous studies.

Limitations of our method

Here we describe several limitations of this work. Our method is based on the SVD of data matrices for the N - channel LFP data. Our synchronization measure is the α parameter of a Pareto density function, obtained via data fitting of the energy distribution of the singular values. Other distributions could be in principle used for the same purpose. Our first candidate was the exponential distribution, but the fitting process did not work well because the SVD energy tails generally showed faster decay than exponential. Other synchronization measures could also be applied on our dataset in order to obtain comparative results, such as autocorrelation function ([105]), synchronization likelihood ([106]) and nearest neighbor phase synchronization ([107]). It would be interesting to use

Table 4.1: Classical and recent works on seizure detection.

Paper	**Delay (sec)**	**Summary of methods**
Osorio et al. [99]	15.5	Wavelets & image processing
Qu et al. [100]	9.35	Nearest - Neighbour Classifier
Saab et al. [61]	10	Seizure probability of EEG sections
Gardner et al. [102]	-7.58	SVM for energy statistics
Aarabi et al. [101]	11	Fuzzy rule system
Kharbouch et al. [103]	5	SVM for spectral properties.
Ahammad et al. [104]	1.76	Wavelet - based features
Our work	7.08	desynchronization threshold

such techniques and look for some analogous for the α - drops, which could also provide alternative detection algorithms.

We have developed a new seizure detection method, but several authors have been working on the problem of seizure prediction, which consists on identifying pre-ictal state sufficiently long before the electro-graphical seizure onset ([49], [108]). Seizure prediction is a difficult problem ([109]), but at the same time more efficient for the development of implantable devices and warning systems. Our α - series method does not show any significant temporal signature before the seizures.

The drop of the α parameter is the cornerstone of our detection algorithm, but such pattern was not found for 4 of 19 seizures in our dataset. Till date, we haven't found a definitive explanation for such lack of dropping profile. We hypothesize that those 4 seizures have their seizure onset zone out of the HPC, where the channels are located. In fact, the drop of the α parameter indicates desynchronization, which might be seen as a sign that the ictal activity has started at some region of the HPC. Therefore, if the drop does not occurs, it could be an indicative that the seizure started at some other region of the brain.

From all 15 seizures where the α drops occurred, 3 of them (S_8, S_{17} and S_{18}) had almost

zero detection rates $g(S)$. Such problem is related to the existence of false positives, which according to our criterion means that the time distance between the onset and detection times was greater than 30 seconds. We could improve our method by applying some machine learning algorithm to classify the profile of the α - series during such false positives and discriminate them from those of the seizures epoch. In this work we also did not compute the false positive rates for each animal, which is the quotient between number of false positives and the total duration time in hours. This quantity is well known in the literature and could be used for comparison with other existing detection algorithms.

We have explored the α - series and detection results for a wide range of window parameters. The chosen discrete intervals were $\Omega_s = [1.5 : 0.5 : 4]$ and $\Omega_g = [0.25 : 0.25 : 1.5]$ for W_s and W_g respectively. Our results indicated $W_s = 4$ and $W = 0.25$ as the best windows size and gap for the detection process. Surprisingly, such values corresponds to our highest window size and lowest window gap. Further analysis could be employed in order to extract optimal parameters from more refined intervals. From the practical point of view, we believe $W_s = 4$ and $W = 0.25$ could be kept as the best choices for our method.

Practical implementation

This work was intended to introduce the theoretical background of a new seizure detection method based on the low-dimensional structure of the LFP data matrices. Future work should address the limitations described above in order to provide the basis for the practical implementation of our methodology. We could also apply the α - series on more TLE datasets to verify our results about desynchronization during seizure onset and calculate the detection parameters with the same training and testing algorithm. A larger dataset would provide a statistically more robust detection method. An other interesting point is about the capability of our method to perform the seizure detection before the 'points of no return', from which the epileptic process will inevitably evolve to a clinical seizure ([109]). Such question is fundamental for designing a reliable and efficient alarming system for seizure detection. Such system would need to alert when the occurrence of an epileptic seizure is imminent and also provide mechanisms for its abortion.

Conclusions

Advances in seizure detection are giving rise to implantable devices able to trigger therapies to prevent or abort epileptic attacks. Here we introduce a simple mathematical tool for analysing synchronization of a neural networks which also provides the basis for an early detection system. We have found time signatures indicating structural changes on hippocampal networks at the onset of seizures. A drop of the synchronization level was our biomarker for the detection algorithm. Detection rates and time delays were evaluated by a training and testing procedure with all seizures for which the desynchronized activity was found. On the other hand, the postictal epoch was characterized by high levels of synchronous activity, thus highlighting the long term effects of the seizure in the hippocampus. The techniques described here can be used for research in the area of synchronization at different neuronal levels.

Chapter 5

Supplementary Information

This chapter contains technical information about our method and also a summary of the α - drops and α - histograms for all seizures.

5.1 Filtering process

In order to eliminate noise keeping high frequency oscillations which are relevant for our purposes, raw electrophysiology data were band pass filtered bellow 600 Hz. We allow such high frequency values in order to capture interesting oscillations such as fast ripples ([110, 111]) even though they are usually found only during the interictal periods, which are not our main focus in this work. Fig. 5.1 describes the main steps of this process: For each channel and for a fixed W_s - seconds time window, we apply a shift - Fast Fourier Transform (FFT) $\mathcal{F}$ on the LFP signal and compute the complex magnitude associated to each frequency f. We then multiply this magnitude by a Gaussian filter function Φ with cutoff [112], which is given by

$$\Phi(f) = \begin{cases} e^{-\kappa f^2} & \text{if } |f| \leq 600\text{Hz} \\ 0 & \text{if } |f| > 600\text{Hz} \end{cases}$$

The filter Φ cuts off frequency oscillations above 600 Hz but also has a smoothing effect high frequencies below this threshold. After this step, we apply the inverse FFT $\mathcal{F}^{-1}$ and obtain the filtered signal.

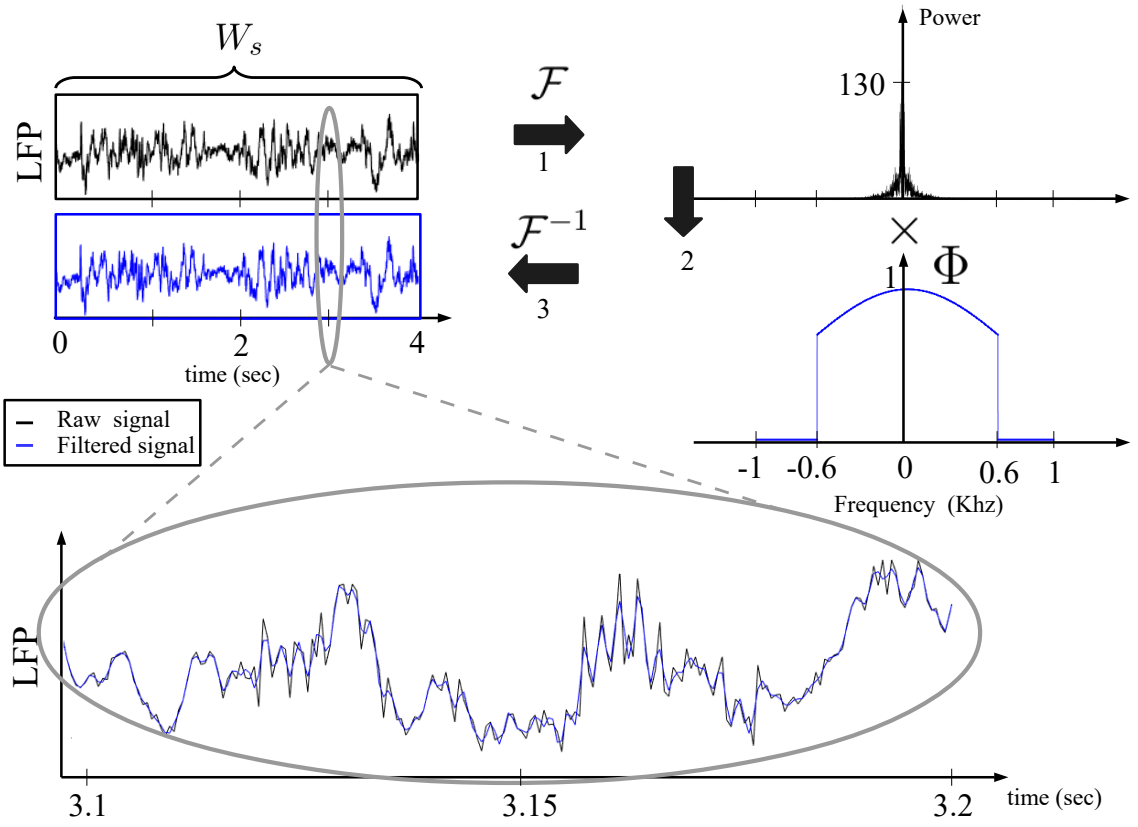

Figure 5.1: **The filtering process**. For a time window of length W_s seconds we pick the raw data of each channel and proceed as follows: 1. We apply the shift - FFT $\mathcal{F}$ and take the power spectral density of the signal. 2. Multiply the power density by the filter function Φ that smooths high frequency oscillations and cuts off those above 600 Hz. 3. With the inverse FFT transform (denoted by $\mathcal{F}^{-1}$) we get the filtered signal. In the bottom of this picture we plot a zoom of both raw and filtered signals.

5.2 SSR of the Pareto fitting process

The goodness of fit in the process of computing the α - series has been analysed. For each seizure S and given W_s and W_g, the α parameter of the Pareto distribution $\rho(x, \alpha)$ has been computed at times t_k for the data matrix $\mathcal{M}(t_k, S)$, where $t_k = W_s + (k-1)W_g$ for all $k = 1, 2, ...k^*$ as defined in section 2.2. The α - series $\{\alpha(t_k)\}_{k=1}^{k^*}$ describes the evolution of synchronization across the LFP traces. The associated *Sum of Square Residuals* (SSR)

time series is given by

$$\mathcal{E}(t_k) = \sum_{i=1}^{N} \Big(\rho(i, \alpha(t_k)) - E(\sigma_i)\Big)^2$$

where $E(\sigma_i)$ is the energy of the singular value σ_i and N is the number of singular values coming from the SVD decomposition of $\mathcal{M}(t_k, S)$. Fig. 5.2 shows an example for S_7, $(W_s, Wg) = (1.5, 0.25)$. In this example, $\mathcal{E}(t_k) = 0.5$, obtained at $t = 1437.75$ and is indicated by the red arrow. We have calculated $\max_{t_k} \mathcal{E}(t_k)$ for all seizures, $W_s \in \Omega_s = [1.5 : 0.5 : 4.0]$ and $W_g \in \Omega_g = [0.25 : 0.25 : 1.5]$. The mean of the maximum values across different W_s and W_g values are shown in table 5.1 for each seizure. This table illustrates the well behavior of the error and consequently the good performance of the Pareto density in the fitting process. The highest value of $\max_{t_k} \mathcal{E}(t_k)$ was found for S_8 and is equal to 0.0736.

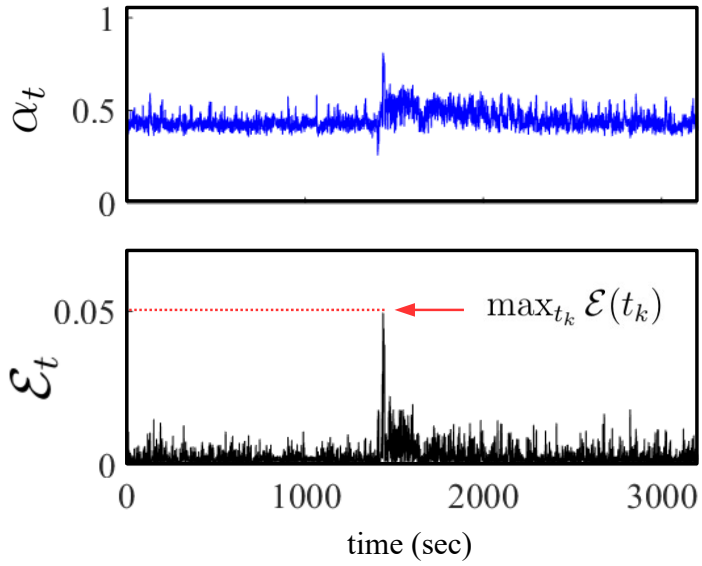

Figure 5.2: **The SSR series**. The α values and their respective SSR (represented by $\mathcal{E}_t$) across the whole α - series. The $\mathcal{E}_t$ is discrete time that gives series gives the squared L_2 error during the fitting process at the times t_k. For this example we indicate by the red arrow the maximum $\max_{t_k} \mathcal{E}(t_k) = 0.5$. Here we have chosen seizure S_7 and $(W_s, W_g) = (1.5, 0.25)$

Table 5.1: SSR Error: average $\max_{t_k} \mathcal{E}(t_k)$ for $W_s \in [1.5 : 0.5 : 4]$ and $W_g \in [0.25 : 0.25 : 1.5]$

Seizure	**average** $\max_{t_k} \mathcal{E}(t_k)$*****
1	0.0262
2	0.0279
3	0.055
4	0.0668
5	0.0245
6	0.0198
7	0.0443
8	0.0736
9	0.0417
10	0.0465
11	0.0319
12	0.0442
13	0.0561
14	0.0169
15	0.0318
16	0.0244
17	0.0259
18	0.0267
19	0.0191

5.3 α - series for all seizures

In this section we show a zoom of all α - drops with the respective seizures LFP trace (channel 1). In total, from the 19 seizures in our dataset, 15 of them have exhibited such behavior of the α - series. In all simulations we have chosen $(W_s, Wg) = (1.5, 0.25)$ to

compute the α - series. This pattern indicates a high degree of desynchronization at the seizure epoch, which is a surprising result. Moreover this time signature has been used for our seizure detection process. On the other hand, 4 seizures (S_1, S_2, S_{14} and S_{19}) haven't showed the α - drop. Fig 5.6 shows a zoom for each of them. The α - series have shown low values during all recording periods. This phenomenon is still unexplained in a reasonable way.

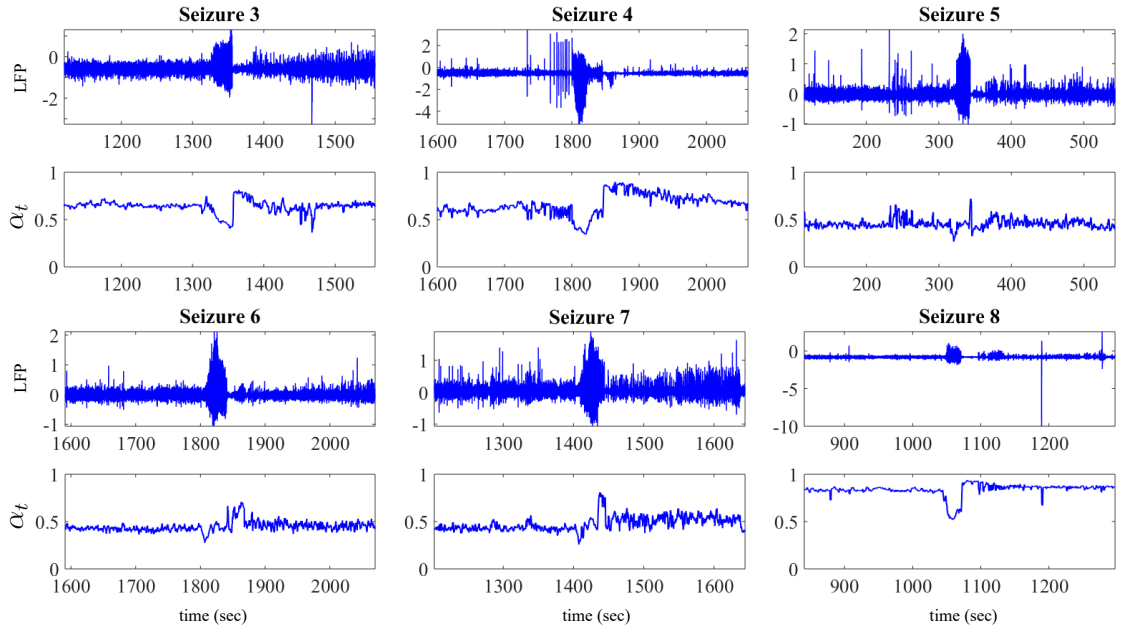

Figure 5.3: α **- series for** S_3**,** S_4**,** S_5**,** S_6**, and** S_8 Highlight of the drop of the α parameter. For these simulations we have assumed $(W_s, W_g) = (1.5, 0.25)$. LFP from Channel 1 in all pictures.

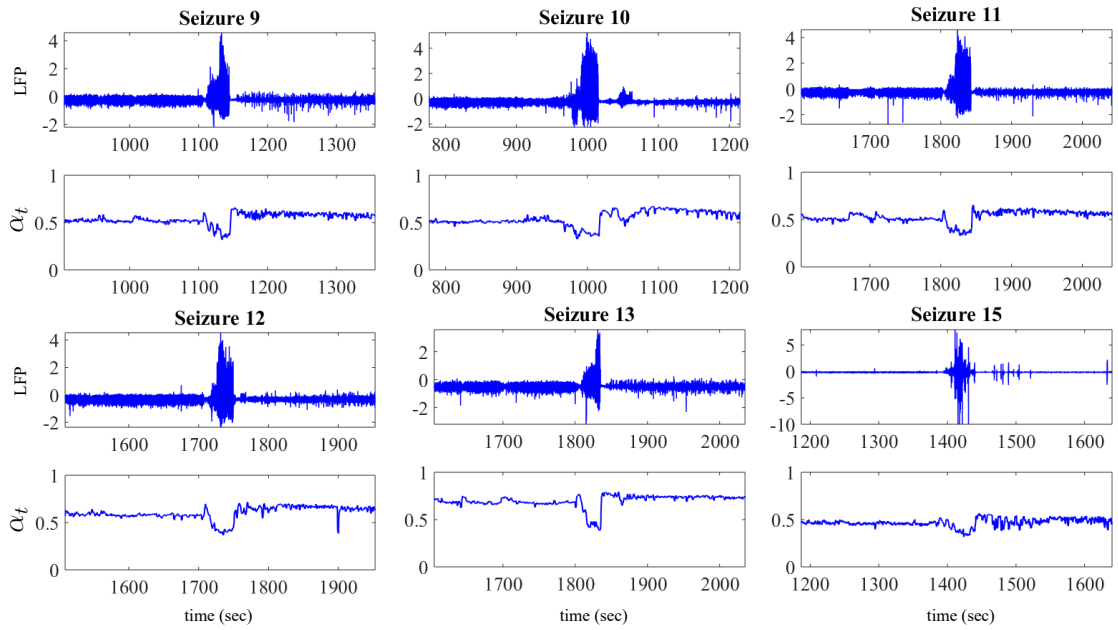

Figure 5.4: α **- series for** S_9**,** S_{10}**,** S_{11}**,** S_{12}**,** S_{13} **and** S_{15}**.** Highlight of the drop of the α parameter. For these simulations we have assumed $(W_s, W_g) = (1.5, 0.25)$. LFP from Channel 1 in all pictures.

