

GREGORY MCDONALD

**MEASURES OF DYNAMICAL CHANGE:  
DYNAMICAL SIMILARITY  
AND  
INFORMATION FLOW**

Mémoire  
présenté  
à la Faculté des études supérieures  
de l'Université Laval  
pour l'obtention  
du grade de maître ès sciences (M.Sc.)

Département de physique, de génie physique et d'optique  
FACULTÉ DES SCIENCES ET DE GÉNIE  
UNIVERSITÉ LAVAL

Avril 2003

# Résumé

Nous avons étudié deux mesures de changement dynamique sur des séquences temporelles, la similarité dynamique (DS) et le flot d'information intégré à base de cumulants (IIF). Des séquences temporelles synthétiques de contrôle ont été produites avec un réseau neuronal asymétrique à délai. Le pouvoir discriminatoire entre différents niveaux de chaoticité du IIF s'est avéré plus fort sur ces séquences. Ce sont les fluctuations statistiques de la DS, par contre, qui étaient les moins importantes sur des données stationnaires. L'influence de bruit gaussien additif sur les deux mesures était négligeable. Nous avons calculé les deux mesures sur toutes les voies de 25 enregistrements EEG de 5 patients souffrant d'épilepsie focale. Des fluctuations significatives préictales de la DS ont été observées pour 8 crises cliniques sur 15 versus 4 sur 15 pour le IIF. Toute crise anticipée par ces fluctuations l'était aussi par des changements spectraux précédant la crise.

---

Greg McDonald  
Candidat

---

Louis-J. Dubé  
Directeur de recherche

# Abstract

Two methods of measuring dynamical change on time series, dynamical similarity (DS) and cumulant-based integrated information flow (IIF), were studied. To compare the measures under controlled conditions, artificial time series of chaotic data qualitatively similar to EEGs were generated using an asymmetrical neural network with delays. IIF was found to discriminate different degrees of chaos better than DS on these data. On stationary data, however, DS fluctuations were smaller. Additive noise had little effect on the performance of either measure. The measures were subsequently computed on all channels of 25 electroencephalographic (EEG) recordings from 5 patients with focal epilepsy. Significant preictal changes in DS were observed before 8 out of 15 clinical seizures versus 4/15 for IIF. Seizures anticipated by these measures could also be anticipated by changes in the power spectral density.

---

Greg McDonald  
Candidate

---

Louis-J. Dubé  
Thesis Advisor

# Acknowledgements

I would first like to express my gratitude to my advisor, Prof. Louis-J. Dubé, for his instruction, guidance and contagious optimism. We have been on opposite sides of the Atlantic for most of the past two years, but he has always managed to remain “present” and actively involved in my project. For setting the bar high, giving me the opportunity to explore my topic freely, and reminding me at just the right times of the playful aspect of our work, I thank him.

Financial support came in large part from a Masters level Research Scholarship generously provided by the FCAR (*Fonds pour la Formation de chercheurs et l'aide à la recherche*). This agency's contribution to the training of new scientists is vitally important and is gratefully acknowledged.

After a morning spent hunting for bugs in my code, trying in vain to find that “magic” threshold that might turn my merely interesting results into truly spectacular ones or tweaking figures, I was so glad to share a few laughs over a coffee with Simon Frédérick. Over these two years and more, my days at Laval have been enriched by the companionship of some fine people: David, Louis-Philippe, Gabrielle, Louis, Eric, to name just a few. Conversations about matters both technical and theoretical with Bernard Doyon and Philippe Després as well as code written by Phil and Frédéric Beaulieu were of great help.

Chapter 2 was the first one I wrote. It was completed in New York City during an “academic retreat” in November, 2002. A special thank-you to Jennifer Milligan and my great friend David Hogg for being wonderful hosts.

My parents, Anne Marie and Gary, taught me to believe in myself and nourished

my interests as a boy with such things as books and computers (let's hear it for the Commodore 64!) and intangibles like trust, as I stayed late at school to attend and later run Science Club meetings. I wish to thank them for that and for ensuring that I had the best possible education available to me. My parents and my wife's, my brother Bob, other family and friends have all been wonderfully positive about my return to university and my pursuit of a career in physics. At times, I am amazed—and always flattered—by the interest they manifest in my studies and research.

Lastly, I could never have done this without the encouragement and support of my wife, Marie-Lou Bois. It is hard to express how much I appreciate her efforts to balance the demands of her own medical residency, the care of our sons Benjamin and Nathan and my need for large blocks of time free of interruption to bring this project to fruition. When the work seems to drag along, she is always able to restore my confidence and get me back on track. After all, what could motivate a guy more than being told he's sexy when he talks physics! *Merci, ma belle résidente.*

# Contents

<b>Résumé</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>Table of Contents</b>	<b>v</b>
<b>List of Tables</b>	<b>viii</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Acronyms</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Dynamical Systems and Dynamical Change . . . . .	1
1.2 Nonlinear Time Series Analysis . . . . .	3
1.3 Anticipation of Epileptic Seizures . . . . .	4
<b>2 Data Sets</b>	<b>8</b>
2.1 Bondarenko’s “Synthetic Brain” . . . . .	8
2.1.1 Artificial Neural Networks with Chaotic Solutions . . . . .	9
2.1.2 Implementation . . . . .	13
2.1.3 Bondarenko Data Sets . . . . .	19

2.1.4	Characterizing the Bondarenko Data . . . . .	22
2.1.4.1	Maximal Lyapunov Exponent . . . . .	23
2.1.4.2	Correlation Dimension . . . . .	26
2.2	EEG Recordings . . . . .	29
2.2.1	Electroencephalography . . . . .	29
2.2.2	Description of the Data Set . . . . .	30
2.2.3	Amplitude-Frequency Features . . . . .	31
<b>3</b>	<b>Dynamical Similarity</b>	<b>38</b>
3.1	Defining the Measure . . . . .	39
3.1.1	Phase Space Reconstruction . . . . .	40
3.1.2	Finding the Optimal Basis: Singular Value Decomposition . . . . .	41
3.1.3	Similarity Index . . . . .	43
3.1.4	Statistical Analysis . . . . .	45
3.2	Experiments on Artificial Data . . . . .	46
3.2.1	The $c$ Series . . . . .	48
3.2.1.1	Methods . . . . .	48
3.2.1.2	Results and Discussion . . . . .	48
3.2.2	The $e$ Series . . . . .	52
3.2.2.1	Methods . . . . .	52
3.2.2.2	Results and Discussion . . . . .	52
3.3	Experiments on EEG Recordings . . . . .	55
3.3.1	Methods . . . . .	58
3.3.2	Results and Discussion . . . . .	62
<b>4</b>	<b>Information Flow</b>	<b>80</b>
4.1	A Formal Measure of Information Flow . . . . .	80
4.2	Cumulant-based Information Flow . . . . .	84
4.3	Surrogate Data . . . . .	88

4.4	Experiments on Artificial Data . . . . .	90
4.4.1	Methods . . . . .	91
4.4.2	Results and Discussion . . . . .	91
4.4.2.1	The $c$ Series . . . . .	91
4.4.2.2	The $e$ Series . . . . .	98
4.5	Experiments on EEG Recordings . . . . .	101
4.5.1	Methods . . . . .	101
4.5.2	Results and Discussion . . . . .	102
4.6	Cross-Channel Information Flow . . . . .	115
<b>5</b>	<b>Closing Discussion and Conclusion</b>	<b>118</b>
5.1	Experiments with Artificial Data . . . . .	119
5.2	Seizure Anticipation from EEG Recordings . . . . .	121
5.3	Conclusion . . . . .	124
	<b>Bibliography</b>	<b>126</b>
<b>A</b>	<b>Elements of Information Theory</b>	<b>134</b>
A.1	Introduction . . . . .	134
A.2	Information and the Entropies . . . . .	135
<b>B</b>	<b>Elements of Probability Theory</b>	<b>138</b>
B.1	Descriptions of a Probability Distribution . . . . .	138
B.2	Expressions of Statistical Independence . . . . .	140
<b>C</b>	<b>Coupling Matrix</b>	<b>142</b>
<b>D</b>	<b>Gap Cataloguing Algorithm</b>	<b>148</b>



# List of Tables

2.1	User-configurable model parameters in <code>Bondarenko</code> class and their default values. . . . .	16
2.2	Periodogram results described. . . . .	34
3.1	Anticipation and free hot zone results on intracranial EEG channels (“Type 2”) using the first reference window, $\sigma_{th} = 5$ , $\tau_{th} = 120$ s and $\tau_g = 0$ . . . . .	69
3.2	Anticipation and free hot zone results on surface EEG channels (“Type 1”) using the first reference window, $\sigma_{th} = 5$ , $\tau_{th} = 120$ s and $\tau_g = 0$ . . . . .	70
3.3	Anticipation and free hot zone results on intracranial EEG channels (“Type 2”) using the first baseline, $\sigma_{th} = 5$ , $\tau_{th} = 120$ s and $\tau_{grace} = 120$ s. . . . .	71
4.1	Positive slope statistics on recordings with seizures. . . . .	108
5.1	Summary of seizure anticipation results. . . . .	123

# List of Figures

1.1	Example of an interictal EEG recording. . . . .	5
1.2	Example of an EEG recording of a clinical seizure. . . . .	6
2.1	Code fragment illustrating features of <code>Bondarenko</code> class. . . . .	17
2.2	Typical output for each of the values of $c$ in the $c$ set. . . . .	20
2.3	Typical output for each of the values of $e$ in the $e$ set. . . . .	21
2.4	Power spectrum of Bondarenko model chaotic output ( $c = 1.0$ ). . . . .	22
2.5	Maximal Lyapunov exponents on artificial TSs vs. $c$ . . . . .	24
2.6	Correlation sum results for each value of $c$ : short TSs. . . . .	28
2.7	Correlation sum results for each value of $c$ : long TSs. . . . .	29
2.8	Schematic representation of major features of the EEG data set. . . . .	32
2.9	Periodogram for intracranial channel in <i>amysez01</i> : preictal increase at high frequencies. . . . .	35
2.10	Periodogram for intracranial channel in <i>bobsez02</i> : rapid preictal fluctuations in intensity. . . . .	35
2.11	Periodogram for intracranial channel in <i>eveawake</i> : highly nonstationary. . . . .	36
2.12	Periodogram for intracranial channel in <i>carlsleep</i> : slow variation. . . . .	36
2.13	Standard deviation of EEG signals on sliding windows . . . . .	37
3.1	Illustration of phase space reconstruction with ICIs. . . . .	41
3.2	Illustration of influence of “grace period” on anticipation times. . . . .	47
3.3	Similarity results on the $c$ series with $\ell_w = 1024$ and $\Delta t = 1.0$ . . . . .	49

3.4	Similarity results on the $c$ series with $\ell_w = 1024$ and $\Delta t = 10.0$ . . . .	50
3.5	Similarity results on the $c$ series with $\ell_w = 8192$ and $\Delta t = 1.0$ . . . .	50
3.6	Similarity results on the $c$ series with $\ell_w = 8192$ and $\Delta t = 10.0$ . . . .	51
3.7	Similarity results on the $e$ series with $\ell_w = 1024$ and $\Delta t = 1.0$ . . . .	53
3.8	Similarity results on the $e$ series with $\ell_w = 1024$ and $\Delta t = 10.0$ . . . .	54
3.9	Similarity results on the $e$ series with $\ell_w = 8192$ and $\Delta t = 1.0$ . . . .	54
3.10	Similarity results on the $e$ series with $\ell_w = 8192$ and $\Delta t = 10.0$ . . . .	55
3.11	Similarity “hot zones” on EEG recordings: $\sigma_{th} = 5$ , $\tau_{th} = 2$ min, $\tau_g = 0$ , first RW. . . . .	65
3.12	Detail from Fig. 3.11: <i>amysez01</i> . . . . .	67
3.13	Detail from Fig. 3.11: <i>carlsez01</i> . . . . .	68
3.14	Similarity “hot zones” on EEG recordings: $\sigma_{th} = 5$ , $\tau_{th} = 2$ min, $\tau_g = 0$ , second RW. . . . .	73
3.15	Detail from Fig. 3.14: <i>carlsez03</i> . . . . .	74
3.16	Similarity “hot zones” on EEG recordings: $\sigma_{th} = 5$ , $\tau_{th} = 2$ min, $\tau_g = 0$ , first RW, cross-recording results for <i>amy</i> . . . . .	75
3.17	Similarity “hot zones” on EEG recordings: $\sigma_{th} = 5$ , $\tau_{th} = 2$ min, $\tau_g = 0$ , first RW, cross-recording results for <i>bob</i> . . . . .	77
3.18	Ratio of threshold crossings in TWs to those in the first RW over first intracranial TS of each recording. . . . .	79
4.1	IIF on $c$ series for $\ell_w = 1024$ and $\Delta t = 1.0$ . . . . .	92
4.2	IIF results on the $c$ series with $\ell_w = 1024$ and $\Delta t = 10.0$ . . . . .	94
4.3	IIF results on the $c$ series with $\ell_w = 8192$ and $\Delta t = 1.0$ with a lag of $\tau = 10$ . . . . .	95
4.4	IIF results on the $c$ series with $\ell_w = 8192$ and $\Delta t = 1.0$ . . . . .	96
4.5	IIF results on the $c$ series with $\ell_w = 8192$ and $\Delta t = 10.0$ . . . . .	97
4.6	IIF results on the $e$ series with $\ell_w = 1024$ and $\Delta t = 1.0$ . . . . .	99
4.7	IIF results on the $e$ series with $\ell_w = 8192$ and $\Delta t = 1.0$ . . . . .	99

4.8	IIF results on the $e$ series with $\ell_w = 1024$ and $\Delta t = 10.0$ . . . . .	100
4.9	IIF results on the $e$ series with $\ell_w = 8192$ and $\Delta t = 10.0$ . . . . .	100
4.10	Strip chart at full scale and log scale of $m_{\text{avg}}(10, 1)$ on channel $LA1$ of $amysez01$ . . . . .	103
4.11	Rescaled IIF results— $m_{\text{avg}}(10, 1)$ , as well as $m_2$ , $m_3$ and $m_4$ —for channel $LA1$ of $amysez01$ . . . . .	104
4.12	Rescaled IIF results for channel $LA3$ of $amysez01$ with surrogates. . .	105
4.13	Rescaled IIF results for channel $LA3$ of $bobsez01$ with surrogates. . .	106
4.14	IIF “hot zones” on EEG recordings: absolute thresholds in $m_{\text{avg}}$ chosen for each patient. . . . .	109
4.15	Percentile rank of baseline standard deviation values $\sigma$ for $m_{\text{avg}}$ for each patient’s recordings. . . . .	111
4.16	Percentile rank of baseline average values $\mu$ for $m_{\text{avg}}$ for each patient’s recordings. . . . .	112
4.17	IIF “hot zones” on EEG recordings: deviations above baseline mean. . . . .	113
4.18	Normal IIF and cross-channel information flow computed on artificial time series generated with coupled Bondarenko models, “brains”, of varying degrees of chaoticity. . . . .	116
C.1	Connection eigenvalues of Bondarenko coupling matrix colour coded by solution type . . . . .	146
C.2	Solution type for each coupling matrix and initial condition combination. . . . .	147

# List of Acronyms

<b>AP</b>	action potential
<b>atu</b>	arbitrary time units
<b>CND</b>	cumulative neighbour distribution
<b>DDE</b>	delay differential equation
<b>DS</b>	dynamical similarity
<b>EEG</b>	electroencephalographic
<b>FFT</b>	fast Fourier transform
<b>GPL</b>	General Public License
<b>ICI</b>	intercrossing interval
<b>IID</b>	independent identically-distributed
<b>IIF</b>	integrated information flow
<b>ISI</b>	interspike interval
<b>ODE</b>	ordinary differential equations
<b>PDF</b>	probability density function
<b>PGM</b>	Portable Grey Map
<b>RC</b>	resistance-capacitance
<b>RK4</b>	fourth-order Runge-Kutta
<b>RNG</b>	random number generator
<b>RW</b>	reference window
<b>SD</b>	standard deviation
<b>SVD</b>	singular value decomposition
<b>TLE</b>	temporal lobe epilepsy

<b>TS</b>	time series
<b>TW</b>	test window
<b>XIIF</b>	cross-channel integrated information flow

# Chapter 1

## Introduction

### 1.1 Dynamical Systems and Dynamical Change

A dynamical system is one whose time evolution is defined in some *phase space*, a vector space every point of which corresponds to a possible state of the system [37]. *Deterministic* systems are those whose time evolution is defined by a set of equations free of any stochastic or random terms (dynamical noise). Given any initial position in the phase space, those equations can in principle be solved to yield the system's subsequent *trajectory* or orbit in phase space [60].

Our topic is measures of dynamical change. It is important to understand what is meant by dynamical change and how this might differ from the idea of nonstationarity. A more formal definition of a dynamical system [30] will clarify the difference. We let  $\mathbf{x} = \mathbf{x}(t) \in \mathbb{R}^n$ , a vector valued function of the independent variable  $t$ , denote a point in an  $n$ -dimensional phase space corresponding to the state of the system at time  $t$ . The evolution of continuous-time dynamical systems is commonly defined by a set of first order differential equations

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, t), \tag{1.1}$$

where the vector field  $\mathbf{F} : U + \mathbb{R} \rightarrow \mathbb{R}^n$  is a smooth function defined on some subset  $U \subseteq \mathbb{R}^n$  of the phase space. The vector field  $\mathbf{F}$  is said to generate a *flow*  $\phi_t : U \rightarrow \mathbb{R}^n$  in phase space.

The behaviour of a system can be understood in terms of the geometrical and topological properties of this flow, which typically depends on the values of one or more parameters and may depend explicitly on time. Guckenheimer and Holmes take this approach in their classic text on dynamical systems theory [30].

Two types of dynamical change can be identified: shifting of the flow itself as a result of changes in the system parameters and movement between distinct regions of a *globally stable flow*. The first type involves change in the equations of motion of the system via the control parameters or, if the equations are nonautonomous, the simple passage of time deforming the flow. In the second type of change, the equations describing the system are unaltered. Movement between multiple attractors of a deterministic chaotic system as a result of perturbations to the system is *not* an example of the latter; the “external force” which perturbs the system is not present in the system initially. A proper example is provided by systems displaying intermittency, i.e. chaotic behaviour interspersed with bursts of near periodicity of varying duration. At a given observational time scale, there may be no way of determining which type of dynamical change a system is undergoing.

A random process is said to be *strict-sense stationary* if its statistical properties are invariant to a shift in the origin of time. *Wide sense stationarity* requires a constant mean and that the autocorrelation depend only on a lag [62]. Stationarity is a precondition of most methods of time series analysis, linear and nonlinear [37]. When this condition cannot be satisfied on an entire series, the method of sliding windows is often used; both measures presented in this thesis use this approach. Within any given window of data, the system is *assumed* to be stationary. This imposes significant constraints on the choice of window length and hence on the statistics of these measures.

A system undergoing dynamical changes of *either* type identified above is at least strict-sense *nonstationary*. The measures we examine here either began life as measures of nonstationarity (similarity) or have been used to *detect* stationarity (infor-



mation flow). As measures of dynamical change, however, we expect more from these measures; we want measures which reflect the *extent* of differences in the flow or between different regions of the flow. This is satisfied by their theoretical underpinning, as the similarity measure explicitly compares reconstructions of the flow and the information flow measure does so implicitly by estimating cumulants of the distribution of states on the flow in any given window.

## 1.2 Nonlinear Time Series Analysis

Experimentally, our access to a system’s dynamics, i.e. the equations governing its evolution in time, is through the portal of measurements. When these measurements are repeated at regular intervals, they form a time series. From a dynamicist’s point of view, this vector of data encodes the entire dynamics of the system—often distorted by some measurement function and invariably contaminated by noise—so that in principal this dynamics can be recovered by the embedding process [74, 80]. This reconstructed phase space is the starting point of most methods of nonlinear time series analysis. It is explicitly so in the dynamical similarity measure we study in Chapter 3 and implicitly in the cumulant-based information flow (Chapter 4). These methods have grown out of the explosion of interest in nonlinear systems, in particular those displaying deterministic chaos. It is the presence of nonlinear terms in the equations of motion which make these systems nonlinear. Additional nonlinearities are often introduced by the measurement function. Special methods, such as those described later involving surrogate data, are required to separate these measurement nonlinearities from those of the underlying system.

We will present two measures of dynamical change, both of which are sensitive to changes in both linear and nonlinear structure underlying time series. First, we examine dynamical similarity, which compares clouds of points in reconstructed phase space to the cloud representing a reference dynamics. Statistical independence between the past and future of a process are reflected in the cumulant-based integrated

information flow, the second measure we consider. Both are *measures* in that the computed quantities allow us to *gauge* the extent of the change rather than just detect whether some change has occurred. In principle, thresholds on the measure and in time can be used to detect significant dynamical changes, but we will see that this is not always easy to implement.

Before applying these measures to real data, experiments on artificial time series allow us to explore the strengths and weaknesses of each measure. An asymmetrically connected Hopfield neural network with delay was used to generate our synthetic time series. In a series of papers by V. Bondarenko, this model was shown to produce solutions qualitatively similar in many respects to EEG recordings.

Throughout the following text, considerable attention is paid to implementation details. Much of what we have to present involves numerical experiments on artificial time series, computation of various measures on real data and automatic data reduction. Algorithmic ambiguity and differences in implementation could have a strong influence on results, so in the interest of reproducibility, we have tried to be as explicit and clear as possible.

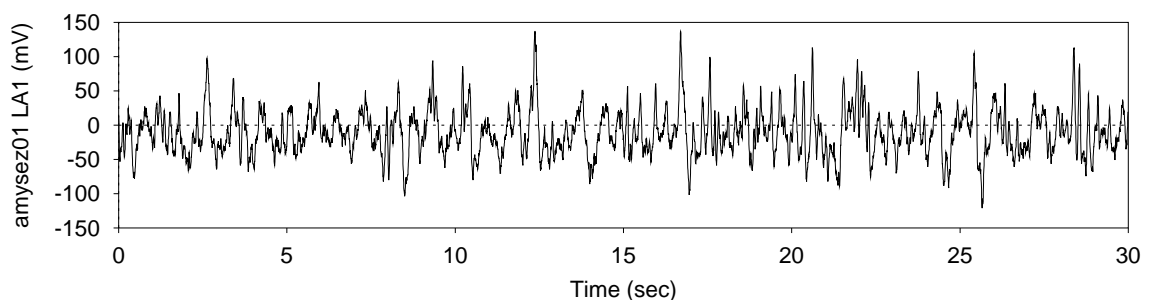
### 1.3 Anticipation of Epileptic Seizures

One direct result of the recent period of intense research into nonlinear systems and methods of nonlinear time series analysis has been a renewal of interest in the problem of epileptic seizure anticipation from EEG recordings [46, 48, 52]. The goal is to identify, from these signals, changes in the brain dynamics which eventually lead to a seizure. Timely detection of these changes could eventually make it possible to steer the brain away from the seizure by weak pharmacological or electrical perturbation of the system. At the very least, patients would be aware of the upcoming seizure, which would greatly alleviate the anxiety of never knowing when a seizure will strike.

Epilepsy is one of the most common neurological diseases, affecting about 1% of the population at some time in their life. It is the propensity to have *seizures*, defined

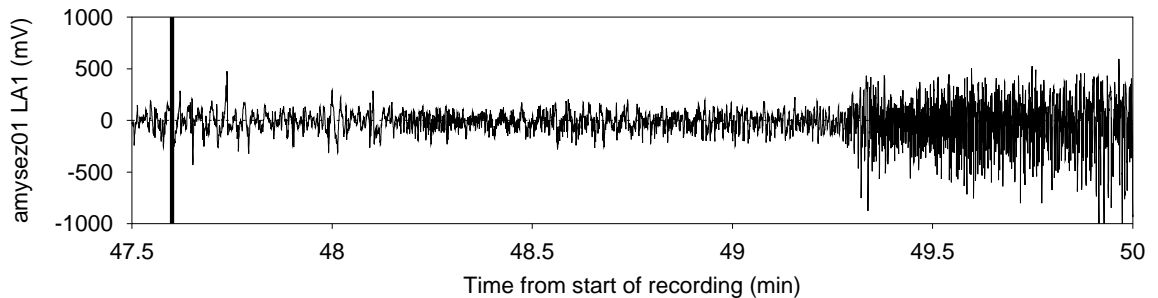
as “transient paroxysms of excessive discharges of neurones” in the brain “causing a clinically discernable event”. If this discharging is only short-lived, it will produce an *interictal spike* in brain activity, observable on the EEG and characteristic of an epileptic brain. In Fig. 1.1, we show a 30 sec segment of EEG recording far from any seizures. Fig. 1.2, on the other hand, shows the onset of a clinical seizure about 45 min later in the same intracranial recording. The transition is fast and the violence of the synchronous discharging that characterizes a seizure is astonishing. Although these discharges are reasonably well understood on a cellular level, the larger-scale mechanisms producing and maintaining seizures have still not been fully elucidated [25]. Doing so is a concurrent goal of those working on seizure anticipation, as the measures being used have the potential to shed light on seizure dynamics. Conversely, a better understanding of these dynamics may allow methods of time series analysis to be tailored to seizure anticipation or suggest promising combinations of measures.

Figure 1.1: This 30 second segment of intracranial EEG recording typifies interictal (between seizure) behaviour. Seemingly random fluctuations are superimposed on slower, larger amplitude oscillations. This particular example was taken from the start of *amysez01*, channel *LA1*.



Again from the dynamicist’s point of view, EEG recordings are assumed to capture the “dynamics of coupled nonlinear interactions between neuronal populations.” [45] The methods of nonlinear time series analysis applied to these signals should thus allow characterization of the underlying dynamical processes and detection of patterns

Figure 1.2: The line running top to bottom in this figure indicates the clinician identified start of a clinical seizure, the first in *amysez01*, channel *LA1*. Prior to that, the recording looks like Fig. 1.1. A characteristic 10 Hz oscillation develops at the onset of the seizure, which we see as a “darkening” of the plot at this time scale. This is followed by the high amplitude fluctuations.



that would not be revealed by traditional linear methods. The situation is greatly complicated in reality by the simultaneous presence of noise, measurement nonlinearities and stochastic processes which may be operating in parallel with deterministic processes. A thorough review of work on the problem of characterizing neurodynamic changes before seizures and seizure anticipation has been written by Le Van Quyen et al. [45]. Anticipation times reported with nonlinear measures have been on the order of 5 min. More recently, significant changes in measures of neuronal synchrony have indicated preictal periods extending 20 min prior to seizures.

On a data set consisting of 25 long multi-channel EEG recordings of intracranial and scalp electrodes on five patients with epilepsy, we have studied the variations in the similarity and information flow measures preceding clinical seizures. Particular attention has been paid to the *specificity* of these variations to the so-called “preictal period”, contrary to almost all existing publications concerning this problem. We compare the performance of the two measures against each other but also against *linear* methods of time series analysis. The question of interest is whether the strengths of these two measures, which are motivated by very different theoretical considerations, are complementary and whether they do reveal pre-seizure dynamical changes that linear measures miss. A positive answer would suggest that the measures be

combined to yield a “better” (higher specificity and sensitivity to preictal dynamical changes) integrated measure for anticipating epileptic seizures.