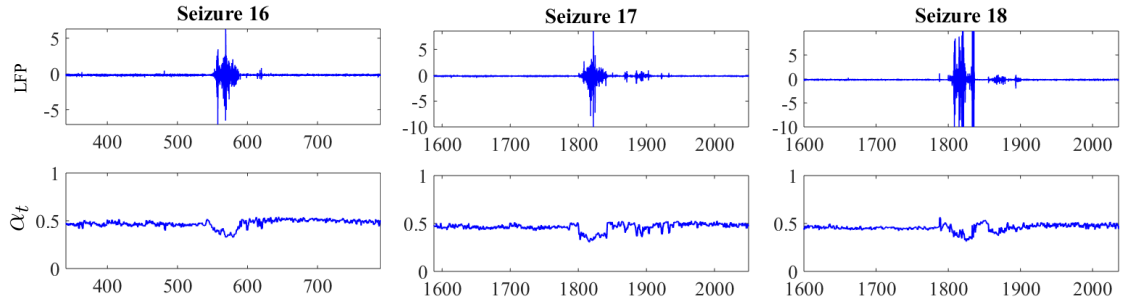

Figure 5.5: α **- series for** S_{16}**,** S_{17} **and** S_{18}**.** Highlight of the drop of the α parameter. For these simulations we have assumed $(W_s, W_g) = (1.5, 0.25)$. LFP from Channel 1 in all pictures.

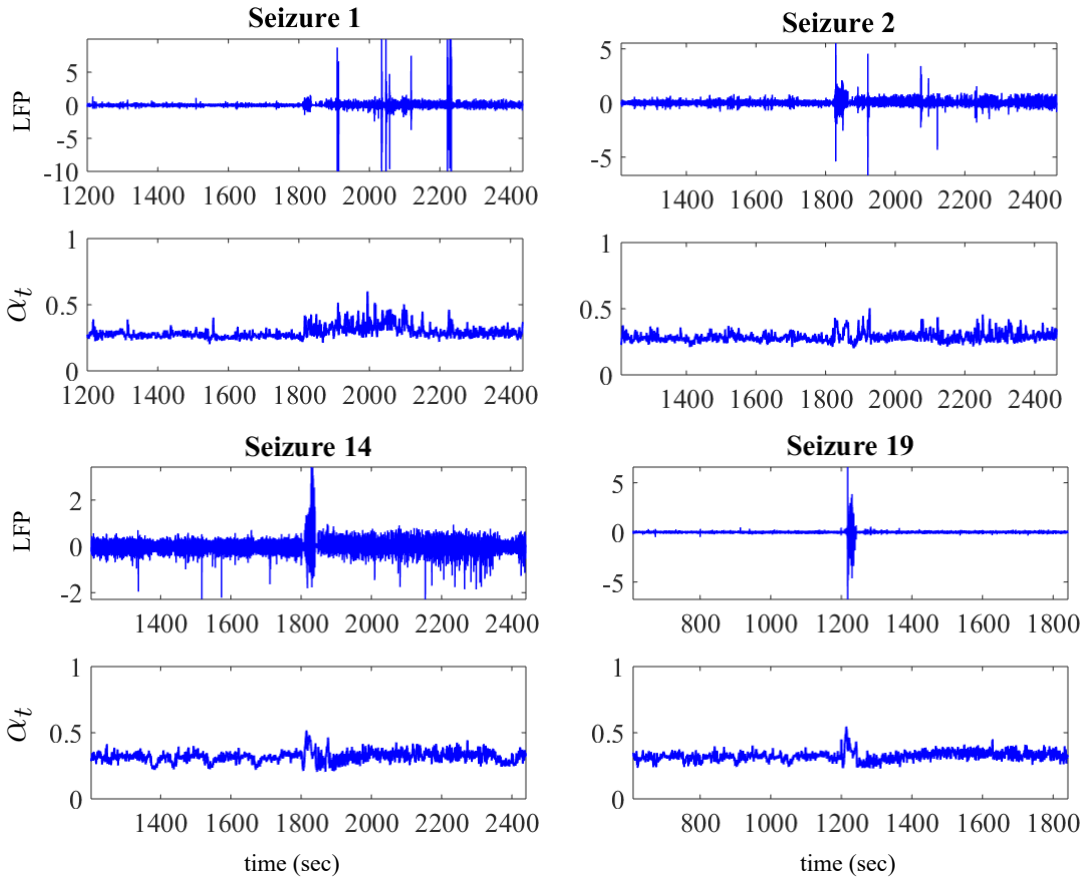

Figure 5.6: α **- series for** S_1**,** S_2**,** S_{14} **and** S_{19}**.** Seizures for which the α parameter does not drop. For these simulations we have assumed $(W_s, W_g) = (1.5, 0.25)$. LFP from Channel 1 in all pictures.

5.4 α - histograms for all seizures

Figs. 5.7, 5.8 and 5.9 shows the histograms of all α - drop seizures for$(W_s, W_g) = (1.5, 0.25)$. We have chosen interval sizes of $T = 1$ and 10 minutes, but there were no 10 minutes time window of LFP data for S_5, S_6, S_{11} and S_{16}. Therefore we have chosen tha maximum interval as possible: $T = 5.2$ for S_5, $T = 3.7$ for S_6, $T = 5.02$ for S_{11} and $T = 8.9$ for S_{16}. For most seizures the average of α values during the post seizure is considerably higher than the average during the pre seizure epoch. This indicates higher synchronization levels in the postictal depression and highlights the long term effects of

the seizures.

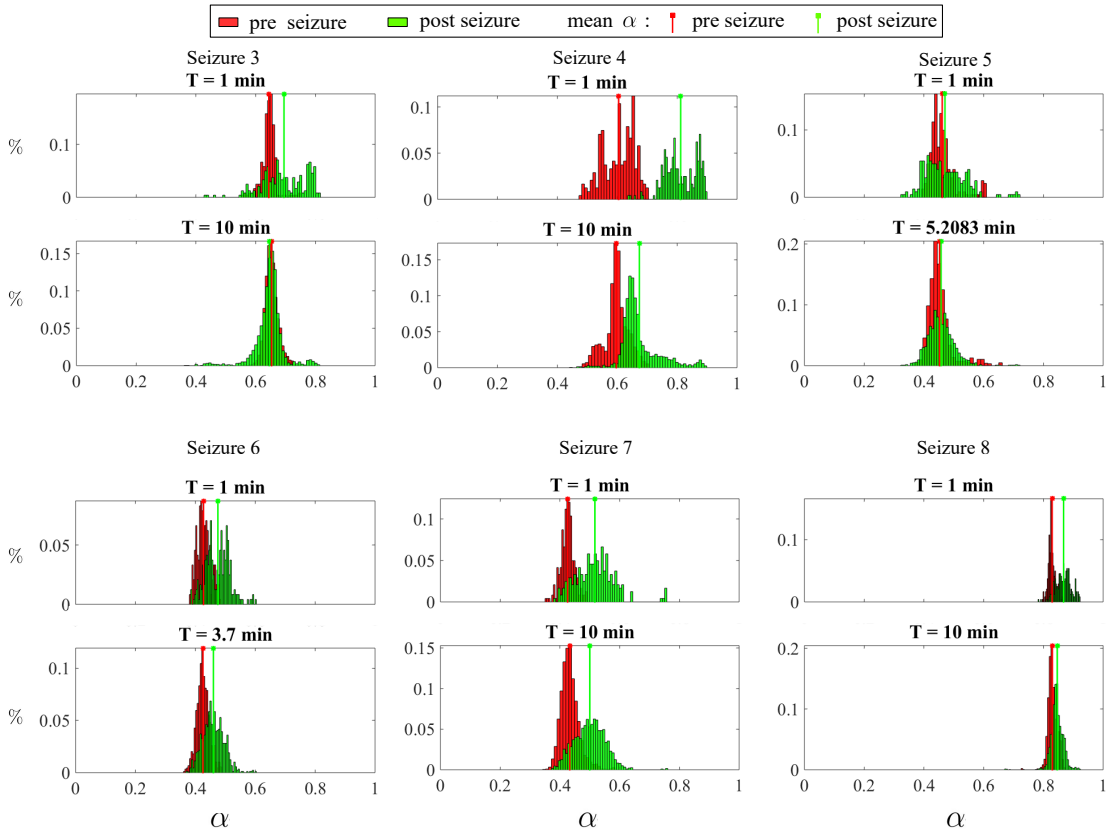

Figure 5.7: α **- histograms for** S_3**,** S_4**,** S_5**,** S_6**,** S_7 **and** S_8**.** The post seizure α - histograms are considerably shifted to the right if compared with the pre seizure ones. The vertical lines represent the mean for both α - histograms. They also show the difference of the α values of the pre and post seizure epochs. This results highlights both short and long term effects of the post seizure depression in the brain. For this example $(W_s, W_g) = (1.5, 0.25)$.

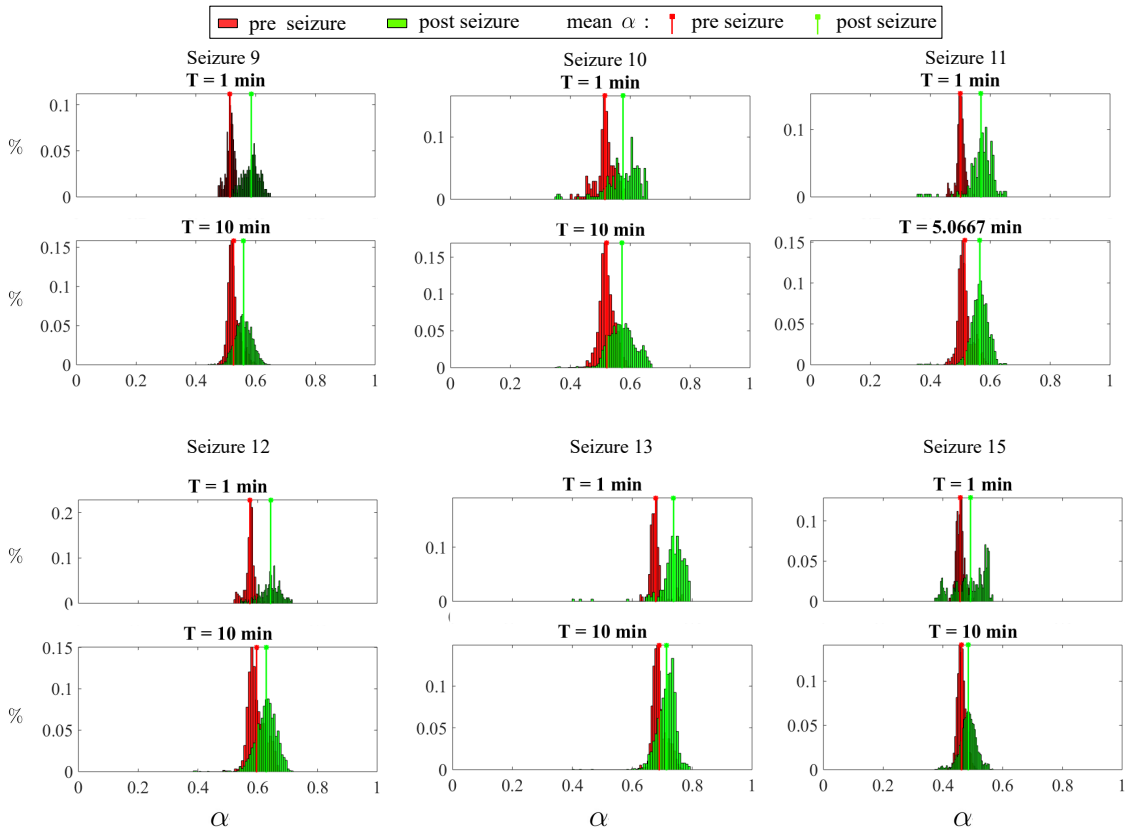

Figure 5.8: α **- histograms for** S_9**,** S_{10}**,** S_{11}**,** S_{12}**,** S_{13} **and** S_{15}**.** For this example $(W_s, W_g) = (1.5, 0.25)$

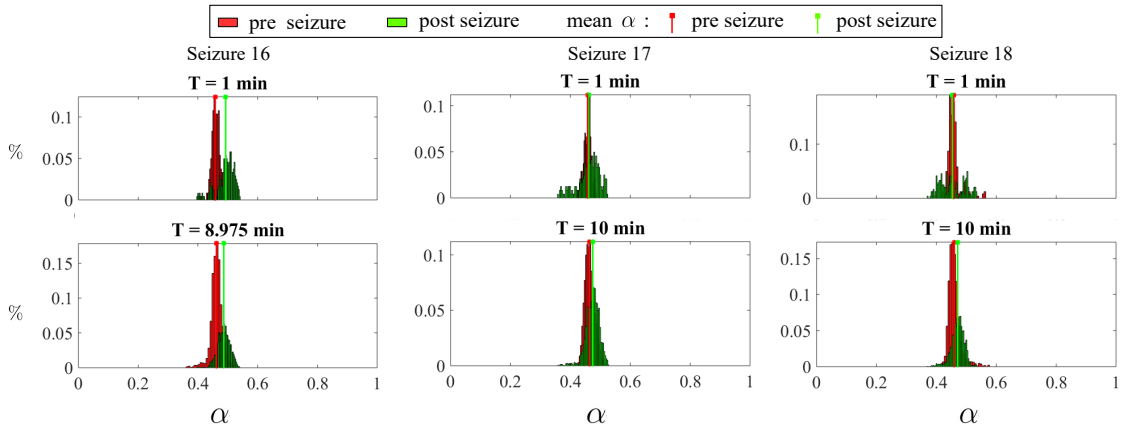

Figure 5.9: α **- histograms for** S_{16}**,** S_{17} **and** S_{18} For this example $(W_s, W_g) = (1.5, 0.25)$

Part III

Data-Driven analysis of Dengue Outbreaks in Brazil: a Critical Assessment of Climate Conditions for Different Capitals.

Chapter 6

Introduction

Dengue Fever is a tropical mosquito-borne viral disease present in more than 110 countries and a current threat to half of the world population [113, 114, 115, 116]. The DENV virus – and the more perilous Chikungunya and Zika virus – are primarily transmitted to humans through infected *Aedes Aegypti* mosquitoes, which were the subject of much debate during the 2016 Olympic Games in Rio de Janeiro. This main disease vector is well adapted to urban environments, which allow viruses to spread easily through cities. Still, regional climate conditions play a critical role in the development of epidemic outbreaks in major urban centers. In this work we analyze temperature and precipitation time series data for Brazilian state capitals and determine critical periods and seasons in which these climate variables might favor the mosquito development cycle and therefore the occurrence of Dengue outbreaks.

The first cases of Dengue in Brazil date from the end of the 19^{th} century, and despite the elimination of the *Aedes Aegypti* in 1955, the mosquito was reintroduced in the country in the 70s. A historically important outbreak occurred in 1981 in Boa Vista, in the state of Roraima, following several outbreaks in Central America involving the DENV-1 and DENV-4 serotypes [117, 118]. Since then, Dengue has become one of the major public health problems in Brazil, with several epidemics reported yearly across the country. While Dengue symptoms are usually limited to fever and muscle/joint pain, some develop more severe forms of the disease such as hemorrhagic fever or shock syndrome. The epidemics

were aggravated with the latest Zika and Chikungunya developments. In fact, 91,387 thousand cases of Zika and 39,017 thousand cases of Chikungunya were reported in 2016 from February to April alone [119], which caught the world's attention just in time for the Olympic Games in Rio de Janeiro. Until recently, Brazilian authorities limited their actions to vector control measures, but a first-generation vaccine may represent a turning point for stopping these epidemics [120, 121, 122].

The proliferation of *Aedes aegypti* and the sustained transmission of Dengue are influenced by a complex, interplay of multi-scale factors such as the circulation of different serotypes [123, 124], the commuting of infected and susceptible humans within a city [125, 126, 127], and the population size of the mosquitoes. There is also a growing body of evidence showing that local climate conditions such as temperature and precipitation may highly influence the development of the mosquitoes throughout different stages of their life cycle [128, 129, 131, 132]. Complicating our understanding is the fact that several regions experienced a nontrivial alternation between periods with and without epidemic outbreaks over the past years, suggesting that the specific critical climate conditions that propitiate the transmission of the disease is heterogeneous and still poorly understood [133, 134, 137, 135, 136].

In this work we analyze climate and epidemiological data from seven major Brazilian cities that in the recent past had years with and without Dengue outbreaks in order to identify critical climate signatures that may have contributed to the epidemic outcomes. Fig. 6.1 is a schematic overview of the work presented. We estimate the influence of climate conditions in different epochs preceding epidemic periods using two data-driven methodologies; the first one is based of the singular value decomposition and exploits the low dimensional structures present in the climate time series [138, 139], and the second one is based of machine learning algorithms for clustering and classification [14, 15] such as Support Vector Machines (SVM) applied to climate variables that are key to the life cycle of the mosquito [140, 141]. A crucial step in our methodology includes the usage of compressed sensing to recover missing data [139, 142, 143, 144, 145] – in a plausible and in a $\mathcal{L}^1$-optimal way – from climate recordings by the National Institute of Meteorology (INMET) [146]. This allow us to explore the link between climate and Dengue in the following

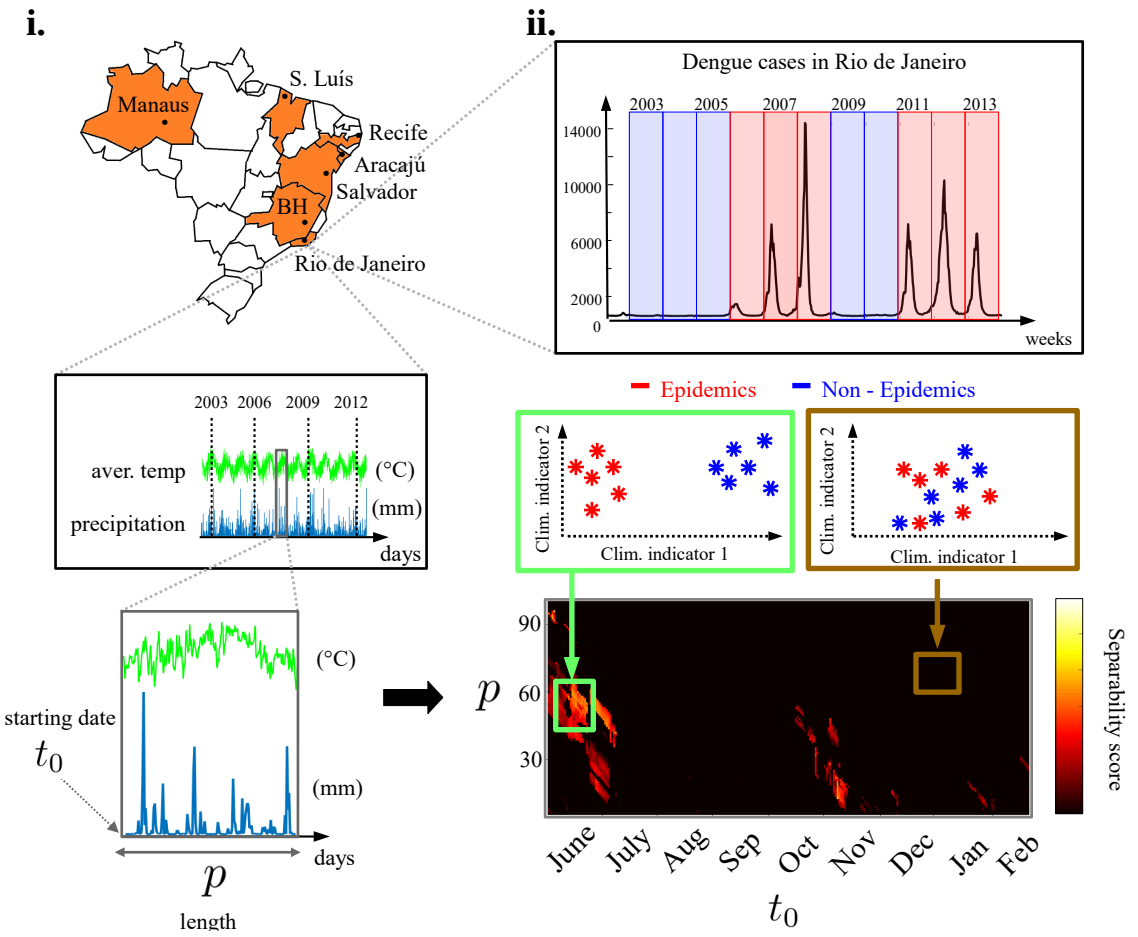

Figure 6.1: **Schematic Overview.** We analyze time series data for climate variables from seven Brazilian state capitals (Aracajú, Belo Horizonte, Manaus, Recife, Rio de Janeiro, Salvador and São Luís) and their connection to Dengue outbreaks. **(i)** Illustrative example showing data from Rio de Janeiro. Two parameters define the epochs in which climate conditions are considered: the starting date t_0 (month/day) and period length p (days). **(ii)** By applying machine-learning algorithms to historical data we locate periods along the year where the separability between epidemic and non-epidemic climate is higher. Keeping track of signature differences at key epochs, that vary from capital to capital, may significantly improve Dengue outbreak forecasting in the upcoming years.

major Brazilian cities: Aracajú, Belo Horizonte, Manaus, Recife, Rio de Janeiro, Salvador and São Luís. For each city, we highlight epochs that are critical for both methodologies.

Surprisingly, there is a strong correlation between Dengue epidemics and favorable climate conditions during winter and spring. This long-term influence is important evidence that the interplay between climate, mosquito populations and Dengue outbreaks are extremely complex. The insights of this work may help tailor public health policies for each different city by increasing vector control measures during neglected critical epochs and ultimately improving the forecasting of Dengue outbreaks – which would allow the public health system to make earlier logistic preparations to better accommodate a large number of patients, or alternatively, mosquito eradication programs can be enacted during the winter and spring months that are known to be associated with epidemic outbreaks.

This part is outlined as follows. In chapter 7, we describe both epidemiological and climate datasets, our techniques for data completion and other details of our analysis. In chapter 8 we present our findings for all seven Brazilian state capitals, emphasizing epochs that are critical for all methods. In chapter 9 we summarize the most important seasons for dengue epidemics in each city, highlighting the long-term impact of climate. We also discuss the limitations of this work, its potential impact for improving early warning systems, and the usage of our methods as a modest outbreak prediction tool.

Chapter 7

Methods

7.1 Description of epidemiological and climate datasets

All epidemiological data utilized in this work were taken from the publicly available datasets of the Brazilian Notifiable Diseases Information System (SINAN, [147]). This includes the total number of Dengue cases per year (from 2002 to 2012) for all Brazilian state capitals. We also include data made available for Rio de Janeiro by the city's hall health department for 2013 [148]. A year is conventionally classified as an *epidemic* year for a given city if the incidence of Dengue is above 100 cases (per 100,000 inhabitants) and classified as a *non-epidemic* year otherwise [149]. In order to find critical climate signatures that may have contributed to the epidemic outcomes, we restrict ourselves to seven state capitals that displayed at least 3 epidemic years and 3 non-epidemic years in the recent past. This allowed us to investigate the correlation between distinct climate conditions and the complicated alternations between years with and without epidemic outbreaks over time. The climate data utilized in this work was obtained from the National Institute of Meteorology (INMET) and included time series for the average temperature and precipitation for the state capitals Aracajú, Belo Horizonte, Manaus, Recife, Salvador, and São Luís (from 1/1/2001 to 12/31/2012) and for Rio de Janeiro (from 1/1/2002 to 12/31/2013).

7.2 Completing missing climate data via compressive sensing

The time series of our selected climate dataset contain episodical gaps on days where variables (temperature and precipitation) were not recorded. To fill in the missing data gaps, we employ two different methods: Compressive sensing [139, 142, 143, 144, 145] and interpolation (see Fig. 7.1 for illustrative examples). For temperature time series data with 2 or more consecutive missing recordings, we use a recently developed compressive sensing method based upon $\mathcal{L}^1$-convex optimization for approximating the missing data [139, 142, 143, 144, 145]. The compressive sensing method attempts to reconstruct a signal from a sparse, sub-sampling of the time series data. In this case, the sparse sub-sampling occurs from the fact that we have missing data. The signal reconstruction problem is nothing more than a large underdetermined system of linear equations. To be more precise, consider the conversion of a time series data to the frequency domain via the discrete cosine transform (DCT)

$$\psi \mathbf{c} = \mathbf{f} \tag{7.1}$$

where $\mathbf{f}$ is the signal vector in the time domain and $\mathbf{c}$ are the cosine transform coefficients representing the signal in the DCT domain. The matrix ψ represents the DCT transform itself. The key observation is that most of the coefficients of the vector $\mathbf{c}$ are zero, i.e. the time series is sparse in the Fourier domain. Note that the matrix ψ is of size $n \times n$ while $\mathbf{f}$ and $\mathbf{c}$ are $n \times 1$ vectors. The choice of basis functions is critical in carrying out the compressed sensing protocol. In particular, the signal must be sparse in the chosen basis. For the example here of a cosine basis, the signal is clearly sparse, allowing us to accurately reconstruct the signal using sparse sampling. The idea is to now sample the signal randomly (and sparsely) so that

$$\mathbf{b} = \phi \mathbf{f} \tag{7.2}$$

where $\mathbf{b}$ is a few (m) random samples of the original signal $\mathbf{f}$ (ideally $m \ll n$). Thus ϕ is a subset of randomly permuted rows of the identity operator. More complicated sampling

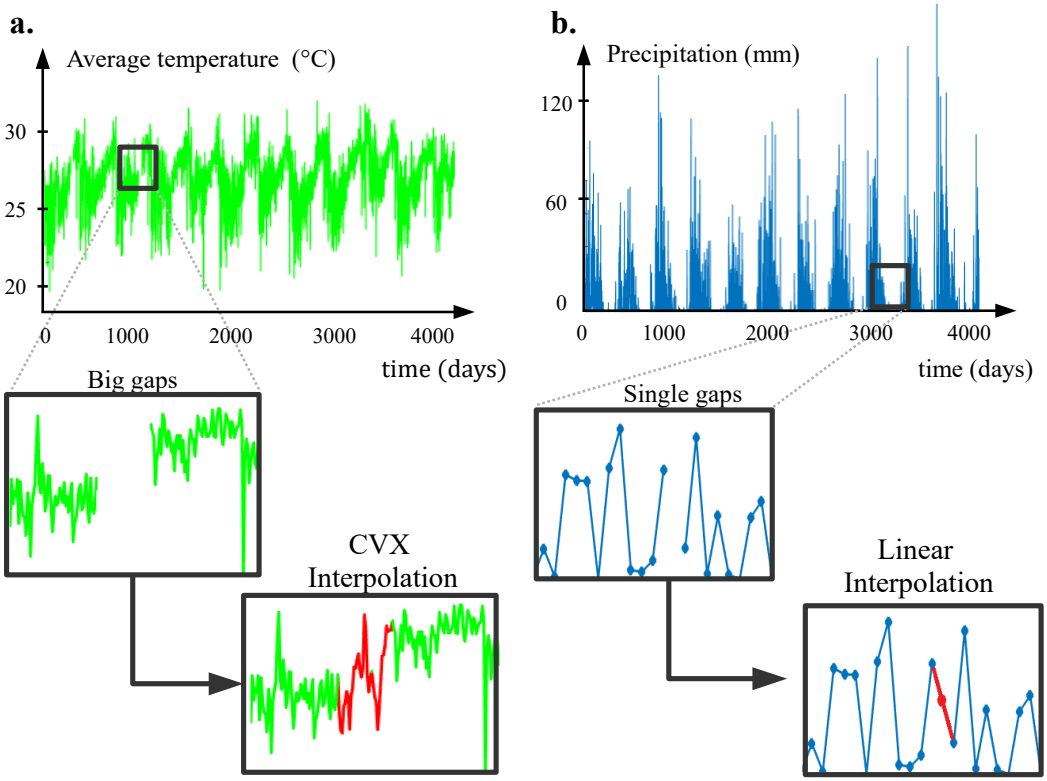

Figure 7.1: **Completing missing data.** The daily measurements of climate variables for Brazilian state capitals from the National Institute of Meteorology (INMET) **a.** We reconstruct larger portions of lacking data with compressed sensing ($\mathcal{L}^1$-convex optimization routines). **b.** Data values at minor holes were estimated by simpler interpolation protocols. The state capitals with intractable missing portions of data were not considered (see appendix) for more details.

can be performed, but this is a simple example that will illustrate all the key features. Note here that $\mathbf{b}$ is an $m \times 1$ vector while the matrix ϕ is of size $m \times n$.

Approximate signal reconstruction can then be performed by solving the linear system

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{7.3}$$

where $\mathbf{b}$ is an $m \times 1$ vector, $\mathbf{x}$ is $n \times 1$ vector and

$$\mathbf{A} = \phi\psi \tag{7.4}$$

is a matrix of size $m \times n$. Here the $\mathbf{x}$ is the sparse approximation to the full DCT coefficient vector. Thus for $m \ll n$, the resulting linear algebra problem is highly underdetermined. The idea is then to solve the underdetermined system using an appropriate norm constraint that best reconstructs the original signal, i.e. the sparsity promoting $\mathcal{L}^1$ is highly appropriate. The signal reconstruction is performed by using

$$\mathbf{f} \approx \psi\mathbf{x}\,. \tag{7.5}$$

If the original signal had exactly m non-zero coefficients, the reconstruction could be made exact (See Ref. [139], Ch. 18).

We applied this technique specifically to the climate series of Rio de Janeiro, Salvador and São Luís. For the other state capitals, we just linearly interpolate the time series whenever a single daily recording is missing. We note that there were intractable large gaps for the INMET precipitation series for Rio de Janeiro, which forced us to use alternative data sources made available by the city's alert system of rain events [150]. See the SI tables for details.

7.3 Defining periods of critical climate conditions for Dengue

In what follows, we investigate the influence of climate conditions on Dengue outbreaks at different periods along the yearly cycle. We let (t_0, p) denote a sampling period of p days starting at the date t_0. Then, for a fixed period, we evaluate a score quantifying the discrepancy between climate conditions in epidemic years and non-epidemic years. See Fig. 6.1 for an illustrative example using data from the city of Rio de Janeiro; we postulate that periods with high climate *separability* between epidemic years (in red) and non-epidemic years (in blue) might be of critical importance to the cycle of the urban mosquito population and consequently, to the occurrence of Dengue outbreaks in the following year. We calculate the separability score of a period using two different methodologies: The first is

based of the Singular Value Decomposition (SVD) [139] and a low dimensional representation of the climate data, while the second is based of a machine learning algorithm know as support vector machine (SVM) [14, 15]. In both cases, the methods highlight potentially critical periods for the occurrence of Dengue. Finally, since Dengue outbreaks in Brazil typically take place between March–May in a given year, we limit the range of (t_0, p) from June, (of the previous year) to May. In Fig. 6.1, we note that there are critical periods in the winter (green box with t_0 in June) that may be critical for the occurrence of Dengue.

7.4 Separability scores from SVD methodology

Fig. 7.2 shows how we select climate data over the same period (t_0, p) for different years and build a corresponding matrix $X(t_0, p)$ that allows for a SVD analysis: (i) We select data from k climate variables over the years, always starting at t_0 and ending p days later. (ii) We stack and normalize the data associated with year l in a block matrix $\mathcal{B}_l(t_0, p)$, for $l = 1, 2, \cdots, N$. (iii) Finally, all blocks are reshaped into column vectors, forming a new matrix $\mathbf{X} = \mathbf{X}(t_0, p)$, which yields

$$\mathbf{X}(t_0, p) = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T(t_0, p). \tag{7.6}$$

The columns of $\mathbf{U}$ – the SVD modes – form an orthogonal basis for the space generated by the columns of X and the projections of the principal components are given by the $\mathbf{\Sigma}\mathbf{V}^T(t_0, p)$ matrix (see Fig. 7.3 **i**).

In our analysis, we project climate data collected over (t_0, p) each year onto a 2-mode plane and label years as epidemic (red) or non-epidemic (blue) according to our outbreak convention (see Fig. 7.3 **ii**). This yields a set of l points (one for each year) and allow us to quantify how separate the blue/red dots are from each other: We consider two convex hulls connecting red/blue vertices and evaluate the distance $\mathcal{H}(t_0, p)$ between them (see Fig. 7.3 **iii a,b**). Finally, we explore a large range of values for t_0 and p to find periods along the yearly cycle in which discrepancies between climate conditions might have contributed to Dengue outbreaks in the following year.

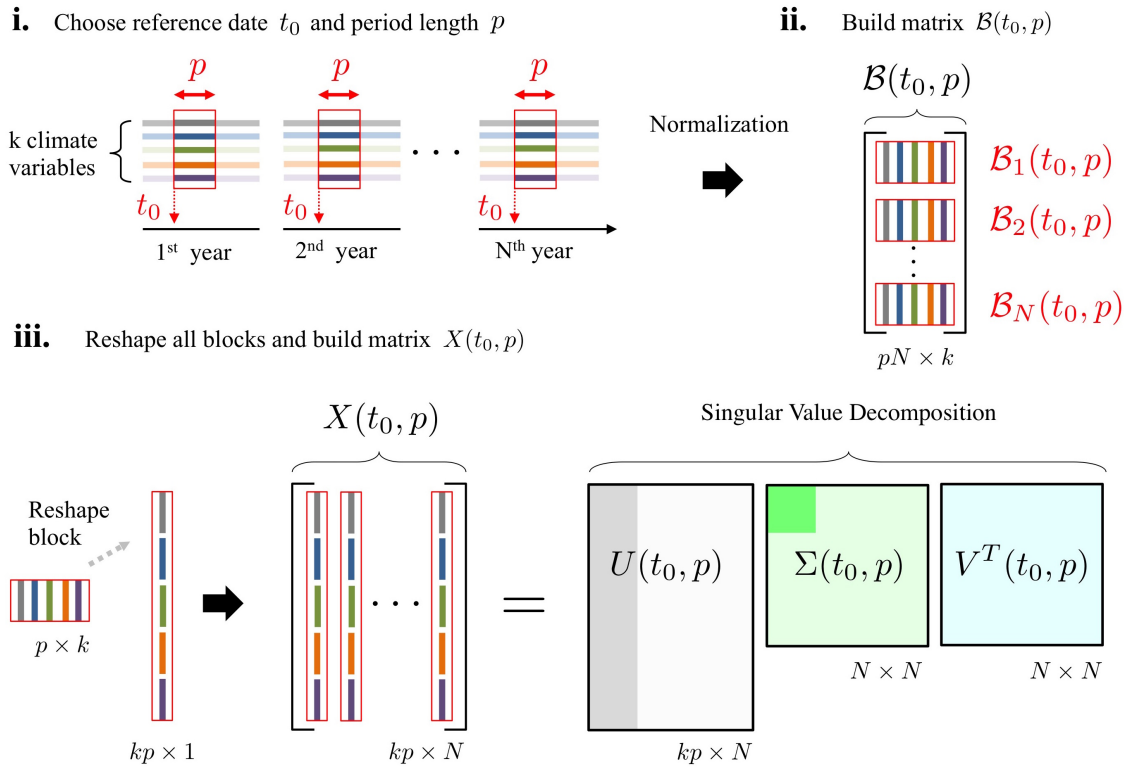

Figure 7.2: **Outline of SVD methodology: Data matrix setup. (i)** We select climate data with the same starting date t_0 and length p across the years $(1, 2 \ldots, N)$. **(ii)** After normalizing each climate variable over the years, we store them in block matrices $\mathcal{B}_j(t_0, p)$, which in turn, are stacked in a matrix $\mathcal{B}(t_0, p)$. **(iii)** Reshape $\mathcal{B}$ into $\mathbf{X}$, where different columns correspond to climate information collected at (t_0, p) in different years. The SVD of $\mathbf{X}$ provides a low-dimensional representation of the internal structure of the data from its most informative (correlated) viewpoint. Our goal is to, based in the historic data, determine specific epochs of the year in which the separability between epidemic and non-epidemic climate is higher.