## Chapter 2

# Data Sets: Artificial and Electroencephalographic

### 2.1 Artificial Time Series: Bondarenko’s “Synthetic Brain”

One of the goals of this thesis is to compare two measures of dynamical change motivated by different theoretical considerations. In order to make this comparison, we need a data set consisting of long time series in which known, quantifiable dynamical changes take place. After all, before any instrument or measure can be effectively used to probe complex systems, it is vital to establish its response—dynamic range, sensitivity and precision—to *known* inputs.

The second objective of our research is to assess the influence of combining multiple, complementary measures on sensitivity and specificity to preictal changes in EEG signals, so it would be preferable to use artificial data that at least *looks* like an EEG. Asymmetric artificial neural networks with delay have been found by Bondarenko to produce “chaos similar to the human EEG” [8] and even “seizures” [11]. This model was retained to generate control data. In the following section, we present the model and describe the data sets produced with it, including some of their nonlinear dynamical characteristics.

### 2.1.1 Artificial Neural Networks with Chaotic Solutions

In the last two decades, a number of researchers have reported evidence of low dimensional chaos in electroencephalographic recordings of human and other animal brains (for example, [3] and [64]). The application of nonlinear methods in some of this work was criticized for lack of rigor and adequate testing against suitable surrogates [81]. Today, many of the pitfalls of nonlinear time series (TS) analysis are much better understood and excellent texts are available to guide the uninitiated [37]. Robust, well-designed implementations of key algorithms are also available, including the public domain Time-Series Analysis (TISEAN) package [38]. But once bitten, twice shy; scientists remain wary of strong claims of deterministic chaos in the brain. The role such a phenomenon might play in the brain continues to be debated [47].

Nevertheless, it is fair to say that some of the interest in chaos in artificial neural networks was motivated by these findings in the brain [8, 34]. Initially, the existence of chaotic solutions was demonstrated in neural network models both with [70] and without time delay [41]. Values obtained for the maximal Lyapunov exponent and various dimensions (fractal, information, etc.) quantified the low-dimensional chaos in single neurons<sup>1</sup> and neural network models [2, 16, 20, 36, 57, 58, 77]. Estimation of these measures is notoriously data intensive, requiring long, relatively noise-free, stationary TS. Unlike the brain, of course, there is no inherent limitation in TS length from computer and electronic circuit simulations of neural networks.

The neural network model investigated by Bondarenko [8–11] is a generalization of the Hopfield analog neural network [33]. The latter consists of a system of highly interconnected neurons with a graded response function (a “sigmoid input-output relation”), which we designate  $f(x)$ . It was argued that the response of *real* neurons and physical devices such as operational amplifiers which might be used in

---

<sup>1</sup>Unless otherwise noted, neurons in this chapter are of the *artificial* variety. These model networks can represent numerous nonneuronal, nonbiological systems, but the neuronal paradigm is traditional and appealing: we all have a brain. That being said, we can dispense with the onerous use of enclosing quotation marks: “neurons”.

electronic circuit implementations of these networks is better modelled by a sigmoid than a simple two-state output. The equation as given by Hopfield is an explicit resistance-capacitance (RC) charging equation, but here we present a dimensionless form obtained by rescaling the time to units of the network relaxation time [51]. Let  $u_i(t)$  denote the state of neuron  $i$  at time  $t$ . A network consisting of  $M$  neurons is described by the system of ordinary differential equations

$$\dot{u}_i(t) = -u_i(t) + \sum_{j=1}^M a_{ij} f(u_j(t)) + e_i(t), \quad i = 1, \dots, M \quad (2.1)$$

where the  $a_{ij}$  are elements of the  $M$ -square matrix of coupling coefficients,  $a_{ij} f(u_j(t))$  represents the input to neuron  $i$  due to the current state of neuron  $j$  and all external inputs are combined in  $e_i(t)$ . Hopfield shows that the collective computational properties of this model are like those found in earlier studies of stochastic models and consequently that functioning circuits with these useful properties can be built out of real components. Moreover, the results add strength to the idea that similar properties may somehow operate in the brain.

Signal propagation between neurons, be they real or electronic, is not instantaneous. Hopfield writes that the stochastic model whose significant behaviours are reproduced in his continuous, deterministic model employs an *asynchronous* algorithm. This means that each neuron is interrogated and its output state updated at random times and independently of the other neurons; the average rate of interrogation is the same for all neurons. Propagation delays in real systems, among other things, are represented by this asynchrony. In omitting these delays from the continuous model, one assumes that their time scale is considerably shorter than the integration time of the network. However, to examine the effects of longer delays on network stability, Marcus and Westervelt introduce an explicit delay  $\tau$  in the response [51] to obtain the following system of delay differential equations (DDEs):

$$\dot{u}_i(t) = -u_i(t) + \sum_{j=1}^M a_{ij} f(u_j(t - \tau)) + e_i(t), \quad i = 1, \dots, M. \quad (2.2)$$



A *symmetrically* interconnected Hopfield network (2.1) will not oscillate. In such a network, the nature—excitatory or inhibitory<sup>2</sup>—and strength of coupling between neurons is the same in both directions, i.e.  $a_{ij} = a_{ji}$ . The same network with delay, however, is found to oscillate if the delay exceeds a certain threshold. The promise of parallel computation by artificial neural networks depends of their stability, so this result is important from an engineering standpoint. The dynamicist sees first and foremost a model greatly enriched by the introduction of a delay. Sompolinsky, Crisanti and Sommers’ [79] finding that in randomly and hence asymmetrically connected Hopfield models ( $a_{ij} \neq a_{ji}$  in general) chaotic solutions appear further enriches the dynamics of this neural network.

Bondarenko’s model combines these elements in an *asymmetrical analog neural network with delay*, the equation for which reads exactly as (2.2) except that there is a provision for different delays  $\tau_j$  on each neuron. Nevertheless, Bondarenko immediately makes the simplifying assumption that  $\tau_j$  equals a constant  $\tau$  for all neurons. The asymmetry of the network is a result of initializing the coupling coefficients  $a_{ij}$  to random deviates on  $[-2, 2]$ . Initial conditions for DDEs are specified as a function

---

<sup>2</sup>Synapses are the main channels for interneuron communication and come in two varieties, excitatory and inhibitory. These terms describe the influence each type of synapse has on the generation of action potentials (APs), the characteristic spikes in voltage which propagate down the axon and stimulate the release of neurotransmitter molecules. The mean soma (cell body) potential must exceed a certain threshold for an action potential (AP) to be produced. Input from excitatory synapses promotes or excites the generation of APs by increasing the mean soma potential; inhibitory input has the opposite effect. More excitatory input results in an increased mean rate of AP generation or “firing” [83].

The graded response function of the Hopfield model describes a sigmoid input-output relationship for the neurons; increasingly positive input pushes the output towards the function’s upper asymptote. Equation (2.1) describes a situation wherein positive coupling  $a_{ij}$  to a positive output from neuron  $j$  will work to increase  $u_i$  and subsequently increase its output. While this fact alone might be enough to justify use of the excitatory/inhibitory classification of interconnections in the model, the fact that the dynamical variables  $u_i$  are readily interpreted as mean soma potentials and the output of the response function as the mean firing rate makes the terminology irresistible.

$\phi_i : [-\tau, 0]$ . In this case, all  $u_i(0)$  are initialized to random deviates on  $[-2, 2]$  and  $u_i(t)$  is taken equal to zero for  $t \in [-\tau, 0[$ . The nonlinear response function of the network is

$$f(x) = c \cdot \tanh(x), \quad (2.3)$$

the coefficient  $c$  being used to make global adjustments to response strength.

The form of the equation stated in [11] changes in a small but significant way from the form presented above (2.2), which was used in Bondarenko's 1997 papers: the sign preceding the sum on  $j$  switches from positive to negative:

$$\dot{u}_i(t) = -u_i(t) - \sum_{j=1}^M a_{ij} f(u_j(t - \tau)) + e_i(t), \quad i = 1, \dots, M. \quad (2.4)$$

*This* is the equation implemented in the **Bondarenko** class (see section 2.1.2 below) with which the control TS were generated (section 2.1.3) and the stability of the model's *chaotic* solutions studied (section 2.1.4 and appendix C). The equation for the Hopfield model and its variants is otherwise always written with a plus before this term [14, 33, 51, 61, 79]. The change is not a typographic error, because the results obtained are consistent with the use of the negative sign. Specifically, maximal Lyapunov exponent values and other nonlinear measures of the solutions are studied as a function of the average of coupling matrix elements

$$e = \frac{1}{M^2} \sum_{i,j=1}^M a_{ij}, \quad (2.5)$$

and nonzero fixed-point and quasiperiodic solutions are reported for  $e < -0.5$  and  $e > 0.5$  respectively. We found that reverting to the conventional plus sign in the model inverts the sign on these results. For  $e \approx 0$ , positive coupling approximately balances negative coupling and because the chosen response function (2.3) is not only sigmoidal but *odd*, the sign in question has no influence on the long-term behaviour of the network. Except where otherwise specified, all numerical results take  $e$  adjusted to equal zero (see Implementation for details).

At this point, we extend the notation to allow for delayed mean field coupling between networks; long-range synchronization between “brain regions” represented by mean-field coupled networks operating in different dynamical regimes can then be investigated. Let  $u_i^{(k)}(t)$  for  $i = 1, \dots, M$  denote the state of the  $M$  neurons of network  $k$  and

$$\langle u^{(k)}(t) \rangle = \frac{1}{M} \sum_{i=1}^M u_i^{(k)}(t) \quad (2.6)$$

that network’s mean field at time  $t$ . If there are  $M'$  networks, the equations for network  $k$  then reads

$$\dot{u}_i^{(k)}(t) = -u_i^{(k)}(t) - \sum_{j=1}^M a_{ij}^{(k)} f(u_j^{(k)}(t - \tau)) + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{M'} T_{i\ell} \langle u^{(\ell)}(t - \tau_{\text{mf}}) \rangle, \quad (2.7)$$

for  $i = 1, \dots, M$  and  $k = 1, \dots, M'$ . This sum on the mean fields replaces the generic external force term  $e_i(t)$  in (2.4). The  $T_{i\ell}$  are internetwork coupling coefficients and a delay  $\tau_{\text{mf}}$  (for “mean field”) is introduced to the coupling; normally one would set  $\tau_{\text{mf}} \gg \tau$ .

### 2.1.2 Implementation

The heart of any implementation of this artificial neural network is necessarily a routine for numerically integrating the model equations (2.4). Taking a cue from previous studies using this model [8–10, 31, 32, 67], we might disingenuously state that the fourth-order Runge-Kutta (RK4) routine `rk4` from *Numerical Recipes in C* [66] modified for double precision was used with a step size of  $h = 0.01$  and sampling interval of  $\Delta t = 1.0$ ...and leave it at that. However, numerical integration of DDEs is by no means straightforward. Use of a classic integration algorithm like RK4 implies certain simplifying approximations which must be made explicit. Furthermore, specific values of  $e$  were used by *adjusting* the elements of the random coupling matrix; we must specify when and how this is done. Finally, the broad

characteristics of the object-oriented approach we use and the use of the *Numerical Recipes* routines within the code should be outlined.

As mentioned earlier, initial conditions for DDEs are specified as a function  $\phi_i : [-\tau, 0]$ , implying a state space of infinite dimension. As the system evolves, so must this functional representation  $\phi_i : [t - \tau, t]$ . All numerical integration schemes necessarily discretize time, so the state space is effectively reduced to finite size. Why do DDEs pose a problem then for RK4, perhaps with adaptive step-size? In our case we have a single, constant delay  $\tau$  in the equations and we assume that it is an integer multiple  $n \in \mathbb{Z}^+$  of a *fixed* step-size  $h$ . The system's present state and relevant past ( $n$  previous steps) can then in principle be stored in an  $(n + 1) \times M$  element array. This would all be fine except that the RK4 algorithm requires evaluation of derivatives at times  $t$ ,  $t + h/2$  and  $t + h$ . Evaluating the derivatives (2.4) at the current time  $t$  and  $t + h$  is not a problem; values of the state variables at  $t - \tau$  and  $t + h - \tau$  are stored in the array. But what of the values at  $t - \tau + h/2$ ? The problem is only further complicated by the use of adaptive steps (variable  $h$ ). What is needed is a combination of a numerical integration method, say some brand of Runge-Kutta algorithm, with an *interpolation* scheme to allow the integrator access to a *continuous* past. This is not what we did, nor did Bondarenko [12].

Rather, we have assumed that for  $h$  sufficiently small, neglecting changes in the  $u_i$  over one integration step will not influence the gross characteristics of network behaviour. Past values of  $u_i$  required for computing the step from  $t$  to  $t + h$  were all approximated to  $u_i(t - \tau)$ , including  $u_i(t + h - \tau)$  although it could have been looked up in the array. This simplifying assumption seemed reasonable but to get a feel for the size of step-wise changes in  $u_i$  and, more importantly, their influence on the sigmoidal response function (2.3), we looked at the distribution of

$$|\Delta u_1| = |u_1(t + h) - u_1(t)|, \quad (2.8)$$

i.e. changes in the state of neuron 1 over one integration step, and

$$\Delta f = c |\tanh(u_1(t + h)) - \tanh(u_1(t))| \quad (2.9)$$

in an  $M = 10$  neuron network over  $5 \times 10^5$  integration steps  $h = 0.01$ , with default values for all other parameters except the output step, which was of course set to  $\Delta t = 0.01$  (cf. Table 2.1). The average fluctuation was  $\overline{|u_1|} = 0.03 \pm 0.03$  (error is standard deviation), but the average change in the response was only  $\overline{|\Delta f|} = 0.007$  or about 0.2% of  $c$ . The normalized distribution of  $|\Delta f|$  went like  $0.033/|\Delta f|$ , so over 80% of the values for  $|\Delta f|$  were less than 0.0005. These results support our view that approximating  $u_i(t + h/2 - \tau)$  and  $u_i(t + h - \tau)$  to  $u_i(t - \tau)$  has negligible impact on the average characteristics of solutions, which are the properties we are concerned with.

Although an integration step of  $h = 0.01$  was used for all subsequent results from this model, a cursory comparison of solutions with  $h = 0.01$  and  $h = 0.001$  was made to ensure that solutions that were chaotic, quasiperiodic or fixed point at  $h = 0.01$  remained so at  $h = 0.001$ . With default values and  $e = -0.8$ , the same fixed point was obtained in both cases. At  $e = 0.8$ , similar quasiperiodic solutions were found for both step sizes, with a very slow drift in relative phase. The duration of one “beat” was about 2400 times the main oscillation period. In both cases, chaotic solutions were found at  $e = 0.0$ , which naturally drifted apart at an approximately exponential rate.

In [11], Bondarenko examines the influence of network “excitability” on solutions, where a rise in positive, and therefore excitatory, coupling coefficients is reflected in a rise in  $e$ , equation (2.5). Again, the coupling coefficients are random deviates on  $[-2, 2]$ . We note that  $e$  as defined is truly the average of matrix elements and not the *expectation value* of the distribution from which they are drawn. That is, a value of 0.2 does not mean random deviates on  $[-1.8, 2.2]$  were used instead. Our approach is to initialize the matrix with random deviates on  $[-2, 2]$ , compute  $e$  on this matrix, calculate the difference between this value and the desired “excitability”  $a'$

$$\Delta a = a' - e \tag{2.10}$$

and offset all matrix elements by this amount. Consequently, the  $a_{ij}$  range in value

Table 2.1: User-configurable model parameters in Bondarenko class and their default values.

<b>Parameter</b>	<b>Default Setting</b>
Number of neurons $M$	10
Global Coupling Strength $c$	3.0
Delay $\tau$	10.0
Average of matrix elements $e$	0.0
RK4 integration step size $h$	0.01
Output step size (sampling interval) $\Delta t$	1.0
Seed for coupling matrix initialization	-1
Seed for state variable initialization	-1
Number of networks coupled to this one	0
Delay in mean field coupling to this network $\tau_{\text{mf}}$	0
Internetwork coupling coefficients	0.0

from  $-2 + \Delta a$  to  $2 + \Delta a$ . In a private communication, Bondarenko confirmed use of this method in his experiments [12]. Using the default network and a few different seeds for the matrix random number generator (RNG), solutions obtained by (a) generating matrices until one within  $10^{-6}$  of  $a'$  was found (brute force), (b) shifting the distribution by  $a'$  (offset distribution) and (c) offsetting the matrix by  $\Delta a$  were not found to differ qualitatively. Fixed point, quasiperiodic and chaotic behaviour was observed on the same ranges of  $e$  using each of the three methods as those reported by Bondarenko. With the exception of the set of TS where  $e$  is explicitly varied, a value of  $e = 0$  was used in all the experiments and TS generation in this thesis.

The model was implemented in object-oriented C++. Each instance of the **Bondarenko** type is a network with its own protected set of parameter values, state variables, RNG seeds and registers, mean field coupling properties and “remembrance of things past”, i.e. the array of past values of the state variables  $u_i^{(k)}$ . Model parameters and their default values are listed in Table 2.1. The network is highly configurable. Multiple instances can naturally coexist in a program and be interconnected. The characteristics and entire history of one network can be copied to another by simple association. A code fragment illustrating some of these features is presented in Figure 2.1. In addition to the many inherent advantages of object-oriented programming

Figure 2.1: Code fragment illustrating features of Bondarenko class.

```

...
Bondarenko brain[2]    //Declare an array of two
                      //Bondarenko class objects
brain[0].setGlobalCoupling( 5.0 ); // Set c (default is 3.0),
                                   // but otherwise use default values
brain[1].setCouplingSeed( -2 ); //For coupling matrix initialization
                                   //(default is -1);

brain[0].initialize();
brain[1].initialize();

    // Networks will have different coupling matrices (because of
    // different seeds) and different values of c in response
    // function.

    // Integrate the system, outputting only the state of
    // brain[0]'s neurons
while( integrating ){
    for( n=0; n<2; n++) brain[n].computeOutputStep();
    brain[0].printState();
    ...
}
brain[1] = brain[0];
    // brain[1] now has the same coupling matrix, current state
    // and remembered past as brain[0]. Their c values still
    // differ, but that is easily remedied...
brain[0].setGlobalCoupling( brain[1].getGlobalCoupling() );
...

```

(communication with the model via a well-defined interface, information hiding and data encapsulation [19], etc.), this approach makes it easy to build multi-network “brains” (i.e.,  $M' > 1$  and  $T_{ij} \neq 0$  for some pair  $(i, j)$ ).

We have said that numerical integration in this implementation is carried out in double precision floating point arithmetic using a modified version<sup>3</sup> of the fourth-order Runge-Kutta algorithm `rk4` [66]. To compute the step from  $t$  to  $t + h$ , the routine requires the current time from zero, current values of the state variables  $u_i(t)$  and their derivatives  $\dot{u}_i(t)$ .<sup>4</sup> Let  $t_h$  and  $\tau_h$  be respectively the time and delay in units of the step size  $h$ ; we require that  $\tau_h$  be integer. In this way, a  $(\tau_h + 1) \times M$  array can hold all the past states we will need for evaluating derivatives. The current state is in row

$$i = t_h \bmod \tau_h, \tag{2.11}$$

and the delayed state required to evaluate derivatives is in row

$$i' = (t_h - \tau_h) \bmod \tau_h = (t_h + 1) \bmod \tau_h. \tag{2.12}$$

Once the step is computed, this is also the row in which it will be stored; in other words, every step writes over the oldest row.

One other *NR* routine is used in the `Bondarenko` class and that is `ran1`, an RNG needed to initialize the coupling matrix and set initial conditions. Use of random numbers is not intensive in this program, but it is always wise to use a reliable RNG (rather than the notoriously poor C library function `rand`).

---

<sup>3</sup>All `float` declarations were changed to `double` in the function definition and double arrays were passed to the function (initialized using `dvector` rather than `vector`).

<sup>4</sup>This evaluation of the  $\dot{u}_i(t)$  before calling the integration routine is easy to forget; realizing I had done so was a humbling experience. Fortunately, it was still early days, so the bug was not too “costly”. Why not have `rk4` handle all four calls to the derivative function? The “driver” is required to make the first call to accommodate use of the routine in adaptive step size schemes. If the driver decides a step is too big, it can call `rk4` again with a smaller step without having to reevaluate these derivatives.



### 2.1.3 Bondarenko Data Sets

Using an  $M = 10$  neuron network, two data sets were generated, the  $c$  set and the  $e$  set. The *same realization* of the coupling matrix was used to produce all of the data; this particular network did show the dependence on these parameters described by Bondarenko. Small changes in even a single coupling coefficient can completely change the *type* of solution obtained [9]; initial conditions leading to long term chaotic behaviour vary enormously between realizations. This variation is explored further in Appendix C. In the  $c$  set, for each of

$$c \in \{1.0, 2.0, 3.0, 5.0, 10.0\}, \quad (2.13)$$

and  $e = 0.0$ , a TS of  $N = 2^{13} = 8192$  samples was recorded generated for each of 25 different sets of initial conditions at two sampling intervals,  $\Delta t = 1.0$  and  $\Delta t = 10.0$ . A similar set of TS, the  $e$  set, was generated for values of

$$e \in \{0.465, 0.475, 0.480, 0.485, 0.486\}, \quad (2.14)$$

with  $c = 3.0$ . All neuronal outputs were recorded. To ensure that a stationary solution had been reached, we discarded the first  $3 \times 10^5$  integration steps. This equals an elapsed time of  $t = 3000$ , slightly longer than Bondarenko's own estimate of time to stationarity,  $t_{st} = 30Nh = 2458$  [8]. Typical signals for  $c$  and  $e$  series are shown in Figures 2.2 and 2.3 respectively.

In [8], the sampling frequency was chosen such that 8-10 samples per period of the highest frequency "significant spectral component" were taken. Figure 2.4 shows the power spectrum of the default network with  $c = 1.0$ ; the three most prominent peaks correspond to periods of about 43, 22 and 15 arbitrary time units (atu). Similar spectra are obtained for other chaos producing values of  $c$ . A sampling interval of  $\Delta t = 2$  would satisfy the spectral requirements; we used  $\Delta t = 1.0$ . This and  $\Delta t = 1.25$  were indeed the (uncited) sampling intervals used by Bondarenko [12]. An interval of  $\Delta t = 10.0$  was used to produce "undersampled" data.

Figure 2.2: Typical output for each of the values of  $c$  in the  $c$  set. A:  $c = 1.0$ , B:  $c = 2.0$ , C:  $c = 3.0$ , D:  $c = 5.0$ , E:  $c = 10.0$ , with  $e = 0.0$  throughout. Shown are neuron 1 outputs over an interval of  $\Delta t = 700$  from the first TS for each  $c$  sampled at  $\Delta t = 1.0$ . Besides an obvious difference in amplitude, it is difficult to distinguish between the signals.

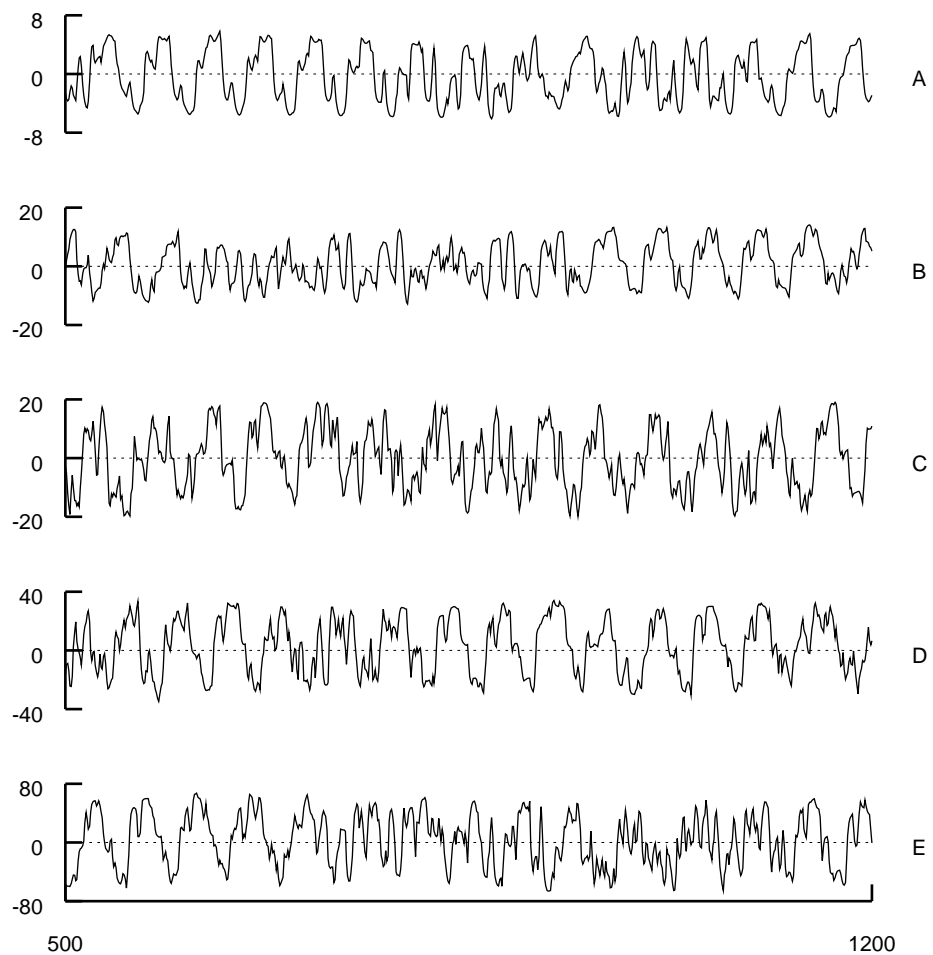


Figure 2.3: Typical output for each of the values of  $e$  in the  $e$  set. A:  $e = 0.465$ , B:  $e = 0.475$ , C:  $e = 0.480$ , D:  $e = 0.485$ , E:  $e = 0.486$ , with  $c = 3.0$  throughout. Shown are neuron 1 outputs from the first TS for each  $e$  sampled at  $\Delta t = 1.0$ . Note that the interval shown is more than twice as long as that in Figure 2.2 in order to show increasing frequency of quasi-periodic bursts with increasing  $e$ .

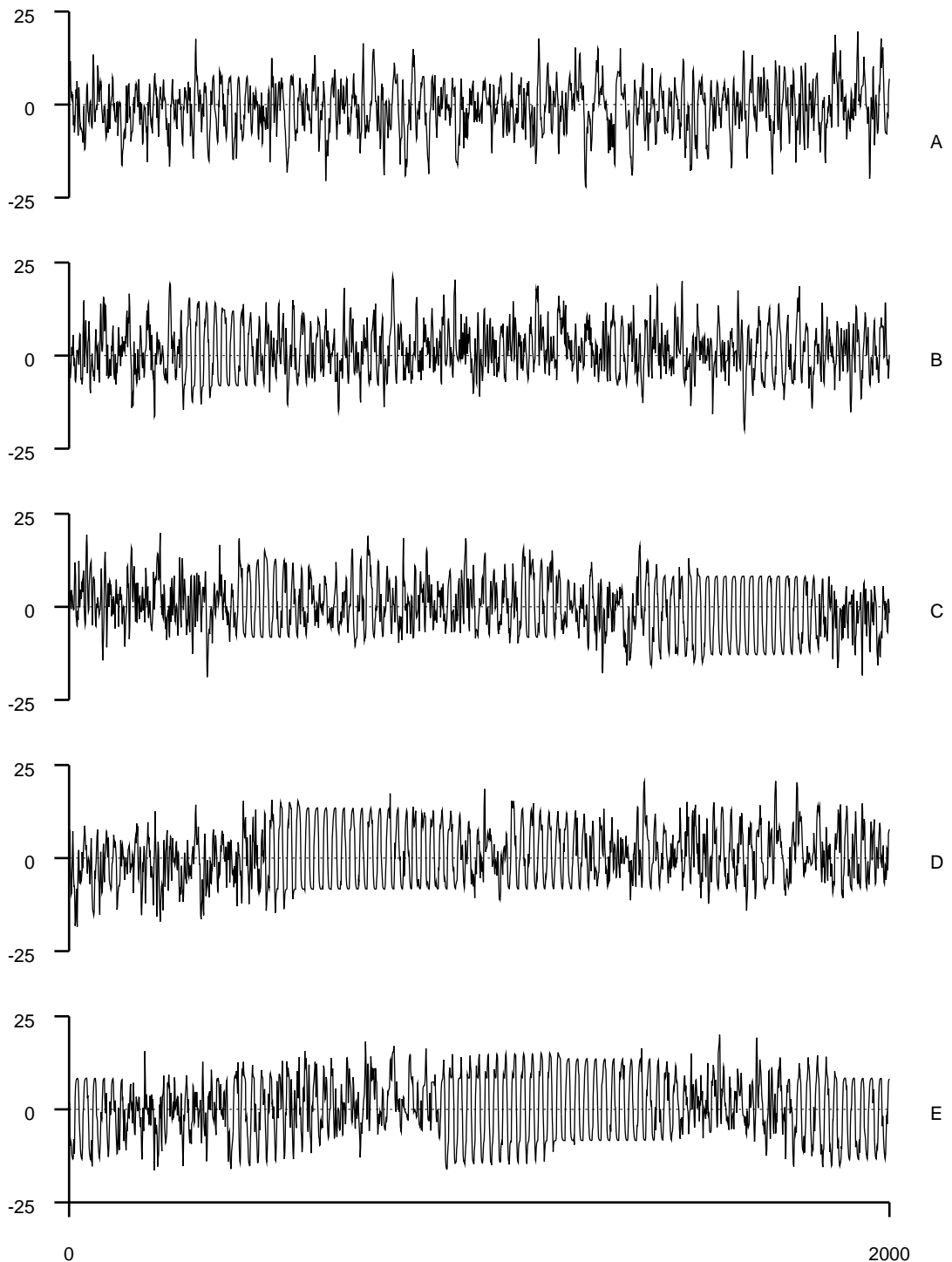
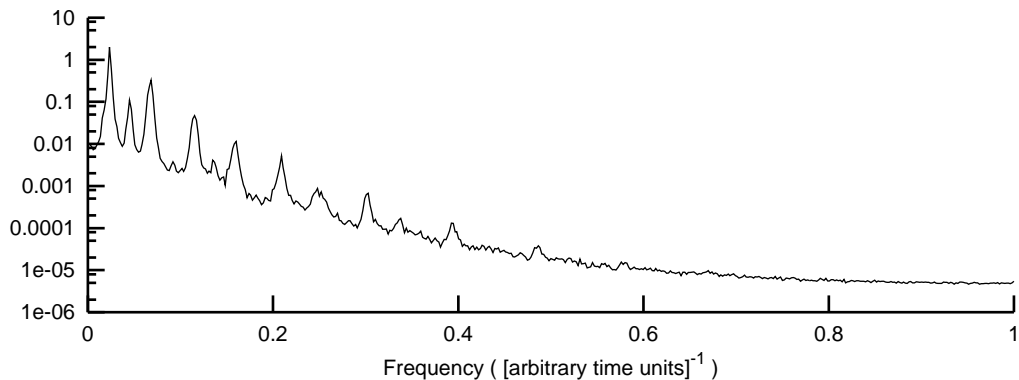


Figure 2.4: Estimated power spectrum of neuron 1 output of Bondarenko model  $M = 10$ ,  $\tau = 10$ ,  $e = 0$  and  $c = 1.0$ . A TS of  $N = 100 \times 1024$  samples, sampled at intervals of  $\Delta t = 0.5$  or  $50 h = 0.01$  RK4 integration steps, was generated and the power spectrum estimated by coherent phase averaging [39, 66] of the fast Fourier transforms of 100 windows of 1024 samples each: the TISEAN program spectrum [38] was used to obtain the spectrum of each window and these spectra were then averaged.



#### 2.1.4 Characterizing the Bondarenko Data

Fixed point, quasi-periodic and chaotic solutions can all be observed in this model. Bondarenko found that signals in the chaotic regime yielded values of correlation dimension and maximal Lyapunov exponents which were in the same range as those obtained from human EEGs. The fact that different values of the correlation dimension were obtained from different neurons for a given set of parameter values was emphasized as being similar to the variation of EEG results across different areas of the brain. On the other hand, most of his results are obtained using *a single realization* of the network and the strong influence that changes in coupling coefficients can have on solutions has been acknowledged. The relationships between  $c$  or  $e$  and nonlinear measures computed on the solutions, then, cannot be taken at face value.

This model has already been used in assessing the discriminating power of measures across (presumably) increasingly chaotic TS [67], but these authors did not report validation of any of the results in Bondarenko's paper. Since changes in precisely these quantities reflect the dynamical changes to which our measures should

be sensitive, however, characterizing the solutions obtained from *our own* network realization is a necessary first step.

The dynamical change across different values of  $e$  for fixed  $c$  is fairly obvious. The frequency and duration of quasi-periodic bursts increase. In fact, these values of  $e$  were chosen for their proximity to a critical value after which solutions are strictly quasi-periodic. That the signals obtained with different values of  $c$  are dynamically different is less evident. This is the series we focus on in this section. Results in this section were once again obtained with the default network (cf. Table 2.1 for parameter values) modified as indicated in each experiment. In other words, these results are for *one* realization of the network, the one used to generate the data set.

#### 2.1.4.1 Maximal Lyapunov Exponent

Exponential divergence of trajectories is a well-known characteristic of chaotic systems. *Lyapunov exponents* quantify this divergence, and although a dynamical system in  $M$ -dimensional state space will have  $M$  exponents  $\lambda_i$  for  $i = 1, \dots, M$ , we will be interested in the *maximal* Lyapunov exponent  $\lambda$  defined as follows [37]. Let  $\mathbf{u}^{(1)}(0)$  and  $\mathbf{u}^{(2)}(0)$  be two initially close points in phase space,  $\|\mathbf{u}^{(1)}(0) - \mathbf{u}^{(2)}(0)\| = \delta_0 \ll 1$ . If  $\delta_t$  denotes the distance at time  $t$  between the two points,  $\delta_t = \|\mathbf{u}^{(1)}(t) - \mathbf{u}^{(2)}(t)\|$ , then  $\lambda$  is determined by

$$\delta_t \simeq \delta_0 e^{\lambda t}, \quad \delta_t \ll 1, \quad t \gg 1. \quad (2.15)$$

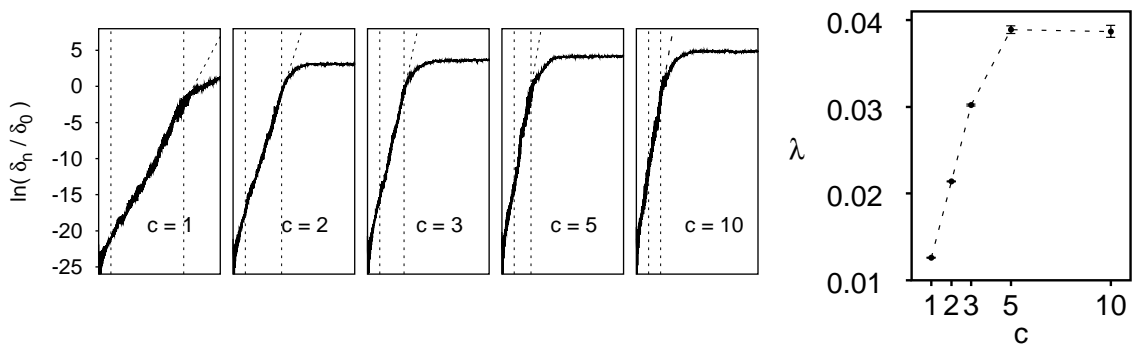
Positive  $\lambda$  means chaos.

We estimated  $\lambda$  as a function of  $c$  in  $M$ -dimensional phase space. For each value of  $c$ , two identical networks,  $\mathbf{u}^{(1)}$  and  $\mathbf{u}^{(2)}$ , were initialized and the first  $\mathbf{u}^{(1)}$  was integrated to  $t = 2500$  (skipping transient solution). Its state was then assigned to  $\mathbf{u}^{(2)}$ . A random perturbation  $\varepsilon u_i^{(2)}$ , where  $\varepsilon$  is a random number<sup>5</sup> between  $10^{-6}$  and  $10^{-14}$  was added to each neuron  $u_i^{(2)}$  of network 2 and the initial separation between

---

<sup>5</sup>Obtained by computing  $10^{(-14+8\xi)}$ , where the  $\xi$  are random deviates on  $[0,1]$ .

Figure 2.5: Growth of  $\ln(\delta_n/\delta_0)$  shown for  $n = 0, \dots, 2500$  versus  $c$ . For details see text. The linear scaling region used for estimating  $\lambda$  is indicated in each case by vertical dotted lines, whereas the sloped dotted line is the resultant fit. Scaling regions were determined by visual inspection of each resultant  $\ln(\delta_n/\delta_0)$  curve. Fluctuations were generally larger on the first 200 points of each curve, so scaling regions begin after this interval and continue up to a point identified as the start of the shoulder on each curve. The slope of the fit to this region along with the fitting error are plotted for each value of  $c$  in the graph on the far right. Fits were done in Gnuplot using an implementation of the non-linear least squares Marquardt-Levenberg algorithm.



the two networks in state space computed using the Euclidian norm

$$\delta_0 = \left[ \sum_{i=1}^M (u_i^{(1)}(0) - u_i^{(2)}(0))^2 \right]^{1/2}. \quad (2.16)$$

Both networks were then integrated and the separation between their positions in state space computed every  $\Delta t = 1.0$ ,

$$\delta_n = \left[ \sum_{i=1}^M (u_i^{(1)}(n \Delta t) - u_i^{(2)}(n \Delta t))^2 \right]^{1/2}. \quad (2.17)$$

The ratio  $\delta_n/\delta_0$  was stored for  $n = 1, \dots, 3000$ . At that point, the state of network 2 was reset to that of network 1, network 2 was randomly perturbed again and the integration of the two networks continued. After 50 such runs, the average value of  $\delta_n/\delta_0$  across all runs for each  $n$  was output to an ASCII file. These results are plotted in Figure 2.5. We find a rise in  $\lambda$  with increasing  $c$  up to a plateau for around  $c = 5$ . For a network of this size ( $M = 10$ ), Bondarenko obtained a similar but nonmonotonic dependence on  $c$ .

The values of  $\lambda$  cited in Bondarenko’s more recent papers [9, 11] differ from our own by about a factor of 2, which is readily attributable to the differences between realizations of the coupling matrix. On the other hand, our results appeared to be *two orders of magnitude smaller* than those in [8], which was initially a cause of some concern. Upon closer rereading of this paper, we realized that we had been overlooking the *units* of  $\lambda$  there, namely  $[\text{s}^{-1}]$ . This is a suitable unit for a Lyapunov exponent computed on a *real* time series, such as an EEG recording. In estimating these exponents on his model data, Bondarenko rescales (“make[s] time normalization”) the results such that “the main frequency of [the] numerical solution is equal to the frequency of the human  $\alpha$ -rhythm (approximately 10 Hz)”, although he neither states the scaling factor explicitly nor shows any representative spectra in the earlier papers. The rescaling allows him, in principle, to compare his model results with those obtained on real EEGs.

In [9, Fig. 1], the main frequency of the solution for an  $M = 10$  neuron network, with  $c = 3.0$ ,  $\tau = 10.0$  and  $e = 0.0$ , is clearly  $\omega = 0.12$  or  $f = \omega/2\pi = 0.019 \text{ atu}^{-1}$ . The period of this oscillation is  $T = 1/f = 52 \text{ atu}$ , which would correspond to a scaling factor of  $0.002 \text{ s/at}$ . In [8], the author gives a result of  $\lambda = 1.5 \text{ s}^{-1}$  for these parameter values, but we do not know whether the realization of the coupling matrix and initial conditions used in this paper was the same as in [9]. The spectra of the time series from our own realization of this network, on the other hand, have strong peaks at around  $f = 0.023$  and  $0.067 \text{ atu}^{-1}$  or  $T = 43$  and  $15 \text{ atu}$  respectively. Using these periods to compute the scaling factor and applying this rescaling to our own results, we obtain values of  $\lambda = 4.5 \text{ s}^{-1}$  for  $T = 15 \text{ atu}$  and  $13.6 \text{ s}^{-1}$  for  $T = 43 \text{ atu}$ . In the end, this “mystery” scaling factor makes it difficult to compare our results to those on Bondarenko’s “normalized” data, although we can see that applying such a rescaling brings us into the same general range for  $\lambda$  as those Bondarenko cites.

### 2.1.4.2 Correlation Dimension

Introduced by Grassberger and Procaccia [29], the correlation dimension  $\nu$  is a measure of attractor geometry suitable for use on objects reconstructed from finite time series contaminated by noise. It can be used as a tool to quantify self-similarity (fractality) in a system when it is already *known to be present* [37]. The number of *excited* degrees of freedom in a system corresponds to the dimension of the attractor on which it lives. Given a set of  $N$  points  $\mathbf{x}_i, i = 1, \dots, N$ , in some vector space and a distance  $\epsilon$ , one computes the correlation sum

$$C(N, \epsilon) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N \Theta(\epsilon - \|\mathbf{x}_i - \mathbf{x}_j\|), \quad (2.18)$$

where  $\Theta$  is the Heaviside step function,

$$\Theta(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0. \end{cases} \quad (2.19)$$

This sum counts the number of pairs  $(\mathbf{x}_i, \mathbf{x}_j)$  separated by less than  $\epsilon$ .  $C(N, \epsilon)$  is expected to scale like a power law,  $C(\infty, \epsilon) \sim \epsilon^\nu$ , in the limit of an infinite amount of data and small  $\epsilon$ . Thus the correlation dimension  $\nu$  is defined as

$$d(N, \epsilon) = \frac{\partial \ln(C(N, \epsilon))}{\partial \ln \epsilon} \quad (2.20)$$

$$\nu = \lim_{\epsilon \rightarrow 0} \lim_{N \rightarrow \infty} d(N, \epsilon). \quad (2.21)$$

When dealing with finite series of real data, the limits in (2.21) will obviously pose a problem! Furthermore, this dimension is an invariant quantity only if a proper embedding is found, so if delay coordinates are used, a correct choice of delay  $\tau_{\text{embed}}$  and embedding dimension  $m$  must be made.

In practice one usually computes  $C(N, \epsilon)$  for a range of embedding dimensions and looks for a linear scaling region in a log-log plot of  $C(N, \epsilon)$  versus  $\epsilon$  over several values of  $m$ . Caution must be exercised when interpreting the results of this algorithm. In particular, Eckmann and Ruelle argue that for a time series of  $N$  samples, regardless



of the precision of the data, the Grassberger-Procaccia algorithm *will not produce dimensions larger than*

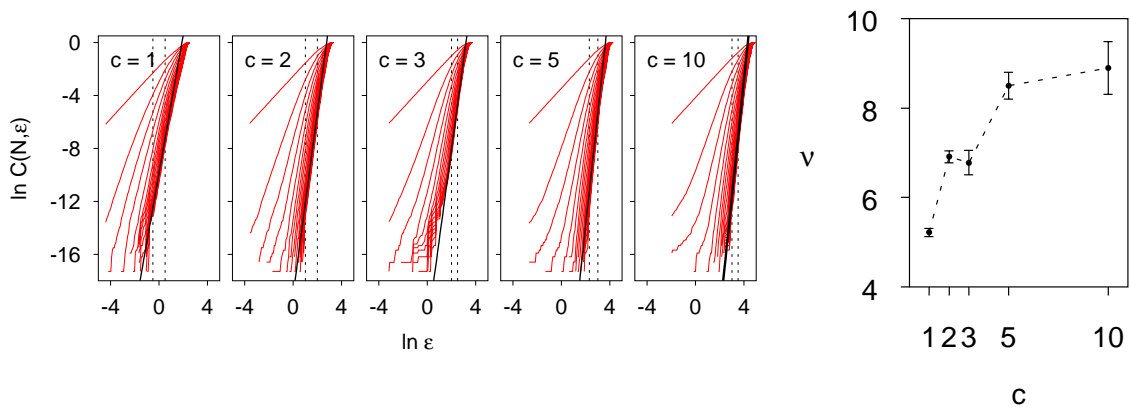
$$\nu_{\max} = \frac{2 \log N}{\log(1/\rho)}, \quad (2.22)$$

where  $\rho = \tilde{\epsilon}/D$ ,  $D$  being the diameter of the reconstructed attractor and  $\tilde{\epsilon}$  the order of magnitude of the “small” distances  $\epsilon$  used in the algorithm [22]. If  $N = 8192$  and  $\rho = 0.1$ , we find using decimal logarithms that  $\nu \leq 7.8$ . Taking  $N = 100000$  and the same value of  $\rho$ , we find  $\nu \leq 10$ . In light of this result, we should not be surprised that values of the correlation dimension plateau at  $\nu \approx 6 - 8$  in Bondarenko’s findings [8], regardless of the independent variable. Furthermore, we should be wary if a linear scaling region has not appeared by embedding dimensions of  $m = 2 \times 8 + 1 = 17$  [80] for the short series or  $m = 21$  for the longer series.

Using an implementation of the Grassberger-Procaccia algorithm with a fast neighbour searching scheme included in the TISEAN software package (called `d2`) [38], we examined the behaviour of  $C(N, \epsilon)$  on our own time series of  $N = 8192$  points sampled at  $\Delta t = 1.0$ . We used embedding dimensions ranging from 1 to 37 (for “high-dimensional outputs” with  $\nu = 6 - 8$ , Bondarenko used  $m = 11$  to 37 in increments of 2 [12]), although it is clear that embedding so little data in so many dimensions will yield poor statistics at low  $\epsilon$ . Furthermore, even on our longest TS, the G-P algorithm will not produce dimensions higher than 21. We did the computation up to  $m = 37$  to remain faithful to Bondarenko’s method, but we required that any scaling region we identified—by visual inspection—be present beginning at  $m \sim 15$ . Pairs separated in time by less than  $\Delta t = 100$  were excluded from the sum to eliminate spurious convergence to a low dimension resulting from autocorrelation [37]. The log-log plot of the correlation sums obtained and “correlation dimension estimate” are shown in Figure 2.6. We find a similar plateau in  $\nu$  around 8 and are similarly skeptical.

Naturally we tried repeating the experiment with longer time series. Bondarenko informs us that he later repeated his calculations on TS of  $N = 10^5$  samples and obtained the same results as with the  $N = 8192$  series [12]. Our own experiments

Figure 2.6: For each  $c$ , correlation sum results  $\ln C(N, \epsilon)$  vs.  $\ln \epsilon$  for embedding dimensions  $m = 1$  to 37 in increments of 2 using neuron 1 signal in first time series in  $\Delta t = 1.0$  data set. Embedding vectors were formed using consecutive samples. The linear scaling region used for linear fitting is indicated with vertical dashed lines in each graph and was determined by visual inspection of the results. In all cases, fitting was performed in this interval on results for  $m = 21$  to 37. Slope and error of fit for each  $c$  are plotted at the far right.

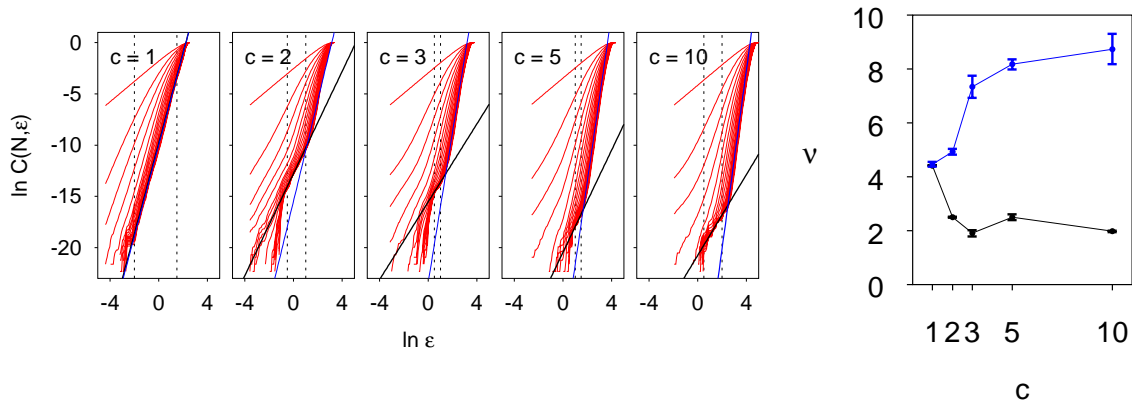


with  $N = 10^5$  TS yielded confounding results (Fig. 2.7).

The long TS appear to be stationary and similar to the shorter series. We examined the long series for  $c = 10.0$  more carefully by computing the average and standard deviation on windows of 8192 samples overlapping by half a window. The average fluctuated slowly with no clear trend in the range between  $+1$  and  $-3$ , while the standard deviation (SD) was always in the  $35 \pm 1$  range. We did not examine the stationarity more closely than that, which may explain the strange results and closer analysis of very long TS would be important in future work.

If we fit on the same interval as in Fig. 2.6 we do obtain similar results (blue line). However, a scaling region which is clearly not just part of an S-curve has appeared; it is particularly striking in the  $c = 2$  results. Linear fitting on these intervals, which are no narrower than those used in the shorter run, yields a very different relationship between  $c$  and  $\nu$  (black line). These results would suggest that the number of active degrees of freedom *deceases* with increasing  $c$ . One can at best conclude that on our realization of this artificial neural network, the behaviour of the correlation sum does

Figure 2.7: Correlation sum results for TS of  $N = 10^5$  points with all other parameters as in Fig. 2.6. Linear fits to the same  $\epsilon$  intervals as in that figure yield similar results (blue lines). However, fits on the newly appearing scaling intervals show a very different trend in  $\nu$  (black lines).



appear to vary with respect to  $c$ . No clear relationship between  $\nu$  and  $c$  could be established, but the presence of significant dynamical differences between the signals which are otherwise very similar to each other and to EEG signals in appearance is all we require for our subsequent experiments. It would be interesting in the future to redo the calculations using data sampled at different rates (maybe our embedding was poor) and series recorded only after discarding a much longer transient.

## 2.2 Experimental Time Series: Epileptic EEG Recordings

### 2.2.1 Electroencephalography

The electroencephalograph is an important tool in neurological diagnosis. Discovered in the 1920's by Hans Berger, electroencephalography is based on measurement of the weak currents produced on the surface of the brain by the aggregate of synchronized activity from hundreds of thousands of neurons or more<sup>6</sup> [1, under "Electroen-

---

<sup>6</sup>It is not the axonal discharges (action potentials) which produce these brain waves but rather fluctuations in the *postsynaptic* membrane potentials between the inside

cephalography”]. Most frequently, electrodes are placed on the scalp and the pattern of electrical voltage produced there by diffusion of these weak currents through the skull are transduced by differential input amplifiers and recorded. Electrodes may also be placed on the brain surface or in contact with deeper structures. These recordings, referred to as intracranial EEG or iEEG, have the advantage of a higher signal-to-noise ratio and greater spatial resolution. In the classic chart recorder, the signals are amplified about a million times in order to drive the instrument’s pens. Today, the signals are often digitized and recorded to magnetic or optical media. Specialized software exists to aid in the analysis of these digitized recordings. Indeed, the interpretation of EEGs is a clinical specialty requiring considerable training and experience.

### 2.2.2 Description of the Data Set

The EEG recordings used in this thesis were kindly provided by Dr. Jean Gotman of the Montreal Neurological Institute. There are five recordings varying in length from four to seven hours from each of five patients undergoing intracranial EEG investigation for medically intractable focal epilepsy. No information on the localization of the epileptic focus was provided with the data. Signals from both intracranial and surface contacts are recorded. Each patient is identified by a unique prefix: *amy*, *bob*, *carl*, *dan* and *eve*. The actual name and sex of the patients was in fact unknown to us. The signals were digitized at 200 Hz and archived on CD-ROM after analysis. Other than the data itself, only the start time of each clinical seizure as marked by a trained clinician was used in this study. To simplify batch calculations on the entire set, the data files were copied to hard disk (filling 12 GBytes) and special routines were written to extract specific channels of data from any given recording.

Of the 15 recordings which capture clinical seizures, only three capture more than one. For each patient, there is one long seizureless recording made while the patient 

---

and outside of dendritic processes.

was awake and another made during sleep. The patients were awake throughout the recordings capturing seizures.