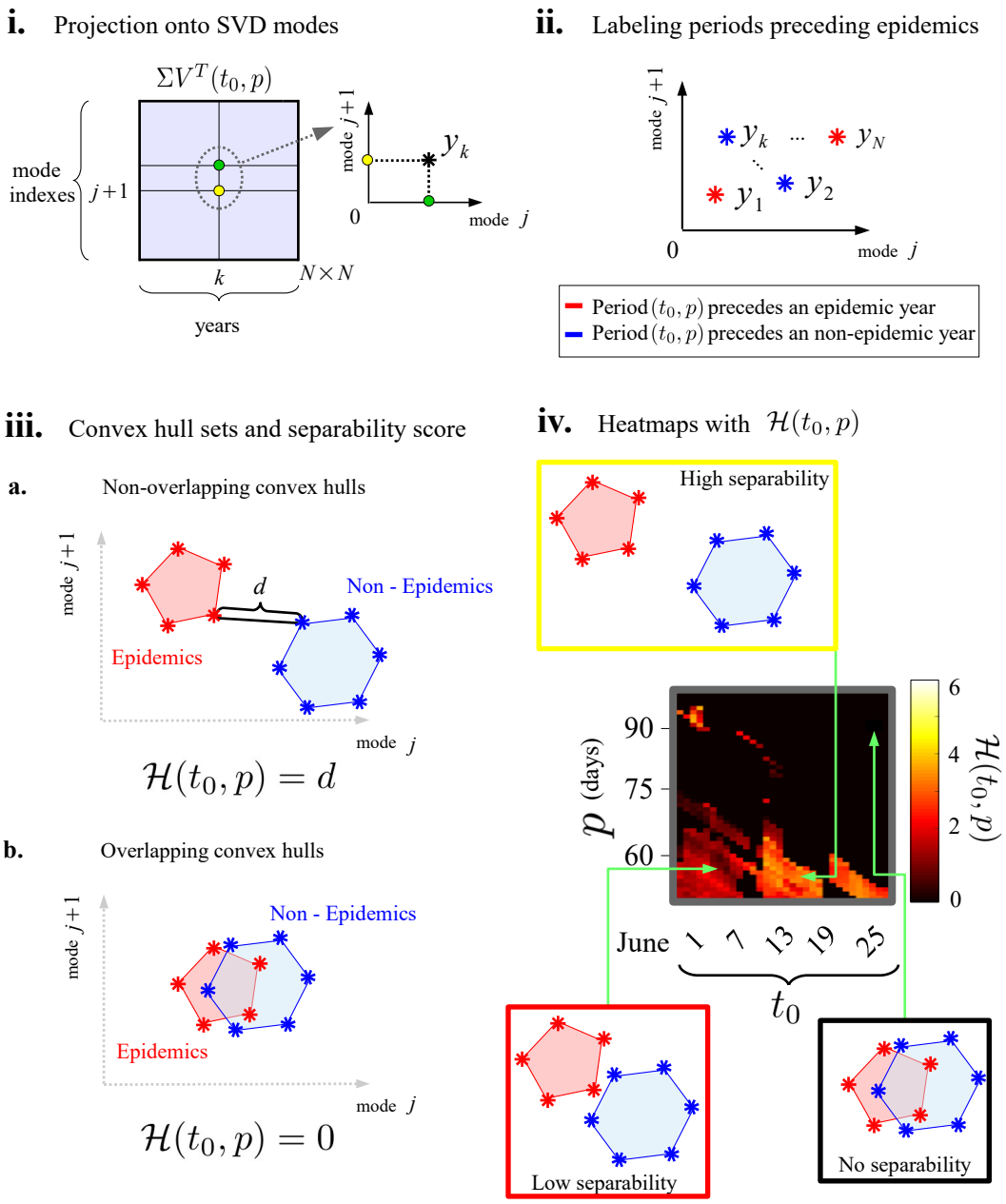

Figure 7.3: **Outline of SVD methodology: Convex Hull analysis.** **(i)** The projection's component of the k-th column of X onto the j-th mode is the (j,k)-element of the matrix ΣV^T. We plot the projection for each year l ($l = 1, 2, \cdots, N$) in the plane spanned by modes j and $j+1$. **(ii)** For each year we color the projections according to epidemic or non-epidemic year criteria. We choose red If the (t_0, p) interval preceded a DF outbreak and blue if it doesn't. **(iii)** We compute the convex hulls for the epidemic and non-epidemic projections set. **a.** If there is no overlapping between the hulls, we calculate the minimum distance between two vertices and set $\mathcal{H}(t_0,p) = d$. **b.** $\mathcal{H}(t_0,p) = 0$ in the case of overlapping hulls. **(iv)** The SVD separability score $\mathcal{H}$ can be obtained for a range of (t_0, p) intervals.

7.5 Separability scores from SVM methodology

Our second separability score for measuring discrepancies between climate conditions in epidemic/non-epidemic years is based on a supervised learning technique for classification. Fig. 7.4 outlines the main steps of our Support Vector Machines (SVM) algorithm: (i) For a fixed (t_0, p) interval, we evaluate two climate indicators – the arithmetic mean of the average temperature $\langle T_j \rangle$ and average frequency of rain events $\langle \delta_j \rangle^{-1}$, where δ_j represents time intervals between consecutive peaks on precipitation data (see Fig. 7.4 **i**). (ii) We label the climate indicators in a 2D plot as an epidemic year (red) or as a non-epidemic year (blue) according to our Dengue outbreak criteria. (iii) We repeat the process for t_0 and p within a rectangular range R in the parameter space. Then, instead of a single point representing year l, we have a collection of red/blue points (dashed ellipses in Fig. 7.4 **iia**). In our simulations, the rectangular range R was 5×6, i.e, spanning 5 consecutive starting dates and 6 consecutive duration lengths. We tried both a linear kernel and a Radial Basis Function (RBF) kernel for the SVM training step on R and cross-validated the climate indicators by sampling 80% of each dataset and testing the accuracy of the predictions in the remaining 20%. Our separability score is ultimately defined as the average classification accuracy after re-sampling and testing data for 100 trials.

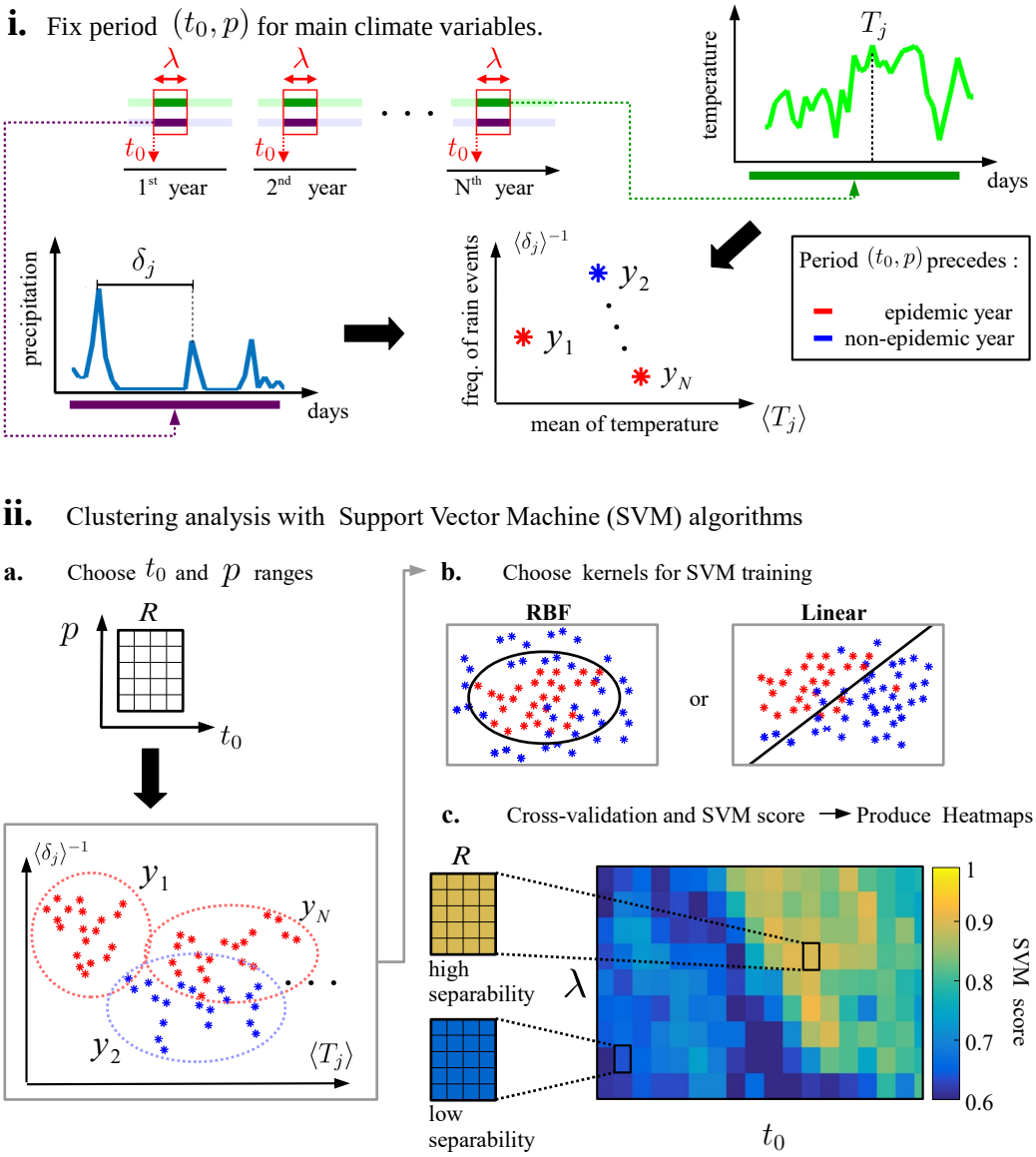

Figure 7.4: **Outline of SVM methodology.** A supervised learning technique for classification: **(i)** We calculate and plot mean of average temperature $\langle T_j \rangle$ and frequency of rain events $\langle \delta_j \rangle^{-1}$ for a fixed (t_0, p) interval of all years, using red and blue colors for periods preceding epidemic and non-epidemic years respectively. **(ii)a.** For each (t_0, p) interval of the rectangle R , we apply (i) to obtain a *cloud* (dashed circles) of points in the plane, for each year. **b.** Linear and RBF kernels are used to execute the SVM train/test and cross validation routines. **c.** The SVM score for R is obtained. We plot $t_0 \times p$ Heatmaps with Regions of High and Low separability scores, which indicates where temperature and precipitation are better correlated with Dengue fever outbreaks.

Chapter 8

Results

8.1 Survey of critical climate conditions for different cities

In this section, we highlight significant differences between climate conditions during epidemic/non-epidemic years for a period starting at day t_0 and duration of p days along the yearly cycle. We postulate that periods with high separability scores might be of critical importance to the cycle of the urban mosquito population and consequently, to the occurrence of Dengue outbreaks in the following year. The values of t_0 range from June 1^{st} to February 21^{st} and the values of p range from 10–100 days, which completely covers plausible periods that may influence Dengue outbreaks. The interpretation of the colormaps presented bellow should be straightforward and we highlight (in green) periods/epochs with high separability for both SVD and SVM methodologies. We restrict our SVD analysis to the five principal modes and fix the color bar for $\mathcal{H}(t_0, p)$ between 0 and 6 (highest value found for all simulations). For the SVM colormaps, we focus our analysis on the highest separability scores and choosing scores above 0.8 for the Linear Kernel. For the RBF kernel, which usually has a better predictive performance, the highlight threshold is 0.95.

Figs. 7.3 and 7.4 demonstrate how assessments and scoring are performed for both the SVD and SVM based methods. In what follows, a detailed evaluation is made for each capital city. Before proceeding to this analysis, however, it is highly informative to

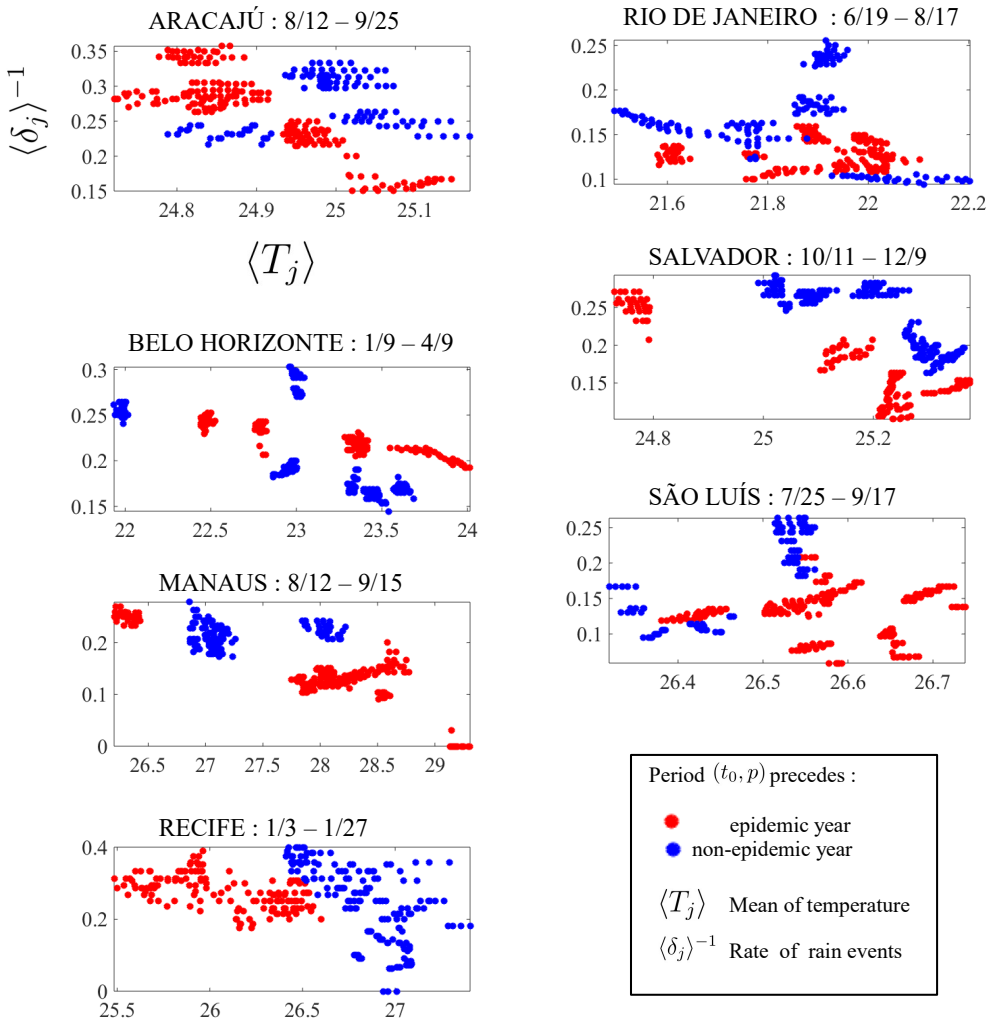

Figure 8.1: **Examples of high separability plots.** For each state Capital we have selected special time windows in which there is a clear separation between climate signatures preceding epidemic and non-epidemic years. Note the distinct separation of the data for each individual city, suggesting that a universal model for climate effects across all cities may be unattainable. The separability of data further suggests that epidemics may be accurately predicted in a given capital six to nine months in advance of the their outbreak. This separability notion is made quantitatively precise by the SVD and SVM separability scores (see text for details).

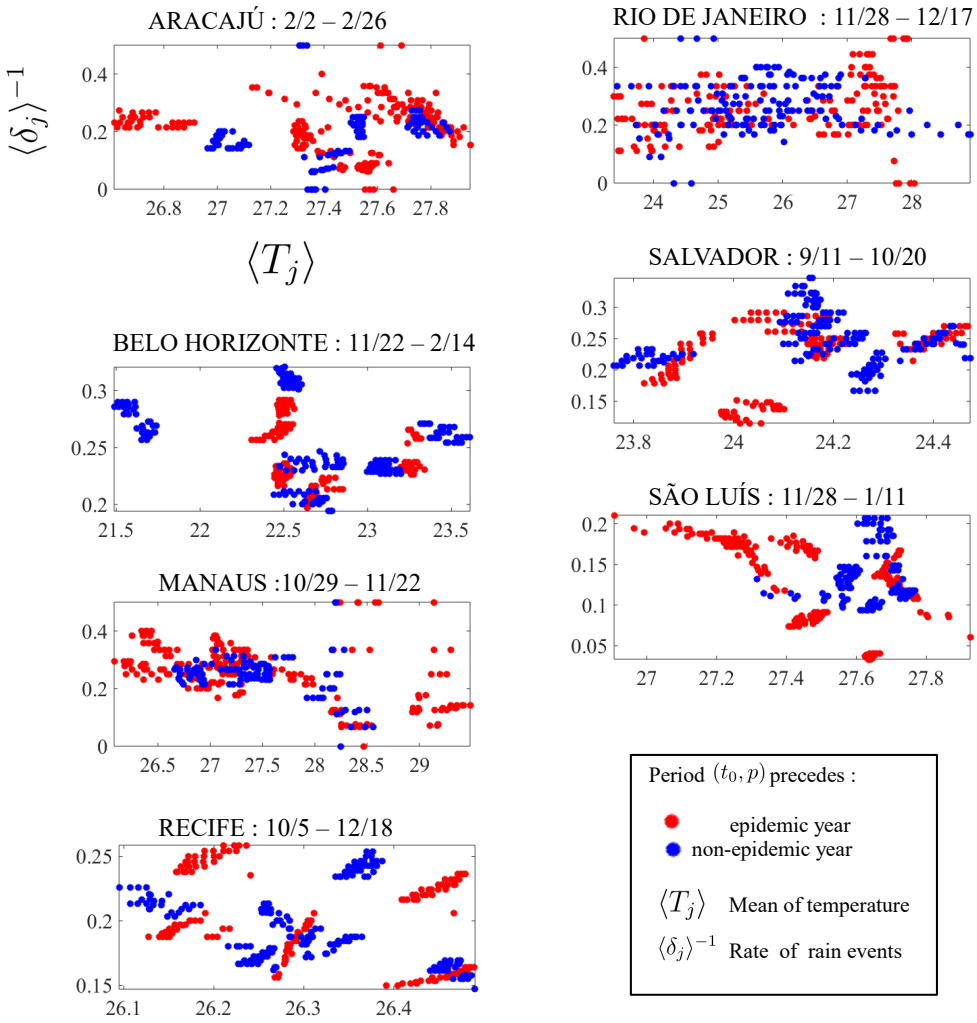

Figure 8.2: **Examples of low separability plots.** Specific time windows in which the epidemic and non-epidemic climate variables seems to be poorly distinguishable, therefore not suitable for Dengue prediction. Unlike Fig. 8.1, the mixing of data suggests poor predictability across all cities. This separability notion is made quantitatively precise by the SVD and SVM separability scores (see text for details).

interpret that a high score or low score achieves for separating epidemic and non-epidemic correlations. Figs. 8.1 and 8.2 demonstrate the clustering of data, or lack thereof, for all cities. In Fig. 8.1, representative data for windows achieving a high correlation score is shown. Remarkably, the red (epidemic) and blue (non-epidemic) dots are well separated

and distinguishable from visual inspection. Indeed, one could easily postulate decision regions which properly identify, months in advance, the oncoming presence of a Dengue epidemic by simply considering the mean temperature and precipitation frequency. Fig. 8.2 shows the data structure when a low correlation score is achieved. Note that in this case, there is significant overlap between the red and blue dots, suggesting that this region for prediction of an epidemic is highly suspect. Figs. 8.1 and 8.2 provide an easily interpretable understanding of the predictive nature of our proposed analysis. It also highlights important and significant differences between the various Brazilian cities. Some cities are on the coast, while others are in the interior, but regardless, each city has a unique pattern of clustering that can be capitalized on in order to provide predictive metrics for epidemic outbreaks. In the figures that follow, a principled analysis is performed for each Brazilian city in order to compute regions that give high scores on the SVD/SVM metrics and provide strong predictive metrics.