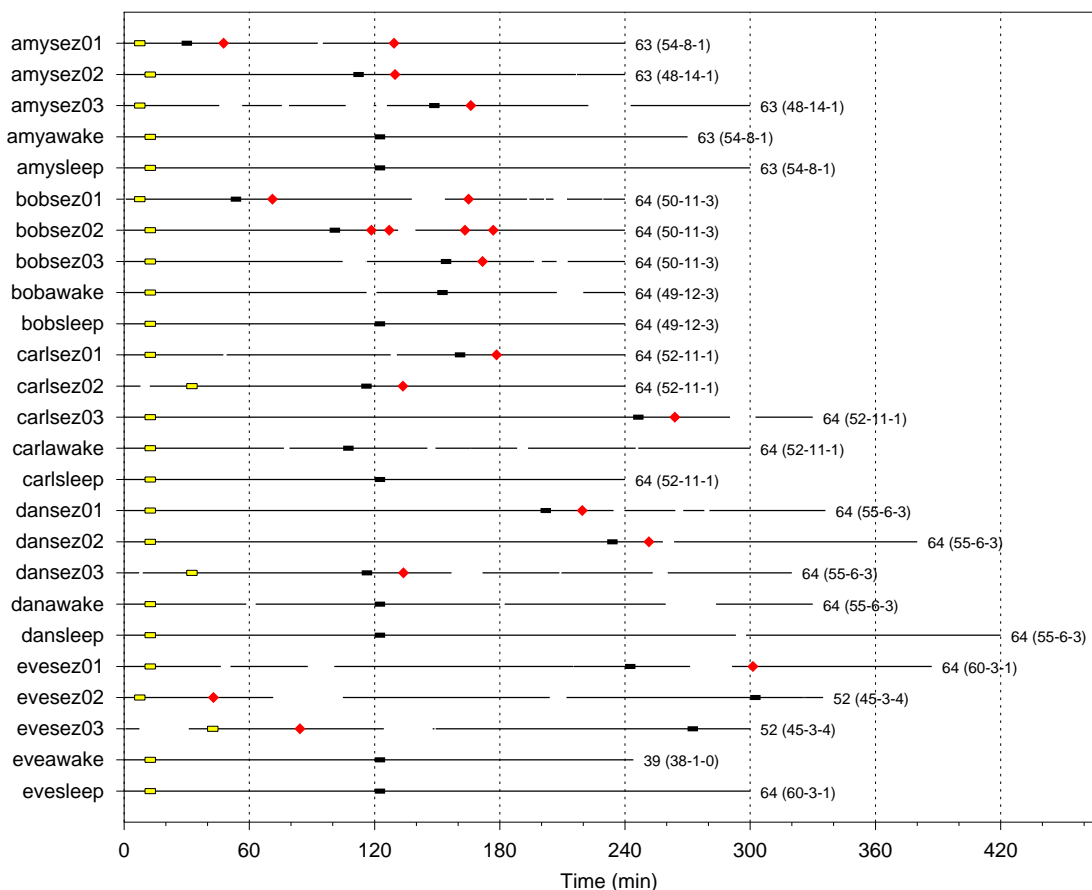
From time to time, a patient needs to be disconnected from the EEG monitor. When this occurs, the system continuously writes “zeroes” on the recording in order to preserve the relative timing of events before and after the disconnection; in other words, the virtual strip chart continues to advance. This leaves gaps in the recording which will naturally cause strong fluctuation in any measure we compute. We would like to be able to ignore these fluctuations systematically in our analysis. For this reason, a program was written to automatically catalogue these gaps based on the algorithm presented in Appendix D. In Fig. 2.8, the duration of each recording, location of its seizures and gaps and information about numbers of each type of signal recorded are presented.

### 2.2.3 Amplitude-Frequency Features

It is never a good idea to undertake a nonlinear analysis of new time series without first familiarizing oneself with the linear structure of these data. Precautions can be taken to help avoid mistaking spurious fluctuations in a measure due to autocorrelations as evidence of nonlinearity, most notably the use of surrogate data. This method is discussed in Chapter 4. As a first step, however, we have computed the periodogram of the first intracranial signal on each recording as well as the standard deviation on a rolling window.

To produce the periodogram, the power spectral density is computed using the fast Fourier transform (FFT) algorithm on nonoverlapping windows of 8192 samples (41 sec) and stored in an array. After the power spectrum on each window is computed, the *entire* array is rescaled to the unit interval. The resulting data was then output in Portable Grey Map (PGM) format. The image was then colorized and its contrast adjusted slightly. Power increases from black to white through the blues and oranges in the final image and variations in total power (i.e. signal amplitude) are easy to

Figure 2.8: Schematic representation of major features of the EEG data set. Labels on the left margin identify each of the 25 recordings in the set, which are represented by a line extending to the right. Gaps in this line correspond to gaps in the recording (see text for details of gap identification) and the total line length indicates the recording length. Red diamonds mark the onset of clinical seizures. Yellow bars outlined in black show the location and length of the first reference windows, and black bars the location and length of the second. These windows are used in computing the similarity measure. To the right of each line is the total number of channels recorded and in parentheses, the breakdown in terms of number of intracranial contacts—surface contacts—ineligible channels. The criteria for channel eligibility are described in §3.3.1.



see. Some of the periodograms are presented in Figs. 2.9 through 2.12, but rather than include all of the figures, we have summarized their main features in Table 2.2. Where pronounced changes in the spectrum precede a seizure, the *approximate* lead time is recorded; this is meant to give a rough impression only. When the level rises in these figures simultaneously at all frequencies, we expect to find an increase in signal amplitude. Examination of Fig. 2.13, which plots the standard deviation results, shows that this is indeed the case.

About a third of the recordings, including the *eve* series and *bobsez02*, show either little variation in the pre-seizure interval (the former) or a lot of variation throughout the interval (the latter). There is, in the results for these series, no evidence of a distinct “preictal phase” when conditions are becoming favorable for a seizure. The performance of the nonlinear measures on these TS will consequently be of particular interest.

Table 2.2: Descriptions of periodograms obtained using the first intracranial signal in each recording and nonoverlapping windows of 8192 samples (41 sec). Times are relative to the start of the recording,  $T_0$ , or relative to the onset of the first clinical seizure,  $T_{S_z}$ . Time from a marked spectral change persisting up to a first seizure is the “anticipation time”  $T_A$  (min).

Recording	$T_A$	Description
amysez01	30	rise in power at higher frequencies followed by overall rise in power (Fig. 2.9)
amysez02	40?	overall drop in power at $T_0 + 25$ min, rising again at $T_{S_z} - 40$ min
amysez03	20	preseizure bloom around 40 Hz
amyawake	—	40 min interval at higher intensity from $\sim T_0 + 120$ min
amysleep	—	overall intensity swells over 60 min interval, drops and begins again
bobsez01	20	drop in intensity at all frequencies at $T_{S_z} - 20$ min
bobsez02	0	rapidly alternating strips of high and low intensity at all frequencies (Fig. 2.10)
bobsez03	5	“ , but alternation stops just prior to first seizure
bobawake	—	same as <i>bobsez02</i>
bobsleep	—	significant change in overall intensity every 10-40 min
carlsez01	15	overall rise in intensity; drops to initial levels at $T_{S_z} + 5$ min
carlsez02	0	strong rise in intensity at seizure onset followed by drop to initial levels and subsequent slow rise to higher overall intensity
carlsez03	100	overall drop in intensity $T_{S_z} - 100$ min with choppy rise to initial levels about $T_{S_z} - 5$ min
carlawake	—	no pronounced changes in power spectrum
carlsleep	—	very uniform distribution; 60 min drop in overall intensity at $T_0 + 100$ min (Fig. 2.12)
dansez01	0	two 20 min bands of lower intensity between start and first seizure, but first 60 min and 30 min prior to seizure very similar
dansez02	20	less power at higher frequencies than other TS; intensity drops for 15 min at $T_0 + 90$ min and again for 20 min prior to Sz
dansez03	0	changes in intensity every 5-40 min in the preictal interval
danawake	—	rapid, strong alternation of intensity at all frequencies
dansleep	—	some 10 min wide bands of lower intensity but little activity overall
evesez01	0	no apparent preseizure activity (figure may be saturated by post seizure levels)
evesez02	0	“
evesez03	0	“
eveawake	—	rapid, strong alternation of intensity at all frequencies (Fig. 2.11)
evesleep	—	uniform; slight intensity fluctuation every 30–60 min



Figure 2.9: Periodogram obtained using the first intracranial signal in recording *amysez01* with nonoverlapping windows of 8192 samples (41 sec). Black lines extending above and below the figure indicate onset times of clinical seizures. Frequencies go from from 0 Hz at the top of the figure to 100 Hz at the bottom (contrary to convention). The figure covers the entire recording, which is 4 hours long. About 30 min prior to the seizure, we observe a rise in power at higher frequencies followed by an overall rise persisting up to the start of the clinical seizure.

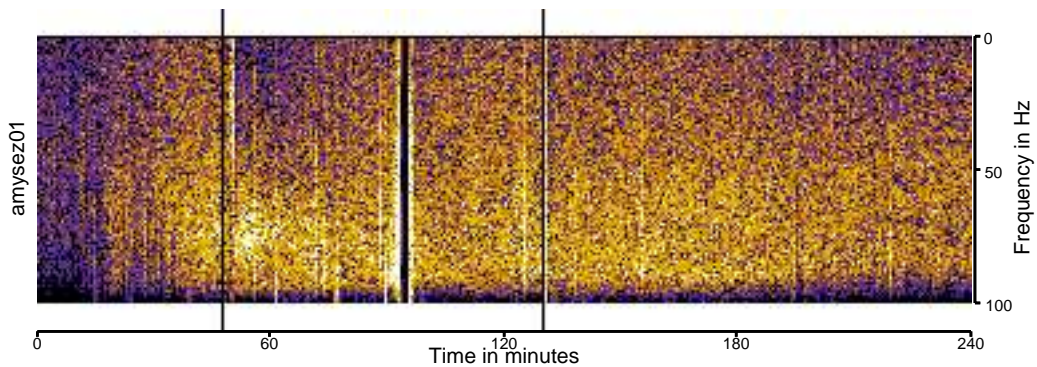


Figure 2.10: Periodogram obtained using the first intracranial signal in recording *bobsez02* with nonoverlapping windows of 8192 samples (41 sec). Rapid fluctuations in overall intensity are observed on the entire interval before the first seizure and, in fact, the first *gap*. Following the gap, the spectrum hardly changes at all. See also caption to Fig. 2.9.

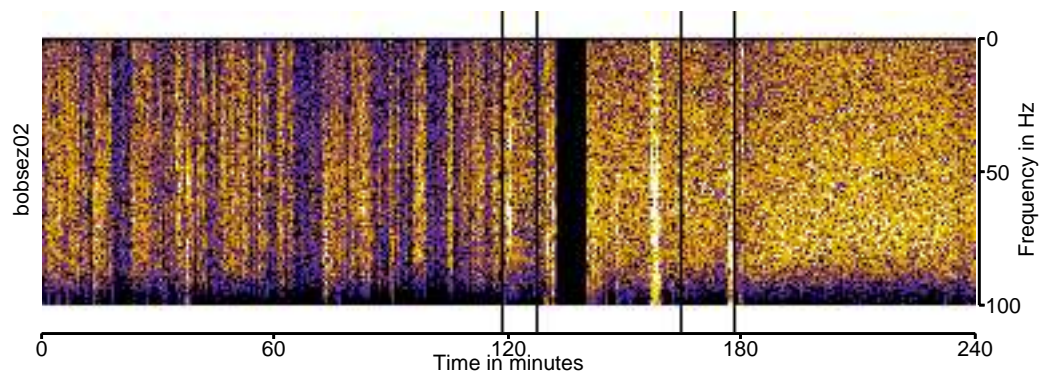


Figure 2.11: Periodogram obtained using the first intracranial signal in recording *eveawake* with nonoverlapping windows of 8192 samples (41 sec). A high level of activity is found in this recording. This characterizes several of the “awake” recordings. See also caption to Fig. 2.9.

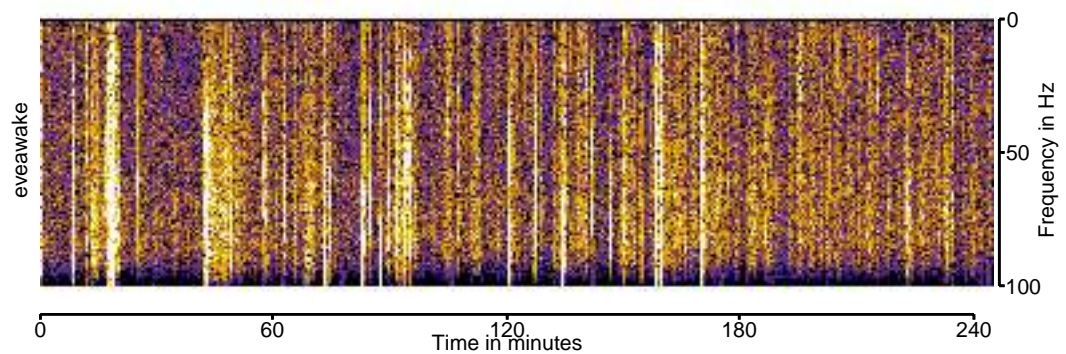


Figure 2.12: Periodogram obtained using the first intracranial signal in recording *carlsleep* with nonoverlapping windows of 8192 samples (41 sec). The “sleep” recordings generally present uniform spectra which vary only slowly in time. See also caption to Fig. 2.9.

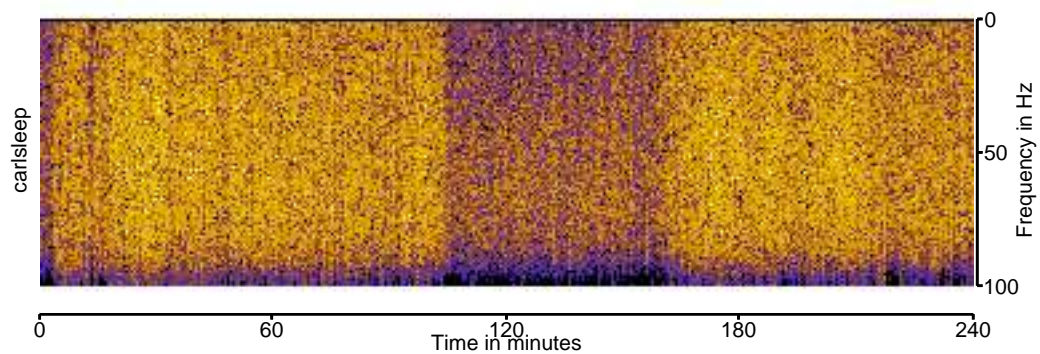
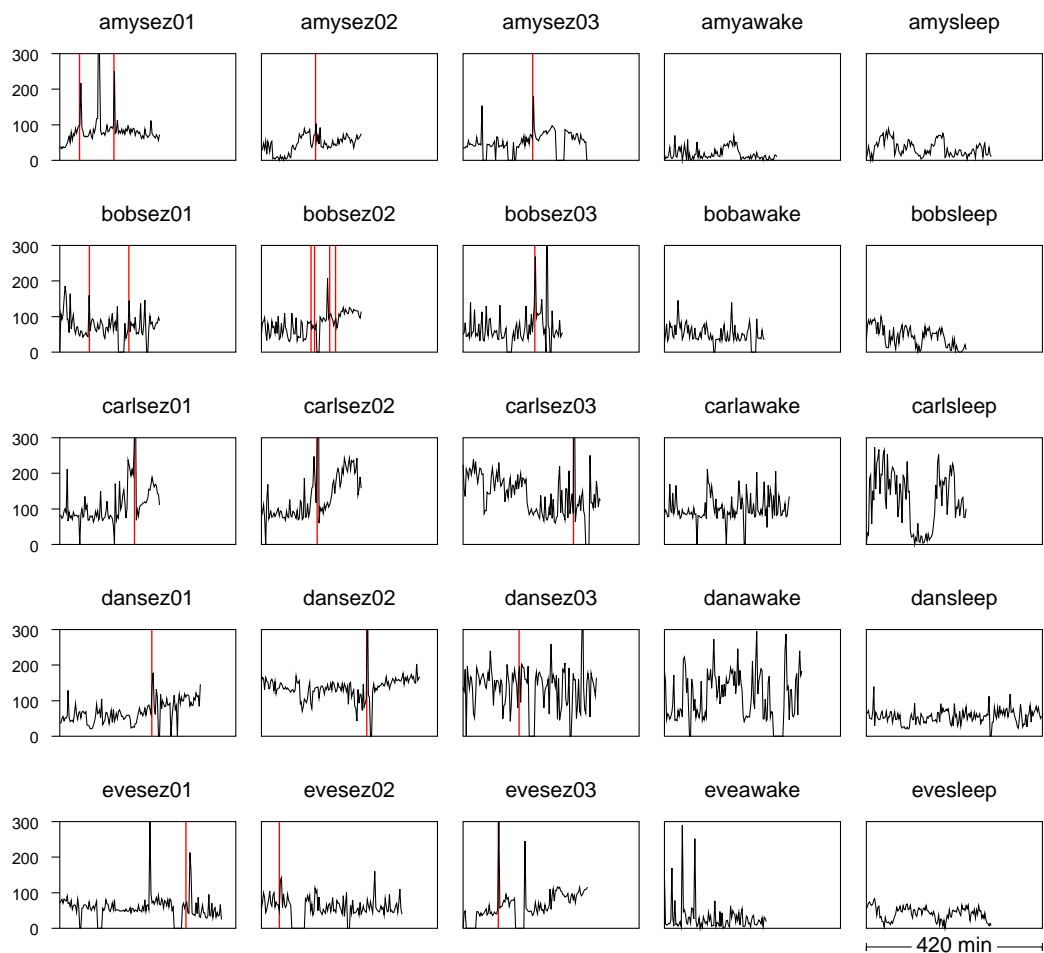


Figure 2.13: These plots show the change in the *standard deviation of the signal*, which reflects signal amplitude, computed on nonoverlapping windows 30 sec in length (60000 samples) using the first intracranial signal in each recording. The full width of each box corresponds to 420 min (i.e. the length of the longest recording, *dansleep*). Red vertical lines mark the start of clinical seizures. We see that several of the seizures (see *amysez01*, *amysez03* and *carlsez01*, for example) are preceded by dramatic increases in SD, i.e. signal amplitude.



## Chapter 3

# Dynamical Similarity

*Dynamically similar processes explore similar regions of phase space.* That is the simple idea behind the nonlinear measure studied in this chapter. That one can compute a meaningful measure of this similarity on a scalar time series is, on the other hand, quite astonishing at first. Like most of the tools of nonlinear time series analysis, this measure is based on dynamical systems theory [37] and begins with phase space reconstruction. The approach, which we call simply dynamical similarity, was pioneered by M. Le Van Quyen and collaborators from the Cognitive Neuroscience and Cerebral Imaging Laboratory at de la Salpêtrière Hospital in Paris specifically as a means of anticipating epileptic seizures from EEG recordings [42]. We will come back to their results in section 3.3.

In the minutes or even hours prior to a seizure, the dynamics of the epileptic brain are believed to change from a “normal” *interictal* state to a *preictal* state in which conditions are put in place for a seizure. This *change* in state is expected to introduce considerable nonstationarity into EEG recordings. Consequently, the Paris group’s starting point was a nonstationarity measure introduced by Manuca and Savit [50].

An important though often implicit assumption of most methods of time series analysis, modelling and prediction is the stationarity of the system. Many a system goes unmodeled for lack of sufficient stationary data, i.e. a lot of it! If one could identify sets of disjoint segments of a time series which *sample systems in dynamically similar states*, then these could be cobbled together and considered as a group to

improve the statistics of these methods. Manuca and Savit proposed to compare the overlap in reconstructed phase space trajectories, using delay coordinate embedding of distinct segments of time series and Grassberger-Procaccia cross-correlation sums, to identify such similarity. In an effort to reduce the computational burden of the method while increasing its discriminating power on short, noisy time series, Le Van Quyen and his collaborators adapted this approach in several important ways. Most notably, intercrossing intervals (ICIs) between positive-going threshold crossings are used for the reconstruction instead of delay coordinates. Although phase space reconstruction using interspike and intercrossing intervals is not addressed explicitly by embedding theorems, such intervals have already been shown to capture much of the interesting dynamics in a variety of systems [35, 63, 72, 73]. Furthermore, Pikovsky *et al.* [65] have shown that the time crossings can be seen as phases of Poincaré sections of the system flow.

After defining the measure in the following section, we will describe several experiments on our artificial and EEG data sets. The goal of these investigations is to allow comparison with the IIF and reveal, if one exists, an optimal combination of the two methods.

### 3.1 Defining the Measure

Given a scalar time series  $S = \{s_1, \dots, s_N\}$  sampled at equal intervals  $\Delta t$ , let  $S_{\text{ref}}$  be a subset of  $S$ , a *window* of length  $\ell_{\text{ref}}$  samples. This window is assumed to sample adequately some reference dynamics of the original system; we wish to measure the similarity between these reference dynamics and those captured in test windows (TWs) on this and possibly other (compatible) TS.

### 3.1.1 Phase Space Reconstruction

The window is first converted to positive-going threshold crossing times, where the threshold is described in [44] and earlier works as the “zero of the signal”. We take this to be identically equal to zero. Another possibility would have been the mean value of  $S$ , which on our data is close to zero anyway. It should be noted that no detrending or preconditioning of any sort was done on any of the data, real or artificial, for our experiments with this measure. Let  $\{t_1, \dots, t_{N_x}\}$  denote the set of these threshold crossing times (for the window in question), which are in units of the sampling interval, such that

$$s_{t_n} < 0 \quad \text{and} \quad s_{t_{n+1}} > 0 \quad (3.1)$$

is satisfied for all  $t_n$ . In other words, no interpolation is done between points either side of the threshold. We assume that  $\Delta t$  is much shorter than the shortest characteristic time scale of the underlying process, in which case this discretization should not influence the results. The thresholding algorithm as described (and indeed implemented) will miss any crossings where points *equal* to zero to within machine precision separate points to either side of the threshold. It would probably have been more prudent, in retrospect, to use  $s_{t_n} \leq 0$  as the first condition in (3.1).<sup>1</sup> We note that the number of crossings  $N_x$  cannot generally be predicted from the length of the window  $\ell_{\text{ref}}$ ; windows of the same size at different positions in  $S$  will in general contain different numbers of crossings.

The set of  $N_x$  times  $\{t_i\}$  are subsequently converted to  $N_x - 1$  intercrossing intervals (ICIs) denoted  $\{I_i\}$  such that  $I_n = t_{n+1} - t_n$ . From these ICIs, we construct  $N_{\text{ref}} = N_x - m$  embedding vectors which we denote  $\{\mathbf{A}_i\}$  where  $\mathbf{A}_n = (I_n, I_{n-1}, \dots, I_{n-m+1})$ ,

---

<sup>1</sup>A quick search for points less than 0.0001 (actually, matching “0.0000”) in the first neuron output in the artificial data  $c$  series produced *no* points out of a possible  $1.024 \times 10^6$ . It is clear that data which would satisfy the condition  $x = 0$  in our programs are similarly rare in the EEG recordings. Small changes in the threshold or to the way in which threshold crossings are identified will have little influence on the series of ICIs derived from any given time series.

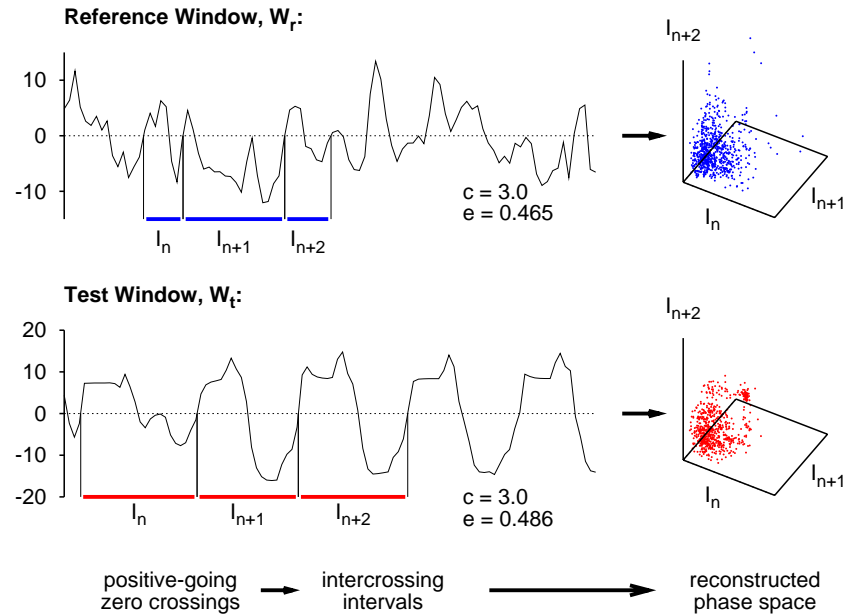


Figure 3.1: A schematic representation of the phase space reconstruction process used in computing the dynamical similarity measure. In this example, the “clouds” of points formed in reconstruction space differ in several obvious ways. The same scale is used in both reference and test plots.

defining an  $m$ -dimensional embedding of the reference dynamics. It is helpful to visualize these vectors as forming a “cloud” of points in the reconstructed phase space. Fig. 3.1 depicts the phase space “clouds” obtained from this process for  $m = 3$  using some of the artificial time series, hence the stated values of  $c$  and  $e$ .

### 3.1.2 Finding the Optimal Basis: Singular Value Decomposition

Application of a singular value decomposition (SVD) of the  $m$ -dimensional embedding space allows further improvement of the dynamical reconstruction for shorter time series [?] and additional nonlinear filtering of noise by “identifying the optimum basis for deployment of the trajectories” [44]. In other words, we project the reconstruction

vectors onto the  $m' < m$  basis vectors corresponding to the largest singular values and in so doing, emphasize the main features of the reconstructed reference dynamics because most of the variance in the vectors  $\{\mathbf{A}_i\}$  is in these directions. Let  $\underline{\underline{\mathbf{A}}}(S_{\text{ref}})$  be the trajectory matrix of the reference window; the  $N_{\text{ref}}$  embedding vectors  $\{\mathbf{A}_i\}$  constructed from  $S_{\text{ref}}$  constitute the rows of this  $m$ -column matrix.

To perform the SVD, we used packaged routines from *Numerical Recipes in C* [66], where we find the following definition of SVD. Any  $N \times m$  matrix  $\underline{\underline{\mathbf{A}}}$  can be written as the product of an  $N \times m$  column-orthogonal matrix  $\underline{\underline{\mathbf{U}}}$ , an  $m \times m$  diagonal matrix  $\underline{\underline{\mathbf{W}}}$  with positive or zero elements and the transpose of an  $m \times m$  orthogonal matrix  $\underline{\underline{\mathbf{V}}}$ , i.e.

$$\underline{\underline{\mathbf{A}}} = \underline{\underline{\mathbf{U}}} \cdot \underline{\underline{\mathbf{W}}} \cdot \underline{\underline{\mathbf{V}}}^T \quad (3.2)$$

For ease of reference, we list the dimensions of these various matrices below:

$$\underline{\underline{\mathbf{A}}} \quad N_{\text{ref}} \times m \quad (3.3)$$

$$\underline{\underline{\mathbf{U}}} \quad N_{\text{ref}} \times m \quad (3.4)$$

$$\underline{\underline{\mathbf{W}}} \quad m \times m \quad (3.5)$$

$$\underline{\underline{\mathbf{V}}} \quad m \times m \quad (3.6)$$

$$\underline{\underline{\mathbf{B}}} \quad m \times m' \quad (3.7)$$

$$\underline{\underline{\mathbf{A}'}} \quad N_{\text{ref}} \times m' \quad (3.8)$$

The diagonal elements of  $\underline{\underline{\mathbf{W}}}$ ,  $\mathbf{w} = \{w_1, \dots, w_m\}$ , are called the *singular values* of  $\underline{\underline{\mathbf{A}}}$ . The routine `svdcmp` returns  $\underline{\underline{\mathbf{U}}}$ ,  $\mathbf{w}$  and  $\underline{\underline{\mathbf{V}}}$  for a given  $\underline{\underline{\mathbf{A}}}$ ; we used a double precision version of this routine, `dsvdcmp`. Subsequently,  $\mathbf{w}$  and  $\underline{\underline{\mathbf{V}}}^T$  (in which the columns are the spanning vectors) are passed to `eigsrt`, which sorts the singular values in decreasing order and the associated columns of  $\underline{\underline{\mathbf{V}}}^T$  correspondingly. The first  $m'$  columns of the sorted  $\underline{\underline{\mathbf{V}}}^T$  then constitute the vectors onto which  $\underline{\underline{\mathbf{A}}}(S_{\text{ref}})$  is projected. If we denote the matrix of these  $m'$  basis vectors as  $\underline{\underline{\mathbf{B}}}(S_{\text{ref}})$  (effectively truncating  $\underline{\underline{\mathbf{V}}}^T$  at  $m'$  columns), the transformed trajectory matrix is obtained by



matrix multiplication:

$$\underline{\underline{\mathbf{A}'}}(S_{\text{ref}}) = \underline{\underline{\mathbf{A}}}(S_{\text{ref}}) \cdot \underline{\underline{\mathbf{B}}}(S_{\text{ref}}). \quad (3.9)$$

### 3.1.3 Similarity Index

We now define a measure of dynamical similarity called the similarity index which gauges the extent of overlap between the points visited in the reconstructed reference dynamics and those in a test segment of length  $\ell_t$  from, typically, the same series  $S$ . We denote this TW by  $S_t$  and proceed as described in §3.1.1 to obtain a trajectory matrix  $\underline{\underline{\mathbf{A}}}(S_t)$  composed of  $N_t$   $m$ -dimensional embedding vectors for dynamics sampled on this window. These vectors are then projected onto the *same*  $m'$  optimal basis vectors  $\underline{\underline{\mathbf{B}}}(S_{\text{ref}})$  as the reference vectors by computing

$$\underline{\underline{\mathbf{A}'}}(S_t) = \underline{\underline{\mathbf{A}}}(S_t) \cdot \underline{\underline{\mathbf{B}}}(S_{\text{ref}}). \quad (3.10)$$

Let  $\mathbf{V}_i(S_\mu)$  denote the  $i^{\text{th}}$  embedding vector for the segment  $S_\mu$  projected onto the optimal basis (i.e. the  $i^{\text{th}}$  row of  $\underline{\underline{\mathbf{A}'}}(S_\mu)$ ) and  $V_i^k(S_\mu)$  the  $k^{\text{th}}$  component of that vector. The similarity index we compute is based on Grassberger-Procaccia correlation sums [29]. The cross-correlation integral of two sets of reconstruction vectors represented by  $\underline{\underline{\mathbf{A}'}}(S_\mu)$  and  $\underline{\underline{\mathbf{A}'}}(S_\nu)$  is defined as

$$C(\underline{\underline{\mathbf{A}'}}(S_\mu), \underline{\underline{\mathbf{A}'}}(S_\nu)) = \frac{1}{N_\mu N_\nu} \sum_{i=1}^{N_\mu} \sum_{j=1}^{N_\nu} \Theta(r - \|\mathbf{V}_i(S_\mu) - \mathbf{V}_j(S_\nu)\|), \quad (3.11)$$

where  $\Theta$  is the Heaviside function defined in (2.19) and  $r$  is the radius of a hyperspherical neighborhood around  $\mathbf{V}_i(S_\mu)$ . In our implementation, the Euclidean norm was used:

$$\|\mathbf{V}_i(S_\mu) - \mathbf{V}_j(S_\nu)\| = \left[ \sum_{k=1}^{m'} (V_i^k(S_\mu) - V_j^k(S_\nu))^2 \right]^{\frac{1}{2}}. \quad (3.12)$$

The cross-correlation integral gives the fraction of pairs of points across the two windows which are separated by  $r$  or less and hence is a measure of overlap.

There are two main considerations in choosing  $\ell_{\text{ref}}$  and  $\ell_t$ . The longer the relatively stationary segment of  $S$  chosen as the reference window, the better will be our representation of the underlying reference dynamics in the reconstructed phase space. For example, a longer window on EEG recordings from an epileptic patient is more likely include interictal spikes. On the other hand, to improve the temporal resolution of the measure, the TWs need to be kept as short as possible. Consequently,  $\ell_{\text{ref}}$  will typically be several times  $\ell_t$ . The computational burden of this method can be significantly reduced without overly compromising the statistics of the measure by downsampling the embedding vectors on the reference dynamics, randomly choosing a fraction  $\ell_t/\ell_{\text{ref}}$  of the available vectors. The more frequently visited regions of phase space of the reference dynamics are presumably better represented in  $\underline{\underline{\mathbf{A}}}'(S_{\text{ref}})$ ; there is a higher density of points in these regions. It stands to reason then that uniform random sampling of these vectors will yield a rarefied version of the reference dynamics cloud, preserving its most significant features.<sup>2</sup> There was concern that using only a single realization of the reference dynamics would either bias the results or increase the size of fluctuations. To be sure, the use of fewer vectors for the reference dynamics influences the statistics of the measure to some extent. However, experiments with averaging the similarity index, defined below, across multiple realizations did not support this concern. Note that  $N'_{\text{ref}} = (\ell_t/\ell_{\text{ref}})N_{\text{ref}}$  does not necessarily equal  $N_t$ .

At this point, it is convenient to let  $\underline{\underline{\mathbf{X}}} \in \underline{\underline{\mathbf{A}}}'(S_{\text{ref}})$  denote this realization of the reference dynamics and  $\underline{\underline{\mathbf{Y}}} = \underline{\underline{\mathbf{A}}}'(S_t)$  the set of embedding vectors for a given TW.

---

<sup>2</sup>Our initial implementation of this measure used a Monte Carlo approach to resample  $\underline{\underline{\mathbf{A}}}'(S_{\text{ref}})$ . A very coarse  $m'$ -dimensional hyperhistogram of this set of vectors was constructed and used to randomly select bins according to the distribution of points among the bins. From each bin selected, one of the vectors contained in the bin was chosen at random and added to the downsampled reference set. Since *outliers* in each dimension effectively determined the size of the bins, the overwhelming majority of points tended to fall in a single bin and hence this bin was chosen most of the time. So in effect, we were just uniformly sampling the reference vectors anyway!

The cross-correlation ratio

$$\gamma[\underline{\underline{\mathbf{X}}}, \underline{\underline{\mathbf{Y}}}] = \frac{C(\underline{\underline{\mathbf{X}}}, \underline{\underline{\mathbf{Y}}})}{\sqrt{C(\underline{\underline{\mathbf{X}}}, \underline{\underline{\mathbf{X}}}) \cdot C(\underline{\underline{\mathbf{Y}}}, \underline{\underline{\mathbf{Y}})}}} \quad (3.13)$$

is defined as the **similarity index** between these sets of vectors, which is to say between the the dynamics represented in these two embeddings.

The *same* realization of the reference dynamics  $\underline{\underline{\mathbf{X}}}$  was used in computing  $\gamma$  on *all* the TWs on a given TS in our implementation. However, when the same  $S_{\text{ref}}$  from one TS was used in conjunction with TWs from each of a series of TS, a different realization of  $\underline{\underline{\mathbf{X}}}$  was used for each TS. For the same reason that averaging  $\gamma$  for a given TW over several realizations of  $\underline{\underline{\mathbf{X}}}$  was found to have little influence on the result, namely that the values of  $\gamma$  for each realization were quite similar, there is no reason to believe that using the same realization for all TS would have had much impact on our findings. Finally, on TWs where the number of reconstruction vectors  $N_t$  did not satisfy the inequality

$$N_t > 0.1 \times (\ell_t/\ell_{\text{ref}}) \times N_{\text{ref}} = 10\% \text{ of } N'_{\text{ref}}, \quad (3.14)$$

the value of  $\gamma$  was taken to be identically equal to zero. This is true also of windows on which no pairs were separated by  $r$  or less, such that  $C(\underline{\underline{\mathbf{Y}}}, \underline{\underline{\mathbf{Y}}}) = 0$ .

### 3.1.4 Statistical Analysis

Let  $\{\underline{\underline{\mathbf{Y}}}_i\}$  be the set of  $q$  trajectory matrices derived from all the nonoverlapping TWs on the reference segment itself. We define the baseline statistics  $\mu$  and  $\sigma$  as the mean and standard deviation of the similarity index computed on these TWs, i.e.

$$\mu = \frac{1}{q} \sum_{i=1}^q \gamma[\underline{\underline{\mathbf{X}}}, \underline{\underline{\mathbf{Y}}}_i], \quad (3.15)$$

and

$$\sigma = \left[ \frac{1}{q} \sum_{i=1}^q [\gamma(\underline{\underline{\mathbf{X}}}, \underline{\underline{\mathbf{Y}}}_i) - \mu]^2 \right]^{\frac{1}{2}}. \quad (3.16)$$

The **normalized deviation** of the similarity index  $\gamma$  is defined as

$$\Sigma = \frac{\gamma - \mu}{\sigma}. \quad (3.17)$$

The statistical significance ( $p$ -value) of fluctuations in  $\gamma$  is determined from the value of  $\Sigma$  relative to a threshold value of  $\sigma_{\text{th}}$  by virtue of Chebyshev's inequality which applies to any statistical distribution of  $\gamma$ ,  $P(|\Sigma| > \sigma_{\text{th}}) < 1/\sigma_{\text{th}}^2 = p$ . For a threshold of  $\sigma_{\text{th}} = 5$ , for example, this says that the null hypothesis for a value of  $|\Sigma|$  above 5 can be rejected at the 96% confidence level.

We now define two *temporal* thresholds which allow us to identify **hot zones** on the TS, intervals on which the statistically significant fluctuations in  $\gamma$  also satisfy certain conditions of persistence in time across multiple TWs. Two conditions are satisfied in any “hot” sequence of TWs. First, the sequence begins with TWs spanning an interval of at least  $\tau_{\text{th}}$  on which  $\Sigma > \sigma_{\text{th}}$ . Second, after these initial TWs, the normalized deviation  $\Sigma$  never falls below  $\sigma_{\text{th}}$  for longer than  $\tau_{\text{g}}$  (“g” is for “grace period”). We call these three values,  $\sigma_{\text{th}}$ ,  $\tau_{\text{th}}$  and  $\tau_{\text{g}}$ , the **threshold parameters**. Fig. 3.2 illustrates their use in determining hot zones.

## 3.2 Experiments on Artificial Data

The purpose of these experiments is to assess the influence of window lengths and sampling frequency on the power of this measure to discriminate different levels of chaoticity (as reflected by increasingly positive maximal Lyapunov exponents) or degrees of intermittency. Ultimately we want to compare the results of this section to those obtained with IIF and to guide us in our choice of parameters in experiments on real data. Two window lengths  $\ell_w = 1024$  and  $\ell_w = 8192$  were used on data of each of the sampling intervals,  $\Delta t = 1.0$  and  $\Delta t = 10.0$ . For reconstruction,  $m = 10$  throughout and  $m' = 4$  (optimal basis vectors).

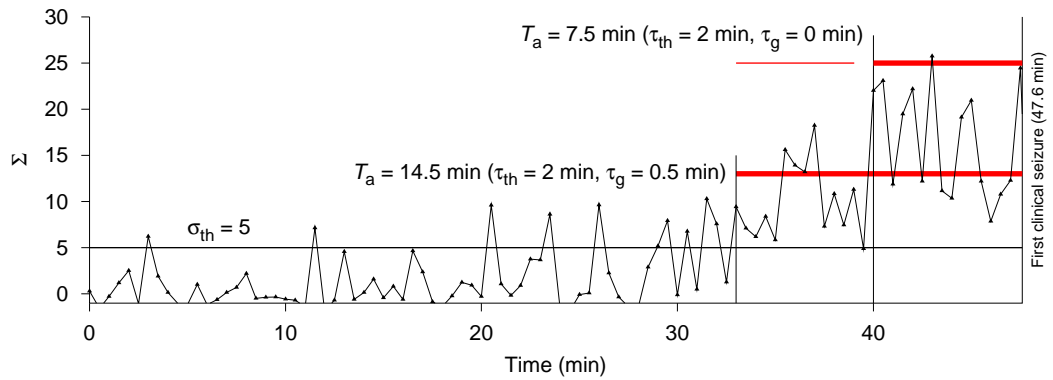


Figure 3.2: A short “grace period” can have a dramatic effect on anticipation times. Shown in this figure is the normalized deviation  $\Sigma$  on channel “LA1” from recording “amysez01” up to the first clinical seizure (first reference window). Using a deviation threshold of  $\sigma_{th} = 5$ , a minimum duration of  $\tau_{th} = 2 \text{ min}$  (4 TWs) *and* a grace period of  $\tau_g = 0.5 \text{ min}$  (1 TW), an anticipation time  $T_a$  of 14.5 min is obtained with no hot zones other than that announcing the seizure. With  $\tau_g = 0$ , on the other hand,  $T_a$  drops to 7.5 min while a 6 min hot zone not directly associated with the seizure now appears.

### 3.2.1 The $c$ Series

#### 3.2.1.1 Methods

For each pair of values  $(c_{\text{ref}}, c_t)$  where  $c_{\text{ref}}, c_t \in \{1.0, 2.0, 3.0, 5.0, 10.0\}$  including  $c_{\text{ref}} = c_t$ , the following calculation was made. Let  $S_i^{\Delta t}(c_j)$  denote the first  $\ell_w$  samples of the  $i^{\text{th}}$  time series sampled at  $\Delta t$  for  $c_j$ . For all pairs  $(i, j)$  where  $i, j = 1, \dots, 25$ , the trajectory matrices  $\underline{\underline{\mathbf{A}}}(S_i^{\Delta t}(c_{\text{ref}}))$  and  $\underline{\underline{\mathbf{A}}}(S_j^{\Delta t}(c_t))$  were constructed, the  $m'$  optimal basis vectors for  $\underline{\underline{\mathbf{A}}}(S_i^{\Delta t}(c_{\text{ref}}))$  were found and both trajectory matrices projected onto this basis, resulting in  $\underline{\underline{\mathbf{A}}}'(S_i^{\Delta t}(c_{\text{ref}}))$  and  $\underline{\underline{\mathbf{A}}}'(S_j^{\Delta t}(c_t))$  respectively. The value  $\gamma[\underline{\underline{\mathbf{A}}}'(S_i^{\Delta t}(c_{\text{ref}})], \underline{\underline{\mathbf{A}}}'(S_j^{\Delta t}(c_t))$  was computed for each pair  $(i, j)$  and the average  $\langle \gamma \rangle$  and standard deviation  $\gamma_{\text{SD}} = \sqrt{\langle \gamma^2 \rangle - \langle \gamma \rangle^2}$  of the resulting  $25^2 = 625$  similarity indices computed. The results for each  $(c_{\text{ref}}, c_t)$  pair were output to an ASCII file. We note that since the reference and test windows were both of length  $\ell_w$ , there was no need to downsample the reference window. For  $c_{\text{ref}} = c_t$  we expect values close to one and hope to find that the difference in results for  $(c_{\text{ref}}, c_{\text{ref}})$  and  $(c_{\text{ref}}, c_t)$  is somehow proportional to the difference in, for example, the maximal Lyapunov exponent found for such values of  $c$ .

#### 3.2.1.2 Results and Discussion

Figures 3.3 to 3.6 show the results for the different combinations of  $\ell_w$  and  $\Delta t$ .

All the  $(c_{\text{ref}}, c_{\text{ref}})$  pairs (in red in the figures) did indeed yield values close to 1 except in Fig. 3.3 ( $\ell_w = 1024, \Delta t = 1.0$ ). There was a lot of variation in those results and even a certain number of zeroes. On the  $\Delta t = 1.0$  results, two distinct subsets of  $c$  are apparent,  $A = \{1\}$  and  $B = \{2, 3, 5, 10\}$ . Results between members of the sets are symmetrical, i.e.  $\langle \gamma(c_{\text{ref}} \in A, c_t \in B) \rangle$  approximately equals  $\langle \gamma(c_{\text{ref}} \in B, c_t \in A) \rangle$ . Furthermore, for all  $(c_{\text{ref}} \in B, c_t \in B)$  pairs, we find  $\langle \gamma \rangle \approx 1$ . Dynamical similarity does not appear to be sensitive to the differences between time series within these sets.

On the data sampled at intervals of  $\Delta t = 10.0$ , the error bars on the short  $\ell_w =$

Figure 3.3: Similarity results on the  $c$  series with  $\ell_w = 1024$  and  $\Delta t = 1.0$ . Shown are  $\langle \gamma \rangle$  and  $\gamma_{SD}$  over all  $25^2$  combinations of artificial times series for each  $(c_{ref}, c_t)$  pair. Results for which  $c_{ref} = c_t$  are in red. Dotted lines divide the figure into horizontal strips, to each of which is associated a  $c_{ref}$  value. The scale bar to the right of the figure indicates the range of values of  $\gamma$ , which here is the full possible range from 0 to 1.

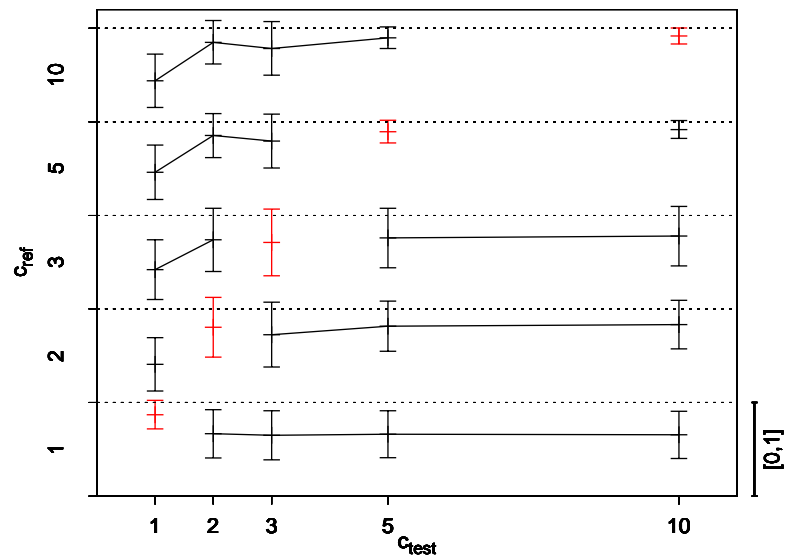


Figure 3.4: Similarity results on the  $c$  series with  $\ell_w = 1024$  and  $\Delta t = 10.0$ . Further details in the caption to Fig. 3.3.

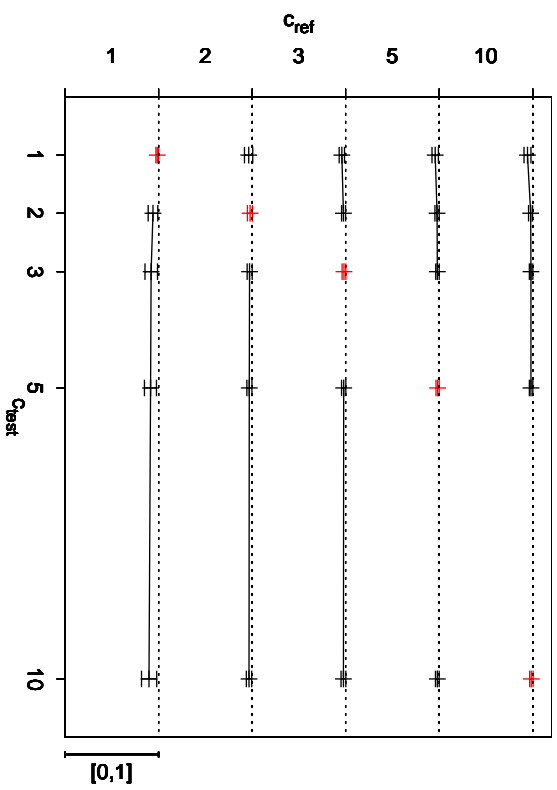


Figure 3.5: Similarity results on the  $c$  series with  $\ell_w = 8192$  and  $\Delta t = 1.0$ . Further details in the caption to Fig. 3.3.

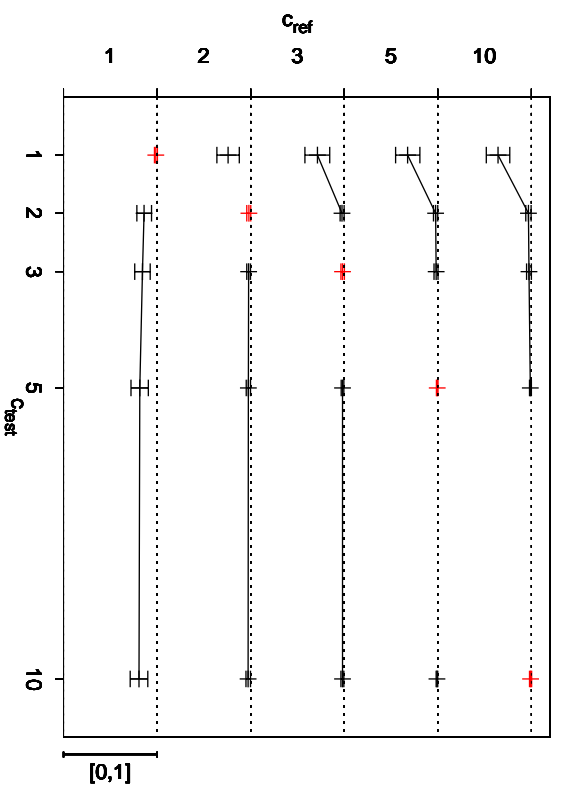
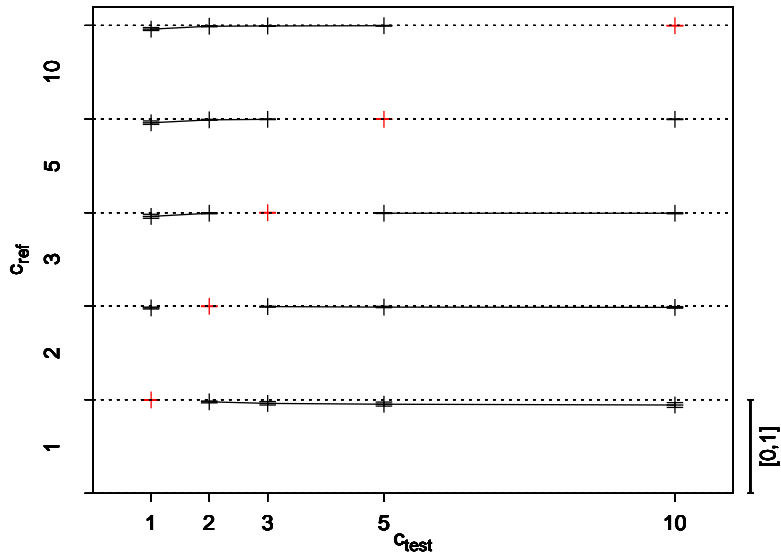




Figure 3.6: Similarity results on the  $c$  series with  $\ell_w = 8192$  and  $\Delta t = 10.0$ . Further details in the caption to Fig. 3.3.



1024 windows are comparable to the  $\ell_w = 8192$  results on the  $\Delta t = 1.0$  data. The distinctions between the  $A$  and  $B$  sets have all but vanished. This suggests that these TS are *undersampled*.

We were also interested in the influence of noise on this measure, but it was not clear how best to introduce such noise. In the end, we looked at additive noise. Let  $\sigma$  be the standard deviation of the time series  $\{x_i\}$ ,  $i = 1, \dots, N$ , to which noise is to be added. A series of zero-mean Gaussian white noise,  $\{\xi_i\}$ , was generated with a standard deviation equal to a fraction  $p$  of  $\sigma$ . A new series was formed by adding the raw signal and the corresponding noise term,  $x_i \rightarrow x_i + \xi_i$ . For each of the time series in the  $c$  set, a single realization of the noisy data was produced. The above experiment was then repeated. For values of  $p < 0.05$ , there was no noticeable difference between the results obtained with the noisy data and those obtained with the raw data. At higher noise levels, there was an overall increase in similarity across all  $(c_{\text{ref}}, c_t)$  pairs and a reduced spread in results. This is to be expected as the original dynamics drowns in the rising noise, so that eventually we are comparing a random cloud of

points to another equally random cloud.

### 3.2.2 The $e$ Series

We recall that in this set of artificial time series, it is the quality of the intermittency which vary; the frequency of occurrence and duration of bursts of periodicity increase with  $e$  (for typical output, see Fig. 2.3).

#### 3.2.2.1 Methods

Once again, for each pair of  $e$  values ( $e_{\text{ref}}, e_t$ ), we found the mean and standard deviation of the similarity results on the  $25^2$  combinations of time series,  $\langle \gamma \rangle$  and  $\gamma_{\text{SD}}$  respectively. Two window lengths,  $\ell_w = 1024$  and  $\ell_w = 8192$  were used for each of the two sampling intervals  $\Delta t = 1.0$  and  $\Delta t = 10.0$ .

#### 3.2.2.2 Results and Discussion

The results are presented in Figures 3.7 to 3.10.

One immediately notices that while a window length of  $\ell_w = 1024$  was sufficient to discern crudely the two sets on the  $c$  series sampled at  $\Delta t = 1.0$  (Fig. 3.3), it is sorely inadequate for discriminating between the members of the  $e$  series (Fig. 3.7). However, the results for the same window length on the “undersampled” data ( $\Delta t = 10.0$ ), Fig. 3.8, reveal that differences between the  $e$  series members were not obscured by an insufficient number of data in the window but rather by the windows *covering too short an interval of time*. This reflects the intermittent nature of these time series; to capture the full dynamics of the system, the reference window must be long enough in time to contain both periodic and chaotic behaviors. Interestingly, the length in time of the windows in Fig. 3.9 is close to that in Fig. 3.8 but the mean differences between  $e$  values are much smaller and all of the values are closer to 1 (note the difference in vertical scale between the two figures). Of course, the size of the error bars is also smaller. When sampled at a higher rate, then, the similarities between

Figure 3.7: Similarity results on the  $e$  series with  $\ell_w = 1024$  and  $\Delta t = 1.0$ . Shown are  $\langle \gamma \rangle$  and  $\gamma_{SD}$  from all  $25^2$  combinations of artificial times series for each  $(e_{ref}, e_t)$  pair. Dotted lines divide the figure into horizontal strips, to each of which is associated an  $e_{ref}$  value. The scale bar to the right of the figure indicates the range of values of  $\gamma$ , which here is the full possible range from 0 to 1.