8.1.1 Rio de Janeiro

Fig. 8.3 shows periods with high separability scores for the city of Rio de Janeiro. Notice that both SVD and SVM methodologies highlight critical epochs during the winter. In fact, there is a good accordance between the projection of climate data to 3^{rd} and 4^{th} SVD modes and the linear SVM kernel for t_0 in June and p around 60 days. This suggests that time series for temperature and precipitation from June to August may be crucial for the occurrence of Dengue outbreak the following year. There is also good accordance between both criteria during the spring, for t_0 starting in October-November and p around 15 days.

8.1.2 São Luís

Fig. 8.4**a.** shows critical periods for São Luís, the state capital of Maranhão. Overall, we found good accordance between separability scores provided by the SVM and SVD methods. The SVD method indicated (for modes 3,4 and 4,5) critical (t_0, p) intervals for t_0 in July and p varying from 30 to 85 days. A similar period was found with the SVM method (using a linear kernel). This match suggests that temperature and rain in late winter and beginning of spring may play an important role in the occurrence of Dengue

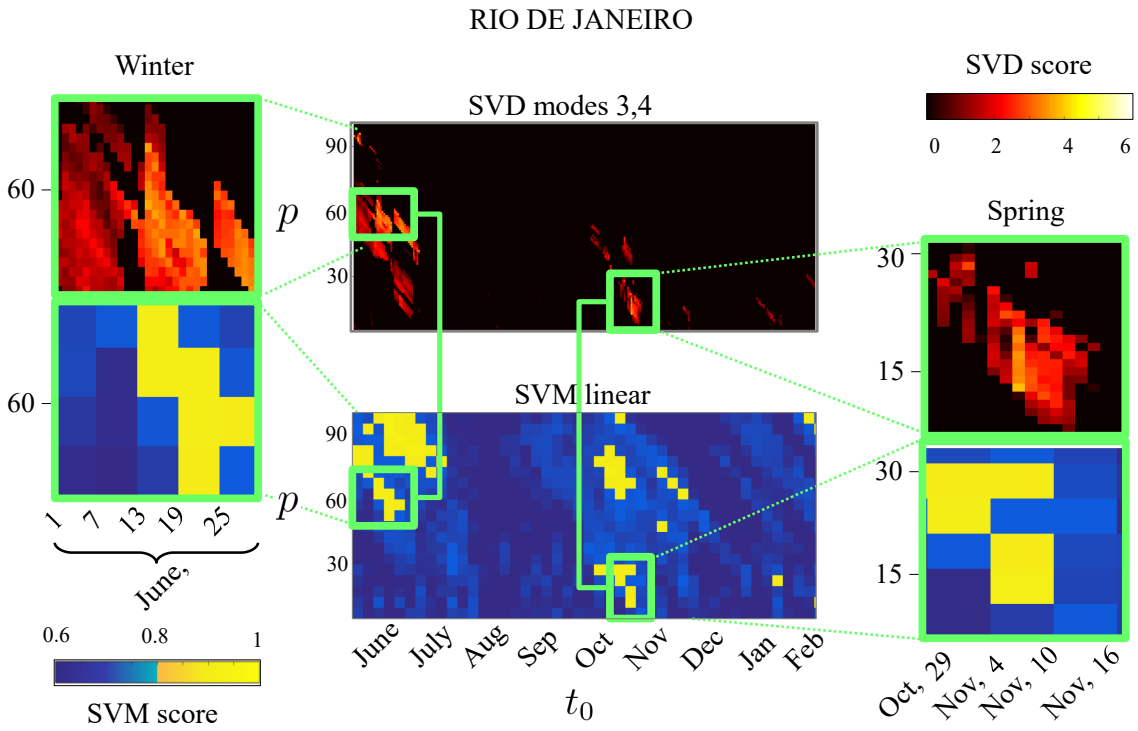

Figure 8.3: **Critical periods for Rio de Janeiro.** There is a good match between the different data-driven methods suggesting that specific climate conditions during winter season may be crucial to Dengue epidemics. Both methods also indicate a critical period of approximately 15 days during spring.

outbreaks. Another critical period indicated by both methods has t_0 in December and duration p around 60 days.

8.1.3 Manaus

The capital of Amazonas has a set of periods with high separability scores in the winter (see Fig. 8.4**b**). In the SVD colormaps (for modes 3,4 and 4,5), the separability score is high for t_0 between June and July and p between 60 and 90 days. This corresponds to the months of June, July and August. This is in good accordance with the scores given by the SVM methodology (using a linear kernel). We also highlight that there is a good match between the methods during a critical period with t_0 lying between July and August and

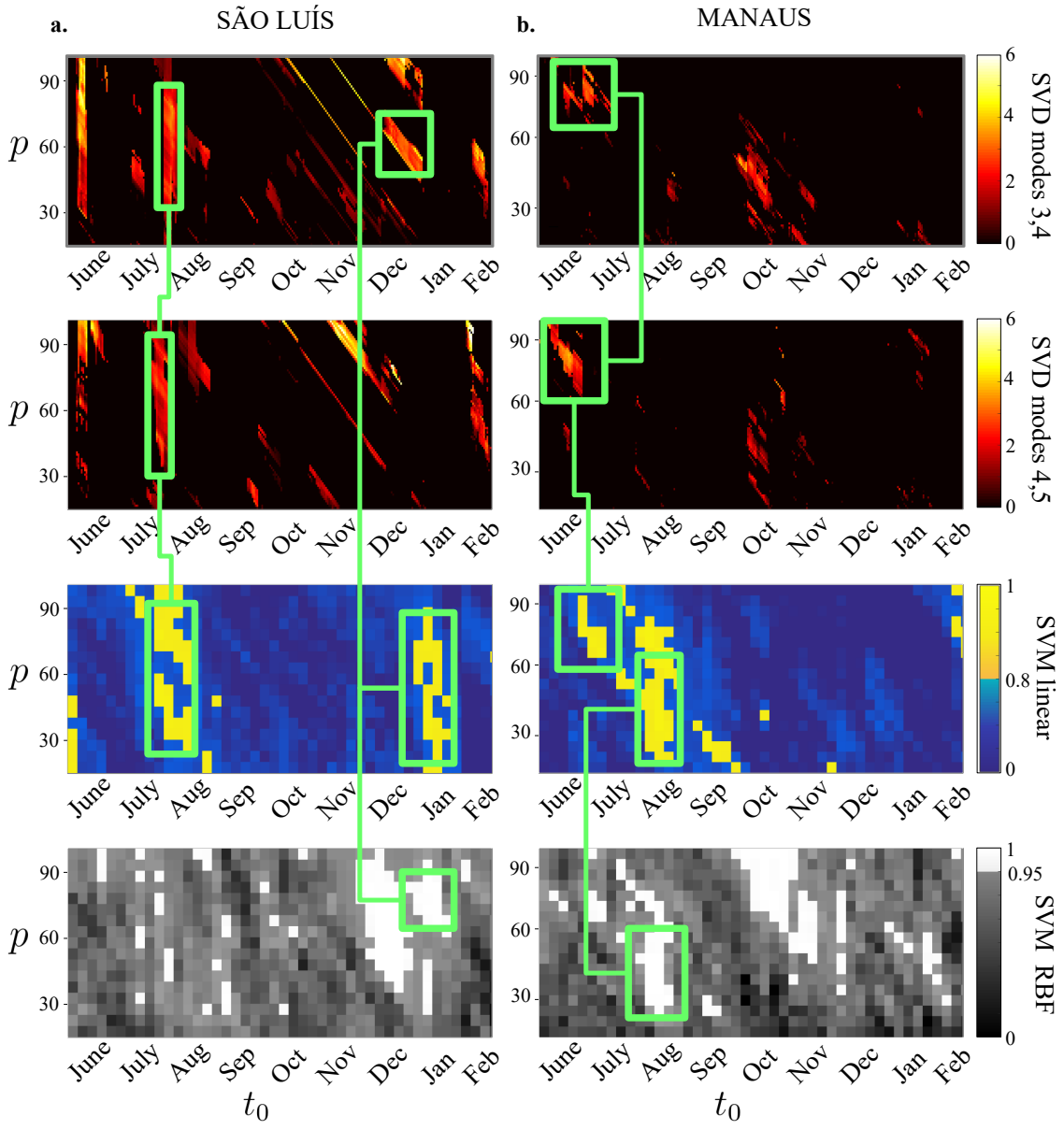

Figure 8.4: **Critical periods for São Luís and Manaus.** The two state capitals in the north of Brazil exhibit good accordance between SVD separability scores (for modes 3,4 and 4,5) and SVM separability scores (for both Linear and RBF kernels). **a.** Temperature and precipitation are correlated with Dengue outbreaks during winter and summer in the case of São Luís. **b.** For Manaus, the correlation is higher during winter and spring.

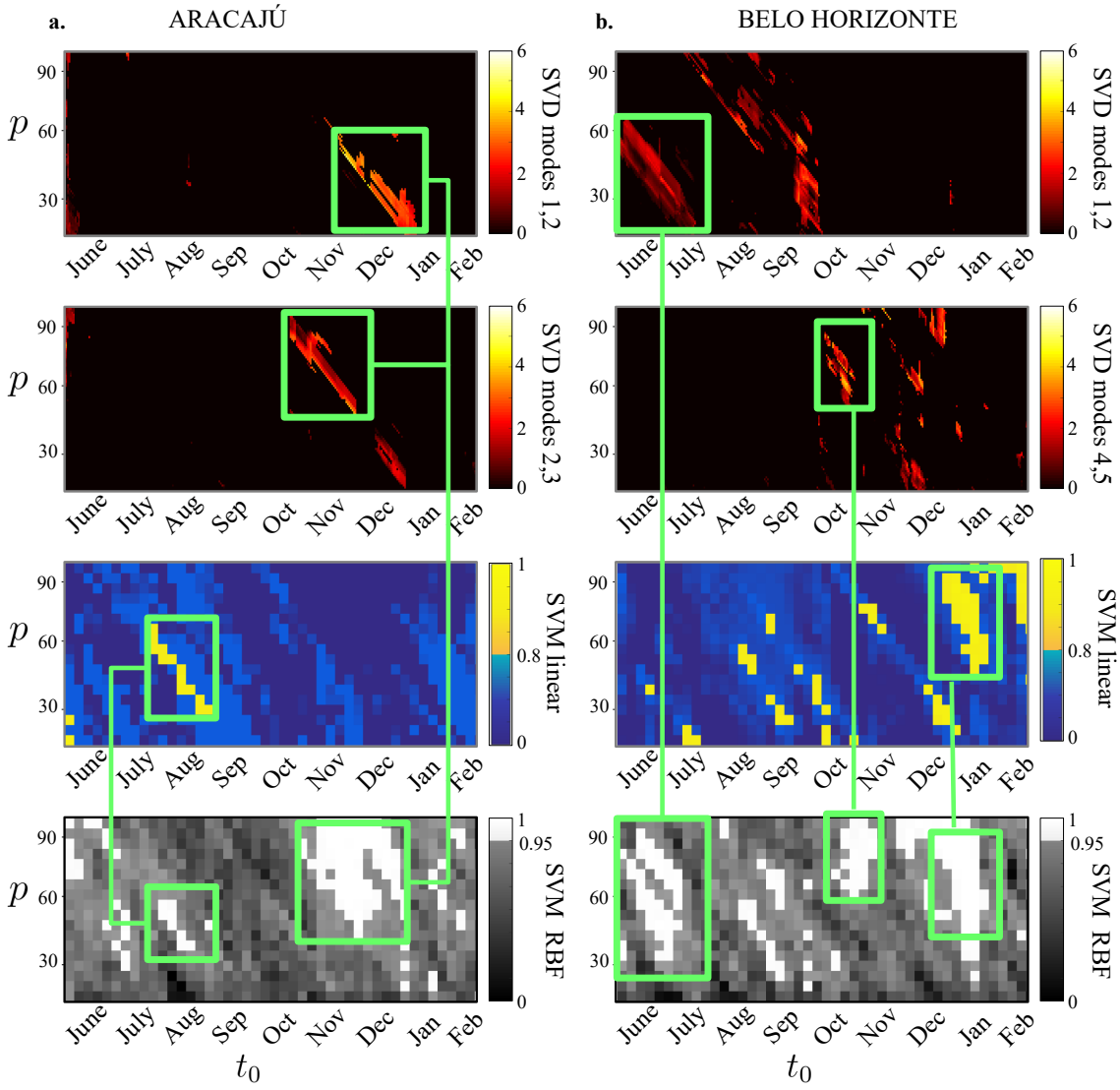

Figure 8.5: **Critical periods for Aracajú and Belo Horizonte.** For these cities, we have found periods with high correlation between climate indicators and Dengue outbreaks during winter, spring and summer. Aracajú (**a.**) and Belo Horizonte (**b.**) are the state capitals of Sergipe and Minas Gerais, located in the northeast and southeast regions of Brazil, respectively.

$p \leq 60$ days, which would also include the first days of spring.

8.1.4 Aracajú

The capital of Sergipe displays high separability scores according to the SVD methodology (for modes 1,2 and 2,3) for periods with t_0 in November – January and period length $p < 60$ days (see Fig. 8.5 **a**). Similar critical periods in the $t_0 \times p$-plane also occur in the SVM-RBF color map. This suggests that the climate conditions during spring and summer are of crucial importance for the occurrence of Dengue in Aracajú. The SVM methodology also highlights critical periods for t_0 in August and p between 30 and 60 days, which would correspond to late winter and/or beginning of spring.

8.1.5 Belo Horizonte

In Fig. 8.5 **b.** we show the highlights for Belo Horizonte, the state capital of Minas Gerais. The SVD color map (for modes 1,2) shows critical regions for t_0 between June–July and p between 30 and 60 days. These (t_0, p)-periods corresponds to the winter season in Brazil. A similar result was found in the SVM - RBF method, but with a larger range of p. There was also a good accordance between SVD scores (for modes 4,5) and SVM (for the RBF kernel) when t_0 is between October and November and p is between 60 and 90 days. These critical periods with high separability scores correspond to the spring season. For the summer period and beginning of the fall (where the epidemic outbreaks usually occur), both SVM kernels indicate critical periods for t_0 between December and January and p between 45 and 90 days.

8.1.6 Recife

The capital of Pernambuco shows high separability scores for both SVD and SVM methodologies during the summer season (see Fig. 8.6**a**). We found critical periods for t_0 between December–January and p varying from 15 to 60 days using the SVD methodology (for modes 2,3) and the SVM methodology (for both Linear and RBF kernels). Both SVD color map (for modes 4,5) and SVM color map (RBF kernel) indicate regions with high separability scores for t_0 between December and January and p around between 60 and 90 day, which would include the first days of the fall season. For winter and spring, the

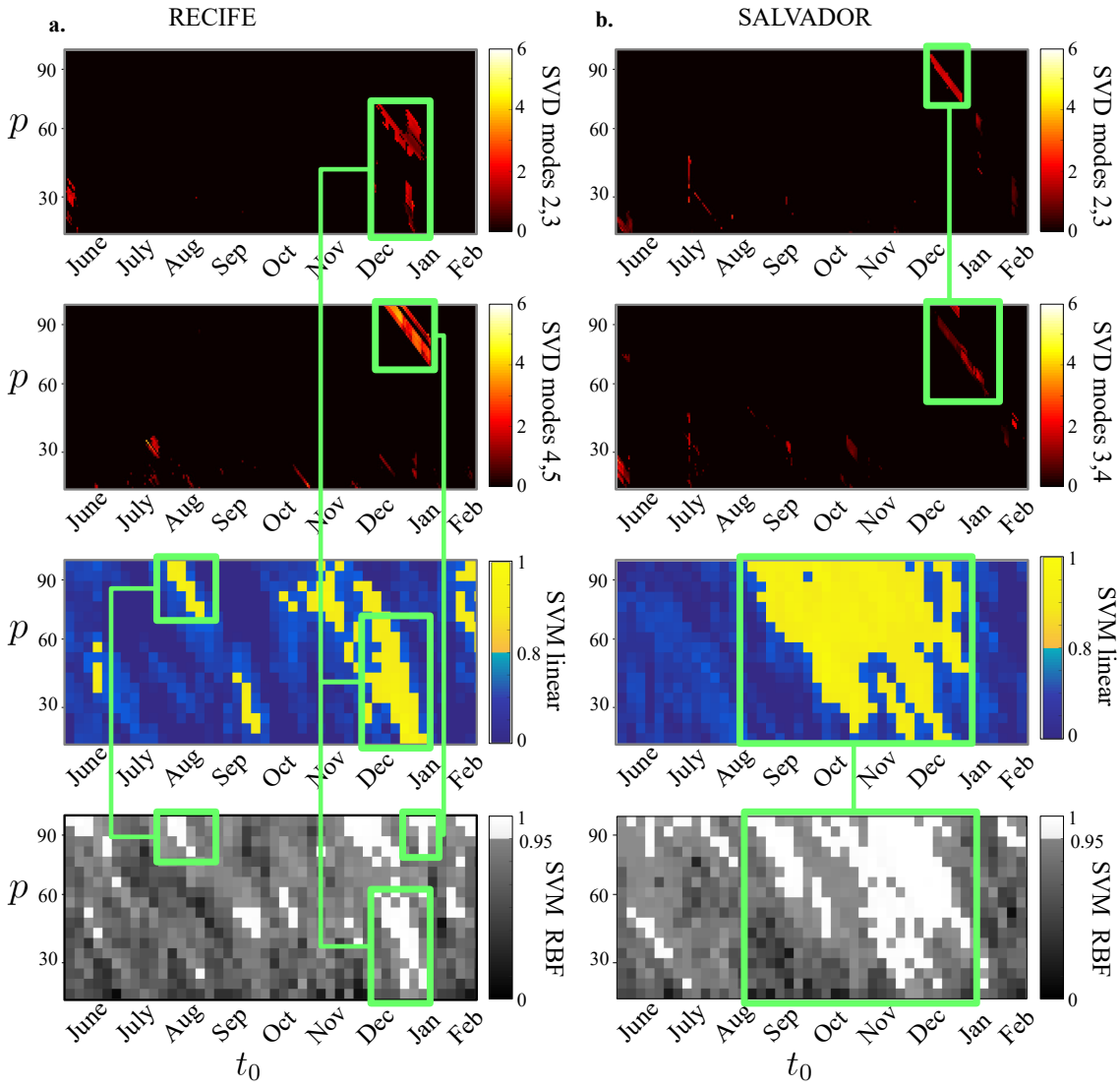

Figure 8.6: **Critical periods for Recife and Salvador.** These two northeast state capitals have exhibited strong correlation between climate signatures and Dengue epidemics, specially during spring and summer. **a.** For Aracajú, we have found accordance between SVD (modes 2,3 and 4,5) and the SVM methods. **b.** For Salvador, SVM methods has shown a good performance by showing big RHS for t_0 between August and December.