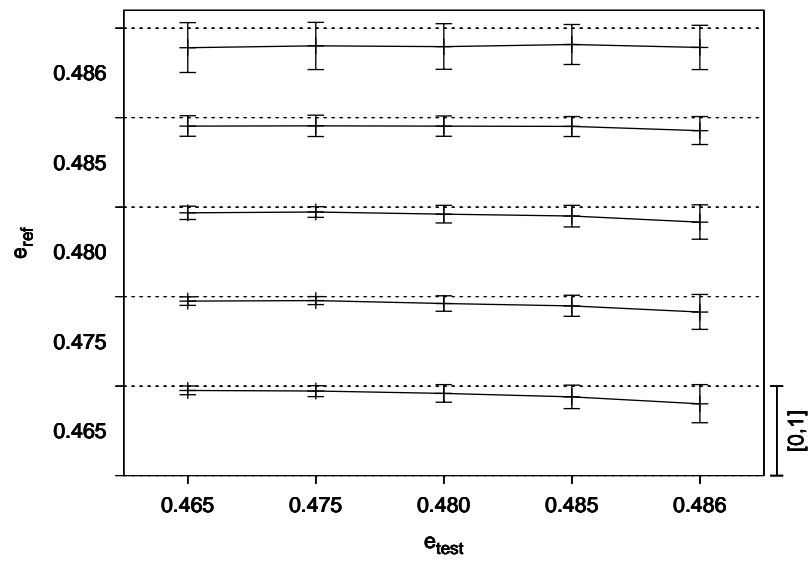


Figure 3.8: Similarity results on the  $e$  series with  $\ell_w = 1024$  and  $\Delta t = 10.0$ . Further details in the caption to Fig. 3.7. Note that here the vertical scale is from 0.0 to 0.75 in each horizontal strip.

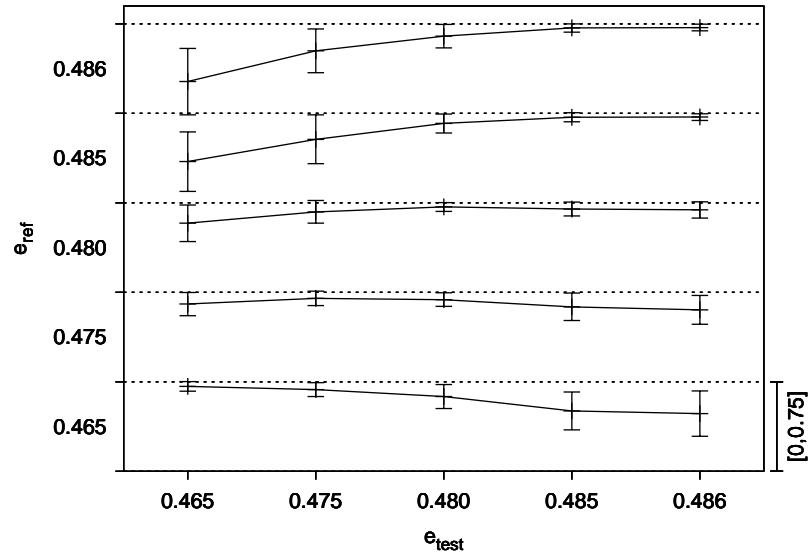


Figure 3.9: Similarity results on the  $e$  series with  $\ell_w = 8192$  and  $\Delta t = 1.0$ . Further details in the caption to Fig. 3.7. Note that here the vertical scale is from 0.75 to 1.0 in each horizontal strip.

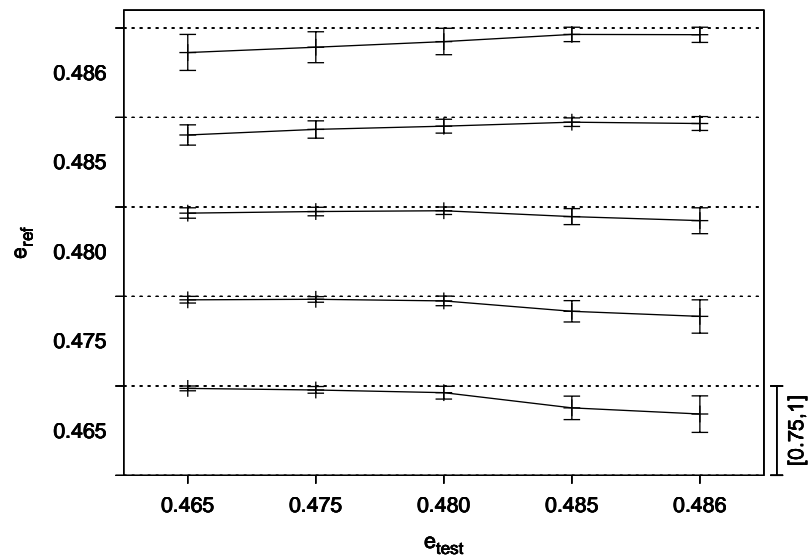
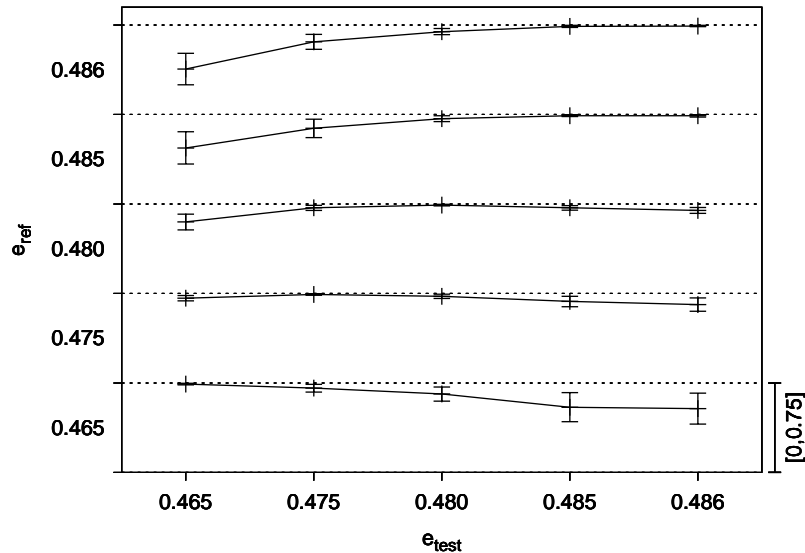


Figure 3.10: Similarity results on the  $e$  series with  $\ell_w = 8192$  and  $\Delta t = 10.0$ . Further details in the caption to Fig. 3.7. Note that here the vertical scale is from 0 to 0.75 in each horizontal strip.



the series become more apparent; we see that the similarity measure is sensitive to this.

Clearly, the best discrimination—largest difference between results for different  $e$  values with the smallest error bars—is obtained with  $\ell_w = 8192$  on the data sampled at intervals of  $\Delta t = 10.0$ , in other words with the most points over the longest time (Fig. 3.10). Since the measure had no discriminating power on the undersampled  $c$  series, however, we conclude that the use of the longer window and the smaller sampling interval  $\Delta t = 1.0$  would be optimal for both the  $c$  and  $e$  sets combined.

### 3.3 Experiments on EEG Recordings

One of the objectives of the following experiments was naturally to compare the anticipation times obtained on the data described in Chapter 2.2 to those found by Michel LeVan Quyen and coauthors [42–44]. We also wanted to test the idea that the reference window (RW) truly captures “normal” interictal dynamics by computing

the similarity between the RW from one recording and the TWs on another from the same patient. The hypothesis is that any “good” RW will be sensitive not only to the changes preceding a seizure on the same recording but also to those on *any* recording from the same patient; this allows for variation in “normality” between individuals. The influence of changes in arousal states between and within recordings must not, however, be neglected.

The focus of these previous studies was mainly on the appearance of hot zones in the  $\sim 5$  min period just before clinical seizure onset. The positive findings were taken as evidence that there is a distinct phase, the *preictal* phase, during which the dynamics of the epileptic brain change significantly in preparation for a seizure. If hot zones appeared long enough before the seizure and *only* before seizures, such an event could trigger any one of a number of methods proposed for preventing the seizure or at the very least warn the person of what’s coming. But hot zones which do not continue *uninterrupted* to a seizure *also* appear, as we can see even in Le Van Quyen’s results. We will call these **free hot zones**, as opposed to **bound hot zones** which overlap the start of a clinical seizure. What exactly the free hot zones mean is uncertain—maybe the brain enters the preictal phase by fits and starts—and this begs the larger question of the *specificity*<sup>3</sup> of any method for anticipating

---

<sup>3</sup>Possible results for any binary diagnostic test can be classified as positive or negative. For example, in the case of an illness, either you have it or you don’t. A positive result on a case that *should* test positive is called a true positive and a negative result on such a case is a false negative. Similarly, a negative result on a case that *should* test negative is a true negative and a positive results a false positive. How does one know what the answer *should* be? Sometimes one has *a posteriori* knowledge. The correct result to a test designed to answer the question “Will patient X die in the next 24 hours?” is relatively easy to get. Just wait a day. Other times, one uses the response of the best available test, the so-called “gold standard”. So the power of a new diagnostic tool, which may be faster or cheaper or less invasive than the gold standard test, is assessed relative to this preexisting tool. The **sensitivity** of the test is defined as the ratio of true positives to the number of true positives plus false negatives, i.e. it measures how often a positive condition is in fact detected by the test. The **specificity**, on the other hand, measures how often a positive response actually indicates what it is supposed to; it is defined as the ratio of true positives

seizures based on the similarity measure, or any measure for that matter. Although the importance of this question is widely recognized, it has never been investigated *per se*.

To start with, it is clear that this issue cannot be addressed even qualitatively when only short recordings with seizures are analyzed. That is the main reason why we are using a data set consisting of recordings several *hours* long, ten of which are free of clinical seizures. In order to illuminate our discussion of specificity, we present figures *showing* all of the hot zones on every channel over the entire duration of each recording, and this for several combinations of threshold parameters. For the same parameter values, the average number and average total duration of free hot zones per channel on each recording will also be tabulated. Three of the recordings capture multiple seizures, however, in tabulating statistics we will never look beyond the first seizure. We have to assume that the recordings begin with the patient in the interictal states. If the recordings had been even longer, we might have required that any segment chosen as a reference window be preceded by two or more seizure-free hours, but we could not afford that luxury.

For later reference, we will summarize the study documented in [44], published in *The Lancet* in 2001 and the methods of which served as a starting point for our own experiments. 26 recordings from 23 patients with temporal lobe epilepsy (TLE) were analyzed. Of these, 18 were 60 – 90 min in duration including 30 – 60 min prior to seizure and were taken from surface electrodes only (between 21 and 27 per recording). The remaining 6 recordings were of 60 min duration including 50 min prior to seizure and were taken from a mixture of surface and intracranial contacts. Threshold parameters were  $\sigma_{\text{th}} = 5$ ,  $\tau_{\text{th}} = 30\text{s}$  (i.e. a single window) and  $\tau_{\text{g}} = 0\text{s}$ : anticipation times  $t_{a\ i}$  for each contact  $i$  were measured from the start of the earliest TW on which  $\Sigma > \sigma_{\text{th}}$  and after which  $\Sigma$  remained at or above  $\sigma_{\text{th}}$  right up to the over true positives plus false positives.

seizure. Only the earliest anticipation time

$$T_a = \max_i t_{ai} \quad (3.18)$$

across all contacts was reported for each recording. For 25 out of the 26 recordings, a nonzero  $T_a$  is reported, with mean 6.9 min and SD 5.9 min. The number of active channels (those on which  $t_{ai} \neq 0$ , by our definition) is indicated graphically for just the one “typical” recording, for which  $T_a = 18$  min is quoted. More than three quarters of the contacts are active and  $\langle t_{ai} \rangle$  is approximately 5 min.

Although the earlier studies [42, 43] used intracranial recordings and the methods and threshold criteria of all the papers differ to a certain extent, the general findings of the three studies are mutually consistent. In [43], the number of “active” channels in each recording is also reported, a result which we feel is quite important. A channel was active if  $\Sigma > 3$  was satisfied for any continuous<sup>4</sup> interval preceding the seizure. On recordings for which  $T_a \neq 0$ , an average of 53% of contacts were active (SD 29%). The variability of results between different recordings for the same patient is noted in all three studies. A significant correlation between the evolution of the measure on both surface and intracranial electrodes was observed in [44]. The authors also noted a tenancy for the state transition, as reflected in values of  $\Sigma$ , to “wax and wane and... occasionally [be] disrupted, in particular, shortly before seizure onset.”

### 3.3.1 Methods

For the measure parameters, we used those quoted in [44]. The reconstruction dimension  $m = 10$  and optimal basis dimension  $m' = 4$  were used, as they had been in the experiments on artificial data presented above. Reference windows were  $\ell_r = 300$  s in length and non-overlapping TWs of  $\ell_t = 30$  s length were used. The neighborhood size  $r$  was set to a level corresponding to 30% of the cumulative distribution of distances between pairs of points or **cumulative neighbour distribution** in the downprojected RW,  $\underline{\underline{\mathbf{A}}}'(S_{\text{ref}})$ .

---

<sup>4</sup>It is possible that  $\tau_g \neq 0$  as it was in [42]. The article is unclear on this point.



Unlike the Paris group’s, our reference windows were chosen without consideration for the presence of artifacts or interictal spikes—of which we had no knowledge—or any systematic evaluation of stationarity on the chosen interval. One way we make up for this is by testing reference windows across recordings for the same patient, because a window which by a combination of the criteria we do use and sheer luck captures the interictal dynamics of that brain in that region and results in a hot zone on one recording, should do so on the other too. In [43] and possibly [44], stationarity was assessed by inspection of the similarity matrix, obtained by computing the similarity index  $\gamma$  of all possible pairs of TW on a TS. *A posteriori*, we assert that the values of  $\sigma$  typically obtained on our reference windows support a claim of sufficient stationarity;  $\sigma$  was typically around 2% of  $\mu$ . On recordings capturing clinical seizures, the main criterion in our placement was distance in time from the first recorded seizure: the first window was placed as early as possible in the recording, the second about 20 min before the first seizure. Adjustments to the position were made to allow for gaps in the recording; a buffer of 10 to 30 min was maintained between the RWs and any gaps as well as the start and end of the recording. The times of all the RWs are illustrated in Fig. 2.8; the RW is chosen at the same time in all channels.

A header file is associated with each of our recordings which, among other things, codifies the channel types and provides a label for each channel. In these recordings, intracranial contacts (“SEEG” for subdural EEG) were type 2 and surface contacts (“EEG”) were type 1; channels of all other types were rejected. Furthermore, channels with labels containing any of the following character sequences (not case sensitive) were also rejected: “ekg”, “emg”, “eog”, “rpar”, “ref”. The remaining channels in each recording were **eligible channels**. On these, we computed the similarity index between each of the two reference windows and consecutive non-overlapping TWs moved over the whole recording.

Using the set of first reference windows only, we also did cross-recording computations for recordings from the same patient. A mapping between channels of different

recordings was automatically constructed on the basis of channel labels. For example, channel 30 in the recording “evesez03” with label “RH2” would be mapped to channel 47 in “evesleep”, which has the same label. Let  $W_i^1(q)$  denote the first reference window on channel  $q$  of recording  $i$  from a given patient and  $S_i(q)$  the entire time series for channel  $q$  in recording  $i$  from the same patient. If a mapping could be established between channel  $q$  in recording  $i$  and some channel  $q'$  in recording  $j$ , then the similarity index was computed between  $W_i^1(q)$  and consecutive non-overlapping TWs on  $S_j(q')$ . We recall that there are five recordings for each patient in the data set; this computation was repeated for all the  $(i, j)$  pairs<sup>5</sup> where  $i, j = 1, \dots, 5$ . The mapping and computation procedure was repeated for each patient separately.

For each combination of RW and time series (TS), the time from the start of the recording at the end of each TW (in numbers of samples) and the value of  $\gamma$  on that window were output to an ASCII file.<sup>6</sup> Two programs were written to reduce the results for a given run (i.e. given set of measure parameters and choice of reference window), one to tabulate<sup>7</sup> anticipation times and other statistics described below and another to generate the figures showing hot zones for a given set of threshold parameters.

For a given set of measure parameters, which is to say the results of a given run on the EEG data set, and values of the thresholds, the tabulation program reduces the results for each recording to the following (in order of column headings on Tables 3.1–

---

<sup>5</sup>Including, in fact, the case  $i = j$ . In this paper, all same-recording results obtained using the first RW were extracted from the set of results for the entire cross-recording batch.

<sup>6</sup>The naming convention for these files was  $[RW\ filename]_{-}[TS\ filename]_{-}[channel\ label].dat$ . So for example, the results of the similarity index computation using a RW from channel “RH1” in recording “carlsez03” on the TS for the identically labelled channel in recording “carlsez01” were written to “carlsez03\_carlsez01\_RH1.dat”.

<sup>7</sup>The tabulation program outputs its results in L<sup>A</sup>T<sub>E</sub>X tabular form which was then inserted as is into this document. The figure program generates a **Gnuplot** script which, when interpreted by the Gnuplot program, produces the PostScript files incorporated herein. Gnuplot is distributed under General Public License (GPL) and is available at <http://www.gnuplot.org>.

3.3):

**Active Channels** are those in which a bound hot zone occurs.

**Eligible Channels** are those whose type and label satisfy the criteria defined above.

**Earliest Anticipation** is the longest anticipation time. The anticipation time on any given active channel is the time in minutes from the *start* of the bound hot zone (i.e. the end of the first TW on which  $\Sigma > \sigma_{\text{th}}$ ) to the seizure (the start of the window in which the seizure begins).

**Mean Anticipation** is the average over all the *active* channels of anticipation times

**Mean Gap** is recorded because on a few recordings, gaps precede the first seizure. If the  $\sigma_{\text{th}}$  and  $\tau_{\text{th}}$  threshold conditions are satisfied at least one minute before the gap, we ignore the gap plus one minute either side of it and continue analyzing the results beyond the gap. If the hot zone containing the gap turns out to be bound (i.e. leads up to a seizure), we including the duration of the gap in this average over all the active channels.

**Length of Segment** that is considered in tabulating free hot zone statistics. For seizure-free recordings, this is the length in hours of the entire recording. Otherwise, it is the length of the recording up to the first seizure.

**Mean Free Hot Zones per Eligible Channel per Hour** is the total number of free hot zones in all *eligible* channels divided by the number of these channels and the length of the segment.

**Mean Total Duration of Free Hot Zones** is the total duration of all the free hot zones on the analyzed segment in all the eligible channels divided by the number of these channels.

Where the error is given, it is the standard deviation.

### 3.3.2 Results and Discussion

The first results we present are those for the threshold parameters used in [44], namely  $\sigma_{\text{th}} = 5$ ,  $\tau_{\text{th}} = 120$  s and  $\tau_{\text{g}} = 0$ , using the first reference window. All of the resulting hot zones, free and bound, are shown in Fig. 3.11 for both intracranial contacts (black) and surface contacts (purple). The purpose of this figure and the others like it is to give an overall appreciation of the behaviour of the measure for a given set of RWs and thresholds on *all* of the recordings, at the expense of resolution. Two of the more interesting plots from Fig. 3.11, *amysez01* and *carlsez01*, are reproduced with enlargement in Figs. 3.12 and 3.13 respectively.

Table 3.1 shows the statistics for the intracranial contacts and Table 3.2 for the surface contacts. The mean nonzero earliest anticipation time on intracranial contacts is 24 min (SD 34 min), but the mean anticipation times are obviously much lower than that. Averaging over the nonzero mean anticipation times, we get a value of 7 min (SD 6 min) for intracranial EEG, which resonates with the Paris group’s findings. Bound hot zones were found in 8 of the 15 first seizures on an average of 26% of eligible intracranial contacts (SD 26%) and results on surface contacts were similar, making allowances for the influence of a much smaller number of eligible contacts on the statistics. Using a lower deviation threshold of  $\sigma_{\text{th}} = 3$ , such that our active channels are comparable to the “involved channels” of [43], we find active channels in 10 of the 15 recordings with seizure, but the mean remains 26% of eligible channels (SD 26%). The hot zone figure plainly shows tremendous intra-patient and inter-patient variation in the behaviour of the measure. We should mention that adjustments in the threshold parameters could produce more optimal performance on this data set in some respect, such as bound hot zones on all first seizures, but we found that this was only at the expense of creating a lot of free hot zones. Small changes in the parameter values did not have a substantial influence on either the qualitative or quantitative results.

Free hot zones are rare in most of the pre-seizure intervals. On a recording like

“amysez02”, the significant deviations clearly precede the seizure, but  $\Sigma$  falls again just before the seizure so these zones are counted as free. This is an example of the “waxing and waning” of the state change observed previously. Adding a short grace period of  $\tau_g = 2$  min increases the number of active intracranial contacts on this recording by 7 to 14 out of 48, as shown in Table 3.3. Whereas a large number of active channels and large mean anticipation time suggest strong anticipation of the seizure in “carlsez03”, the figure reveals a seemingly random spatial distribution of bound hot zones and a distinct strip of free zones long before the seizure (the full width of the box corresponds to 420 min). Taking these considerations into account, a claim for anticipation on the basis of fluctuations in this measure seems justified only on the recordings of patient “amy” and “carlsez01” as well as the seizures in “bobsez02” other than the first. Hot zones mark at least the *beginning* of almost all the seizures (i.e. detection rather than anticipation of the seizure) and interestingly, seem to be dying out towards the end of the recording. This reflects a return to interictal dynamics. The “amy” recordings again and “evesez01” and “evesez02” are the most striking examples of this. Looking down the column of “awake” recordings, free hot zones are few and, when they occur, of relatively short duration. The “sleep” recordings are quite the opposite.

For the same values of the threshold parameters, the second reference windows yielded results generally consistent with the first but poorer anticipation, as Fig. 3.14 shows. We have not included a table for this experiment; suffice it to say that there are active channels in only 6 recordings and that these represent less than 10% of the eligible channels in every case except “amysez01”. With the RWs generally much closer to first seizure onset, it is not surprising to find a slight increase in free hot zones well before the seizures. Free hot zones remain sparse on the “awake” recordings and dense on “sleep”. The most striking difference is in “bobsez03”, which goes from a few unconvincing bound hot zones to none but with a nonetheless clear demarcation between the hour directly before the seizure and the rest of the recording.

We have included figures showing the results of cross-recording calculations for patients “amy” and “bob”, Figs. 3.16 and 3.17 respectively. In the former, the similarity between results on a given recording for *all* the reference windows is remarkable. Even the RW from “amysleep” gives essentially the same results. The abundance of hot zones in the “amysez03” column, many spanning the entire length of the recording, tells us this particular RW is somehow “abnormal”. Nonetheless, results on “amysez03” in the region from which the abnormal RW is taken with all the other RWs reveal nothing. This is one of the best examples of the asymmetry introduced in the measure by projection onto *different optimal bases*. Similarly, the results using the “bobsleep” RW and the other RWs on “bobsleep” reveal that the “bobsleep” RW is defective in some way; it does not appear to sample any sort of normal dynamics for this patient. The observations demonstrate the value of cross-recording trials of any given RW, when the discriminating power of the measure depends strongly on this window’s representation of the reference dynamics.

Finally, while computing the similarity index on the various recordings, we noticed a lot of variation in the *number of crossings* in the reference window segment and test window segment and wondered if a trend in the number of crossings in the TW relative to the RW alone might be enough to anticipate the seizures. Such a trend could be caused by a rise in power to higher frequency spectral components or an increased frequency of interictal spiking. In other words, it would show up under linear time-series analysis. As a simple test, we logged the number of crossings in the RW segment,  $N_x(S_{\text{ref}})$ , and the number in each of the TWs,  $N_x(S_t)$ , for the first intracranial channel of each recording. In Fig. 3.18, we have plotted time series of the ratio  $\gamma_x = N_x(S_t)/N_x(S_{\text{ref}})$  obtained and indicated the start of each clinical seizure. A downward trend in  $\gamma_x$  just before the first seizure is clearly seen in the “amysez” recordings. Very pronounced pre-seizure rises in the ratio are seen in “bobsez01” and “carlsez03”. The strong fluctuations on “bobsleep” are also noteworthy. These results shed light on the “defectiveness” of the “bobsleep” RW; the whole TS would appear

to be exceptionally nonstationary. Furthermore, it helps to explain why the results on the “amy” seizures were so consistently good—these seizures are likely the easiest to anticipate by any method, linear or nonlinear.

---

Figure 3.11 (*following page*): **Overall trends** in similarity values can be appreciated from this figure, which shows the hot zones identified in all eligible channels from the results using the *first* reference window on all recordings and thresholds of  $\sigma_{th} = 5$ ,  $\tau_{th} = 120$  s (i.e. four TWs) and  $\tau_g = 0$ . Results from intracranial EEG channels are in black, surface EEG is in purple. Red lines mark the start of clinical seizures. Blue bars along the bottom show the location and extent of gaps in the recording. The results *on the RW* are *reproduced* at the start of each signal, which adds five minutes to the recording. Black tic marks on the bottom edge delimit the length of the recording plus five minutes, the full width of the box corresponding to 425 min (i.e. the length of the longest recording plus five minutes). Black tic marks on the left-hand border indicate the extent of the signals for that recording; a box can accommodate a maximum of 64 signals. Generally, channels in the top half of this range were in the right hemisphere and those in the bottom the left. Ineligible channels are marked with a green tic on the left (this mark overwrites the black mark in some cases). Those on which the similarity algorithm failed for any reason are marked with a green tic on the right.