SVM methods (for both linear and RBF kernels) find critical periods for t_0 in August and p around 90 days.

8.1.7 Salvador

For the state capital of Bahia (see Fig. 8.6 **b**), the SVM separability scores are high for t_0 between August and December (for both Linear and RBF kernels) and for all values of length p. This suggests that spring and summer are crucial for the development of Dengue epidemics in Salvador. We also highlighted that SVD separability scores (for modes 2,3 and 3,4) are high for t_0 between December and January and p between 60 and 90 days, which would also correspond to the summer season.

Chapter 9

Discussion

In this work, we developed data-driven methods to identify in a systematic manner a set of critical periods in the annual cycle in which climate conditions may play a significant role in the development of Dengue outbreaks the following year. For a fixed time period starting at t_0 and lasting p days, we evaluate separability scores between the climate conditions on epidemic/non-epidemic years. We postulate that the periods where these climate conditions differ most might be crucial for the development of the life cycle of the mosquito population, and consequently, to Dengue outbreaks. The separability scores were calculated following two different methods. The first one is based on dimensionality reduction of data via Singular Value Decomposition (SVD) and the second one on the machine learning classification algorithm known as Support Vector Machines (SVM). We applied these methods to temperature and precipitation time series data for seven state capitals in Brazil where there was a significant alternation between epidemic and non-epidemic years in the recent past. Both methods indicated critical periods with remarkable agreement. The analysis of this particular dataset was only made possible due to the successful application of compressed sensing techniques to plausibly complete missing data. In fact, the cities of Rio de Janeiro, Salvador, and São Luís had the larger gaps in their daily recording of climate variables that were circumvented using compressive sensing.

Long-term effects

After localizing the critical periods with high separability scores between epidemic/non-epidemic climate conditions we were able to find which seasons were crucial for the development of Dengue outbreaks at each city. See Table 9.1 for a summary of the results. We obtained strong evidence that the climate influence on epidemics varies significantly from place to place [133, 151, 152], and thus rejecting simplistic or universal explanations involving temperature and rain precipitation in urban centers. We found a high correlation between critical climate signatures during the winter season and the occurrence of outbreaks in Aracajú, Belo Horizonte, Manaus, Rio de Janeiro, and São Luís.

Several works report and quantify how climate influence the mosquito development on a weekly scale [153, 154, 155]. We suggest that climate conditions may have long-term effects as well, occurring even months before the outbreaks. As a consequence, intensifying mosquito control campaigns during the winter season may prove an interesting epidemic control strategy, especially due to the smaller size of the vector populations during that period. In Brazil, the national and local campaigns are usually restricted to spring and summer periods [156, 157]. In fact, the Brazilian government announced that a special task force for fighting mosquitos was to be formed November 3^{rd}, 2016 [158]. We believe this starting date to be too late since critical climate conditions were detected in some cities even 9 months prior to epochs with higher Dengue incidence.

Assisting early warning systems

A number of early warning systems are available for calculating the risk of Dengue epidemics taking climate factors into account [135, 136, 159, 160, 161, 162]. In this sense, our methodology offers an additional set of key periods that may assist current warning systems or serve as basis to a new model focusing on climate signatures of epidemic years. Fig. 9.1 illustrates how this could be achieved using SVM linear kernel separability scores: (i) We would train climate data projected onto the $\langle T_j \rangle \times \langle \delta_j \rangle^{-1}$ plane and divide it into two regions referring to epidemic (red) or non-epidemic (blue) data. (ii) Once we obtain temperature and precipitation measurements for the following year we can quantify the

Table 9.1: **Summary of most important seasons for Dengue outbreaks.**

Capital	**Winter**	**Spring**	**Summer**	**Fall (DF)**
Aracajú	x		x	
Belo Horizonte	x	x	x	
Manaus	x			
Recife		x	x	x
Rio de Janeiro	x	x		
Salvador		x	x	
São Luís	x		x	

Remark: peaks of Dengue Fever outbreaks happen typically during the fall (March – May).

fraction of climate data falling into each region and use it to forecast Dengue the following year. (iii) After the Dengue outcome is known for that year, we can append that data to our set and retrain the classifier line between epidemic/non-epidemic regions. This should improve, at least in theory, the precision of future forecasts and our understanding of critical climate signatures.

Limitations of our methodology

There are several limitations to our work and all of our results must be interpreted with caution and parsimony. Ultimately, we are only *suggesting* that temperature and precipitation discrepancies in key epochs affected mosquito populations in a critical way, and despite the plausible hypotheses, they were not yet directly measured or reported satisfactory by field studies. Moreover, we didn't consider several other factors believed to be important for explaining Dengue dynamics in details, such as: **(i)** Circulation of different strains of the dengue virus [123, 124, 163, 164]; once the cross-immunity wanes with time, the introduction of new DENV serotypes may affect an entire population. **(ii)** Human mobility within and among the cities [125, 126, 127, 165, 166]; the lifetime movement range of an Ae.Aegypti mosquito is typically less than a kilometer and the spread of Dengue

through an urban area is most likely driven by everyday human movement [167, 168]. In fact, humans act as vectors between relatively localized mosquito populations and might change their commute strategies based on climate factors as well. **(iii)** Human demographic dynamics: Lower death rates may increase the longevity of immune individuals and lower birth rates may decrease the number of susceptible individuals. Such fluctuations over the years may change the magnitude of the infections [169, 170]. **(iv)** Global warming and global climate changes: Several studies examined for instance the influence of El Niño Southern Oscillation (ENSO) in Dengue incidence. While some argue that ENSO is behind the synchronization of Dengue epidemics and traveling waves of infections, others dismiss it as a minor factor [171, 172, 173, 174]. In any case, global climate changes are likely to affect local climate conditions and consequently, Dengue transmission. **(v)** Our methodological limitations constrained the analysis to cities that experienced at least three epidemic and three non-epidemic years in the recent past. Poorly recordings of climate data also prevented us from including two additional state capitals to our dataset (see Appendix and SI for details). In the future, we expect to extend our analysis to other state capitals. Finally, at this stage, we outline a modest prediction system for Dengue outbreaks using our methods only as a proof of concept, leaving detailed forecasting (as done in [135, 136]) for future works.

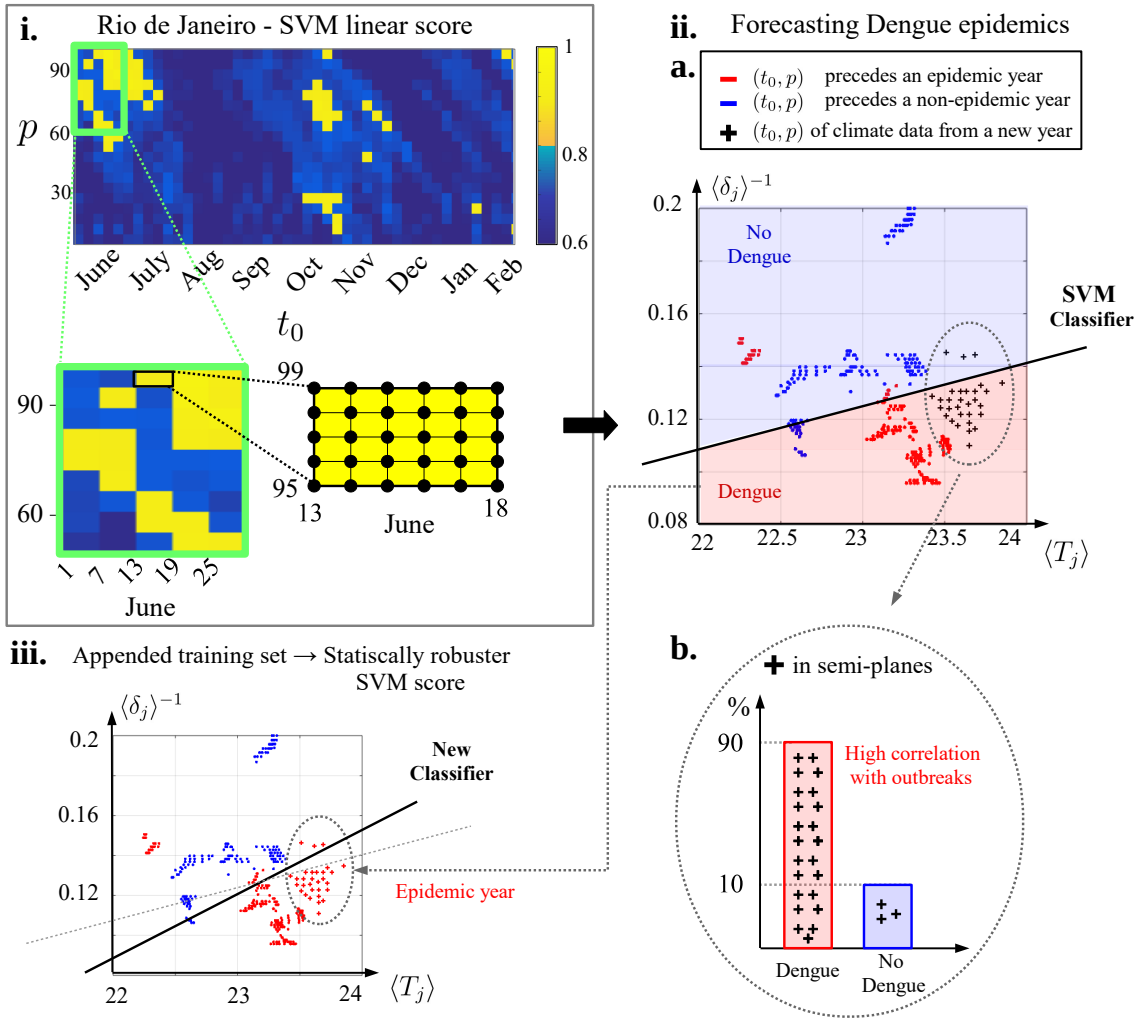

Figure 9.1: **Forecasting Dengue Outbreaks and appending data for further analysis.** Example for the SVM-Linear methodology on climate data from Rio de Janeiro. **(i)** We choose a high scored (t_0, p)-rectangle, for which we plot the climate indicators with their respective colors. **(ii)** We apply a SVM training algorithm on this 2D-dataset. **a.** A classifier line can be drawn and two semi-planes (Dengue and No-Dengue) are obtained. **b.** With data from a new year for the same (t_0, p) periods (black crosses), we can compute the percentage of indicators that falls into each of those semi-planes. Therefore we are able to estimate the correlation between new and previous climate data with respect to Dengue epidemics. **(iii)** Depending on the classification of the new year as epidemic or not, the new data is colored red or blue to become part of a new SVM-training set. This procedure will give a more accurate information about the importance of the chosen (t_0, p)-rectangle on Dengue prediction.

Conclusion

Epidemic control of Dengue, Zika and Chikungunya is one of the most urgent public health challenges in a globalized world, and their effects have dramatic societal consequences in large tropical countries such as Brazil. A better understanding of the multi-scale and long terms effects of climate conditions on the development of *Aedes Aegypti* populations is crucial for improving the timing of vector-control efforts and other policies. In this sense, this work adds a new piece to the complex puzzle that is the development of mosquito populations in dynamic urban environments under variable climate conditions. Figs. 8.1 and 8.2 illustrates the power of our methodology. We have found two specific parameters – mean of temperature and frequency of precipitation – that may be crucial for Dengue prediction in Brazil.

Not only are the analytic metrics developed in this manuscript predictive, they are also easily interpretable in terms of simple to acquire measurement proxies of temperature and precipitation. Moreover, in all cities considered where data was readily available, distinct and separable climate patterns were shown which suggest that accurate prediction of epidemics can be achieved in the winter preceding the outbreak. This suggests that many of the eradication strategies should be performed well in advance of the summer months where the epidemic is manifest.

Table 9.1 summarizes the potential for predictive success of Dengue outbreaks. Remarkably, in almost all cities, aside from Recife and São Luís, a prediction can be made approximately six to nine months in advance of the epidemic outbreak. And aside from Manaus, all cities offer multiple windows of opportunity for forecasting the Dengue levels during the annual cycle. Interestingly, the summer in Rio de Janeiro offers little insight into this matter, since data of years with and without Dengue are qualitatively similar from a climate perspective. Yet public strategies have typically been enacted and decided during this time period, which is both too late and does not leverage the predictive capabilities of the climate data. We conjecture that the winter months are critical for establishing the ideal breeding conditions, through temperature and frequency of precipitation, which ultimately determine the size of the *Aedes Aegypti* population. This suggests that disrupt-

ing the breeding cycle six to nine months in advance may be a robust strategy for vector control. For instance, in Rio de Janeiro, if during the winter months the rain frequency is approximately once per week and the average temperature is approximately 22 Celsius (See Fig. 8.1), then it is highly likely that an epidemic will occur, thus requiring an intervention strategy 9 months in advance.

The work also highlights that the patterns allowing for predictive success are quite distinct from city to city. Fig. 8.1 demonstrates that a simple, universal rule about climate effects may be hard to achieve. Indeed, data on the seven cities demonstrate a remarkably heterogeneous range of behaviors despite each individual city giving rise to clear prediction windows. This is largely to be expected as climatic effects, such as proximity to ocean, jungle, forest, dense populations, etc. will all play a significant role in how precipitation and temperature favorably or unfavorably effects the growth of the disease vector *Aedes Aegypti*.

Chapter 10

Supporting Information

10.1 Details about the choice of the seven capitals

As explained in our methods chapter, we chose state capitals that had at least 3 years with Dengue Epidemics (DE) and at least 3 years without DE in the recent past. The following 9 state capitals passed this criterium: Aracajú, Belo Horizonte, Cuiabá, João Pessoa, Manaus, Recife, Rio de Janeiro, Salvador and São Luís. We completed missing data through linear interpolation and/or usage of alternative sources for precipitation time series given that the CVX routine does not work well for episodical data events. From the 9 state capitals, the following 6 had only single precipitation gaps: Aracajú, Belo Horizonte, Manaus, Recife, Salvador and São Luís. The cities of Cuiabá, João Pessoa and Rio de Janeiro had big missing data epochs. For Rio de Janeiro we found an alternative source of precipitation data, but the other two state capitals had to be discarded from our analysis.

10.2 Epidemic / non-epidemic years and missing Climate data for each chosen state capital

We provide tables with estimated population, total number of Dengue cases, incidence per $100,000$ inhabitants, and details of our climate data completing protocols (if any). For

Rio de Janeiro, we consider the time period from 2003 to 2013 and we use epidemic data from municipal webpage ([1]). For the other state capitals, the analyzed period ranges from 2002 to 2012 and data was collected from the Ministry of Health's Notifiable Diseases Information System. ([2])

Table 10.1: **Aracajú.**

Year	**Pop.**	**Cases**	**Incidence**
2002	473,991	1,933	407.81
2003	479,767	1,301	271.17
2004	491,898	166	33.75
2005	498,619	271	54.35
2006	505,286	355	70.26
2007	520,303	728	139.92
2008	536,785	10,702	1,993.72
2009	544,039	1,232	226.45
2010	571,149	302	52.88
2011	579,563	1,399	241.39
2012	587,701	2,656	451.93

Incidence = Cases per 100,000 inhabitants. Single gaps of missing climate data were filled by linear interpolation; temperature on 12/21/2006 and precipitation on 7/24/2006.

[1] //www.rio.rj.gov.br/web/sms/dengue

[2] http://www.portalsinan.saude.gov.br

Table 10.2: **Belo Horizonte.**

Year	**Pop.**	**Cases**	**Incidence**	**Temp (L.I)**	**Precip (L.I)**
2001	–	–	–	8/9	8/9
2002	2,284,468	4,749	207.88	8/31	8/31
2003	2,305,812	1,800	78.06	–	–
2004	2,350,564	472	20.08	–	–
2005	2,375,329	149	6.27	–	–
2006	2,399,920	872	36.33	–	–
2007	2,412,937	5278	218.74	12/31	12/31
2008	2,434,642	12,967	532.60	1/1 11/21	1/1 11/21
2009	2,452,617	14,494	590.96	12/12	12/12
2010	2,375,151	52,315	2,202.60	–	–
2011	2,385,640	1,749	73.31	–	–
2012	2,395,785	635	26.50	–	–

Incidence = Cases per 100,000 inhabitants. L.I stands for Linear Interpolation.

Table 10.3: **Manaus.**

Year	**Pop.**	**Cases**	**Incidence**
2002	1,488,805	1,855	124.60
2003	1,527,314	3,731	244.29
2004	1,592,555	789	49.54
2005	1,644,690	915	55.63
2006	1,688,524	495	29.32
2007	1,646,602	1,989	120.79
2008	1,709,010	5,975	349.62
2009	1,738,641	623	35.83
2010	1,802,014	3,748	207.99
2011	1,832,424	54,342	2,965.58
2012	1,861,838	3,703	198.89

Incidence = Cases per 100,000 inhabit. Single gaps of missing climate data were filled by linear interpolation; temperature on 12/23/2005 and precipitation on 2/11/2005.

Table 10.4: **Recife.**

Year	**Pop.**	**Cases**	**Incidence**	**Temp (L.I)**	**Precip(L.I)**
2001	–	–	–	–	–
2002	1,449,135	42,791	2,952.86	–	–
2003	1,461,320	449	30.73	–	–
2004	1,486,869	241	16.21	–	–
2005	1,501,008	830	55.30	–	–
2006	1,515,052	1,443	95.24	11/4 12/2	–
2007	1,533,580	1,503	98.01	–	–
2008	1,549,980	4,771	307.81	4/28	–
2009	1,561,659	578	37.01	4/30 7/31 11/19	–
2010	1,537,704	11,494	747.48	9/8	–
2011	1,546,516	5,471	353.76	–	–
2012	1,555,039	11,444	735.93	1/1 5/2 6/14 8/14	1/1 5/2 8/14

Incidence = Cases per 100,000 inhabitants. L.I stands for Linear Interpolation.

Table 10.5: **Rio de Janeiro.**

Year	**Pop.**	**Cases**	**Incidence**	**Temp (CVX)**	**Precip (subst)**
2002	–	–	–	8/31	–
2003	5,974,081	1,610	26.95	3/1 – 3/2 6/20 – 6/30	6/20 – 6/30
2004	6,051,399	607	10.03	–	–
2005	6,094,183	980	16.08	–	–
2006	6,136,652	14,435	235.23	–	12/13 – 12/31
2007	6,093,472	26,507	435.01	1/1 – 2/1	1/1 – 1/10
2008	6,161,047	110,861	1799.39	–	–
2009	6,186,710	2,961	47.86	2/11	–
2010	6,320,446	3,000	47.47	–	–
2011	6,355,949	78,645	1237.34	–	–
2012	6,390,290	137,505	2151.78	12/8 12/26 – 12/27	–
2013	6,429,923	66,278	1030.77	6/13 – 6/21	6/14 – 6/19

Incidence = Cases per 100,000 inhabitants. Number of Dengue cases in 2013 taken from the City's hall health department, because data from SINAN is not available for that year. For the larger gaps of missing data on precipitation time series, we have used data of the *Alerta Rio* system from Saúde neighborhood, the closest to the Santos Dumont airport where INMET's rain collectors are located.

Table 10.6: **Salvador.**

Year	**Pop.**	**Cases**	**Incidence**	**Temp (CVX)**	**Precip (L.I)**
2001	–	–	–	–	–
2002	2,520,504	26,838	1,064.79	10/9 – 10/21	–
2003	2,556,429	908	35.52	–	–
2004	2,631,831	154	5.85	–	–
2005	2,673,560	270	10.10	10/21 – 10/31	
2006	2,714,018	377	13.89	–	–
2007	2,892,625	1,349	46.64	10/6 – 10/7	10/7
2008	2,948,733	2,476	83.97	–	–
2009	2,998,056	6,819	227.45	6/9 12/27	–
2010	2,675,656	6,159	230.19	–	–
2011	2,693,606	5,321	197.54	–	–
2012	2,710,968	5,161	190.37	–	–

Incidence = Cases per 100,000 inhabitants. L.I stands for Linear Interpolation.

Table 10.7: **São Luís .**

Year	Pop.	Cases	Incidence	Temp (CVX)	Precip (L.I)
2001	–	–	–	10/1 – 10/31 11/14 11/21	–
2002	906,567	448	49.42	4/30	–
2003	923,526	567	61.40	9/5 – 9/26 9/28 – 10/10	–
2004	959,124	154	16.06	–	–
2005	978,824	2,580	263.58	–	–
2006	998,385	1,395	139.73	–	–
2007	957,515	3,827	399.68	–	–
2008	986,826	1,183	119.88	–	–
2009	997,098	100	10.03	–	5/31
2010	1,014,837	2,731	269.11	–	–
2011	1,027,430	5,229	508.94	10/20	–
2012	1,039,610	1,315	126.49	6/8 – 6/9 6/12 – 6/13 7/24 – 7/25 7/29	–

Incidence = Cases per 100,000 inhabitants. L.I stands for Linear Interpolation.

Bibliography

[1] Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. science. 2000; 290(5500), 2323-2326.

[2] Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. Springer, Berlin: Springer series in statistics. 2001

[3] Hey T, Tansley S, Tolle KM. The fourth paradigm: data-intensive scientific discovery Redmond. 2009; WA: Microsoft research. (Vol. 1)

[4] Samuel, AL. Some studies in machine learning using the game of checkers. IBM Journal of research and development. 1959; 3(3), 210-229.

[5] Abu-Mostafa YS, Magdon-Ismail M, Lin HT. Learning from data. (Vol. 4). New York, NY, USA:: AMLBook. 2012.

[6] Brunton SL, Proctor JL, Kutz JN. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proceedings of the National Academy of Sciences. 2016; 113(15), 3932-3937.

[7] Broome BM, Jayaraman V, Laurent G. Encoding and decoding of overlapping odor sequences. Neuron. 2006; 51(4), 467-482.

[8] Cohen MR, Maunsell JH. A neuronal population measure of attention predicts behavioral performance on individual trials. Journal of Neuroscience. 2010; 30(45), 15241-15253.

[9] Dimensionality reduction for large-scale neural recordings. Nature neuroscience. 2014; 17(11), 1500-1509.

[10] Ju H, Brasier AR. Variable selection methods for developing a biomarker panel for prediction of dengue hemorrhagic fever. BMC research notes. 2013; 6(1), 365.

[11] Frasca M, Rizzo A, Gallo L, Fortuna L and Porfiri M. Dimensionality reduction in epidemic spreading models. EPL (Europhysics Letters). 2015; 111(6), 68006.

[12] Tenenbaum JB, De Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. science. 2000; 290(5500), 2319-2323.

[13] Destexhe A, Bedard C. Local field potential. Scholarpedia.2013; 8(8), 10713.

[14] Murphy KP. Machine learning: a probabilistic perspective. MIT press; 2012 Sep 7.

[15] Bishop CM. Pattern recognition. Machine Learning. 2006;128.

[16] Annegers JF, Grabow JD, Groover RV, Laws ER, Elveback LR, Kurland L T. Seizures after head trauma : A population study Neurology. 1980; 30(7), 683-683.

[17] Delanty N, Vaughan CJ, French JA Medical causes of seizures The Lancet. 1998;352(9125), 383-390.

[18] Asikainen I, Kaste M, Sarna S. Early and late post traumatic seizures in traumatic brain injury rehabilitation patients: Brain injury factors causing late seizures and influence of seizures on long-term outcome. Epilepsia. 1999; 40(5), 584-589.

[19] World Health Organization. Epilepsy. WHO Factsheet. February 2017; Avaiable from: http://www.who.int/mediacentre/factsheets/fs999/en/

[20] De Boer HM, Mula M, Sander JW The global burden and stigma of epilepsy. Epilepsy & behavior. 2008; 12(4), 540-546.

[21] Carney PR, Myers S, Geyer JD. Seizure prediction: methods. Epilepsy & Behavior. 2011 22, S94-S101.

[22] Téllez-Zenteno JF, Dhar R, Wiebe S. Long-term seizure outcomes following epilepsy surgery: a systematic review and meta-analysis. Brain. 2005; 128(5), 1188-1198.

[23] Epilepsy Foundation. Treating Seizures and Epilepsy : Surgery. Avaiable from : http://www.epilepsy.com/learn/treating-seizures-and-epilepsy/surgery

[24] Jerger KK, Netoff TI, Francis JT, Sauer T, Pecora L, Weinstein SL, Schiff SJ Early seizure detection. Journal of Clinical Neurophysiology. 2001; 18(3), 259-268.

[25] Ramgopal S, Thome-Souza S, Jackson M, Kadish NE, Fernández IS, Klehm J et al. Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. Epilepsy & behavior.2014; 37, 291-307.

[26] Chan AM, Sun FT, Boto EH, Wingeier BM. Automated seizure onset detection for accurate onset time determination in intracranial EEG. Clinical Neurophysiology.2008; 119(12), 2687-2696.

[27] Mirowski P, Madhavan D, LeCun Y, Kuzniecky R. Classification of patterns of EEG synchronization for seizure prediction. Clinical neurophysiology.2009; 120(11), 1927-1940.

[28] Kaufman, PY. Electrical phenomena in the cerebral cortex. Obozrenie Psikhiatrii, Nevrologii i Eksperimentalnoy Psikhologii (St. Petersburg).1912; 7-8.

[29] Pravdich-Neminsky VV. Ein Versuch der registrierung der elektrischen gehirnerscheinungen. Zbl Physiol.1913; 27, 951-960.

[30] Nunez PL, Srinivasan R. Electroencephalogram. Scholarpedia. 2007; 2(2), 1348.

[31] Berger H. Uber das Elektrenkephalogramm des Menschen. European Archives of Psychiatry and Clinical Neuroscience.1932; 97(1), 6-26.

[32] Berger H. Uber das Elektroenzephalogram des Menschen. Archiv für Psychiatrie und Nervenkrankheiten.1933; vol. 100, pp. 301–320.

[33] Gibbs F, Davis H, Lennox WG. The electroencephalogram in epilepsy and in conditions of impaired consciousness. Archives of Neurology and Psychiatry.1935; vol. 34, no. 6, pp. 1133–1148.

[34] Gibbs F, Davis H, Lennox WG. The electro-encephalogram in diagnosis and in localization of epileptic seizures. Archives of Neurology & Psychiatry.1936; 36(6), 1225-1235.

[35] Gibbs F, Davis H, Lennox WG. Epilepsy: a paroxysmal cerebral dysrhythmia. Brain: A Journal of Neurology.1937; vol. 60, no. 4, pp. 377–388.

[36] Gibbs F, Gibbs E. Atlas of Electroencephalography. Boston City Hospital, Oxford, UK. 1941

[37] Magiorkinis E, Diamantis A, Sidiropoulou K, Panteliadis C. Highights in the history of epilepsy: the last 200 years. Epilepsy research and treatment.2014;

[38] Penfield W , Jasper H. Epilepsy and the Functional Anatomy of the Human Brain. Little Brown , Boston .1954.

[39] Brown R, Kocarev L. A unifying definition of synchronization for dynamical systems. Chaos 10: 344–349, 2000.

[40] Carmeli C, Knyazeva MG, Innocenti GM, De Feo O. Assessment of EEG synchronization based on state-space analysis. Neuroimage 25: 339–354, 2005.

[41] Le Van Quyen M, Soss J, Navarro V, Robertson R, Chavez M, Baulac M, Martinerie J. Preictal state identification by synchronization changes in longterm intracranial EEG recordings. Clin Neurophysiol 116: 559–568, 2005.

[42] Mormann F, Lehnertz K, David P, Elger CE. Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. Physica D 144: 358–369, 2000.

[43] Varela F, Lachaux JP, Rodriguez E, Martinerie J. The brainweb: phase synchronization and large scale integration. Nat Rev Neurosci 2: 229–239,2001.

[44] Allefeld, C, Müller, M, Kurths J. Eigenvalue decomposition as a generalized synchronization cluster analysis. International Journal of Bifurcation and Chaos.2007; 17(10), 3493-3497.

[45] Muller M, Baier G. Detection and characterization of changes of the correlation structure in multivariate time series. Phys Rev E 71: 046116, 2005.

[46] Muller M, Lopez Y, Rummel C, Baier G, Galka A, Stephani U, Muhle H. Localized short-range correlations in the spectrum of the equal-time correlation matrix. Phys Rev E 74: 041119, 2006.

[47] Le Van Quyen M, Martinerie J, Navarro V, Baulac M, Varela FJ. Characterizing neurodynamic changes before seizures. Journal of Clinical Neurophysiology. 2001; 18(3), 191-208.

[48] Netoff TI, Schiff SJ. Decreased neuronal synchronization during experimental seizures. Journal of Neuroscience. 2002; 22(16), 7297-7307.

[49] Mormann F, Kreuz T, Andrzejak RG, David P, Lehnertz K, Elger CE. Epileptic seizures are preceded by a decrease in synchronization Epilepsy research.2003; 53(3), 173-185.

[50] Chávez, M., Le Van Quyen, M., Navarro, V., Baulac, M., Martinerie, J. Spatio-temporal dynamics prior to neocortical seizures: amplitude versus phase couplings. IEEE Transactions on Biomedical Engineering. 2003; 50(5), 571-583.

[51] Wendling F, Bartolomei F, Bellanger JJ, Bourien J, Chauvel P. Epileptic fast intracerebral EEG activity: evidence for spatial decorrelation at seizure onset. Brain. 2003; 126(6), 1449-1459.

[52] Schiff SJ, Sauer T, Kumar R, Weinstein SL. Neuronal spatiotemporal pattern discrimination: the dynamical evolution of seizures. Neuroimage. 2005; 28(4), 1043-1055.

[53] Li Y, Fleming IN, Colpan ME, Mogul DJ. Neuronal desynchronization as a trigger for seizure generation. IEEE Transactions on Neural Systems and Rehabilitation Engineering.2008; 16(1), 62-73.

[54] Wang CH, Hung CP, Chen MT, Shih YH, Lin YY. Hippocampal desynchronization of functional connectivity prior to the onset of status epilepticus in pilocarpine-treated rats. PloS one.2012; 7(6), e39763.

[55] Jiruska P, De Curtis M, Jefferys JG, Schevon CA, Schiff SJ, Schindler K. Synchronization and desynchronization in epilepsy: controversies and hypotheses. The Journal of physiology.2013; 591(4), 787-797.

[56] Gotman J. Automatic recognition of epileptic seizures in the EEG. Electroencephalography and clinical Neurophysiology.1982; 54(5), 530-540.

[57] Gotman, J. Seizure recognition and analysis. Electroencephalography and clinical neurophysiology.1984; Supplement, 37, 133-145.

[58] Gotman J. Automatic seizure detection: improvements and evaluation. Electroencephalography and clinical Neurophysiology.1990; 76(4), 317-324.

[59] Tzallas AT, Tsalikakis DG, Karvounis EC, Astrakas L, Tzaphlidou, M, Tsipouras MG, et al. Automated epileptic seizure detection methods: a review study. INTECH Open Access Publisher.2012;

[60] Khan YU, Gotman J. Wavelet based automatic seizure detection in intracerebral electroencephalogram. Clinical Neurophysiology.2003; 114(5), 898-908.

[61] Saab ME, Gotman J. A system to detect the onset of epileptic seizures in scalp EEG. Clinical Neurophysiology.2005; 116(2), 427-442.

[62] Temko A, Thomas E, Marnane W, Lightbody G, Boylan G. EEG-based neonatal seizure detection with support vector machines. Clinical Neurophysiology.2011; 122(3), 464-473.

[63] Ubeyli ED, Güler INAN. Statistics over Lyapunov exponents for feature extraction: electroencephalographic changes detection case. World Academy of Science, Engineering and Technology.2007; 2, 624-628.

[64] Chandaka S, Chatterjee A, Munshi S. Cross-correlation aided support vector machine classifier for classification of EEG signals. Expert Systems with Applications. 2009; 36(2), 1329-1336.

[65] Guo L, Rivero D, Pazos A. Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks. Journal of neuroscience methods.2010; 193(1), 156-163.

[66] Ghosh-Dastidar S, Adeli H, Dadmehr N. Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. IEEE Transactions on Biomedical Engineering.2008; 55(2), 512-518.

[67] Ghosh-Dastidar S, Adeli H, Dadmehr N. Mixed-band wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection. IEEE transactions on biomedical engineering.2007; 54(9), 1545-1551.

[68] Tang Y, Durand DM. A tunable support vector machine assembly classifier for epileptic seizure detection. Expert systems with applications.2012; 39(4), 3925-3938.

[69] Webber WRS, Lesser RP, Richardson RT, Wilson K. An approach to seizure detection using an artificial neural network (ANN). Electroencephalography and clinical Neurophysiology.1996; 98(4), 250-272.

[70] Guo L, Rivero D, Dorado J, Munteanu CR, Pazos A. Automatic feature extraction using genetic programming: An application to epileptic EEG classification. Expert Systems with Applications.2011; 38(8), 10425-10436.

[71] Iscan Z, Dokur Z, Demiralp T. Classification of electroencephalogram signals with combined time and frequency features. Expert Systems with Applications.2011; 38(8), 10499-10505.

[72] Chua KC, Chandran V, Acharya R, Lim CM. Automatic identification of epilepsy by HOS and power spectrum parameters using EEG signals: A comparative study. Engineering in Medicine and Biology Society.2008; EMBS 2008. 30th Annual International Conference of the IEEE (pp. 3824-3827). IEEE.

[73] Henke K, Buck A, Weber B, Wieser HG. Human hippocampus establishes associations in memory. Hippocampus. 1997; 7(3), 249-256.

[74] O'keefe J, Nadel L. The hippocampus as a cognitive map. Oxford: Clarendon Press.1978;

[75] Das SR, Mechanic-Hamilton D, Korczykowski M, Pluta J, Glynn S, Avants et al. Structure specific analysis of the hippocampus in temporal lobe epilepsy. Hippocampus. 2009; 19(6), 517.

[76] Epilepsy Foundation. Temporal Lobe Epilepsy. Avaiable from http://www.epilepsy.com/learn/types-epilepsy-syndromes/temporal-lobe-epilepsy

[77] Dreifuss FE, Martinez-Lage M, Johns RA. Proposal for classification of epilepsies and epileptic syndromes. Epilepsia. 1985; 26(3), 268-278.

[78] Kutz JN. Data-Driven Modeling & Scientific Computation: Methods for Complex Systems & Big Data. Oxford University Press; 2013.

[79] Queiroz CM, Gorter JA, da Silva FHL, Wadman WJ. Dynamics of evoked local field potentials in the hippocampus of epileptic rats with spontaneous seizures. Journal of neurophysiology.2009; 101(3), 1588-1597.

[80] Cavalheiro EA The pilocarpine model of epilepsy. The Italian Journal of Neurological Sciences. 1995; 16.1-2 33-37.

[81] Cavalheiro EA, Santos NF, Priel MR. The pilocarpine model of epilepsy in mice. Epilepsia. 1996; 37(10), 1015-1019.

[82] Curia G, Longo D, Biagini G, Jones RS, Avoli M. The pilocarpine model of temporal lobe epilepsy. Journal of neuroscience methods. 2008; 172(2), 143-157.

[83] Leite JP, Garcia-Cairasco N, Cavalheiro EA. New insights from the use of pilocarpine and kainate models. Epilepsy research.2002; 50(1): 93-103.