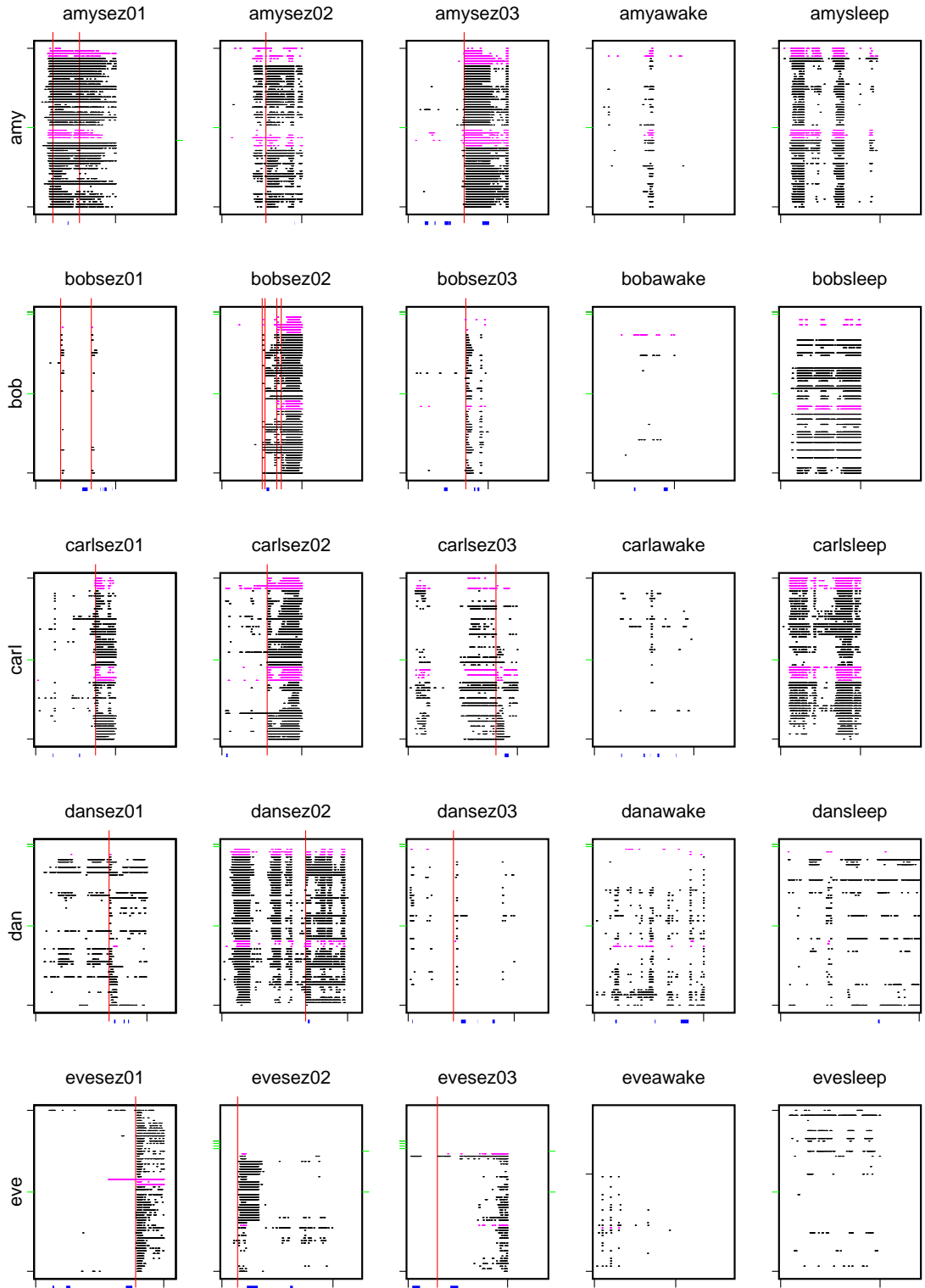




Figure 3.12: Detail from Fig. 3.11: *amysez01*. We recall that the two black ticks at the bottom of the graph mark the start and end of the recording, which here is 4 hours long. We have added a time scale for convenience. Bound hot zones are found in 47 intracranial EEG channels with an average anticipation of 9.8 min and 5 surface channels (mean anticipation of 13.5 min). The hot zones generally persist with intermittent breaks between the two seizures, after which we see a gradual return to a dynamics similar to the reference state.

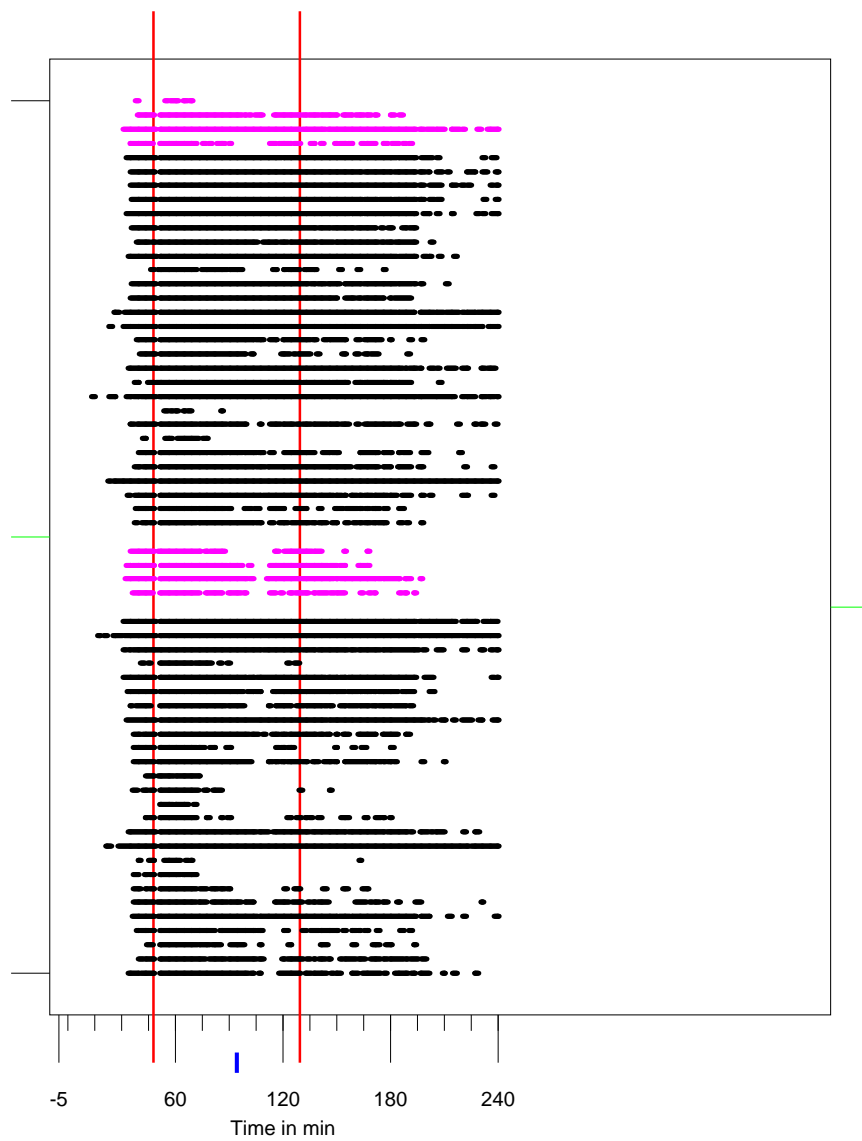


Figure 3.13: Detail from Fig. 3.11: *carlsez01*. Scattered free hot zones precede the seizure, including a number that occur more or less simultaneously in several channels shortly after a gap (indicated by the blue mark at the bottom of the graph). Bound hot zones are found in 18 intracranial channels (mean anticipation time of 7.1 min) and only 2 surface channels. We notice that certain *groups* of channels have hot zones coinciding with the seizure which end very shortly after seizure onset, whereas the hot zones on other groups of channels persist almost to the end of the recording. In other words, we observe *spatial* differences in the dynamics on this recording.

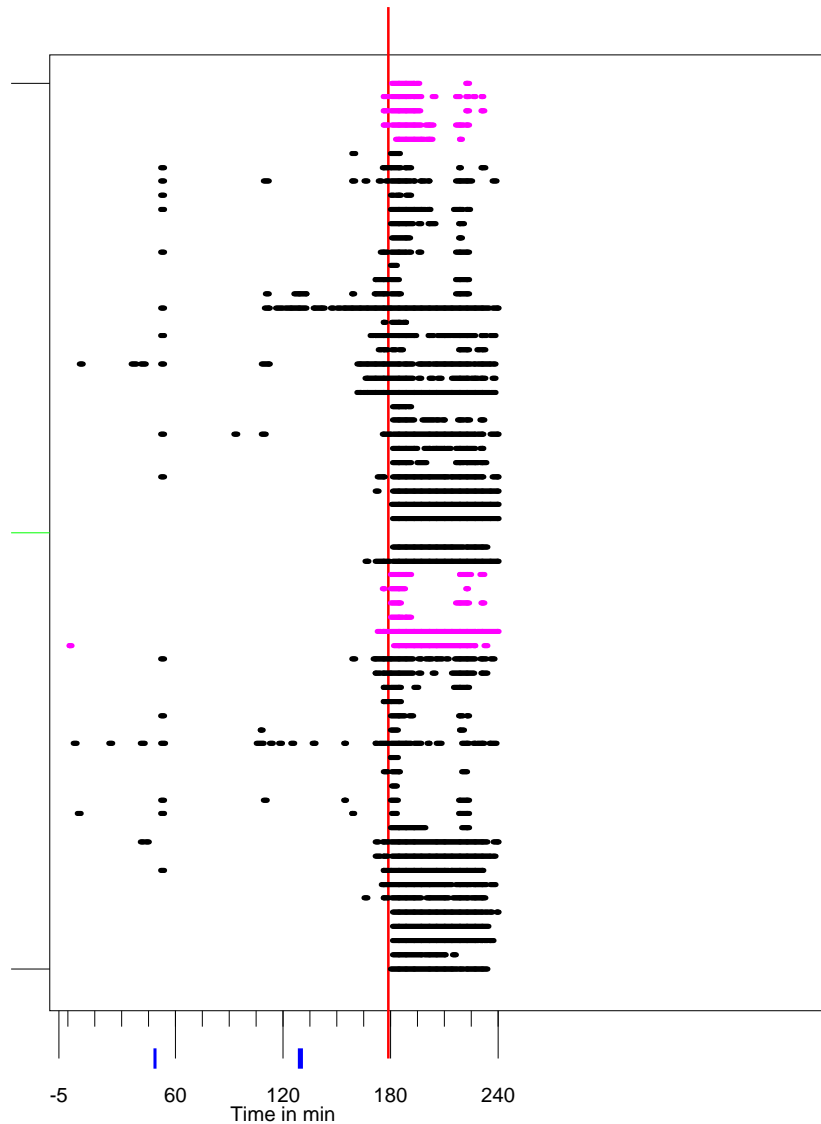


Table 3.1: Anticipation and free hot zone results on intracranial EEG channels (“Type 2”) using the first reference window,  $\sigma_{th} = 5$ ,  $\tau_{th} = 120$  s and  $\tau_g = 0$ .

Recording	Active Channels	Eligible Channels	Earliest Anticipation <i>min</i>	Mean Anticipation (Mean Gap) <i>min</i>	Length of Segment <i>hr</i>	Mean Free Hot Zones per Eligible Channel per hour	Mean Total Duration of Free Hot Zones <i>min</i>
<b>amysez01</b>	47	53	23.0	$9.8 \pm 5.4$ (0.0)	0.8	$1.08 \pm 0.87$	$6.0 \pm 6.8$
<b>amysez02</b>	7	48	3.0	$2.2 \pm 0.5$ (0.0)	2.2	$0.97 \pm 1.15$	$9.4 \pm 8.3$
<b>amysez03</b>	8	48	7.0	$3.1 \pm 2.1$ (0.0)	2.8	$0.12 \pm 0.28$	$0.7 \pm 2.0$
amyawake	.	54	.	.	4.5	$0.31 \pm 0.51$	$2.2 \pm 2.8$
amysleep	.	54	.	.	5.0	$1.55 \pm 1.69$	$19.8 \pm 11.3$
<b>bobsez01</b>	0	50	—	—	1.2	$0.03 \pm 0.24$	$0.2 \pm 1.4$
<b>bobsez02</b>	0	50	—	—	2.0	$0.01 \pm 0.07$	$0.0 \pm 0.3$
<b>bobsez03</b>	0	50	—	—	2.9	$0.06 \pm 0.32$	$0.3 \pm 1.3$
bobawake	.	49	.	.	4.0	$0.09 \pm 0.36$	$0.4 \pm 1.9$
bobsleep	.	49	.	.	4.0	$1.70 \pm 2.25$	$34.2 \pm 32.4$
<b>carlsez01</b>	18	52	25.5	$7.1 \pm 6.3$ (0.0)	3.0	$0.47 \pm 0.87$	$2.2 \pm 3.7$
<b>carlsez02</b>	4	52	7.0	$3.4 \pm 2.1$ (0.0)	2.2	$0.47 \pm 1.09$	$3.9 \pm 10.8$
<b>carlsez03</b>	15	52	113.0	$20.9 \pm 37.3$ (0.0)	4.4	$1.25 \pm 1.60$	$12.2 \pm 13.2$
carlawake	.	52	.	.	5.0	$0.22 \pm 0.54$	$1.2 \pm 2.6$
carlsleep	.	52	.	.	4.0	$3.49 \pm 3.83$	$36.0 \pm 19.0$
<b>dansez01</b>	4	55	9.5	$4.1 \pm 3.2$ (0.0)	3.7	$0.79 \pm 1.54$	$9.7 \pm 18.0$
<b>dansez02</b>	4	55	8.5	$4.2 \pm 2.6$ (0.0)	4.2	$2.00 \pm 2.21$	$30.8 \pm 11.2$
<b>dansez03</b>	0	55	—	—	2.2	$0.26 \pm 0.49$	$1.8 \pm 3.0$
danawake	.	55	.	.	5.5	$1.00 \pm 1.47$	$5.8 \pm 7.6$
dansleep	.	55	.	.	7.0	$0.57 \pm 1.14$	$5.6 \pm 12.9$
<b>evesez01</b>	0	60	—	—	5.0	$0.06 \pm 0.28$	$0.4 \pm 2.2$
<b>evesez02</b>	0	44	—	—	0.7	$0.10 \pm 0.45$	$0.4 \pm 1.9$
<b>evesez03</b>	0	44	—	—	1.4	$0.02 \pm 0.11$	$0.8 \pm 5.5$
eveawake	.	38	.	.	4.1	$0.32 \pm 0.50$	$1.3 \pm 1.7$
evesleep	.	60	.	.	5.0	$0.42 \pm 1.01$	$3.9 \pm 11.6$

Table 3.2: Anticipation and free hot zone results on surface EEG channels (“Type 1”) using the first reference window,  $\sigma_{th} = 5$ ,  $\tau_{th} = 120$  s and  $\tau_g = 0$ .

Recording	Active Channels	Eligible Channels	Earliest Anticipation <i>min</i>	Mean Anticipation (Mean Gap) <i>min</i>	Length of Segment <i>hr</i>	Mean Free Hot Zones per Eligible Channel per hour	Mean Total Duration of Free Hot Zones <i>min</i>
<b>amysez01</b>	5	8	17.0	$13.5 \pm 2.7$ (0.0)	0.8	$0.62 \pm 0.41$	$7.7 \pm 10.6$
<b>amysez02</b>	2	14	7.0	$5.5 \pm 1.5$ (0.0)	2.2	$1.02 \pm 1.24$	$8.0 \pm 9.0$
<b>amysez03</b>	0	14	—	—	2.8	$0.33 \pm 0.54$	$1.8 \pm 2.6$
amyawake	—	8	N/A	N/A	4.5	$0.61 \pm 0.81$	$4.4 \pm 3.2$
amysleep	—	8	N/A	N/A	5.0	$2.10 \pm 2.14$	$33.7 \pm 5.9$
<b>bobsez01</b>	0	11	—	—	1.2	$0.00 \pm 0.00$	$0.0 \pm 0.0$
<b>bobsez02</b>	0	11	—	—	2.0	$0.05 \pm 0.15$	$0.2 \pm 0.5$
<b>bobsez03</b>	0	11	—	—	2.9	$0.10 \pm 0.23$	$0.3 \pm 0.8$
bobawake	—	12	N/A	N/A	4.0	$0.21 \pm 0.72$	$1.5 \pm 4.8$
bobsleep	—	12	N/A	N/A	4.0	$0.96 \pm 1.65$	$12.6 \pm 21.7$
<b>carlsez01</b>	2	11	3.0	$3.0 \pm 0.0$ (0.0)	3.0	$0.15 \pm 0.26$	$0.6 \pm 1.0$
<b>carlsez02</b>	1	11	1.5	$1.5 \pm 0.0$ (0.0)	2.2	$0.94 \pm 1.82$	$7.1 \pm 13.8$
<b>carlsez03</b>	2	11	89.5	$50.0 \pm 39.5$ (0.0)	4.4	$0.97 \pm 1.34$	$8.6 \pm 11.7$
carlawake	—	11	N/A	N/A	5.0	$0.00 \pm 0.00$	$0.0 \pm 0.0$
carlsleep	—	11	N/A	N/A	4.0	$3.66 \pm 3.65$	$45.0 \pm 13.3$
<b>dansez01</b>	0	6	—	—	3.7	$0.05 \pm 0.11$	$0.2 \pm 0.3$
<b>dansez02</b>	0	6	—	—	4.2	$1.95 \pm 1.92$	$28.8 \pm 9.6$
<b>dansez03</b>	0	6	—	—	2.2	$0.15 \pm 0.36$	$0.9 \pm 1.9$
danawake	—	6	N/A	N/A	5.5	$0.85 \pm 1.22$	$4.5 \pm 5.2$
dansleep	—	6	N/A	N/A	7.0	$0.17 \pm 0.30$	$0.7 \pm 1.0$
<b>evesez01</b>	1	3	84.5	$84.5 \pm 0.0$ (22.5)	5.0	$0.00 \pm 0.00$	$0.0 \pm 0.0$
<b>evesez02</b>	0	2	—	—	0.7	$0.00 \pm 0.00$	$0.0 \pm 0.0$
<b>evesez03</b>	0	2	—	—	1.4	$0.00 \pm 0.00$	$0.0 \pm 0.0$
eveawake	—	1	N/A	N/A	4.1	$0.74 \pm 0.71$	$3.1 \pm 0.0$
evesleep	—	3	N/A	N/A	5.0	$0.00 \pm 0.00$	$0.0 \pm 0.0$

Table 3.3: Anticipation and free hot zone results on intracranial EEG channels (“Type 2”) using the first baseline,  $\sigma_{\text{th}} = 5$ ,  $\tau_{\text{th}} = 120$  s and  $\tau_{\text{grace}} = 120$  s.

Recording	Active Channels	Eligible Channels	Earliest Anticipation <i>min</i>	Mean Anticipation (Mean Gap) <i>min</i>	Length of Segment <i>hr</i>	Mean Free Hot Zones per Eligible Channel per hour	Mean Total Duration of Free Hot Zones <i>min</i>
<b>amysez01</b>	51	53	34.5	$13.4 \pm 6.3$ (0.0)	0.8	$0.00 \pm 0.00$	$0.0 \pm 0.0$
<b>amysez02</b>	14	48	7.0	$3.0 \pm 1.4$ (0.0)	2.2	$0.38 \pm 0.43$	$11.7 \pm 8.4$
<b>amysez03</b>	12	48	11.5	$5.0 \pm 2.5$ (0.0)	2.8	$0.05 \pm 0.23$	$0.9 \pm 4.3$
amyawake	.	54	.	.	4.5	$0.17 \pm 0.27$	$3.3 \pm 4.1$
amysleep	.	54	.	.	5.0	$0.72 \pm 0.78$	$23.3 \pm 12.8$
<b>bobsez01</b>	0	50	—	—	1.2	$0.03 \pm 0.24$	$0.4 \pm 2.8$
<b>bobsez02</b>	0	50	—	—	2.0	$0.01 \pm 0.07$	$0.2 \pm 1.2$
<b>bobsez03</b>	2	50	3.5	$3.2 \pm 0.2$ (0.0)	2.9	$0.04 \pm 0.25$	$0.4 \pm 2.3$
bobawake	.	49	.	.	4.0	$0.05 \pm 0.19$	$0.8 \pm 3.7$
bobsleep	.	49	.	.	4.0	$0.53 \pm 0.68$	$39.3 \pm 35.6$
<b>carlsez01</b>	27	52	62.5	$9.3 \pm 11.3$ (0.2)	3.0	$0.28 \pm 0.57$	$1.9 \pm 4.4$
<b>carlsez02</b>	4	52	85.0	$32.9 \pm 34.3$ (0.0)	2.2	$0.27 \pm 0.51$	$3.6 \pm 7.6$
<b>carlsez03</b>	16	52	113.0	$66.3 \pm 41.5$ (0.0)	4.4	$0.42 \pm 0.55$	$11.0 \pm 10.9$
carlawake	.	52	.	.	5.0	$0.17 \pm 0.37$	$1.9 \pm 4.7$
carlsleep	.	52	.	.	4.0	$0.94 \pm 0.99$	$50.2 \pm 24.2$
<b>dansez01</b>	10	55	68.0	$46.5 \pm 28.2$ (0.0)	3.7	$0.29 \pm 0.51$	$9.2 \pm 15.6$
<b>dansez02</b>	10	55	33.5	$12.1 \pm 8.5$ (0.0)	4.2	$0.83 \pm 0.88$	$34.9 \pm 13.0$
<b>dansez03</b>	0	55	—	—	2.2	$0.24 \pm 0.44$	$2.8 \pm 4.6$
danawake	.	55	.	.	5.5	$0.73 \pm 0.95$	$9.3 \pm 11.9$
dansleep	.	55	.	.	7.0	$0.29 \pm 0.50$	$7.9 \pm 16.0$
<b>evesez01</b>	1	60	4.0	$4.0 \pm 0.0$ (0.0)	5.0	$0.04 \pm 0.18$	$0.6 \pm 3.9$
<b>evesez02</b>	2	44	13.5	$9.8 \pm 3.8$ (0.0)	0.7	$0.03 \pm 0.21$	$0.3 \pm 1.7$
<b>evesez03</b>	0	44	—	—	1.4	$0.02 \pm 0.11$	$0.9 \pm 6.0$
eveawake	.	38	.	.	4.1	$0.32 \pm 0.48$	$1.8 \pm 2.7$
evesleep	.	60	.	.	5.0	$0.18 \pm 0.41$	$6.2 \pm 15.7$



Figure 3.14: shows the hot zones identified in all eligible channels from the *second* reference window results on all recordings for thresholds of  $\sigma_{th} = 5$ ,  $\tau_{th} = 2$  min (i.e. four TWs) and  $\tau_g = 0$ . See Fig. 3.11 caption for details.

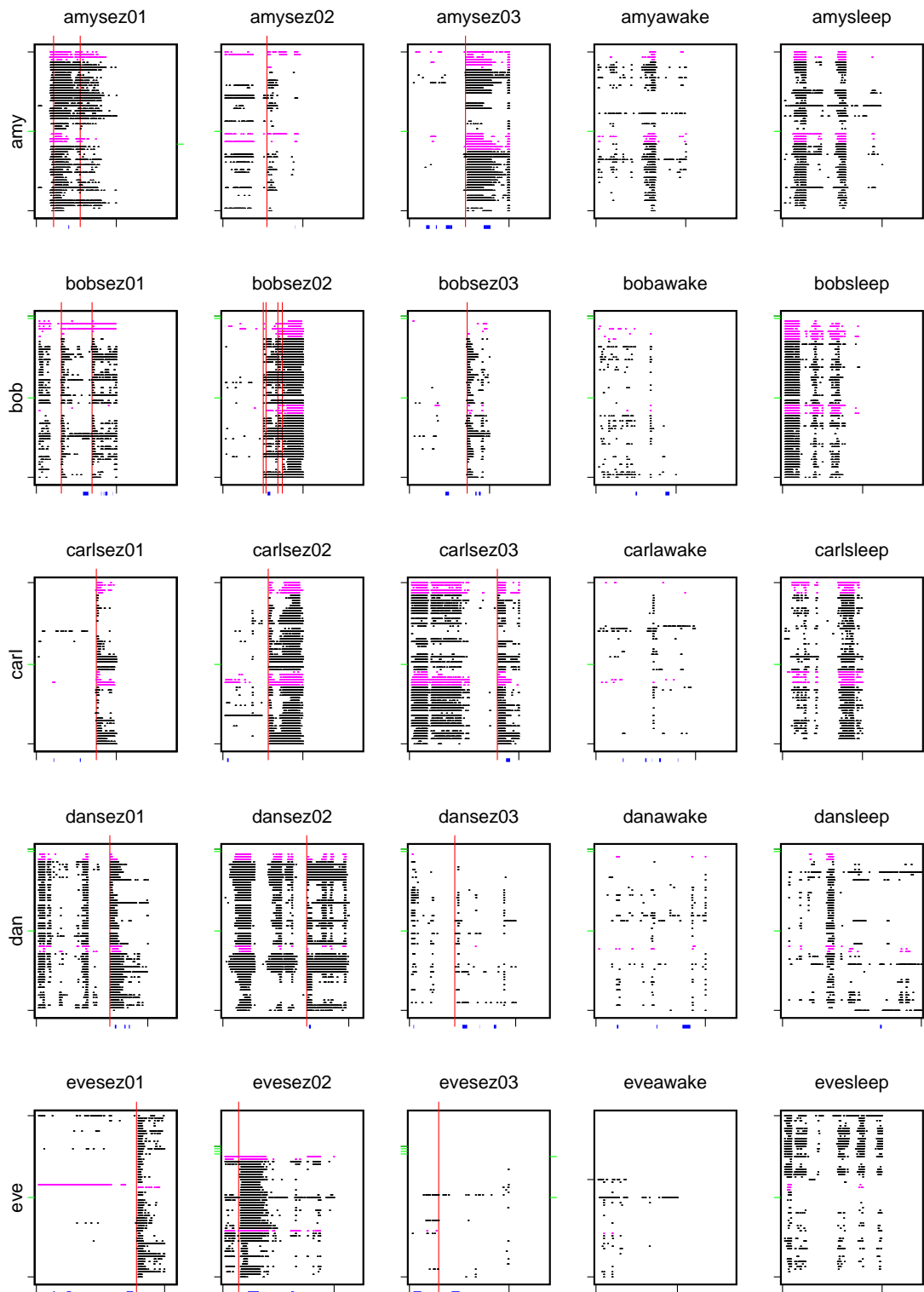
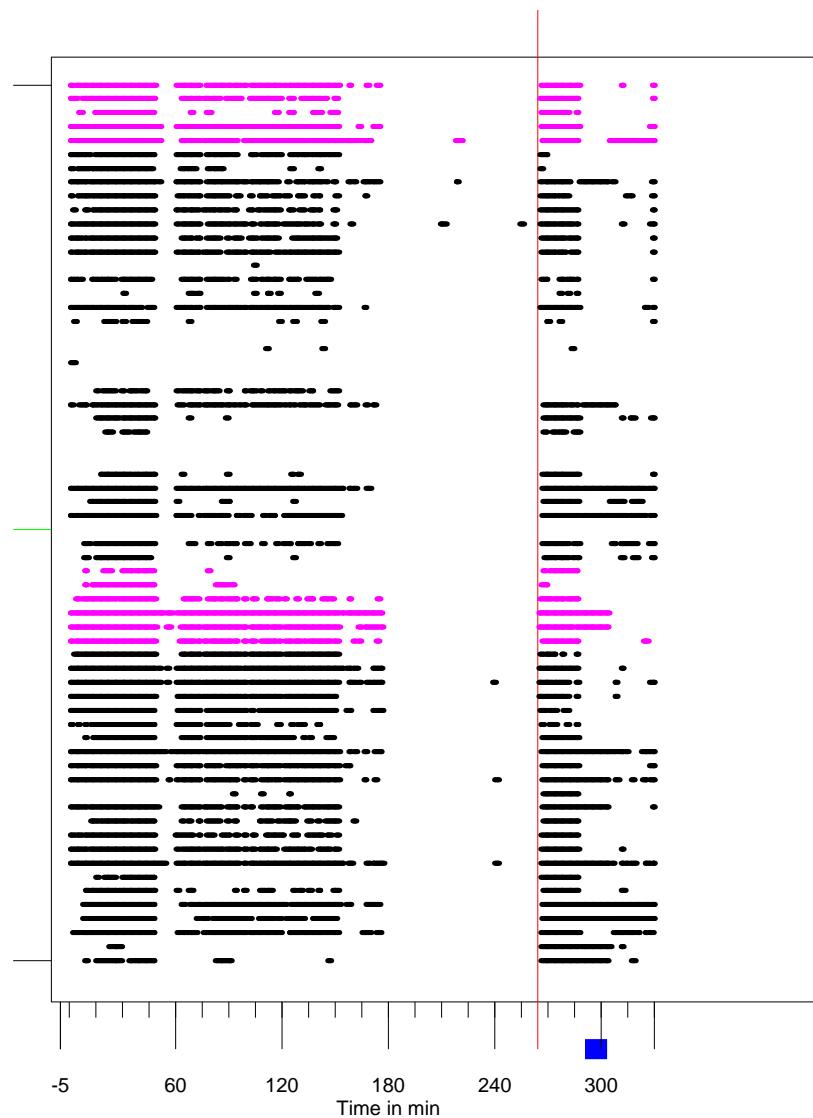


Figure 3.15: Detail from Fig. 3.14: *carlsez03*. The second RW was located 20 min prior to the clinical seizure on this recording. It is interesting to note that this RW appears to be *in* the preictal region, with the interictal and ictal regions showing strong dissimilarity from the band just before the seizure. Some quick tests were done where the RW was placed *immediately* before the seizures and we looked for *dissimilarity* from this region, but this approach did not appear to improve the performance of the measure.