[84] Arnold BC. Pareto distributions. John Wiley & Sons, Ltd. 2015

[85] Lawson CL & Hanson RJ. Solving least squares problems. Society for Industrial and Applied Mathematics. 1995

[86] Higham DJ. An algorithmic introduction to numerical simulation of stochastic differential equations SIAM review.2001; 43(3): 525-546.

[87] Le Van Quyen M, Foucher J, Lachaux JP, Rodriguez E, Lutz A, Martinerie J et al Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony. Journal of neuroscience methods.2001; 111(2), 83-98.

[88] Schindler K, Elger CE, Lehnertz K. Increasing synchronization may promote seizure termination: evidence from status epilepticus. Clinical neurophysiology. 2007; 118(9), 1955-1968.

[89] Jokeit H, Ebner A. Long term effects of refractory temporal lobe epilepsy on cognitive abilities: a cross sectional study Journal of Neurology, Neurosurgery & Psychiatry.1999; 67(1), 44-50.

[90] Sanchez RM, Jensen FE. Maturational aspects of epilepsy mechanisms and consequences for the immature brain. Epilepsia. 2001; 42(5), 577-585.

[91] Hermann B, Seidenberg M, Bell B, Rutecki P, Sheth R, Ruggles K et al. The Neurodevelopmental Impact of Childhood-onset Temporal Lobe Epilepsy on Brain Structure and Function. Epilepsia. 2002; 43(9), 1062-1071.

[92] Cilio MR, Sogawa Y, Cha BH, Liu X, Huang LT, Holmes GL. Long-term Effects of Status Epilepticus in the Immature Brain Are Specific for Age and Model Epilepsia.2003; 44(4), 518-528.

[93] De Lanerolle NC, Kim JH, Robbins RJ, Spencer, DD. Hippocampal interneuron loss and plasticity in human temporal lobe epilepsy. Brain research. 1989 495(2), 387-395.

[94] Houser CR. Granule cell dispersion in the dentate gyrus of humans with temporal lobe epilepsy. Brain research. 1990; 535(2), 195-204.

[95] Pitkänen A, Sutula TP. Is epilepsy a progressive disorder? Prospects for new therapeutic approaches in temporal-lobe epilepsy. The Lancet Neurology. 2002; 1(3), 173-181.

[96] Alonso-Nanclares L, DeFelipe J. Alterations of the microvascular network in the sclerotic hippocampus of patients with temporal lobe epilepsy. Epilepsy & Behavior. 2014; 38, 48-52.

[97] Doucet GE, Sharan A, Pustina D, Skidmore C, Sperling MR, Tracy JI. Early and late age of seizure onset have a differential impact on brain resting-state organization in temporal lobe epilepsy. Brain topography.2015; 28(1), 113-126.

[98] Yasuda CL, Chen Z, Beltramini GC, Coan AC, Morita ME, Kubota B, et al. Aberrant topological patterns of brain structural network in temporal lobe epilepsy. Epilepsia. 2015; 56(12), 1992-2002.

[99] Osorio I, Frei MG, Wilkinson SB. Real-time automated detection and quantitative analysis of seizures and short-term prediction of clinical onset. Epilepsia. 1998; 39(6), 615-627.

[100] Qu H, Gotman J. A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: possible use as a warning device. IEEE transactions on biomedical engineering. 2002; 44(2), 115-122.

[101] Aarabi A, Fazel-Rezai R, Aghakhani Y. A fuzzy rule-based system for epileptic seizure detection in intracranial EEG. Clinical Neurophysiology. 2009; 120(9), 1648-1657.

[102] Gardner AB, Krieger AM, Vachtsevanos G, Litt B. One-class novelty detection for seizure analysis from intracranial EEG. Journal of Machine Learning Research. 2006; 7(Jun), 1025-1044.

[103] Kharbouch A, Shoeb A, Guttag J, Cash SS. An algorithm for seizure onset detection using intracranial EEG. Epilepsy & Behavior. 2011; 22, S29-S35.

[104] Ahammad N, Fathima T, Joseph P. Detection of epileptic seizure event and onset using EEG. BioMed research international. 2014; Article ID 450573.

[105] Bettus G, Wendling F, Guye M, Valton L, Régis J, Chauvel P et al. Enhanced EEG functional connectivity in mesial temporal lobe epilepsy. Epilepsy research.2008; 81(1), 58-68.

[106] Altenburg J, Vermeulen RJ, Strijers RL, Fetter WP, Stam CJ. Seizure detection in the neonatal EEG with synchronization likelihood. Clinical neurophysiology. 2003; 114(1), 50-55.

[107] van Putten MJ. Nearest neighbor phase synchronization as a measure to detect seizure activity from scalp EEG recordings. Journal of clinical neurophysiology. 2003; 20(5), 320-325.

[108] Litt B, Esteller R, Echauz J, D'Alessandro M, Shor R, Henry T, et al. Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. Neuron. 2001; 30(1), 51-64.

[109] Mormann F, Andrzejak RG, Elger CE, Lehnertz K. Seizure prediction: the long and winding road. Brain. 2007; 130(2), 314-333.

[110] Bragin A, Wilson CL, Engel J. Spatial stability over time of brain areas generating fast ripples in the epileptic rat. Epilepsia. 2003; 44(9), 1233-1237.

[111] Foffani G, Uzcategui YG, Gal B, de la Prida LM. Reduced spike-timing reliability correlates with the emergence of fast ripples in the rat epileptic hippocampus. Neuron. 2007; 55(6), 930-941.

[112] Blinchikoff H, Krause H. Filtering in the time and frequency domains. The Institution of Engineering and Technology; 2001

[113] Ranjit S, Kissoon N Dengue hemorrhagic fever and shock syndromes. Pediatr. Crit. Care Med. 2011; 12 (1): 90-100. doi:10.1097/PCC.0b013e3181e911a7. PMID 20639791.

[114] Gubler DJ Dengue and dengue hemorrhagic fever. Clin. Microbiol. Rev.1998; 11 (3): 480-496. PMC 88892. Freely accessible. PMID 9665979.

[115] World Health Organization. Dengue and severe dengue. WHO Factsheet No 117. Geneva. 2015; Avaiable from: http://www.who.int/mediacentre/factsheets/fs117/en/

[116] Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013; 496, 504-507.doi: 10.1038/nature12060. pmid:23563266

[117] Figueredo LTM. Dengue in Brazil : Past, Present and Future Perspectives. Dengue Bulletin. 2003; 27, 25-33.

[118] Fares RC, Souza KP, Añez G, Rios M. Epidemiological Scenario of Dengue in Brazil. Biomed research international. 2015

[119] Brazilian Ministry of Health. Epidemiological Bulletin (in Portuguese). 2016. Avaiable from: http://portalsaude.saude.gov.br/images/pdf/2016/abril/26/2016-014—Dengue-SE13-prelo.pdf

[120] World Health Organization. Weekly epidemiological record Dengue Vaccine : WHO position paper - July 2016. Avaiable from: http://www.who.int/wer/2016/wer9130.pdf?ua=1

[121] Rothman AL, Ennis FA. Dengue Vaccine : The Need, the Challenges and Progress. Journal of Infectious Diseases. 2016; jiw068.

[122] Pitisuttithum P, Bouckenooghe A. The first licensed dengue vaccine: an important tool for integrated preventive strategies against dengue virus infection. Expert review of vaccines.2016; Forthcoming

[123] Rabaa MA, Simmons CP, Fox A, Le MQ, Nguyen TTT, et al. Dengue virus in sub-tropical northern and central Viet Nam: Population immunity and climate shape patterns of viral invasion and maintenance. PLoS Negl Trop Dis. 2013; 7 (12) e2581 doi:10.1371/journal.pntd.0002581.

[124] Raghwani J, Rambaut A, Holmes EC, Hang VT, Hien TT, et al. Endemic dengue associated with the co-circulation of multiple viral lineages and localized density-dependent transmission. PLoS Pathog. 2011 7: e1002064.

[125] Adams B, Kapan DD. Man bites mosquito: understanding the contribution of human movement to vector-borne disease dynamics. PloS one.2009 4(8), e6763.

[126] Stoddard ST, Morrison AC, Vazquez-Prokopec GM, Paz Soldan V, Kochel TJ, et al. The role of human movement in the transmission of vector-borne pathogens PLoS Negl Trop Dis. 2009. 3: e481.

[127] Stolerman, LM, Coombs D, Boatto S. SIR-Network Model and Its Application to Dengue Fever. SIAM Journal on Applied Mathematics SIAM Journal on Applied Mathematics. 2015. 75(6), 2581-2609.

[128] Watts DM, Burke DS, Harrison BA, Whitmire RE, Nisalak A. Effect of temperature on the vector efficiency of Aedes aegypti for dengue 2 virus. Am J Trop Med Hyg. 1987. 36: 143–152.

[129] Yang HM, Macoris MLG, Galvani KC, Andrighetti MTM, Wanderley DMV. Assessing the effects of temperature on the population of Aedes aegypti, the vector of dengue Epidemiol Infect. 2009. 137: 1188–1202.

[130] Foo LC, Lim TW, Lee HL, Fang R. Rainfall, abundance of Aedes and dengue infection in Selangor, Malaysia Southeast Asian J Trop Med Pub Health. 1985. 16: 560–568.

[131] Honório NA, Castro MG, Barros FSM, Magalhães MAFM, Sabroza PC The spatial distribution of Aedes aegypti and Aedes albopictus in a transition zone, Rio de Janeiro, Brazil. Cad Saúde Pública. 2009. 25:1203–1214.

[132] Hopp MJ, Foley JA Global-scale relationships between climate and the Dengue fever vector, Aedes aegypti. Clim Change. 2001. 48: 441–463.

[133] Adde A, Roucou P, Mangeas M, Ardillon V, Desenclos J-C, Rousset D, et al. Predicting Dengue Fever Outbreaks in French Guiana Using Climate Indicators. PLOS Negl Trop Dis. 2016;10: e0004681. doi: 10.1371/journal.pntd.0004681. pmid:27128312

[134] Hii YL , Huaiping Zhu, Nawi Ng, Lee Ching Ng, Joacim Rocklöv. Forecast of Dengue Incidence Using Temperature and Rainfall PloS Negl Trop Dis. 2012;(11):e1908

[135] Buczak AL, Koshute PT, Babin SM, Feighner BH, Lewis SH. A data-driven epidemiological prediction method for dengue outbreaks using local and remote sensing data. BMC Medical Informatics & Decision Making 2012; 12:124.

[136] Buczak AL, Baugher B, Babin SM, Ramac-Thomas LC, Guven E, Elbert Y, et al. Prediction of high incidence of dengue in the Philippines. PLOS Neglected Tropical Diseases, 2014; 8:24:e2771.

[137] Hii YL , Huaiping Zhu, Nawi Ng, Lee Ching Ng, Joacim Rocklöv. Forecast of Dengue Incidence Using Temperature and Rainfall PloS Negl Trop Dis. 2012;(11):e1908

[138] Golub G, Kahan W Calculating the Singular Values and Pseudo-Inverse of a Matrix Journal SIAM Numerical Analysis. 1965; Series B, 2(2), 205-224.

[139] Kutz JN. Data-Driven Modeling & Scientific Computation: Methods for Complex Systems & Big Data. Oxford University Press; 2013.

[140] Cortes C, Vapnik V. Support-vector networks. Machine learning.1995. 20(3), 273-297.

[141] Burges CJ. A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery 2.2 .1998. 121-167.

[142] Candés EJ and Wakin MB. An introduction to compressive sampling. IEEE signal processing magazine 25.2 (2008): 21-30.

[143] Gemmeke JF, Van Hamme H, Cranen B, Boves L. Compressive sensing for missing data imputation in noise robust speech recognition. IEEE Journal of selected topics in Signal Processing 4.2 (2010): 272-287.

[144] Stankovic L, Stankovic S, and Amin M. Missing samples analysis in signals for applications to L-estimation and compressive sensing. Signal Processing 94 (2014): 401-408.

[145] Zhang Y When is missing data recoverable?. Technical Report, 2006.

[146] Brazilian National Institute of Meteorology (INMET) Temperature and precipitation time series. Avaiable from (website in Portuguese): http://www.inmet.gov.br/portal/

[147] Brazilian National Surveillance System (SINAN) Total number of Dengue cases in state capitals Avaiable from (website in Portuguese): http://portalsinan.saude.gov.br/dados-epidemiologicos-sinan

[148] Health department, city Hall. Cases of Dengue registered in Rio de Janeiro. Avaiable from (website in Portuguese): http://www.rio.rj.gov.br/web/sms/dengue

[149] Böhm AW, Costa CDS, Neves RG, Flores TR, Nunes BP. Dengue incidence trend in Brazil, 2002-2012. Epidemiologia e Serviços de Saúde.2016; 25(4), 725-733.

[150] Alert system of rain events, city Hall. Precipitation time series of Rio de Janeiro. Avaiable from (website in Portuguese): http://alertario.rio.rj.gov.br/

[151] Liao CM, Huang TL, Lin YJ, You SH, Cheng YH, Hsieh NH, hen WY Regional response of dengue fever epidemics to interannual variation and related climate variability. Stochastic Environmental Research and Risk Assessment. 2015; 29(3), 947-958.

[152] Johansson MA, Dominici F, Glass GE Local and global effects of climate on dengue transmission in Puerto Rico. PLoS Negl Trop Dis. 2009; 3(2), e382.

[153] Pessanha JEMP, Caiaffa WT, Almeida MCDM, Brandao ST, Proietti FA Diffusion pattern and hotspot detection of dengue in Belo Horizonte, Minas Gerais, Brazil. .Journal of tropical medicine. 2012.

[154] Honório, NA, Codeço CT, Alves FC, Magalhães MDA, Lourenço-de-Oliveira R. Temporal distribution of Aedes aegypti in different districts of Rio de Janeiro, Brazil,

measured by two types of traps. Journal of Medical Entomology. 2009; 46(5), 1001-1014.

[155] Dibo MR, Chierotti AP, Ferrari MS, Mendonça AL, Chiaravalloti Neto, F. Study of the relationship between Aedes (Stegomyia) aegypti egg and adult densities, dengue fever and climate in Mirassol, state of São Paulo, Brazil. Memorias do Instituto Oswaldo Cruz. 2008; 103(6), 554-560.

[156] Brazilian Ministry of Health. Promotion of national mobilization effort against Aedes Aegypti in 2013 (in Portuguese). Avaiable from: http://www.brasil.gov.br/saude/2013/11/governo-lanca-nova-campanha-de-mobilizacao-contra-dengue

[157] Brazilian Ministry of Health. Promotion of national mobilization effort against Aedes Aegypti in 2016 (in Portuguese). Avaiable from: http://www.brasil.gov.br/governo/2016/02/dilma-visita-rio-de-janeiro-no-dia-nacional-de-mobilizacao-zika-zero

[158] Brazilian Ministry of Health. Promotion of national mobilization effort against Aedes Aegypti for 2017. Avaiable from: http://www.brazilgovnews.gov.br/news/2016/11/government-promotes-national-mobilisation-effort-against-aedes-aegypti

[159] Lowe R, Bailey T, Stephenson D, Jupp T, Graham R, Coelho CA et al. Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. Computers & Geosciences. 2011; 37(3), 371-381.

[160] Lowe R, Bailey T, Stephenson D, Jupp T, Graham R, et al. The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. Statist Med. 2012; 32: 864–883. doi: 10.1002/sim.5549

[161] Racloz V, Ramsey R, Tong S, Hu W. Surveillance of dengue fever virus: a review of epidemiological models and early warning systems. PLoS Negl Trop Dis. 2012; 6(5), e1648.

[162] Johansson MA, Reich NG, Hota A, Brownstein JS, Santillana M. Evaluating the performance of infectious disease forecasts: A comparison of climate-driven and seasonal dengue forecasts for Mexico. Nature. 2016; Scientific Reports, 6.

[163] Díaz FJ, Black WC, Farfán-Ale JA, Loroño-Pino MA, Olson KE, Beaty BJ. Dengue virus circulation and evolution in Mexico: a phylogenetic perspective. Archives of medical research. 2006; 37(6), 760-773.

[164] Adams B, Holmes EC, Zhang C, Mammen MP, Nimmannitya S, Kalayanarooj S, Boots M. Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in Bangkok. Proceedings of the National Academy of Sciences. 2006; 103(38), 14234-14239.

[165] Wesolowski A, Qureshi T, Boni MF, Sundsøy PR, Johansson MA, Rasheed SB, et al. Impact of human mobility on the emergence of dengue epidemics in Pakistan. Proceedings of the National Academy of Sciences. 2015; 112(38), 11887-11892.

[166] Barmak DH, Dorso CO, Otero M, Solari HG. Dengue epidemics and human mobility. Physical Review E. 2011; 84(1), 011901.

[167] Harrington LC, Scott TW, Lerdthusnee K, Coleman RC, Costero A, Clark GG, et al. Dispersal of the dengue vector Aedes aegypti within and between rural communities. The American journal of tropical medicine and hygiene. 2005; 72(2), 209-220.

[168] de Castro Medeiros LC, Castilho CAR, Braga C, de Souza W V, Regis L, Monteiro AMV. Modeling the dynamic transmission of dengue fever: investigating disease persistence. PLoS Negl Trop Dis. 2011; 5(1), e942.

[169] Cummings DA, Iamsirithawor S, Lessler JT, McDermott A, Prasanthong R, Nisalak A, et al. The impact of the demographic transition on dengue in Thailand: insights from a statistical analysis and mathematical modeling. PLoS Med.2009; 6(9), e1000139.

[170] Mondini A, Chiaravalloti-Neto F. Spatial correlation of incidence of dengue with socioeconomic, demographic and environmental variables in a Brazilian city. Science of the Total Environment. 2008; 393(2), 241-248.

[171] Johansson MA, Cummings DA, Glass GE. Multi year climate variability and dengue—El Nino southern oscillation, weather, and dengue incidence in Puerto Rico, Mexico, and Thailand: a longitudinal data analysis. PLoS Med.2009; 6(11), e1000168.

[172] Cazelles B, Chavez M, McMichael AJ, Hales S. Nonstationary influence of El Nino on the synchronous dengue epidemics in Thailand. PLoS Med. 2005; 2(4), e106.

[173] Banu S, Guo Y, Hu W, Dale P, Mackenzie JS, Mengersen K, et al. Impacts of El Niño Southern Oscillation and Indian Ocean Dipole on dengue incidence in Bangladesh. Scientific report.2015; 5.

[174] Naish S, Dale P, Mackenzie JS, McBride J, Mengersen K, Tong S. Climate change and dengue: a critical and systematic review of quantitative modelling approaches. BMC infectious diseases. 2014; 14(1), 1.