---

Figure 3.16 (*following page*): shows the hot zones in all eligible channels from the **cross-recording** results for patient “amy”. Thresholds of  $\sigma_{th} = 5$ ,  $\tau_{th} = 2$  min (i.e. four TWs) and  $\tau_g = 0$  were used. Green tic marks to the right indicate channels for which no mapping could be established with the reference window series. Column headings indicate the source of the RW and row labels identify the source of the TWs. Otherwise the plots read as in figure 3.11.



---

Figure 3.17 (*following page*): shows the hot zones in all eligible channels from the **cross-recording** results for patient “bob”. Thresholds of  $\sigma_{\text{th}} = 5$ ,  $\tau_{\text{th}} = 2 \text{ min}$  (i.e. four TWs) and  $\tau_g = 0$  were used. Green tic marks to the right indicate channels for which no mapping could be established with the reference window series. Column headings indicate the source of the RW and row labels identify the source of the TWs. Otherwise the plots read as in figure 3.11.

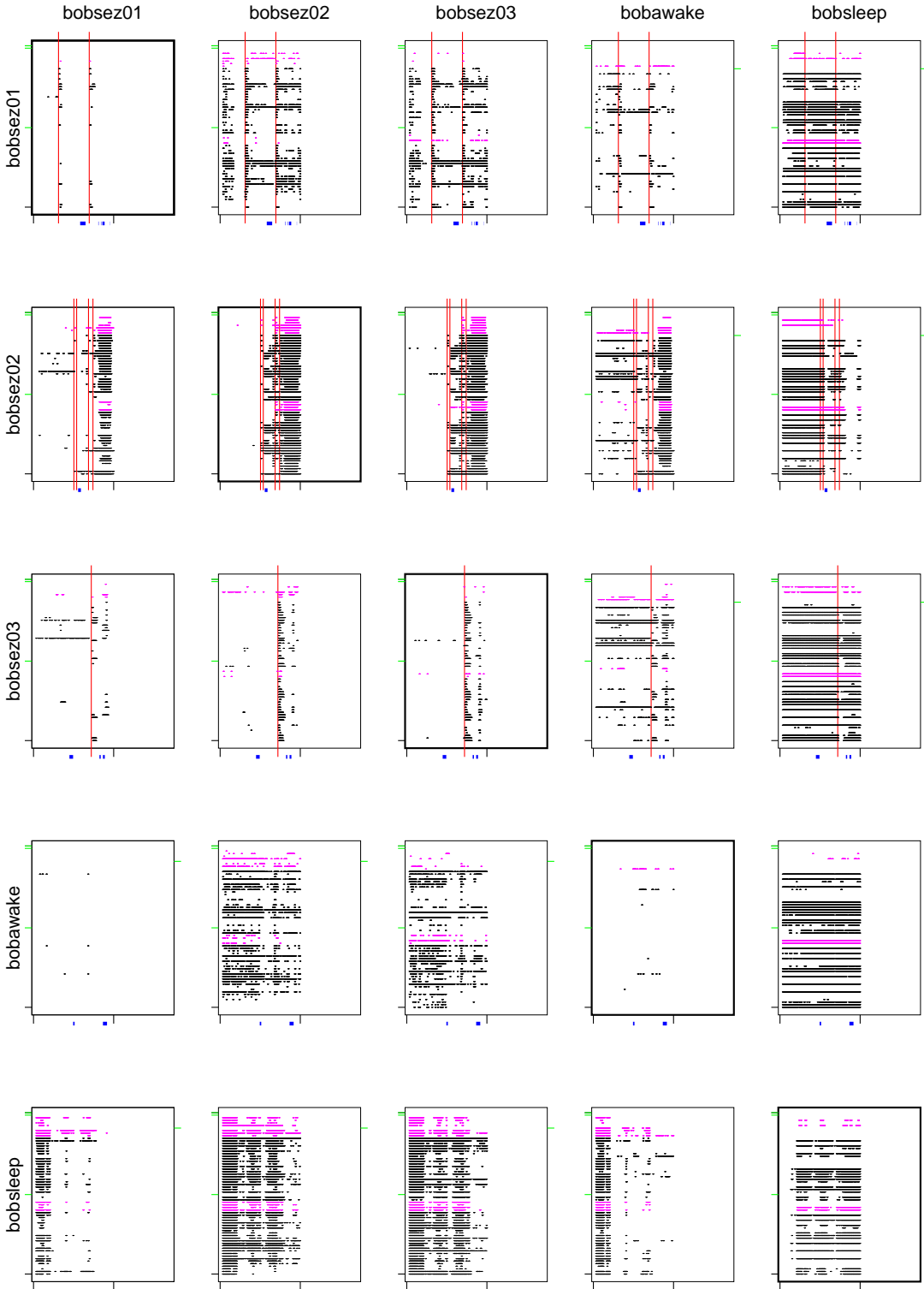
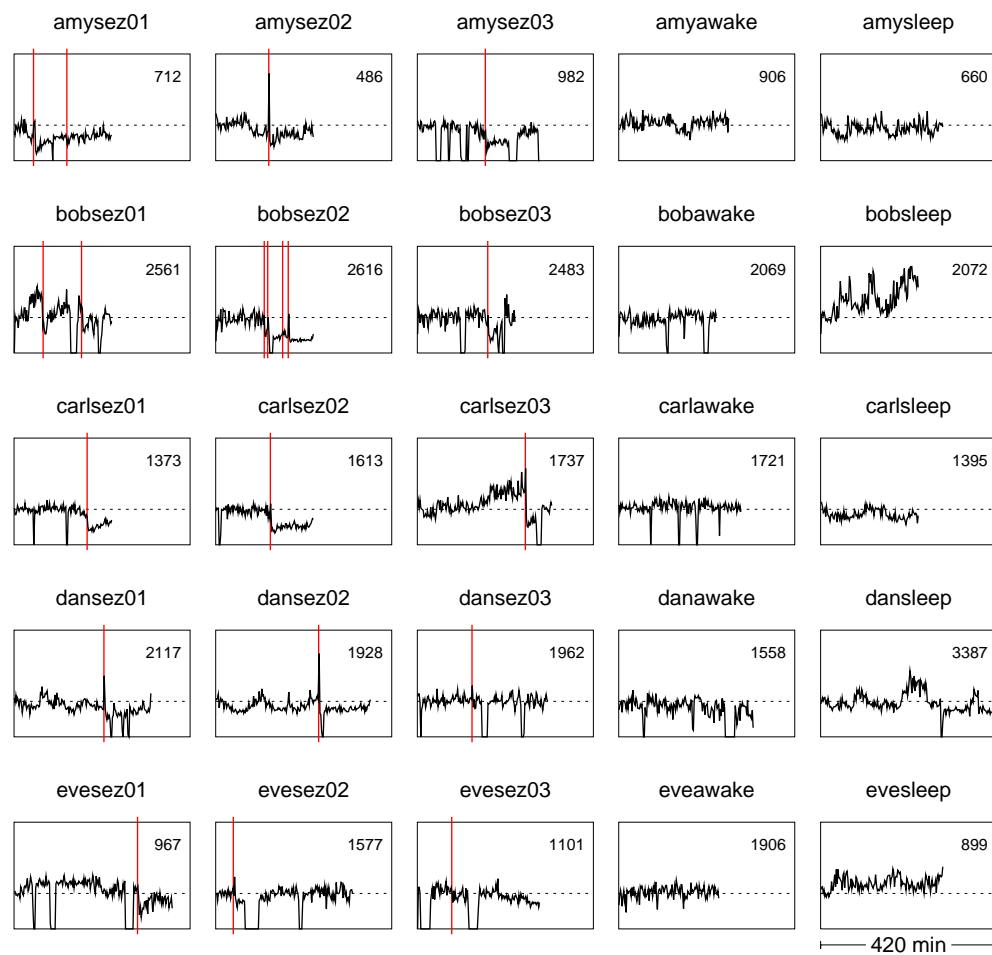


Figure 3.18: Time series of  $\gamma_x$ , the ratio of threshold crossings in the test window to those in the first reference window on the first intracranial contact of each recording. Vertical scale is from 0 to 2, with the horizontal dotted line marking  $\gamma_x = 1$ . The value of  $N_x(S_{\text{ref}})$  on each channel is shown in the upper right corner of each plot. Red lines indicate the start of clinical seizures.



# Chapter 4

## Information Flow

The dynamical similarity measure studied in the previous chapter gauges the degree of overlap between the distribution of states in reconstructed phase space and a *reference* distribution. It is thus, by design, a *relative measure* of dynamical change. In this chapter, our focus shifts to an *absolute* measure of the information flow in a stationary process. We begin with a definition of and derivation of a formal expression for the information flow of a dynamical system in the framework of symbolic dynamics and information theory. A brief overview of the relevant concepts from the latter can be found in Appendix A. The practical limitations of this approach will be discussed and an alternative *cumulant*-based approach introduced by Deco, Schittenkopf and Schürmann [18] will be developed in § 4.2. We then study the behaviour of the cumulant-based measure on our set of artificial time series and its performance in anticipating epileptic seizures from EEG recordings.

### 4.1 A Formal Measure of Information Flow

Information flow on a stationary system can be understood as the extent to which the past of a process conditions its future or the degree to which information about past states “flows” forward to play a role in determining that future. This applies equally well to deterministic and stochastic systems. For example, in a first-order Markovian stochastic process, the future depends only on the *present* state of the system; no

information flows from the past to the future. Equivalently, it measures our residual uncertainty about the future of a process given perfect knowledge of the process' past. Intuitively, information flow decreases with increasing system complexity.

Let us reiterate that in this thesis, we are interested in dynamical systems and measuring changes in those systems from time series. The bridge between dynamical systems theory and probabilistic information theory [15, 17] is provided by symbolic dynamics [60]. More specifically, it was Kolmogorov [40] who applied Shannon's information characterization to ergodic theory. The following derivation follows the same lines as that given by Beaulieu [4] and Deco and Obradovic [17].

Our knowledge of any given real dynamical system is frequently limited to a time series of measurements of the system  $x_1, x_2, \dots$  which, for simplicity, we will say are made at equal intervals. An *embedding* in  $m$ -dimensions of this data, possibly using delay coordinates to form reconstruction vectors such that  $\mathbf{r}_n = (x_n, x_{n+\tau}, \dots, x_{n+(m-1)\tau})$ , yields a series of vectors  $\mathbf{r}_1, \mathbf{r}_2, \dots$  which describe a trajectory in the embedding space. Takens' embedding theorem [74, 80] tells us that an embedding dimension  $m$  and delay  $\tau$  can be found such that there will be a one-to-one correspondence between the reconstruction vectors and states in the system's true phase space. Furthermore, dynamical invariants such as Lyapunov exponents are the same on this embedding as on the true attractor; topological and geometrical properties are preserved in the embedding. Symbolic dynamics provide an alternate representation of this reconstructed dynamics.

Let  $R$  be the region of the embedding space explored by the system; it contains all of the system's (possibly infinite) set of possible states. In terms of ergodic theory, if  $\mu$  is an ergodic invariant probability measure for the dynamical system, then

$$\int_R \mu(\mathbf{x}) \, d\mathbf{x} = 1.$$

We assume that the system is stationary and ergodic such that given a sufficient quantity of data, an invariant measure on  $R$  can be estimated. A covering of this

space by  $N$  disjoint subregions  $R_i$  is called a *partition* of  $R$ , which we label  $\beta$ :

$$\beta = \{R_i\}_{i=1}^N, \quad \bigcup_{i=1}^r R_i = R \quad \text{and} \quad R_i \cap R_j = \emptyset \quad \text{if} \quad i \neq j. \quad (4.1)$$

With each subregion  $R_i$  is associated a symbol  $\sigma_i$ , thus forming an alphabet  $\mathcal{A}$  of  $N$  symbols.<sup>1</sup> This allows us to translate the series of reconstruction vectors into a string of symbols  $s_1, s_2, \dots$ , called the *itinerary* of the system, such that if  $\mathbf{r}_n \in R_i$  then  $s_n = \sigma_i$ .

A common convention (which we adopt throughout this thesis) for designating random variables and their possible values is readily and advantageously adapted to the symbolic dynamics, though this is not meant to imply that the itinerary is necessarily random. We let the capital letter  $A$  represent a symbol *chosen* randomly from anywhere on the itinerary (the event, in the parlance of probability theory) and  $a$  a possible value for that symbol (a possible outcome of the event), where  $a \in \mathcal{A}$ . The probability of observing symbol  $a$  is  $p(a)$ <sup>2</sup>. A block of  $n$  consecutive symbols beginning at a random point on the itinerary is denoted by  $A_n, \dots, A_1$  and the joint probability of a particular sequence  $a_n, \dots, a_1$  is just  $p(a_n, \dots, a_1)$ . Lastly, the conditional probability of finding the symbol  $a_{n+r}$  a time  $r$  steps after observing a given block of  $n$  symbols is written  $p(a_{n+r}|a_n, \dots, a_1)$ .

A definition of the entropy of the itinerary *for a given partition*  $\beta$  follows immediately from (A.4):

$$H^\beta(A) = -\sum_{a \in \mathcal{A}} p(a) \log_2 p(a). \quad [\text{bits}] \quad (4.2)$$

Similarly the block entropy of the symbolic dynamics will be

$$H^\beta(A_n, \dots, A_1) = -\sum_{a_n, \dots, a_1 \in \mathcal{A}} p(a_n, \dots, a_1) \log_2 p(a_n, \dots, a_1) \quad [\text{bits}] \quad (4.3)$$

---

<sup>1</sup>Nothing prevents the space  $R$  from being infinite. However, to ensure that the entropies we will subsequently calculate do not diverge, the alphabet  $\mathcal{A}$  must be finite. In the case of infinite  $R$ , then, the phase space will always be partitioned into some finite set of subspaces *plus everything else*.

<sup>2</sup>These are probability mass functions, not density functions, because the space of possible outcomes is discrete.



and the conditional entropy associated with predicting the symbol  $r$  steps after a block of  $n$  known symbols or outcomes is

$$H^\beta(A_{n+r}|A_n, \dots, A_1) = -\sum_{a_{n+r}, a_n, \dots, a_1 \in \mathcal{A}} p(a_{n+r}|a_n, \dots, a_1) \log_2 p(a_{n+r}|a_n, \dots, a_1). \quad [\text{bits}] \quad (4.4)$$

As our knowledge of the system's past increases to include the entire past, such that the block size  $n \rightarrow \infty$ , the conditional entropy (4.4) for single-step prediction ( $r = 1$ ) tends toward the point entropy

$$h^\beta = \lim_{n \rightarrow \infty} [H^\beta(A_{n+1}, A_n, \dots, A_1) - H^\beta(A_n, \dots, A_1)] \quad (4.5)$$

$$= \lim_{n \rightarrow \infty} H^\beta(A_{n+1}|A_n, \dots, A_1). \quad (4.6)$$

#### point entropy

This represents the uncertainty of the next step in the itinerary, using a given partition  $\beta$ , given perfect knowledge of the infinite past. Since it is the amount by which our uncertainty of the *whole* trajectory *grows* from a given initial condition ( $A_1 \in \mathcal{A}$ ) with each additional step, it can also be seen as an entropy rate. To ensure that the choice of partition  $\beta$  will not influence our characterization of the dynamics, however, we take the supremum of (4.5) over all possible partitions,

$$h_{\text{KS}} = \sup_{\beta} h^\beta. \quad \text{Kolmogorov-Sinai entropy} \quad (4.7)$$

Any partition on which (4.7) is satisfied is called a **generating partition**. The resulting measure  $h_{\text{KS}}$  is called the metric entropy or Kolmogorov-Sinai (KS) entropy. This is the formal measure we were seeking.

KS entropy describes the average *flow of information* from smaller to larger scales in that it gives the rate at which bits of information are lost [37]. For example, the KS entropy of the Bernoulli shift map<sup>3</sup> can be found analytically to be  $h_{\text{KS}} = \log_2 2 =$

---

<sup>3</sup>The Bernoulli shift map  $x_{n+1} = 2x_n \bmod 1$  is a deterministic mapping of the unit interval onto itself twice (the first-return map looks like saw teeth). Applying the

1[bit], i.e. one bit is lost every iteration. Use of a partition with a finite number of subspaces limits the precision of our knowledge of a system's initial conditions in the same way as the finite precision of a measuring instrument. The uncertainty in our prediction of the system's future increases as the limited number of bits of precision are used up at a rate given on average by  $h_{\text{KS}}$ .

The KS entropy of a periodic system is zero and that of a stochastic process divergent. The values for different chaotic systems fall in the broad interval between these extremes, which makes  $h_{\text{KS}}$  a useful measure for discriminating between different degrees of chaos...in principle. In practice, there are two formidable obstacles to computing this measure on real data: finding a generating partition and taking the  $n \rightarrow \infty$  limit [37]. One strategy for getting around the first obstacle is to divide phase space into hypercubes of size  $\epsilon$ . It is impossible to take the supremum over all partitions and the alternative of taking the limit  $\epsilon \rightarrow 0$  is meaningless on real data, so instead we seek a *scaling region* in  $\epsilon$  on which the estimated value of  $h_{\text{KS}}$  is more or less constant. The bigger problem is the need for an infinitely long time series of stationary data. These are hard to come by. The available, relatively stationary data must at least be “long enough” to provide sufficient statistics. Among measures used in non-linear time series analysis, the KS entropy is notoriously data hungry; shorter series will often suffice for estimating dimensions or Lyapunov exponents while being inadequate for calculating entropies.

## 4.2 Cumulant-based Information Flow

A different approach to the study of information flow on a stationary process, which dispenses with symbolic dynamics and many of the attendant complications of dy-  


---

 map to a binary representation of an *irrational* number—necessarily an infinite string of 1's and 0's—results in loss of the leftmost (most significant) bit and a *shifting* of all other bits to the left. Consequently, initially insignificant bits gain in importance on each iteration; this is an example of sensitive dependence on initial conditions of chaotic systems.

namical reconstruction and partitioning, has been introduced by Deco, Schittenkopf and Schürmann [18]. The method is well suited to use on real time serial data. The basic idea is to gauge the statistical independence between the past and future of the underlying process using estimated values of *cumulants*, which are coefficients in the series of expansion of the cumulant characteristic function describing the process. In what follows, we assume familiarity with a number of concepts from probability theory in defining the cumulant-based measure of information flow. An brief overview of this material is found in Appendix B as well as references cited therein.

In a time series obtained by sampling a *stationary* process, a block of  $n$  consecutive points and a point located  $r$  samples later is completely described by a vector of continuous random variables,

$$\mathbf{X} = (X_1, \dots, X_n, X_{n+r}), \quad (4.8)$$

the distribution of which is given by the probability density function  $p(\mathbf{X})$ . If the block of  $n$  points (the past) and the point  $r$  steps later (the future) are statistically independent, then the following equation is satisfied:

$$p(\mathbf{X}) = p(X_1, \dots, X_n) p(X_{n+r}), \quad (4.9)$$

where  $p(X_1, \dots, X_n)$  and  $p(X_{n+r})$  are marginal density functions. Let

$$\Psi(\boldsymbol{\alpha}) = \Psi(\alpha_1, \dots, \alpha_n, \alpha_{n+r}) \quad (4.10)$$

be the cumulant characteristic function associated with  $p(\mathbf{X})$ . If (4.9) is satisfied, then it is also true that

$$\Psi(\boldsymbol{\alpha}) = \Psi(\alpha_1, \dots, \alpha_n) + \Psi(\alpha_{n+r}). \quad (4.11)$$

We let  $\mathcal{K}_{\ell_1 \dots \ell_q}$  denote the order- $q$  cumulants, where the indices  $\ell_i \in \{1, \dots, n, n+r\}$  (examples of this notation are found in Appendix B). The statistical independence relation (4.11) implies that

$$\mathcal{K}_{\ell_1 \dots \ell_q} = 0 \quad \text{if} \quad \underbrace{\exists \ell_i \in [1, n] \quad \text{and} \quad \exists \ell_j = n+r \quad \text{where} \quad i, j \in [1, q]}_{\text{constraint on indices in (4.13)}}. \quad (4.12)$$

and this for all  $q$ , i.e. all orders of cumulants. This observation suggests that we construct a measure

$$\sum_{q=1}^{\infty} \sum'_{\ell_1, \dots, \ell_q} \mathcal{K}_{\ell_1 \dots \ell_q}^2 \quad (4.13)$$

where the prime indicates that the second sum is constrained to indices satisfying condition in (4.12). Under conditions of statistical independence, this strictly positive quantity equals zero. In order to study the information flow from the entire past to the future, one would in principle have to consider this quantity in the  $n \rightarrow \infty$  limit.

Assuming the process is ergodic, the moments and hence the cumulants can be estimated using time averages *given sufficient data*. Because the computational cost of the measure increases rapidly with block length  $n$  and cumulant order  $q$ , in practice we must limit the former and truncate the sum on  $q$  in (4.13). Throughout this study, we take  $n = 10$ . Deco et al. [18] suggest truncation after  $q_{\max} = 4$ , i.e. estimating cumulants up to and including fourth order. The  $\mathcal{K}_{\ell_1 \dots \ell_m}$  are invariant under permutation of the indices  $\ell_i$ , eg.  $\mathcal{K}_{12} = \mathcal{K}_{21}$ . In the set of what we shall call *unique cumulants*, no two cumulants are equivalent under permutation of indices. The computational burden can thus be further reduced by summing only on these unique cumulants. The constraint on (4.13) can be made explicit by letting the last index always equal  $n' = n + r$ . Finally, a greater balance in the influence of each order of cumulants on the measure can be achieved by dividing the contribution of that order by the number of cumulants in the sum.

When we incorporate these various considerations, the proposed measure takes the following form:

$$m(n, r) = \underbrace{\frac{1}{N_2} \sum_{\ell_1=1}^n \mathcal{K}_{\ell_1 n'}^2}_{\text{Second cumulants}} + \underbrace{\frac{1}{N_3} \sum_{\ell_1=1}^n \sum_{\ell_2=\ell_1}^{n, n'} \mathcal{K}_{\ell_1 \ell_2 n'}^2}_{\text{Third cumulants}} + \underbrace{\frac{1}{N_4} \sum_{\ell_1=1}^n \sum_{\ell_2=\ell_1}^{n, n'} \sum_{\ell_3=\ell_2}^{n, n'} \mathcal{K}_{\ell_1 \ell_2 \ell_3 n'}^2}_{\text{Fourth cumulants}}, \quad (4.14a)$$

or

$$m(n, r) \equiv m_2(n, r) + m_3(n, r) + m_4(n, r), \quad (4.14b)$$

where  $m_2(n, r)$  denotes the weighted contribution of the second cumulants to the measure, i.e. the first term in (4.14a),  $m_3(n, r)$  that of the third cumulants and so on. The  $N_q$  are the number of unique cumulants of order  $q$  that are zero-valued under statistical independence between the past block of  $n$  samples and the sample  $r$  steps in the future:

$$N_2(n) = n, \quad (4.15)$$

$$N_3(n) = \frac{n}{2}(n + 3), \quad (4.16)$$

$$N_4(n) = \frac{n}{6}(n^2 + 6n + 11). \quad (4.17)$$

By applying the transformation

$$x \rightarrow \frac{x - \bar{x}}{\sigma} \quad (4.18)$$

to the data on a given window, where  $\bar{x}$  and  $\sigma$  are respectively the mean and standard deviation of the data in that window, we should obtain zero-mean random variables ( $\langle X \rangle = 0$ ) and render the measure insensitive to mere changes of scale. It also makes the measure dimensionless. The transformation of variables allows us to use the relatively simple equations (B.9)-(B.12) to compute the cumulants up to fourth order in term of the moments up to the same order.

It has been shown that various dynamical processes can be distinguished on the basis of their  $m(n, r)$  vs.  $r$  curve. Although ideal deterministic chaotic processes “remember” their semi-infinite past—again, sensitive dependence on initial conditions—the presence of noise in real systems will limit the information flow from the past. The rate at which the information flow dies off with increasing look-ahead depends on the Lyapunov time of the chaotic system (inverse of the maximal Lyapunov exponent). Markov chains of differing lengths can also be distinguished on the basis of the look-ahead curve. However, for the purposes of time series analysis using the method

of sliding windows, it is more practical to have a scalar measure of information flow. Formally then, we compute

$$m_{\text{avg}}(n, r_{\text{max}}) = \frac{1}{r_{\text{max}}} \sum_{r=1}^{r_{\text{max}}} m(n, r). \quad (4.19)$$

Because of this sum over the look-aheads, we call the measure cumulant-based *integrated* information flow. In fact, we use  $r_{\text{max}} = 1$  throughout our experiments. In his experiments on TSs generated with various 1 and 2-D maps, Beaulieu [4] generally found that a look-ahead of one was sufficient to detect nonstationarities introduced on these TSs. We have also found in our brief investigations that use of  $r_{\text{max}}$  greater than 1 does not increase the discriminating power of this measure of the type of data we will be considering; it only increases the computational cost of the measure!

### 4.3 Surrogate Data

The measure of information flow defined in the previous section is, by virtue of its dependence on higher order cumulants, sensitive to differences in nonlinear structure across multiple stationary time series. This is not to say that differences in *linear* structure may not overwhelm the nonlinearities in  $m_{\text{avg}}$ . When the method of sliding windows is used to detect dynamical changes present in a time series, window lengths are chosen such that we can assume “sufficient” stationarity on any given window. This allows us—along with an assumption of ergodicity—to estimate the cumulants of the underlying process.

If the measure is found to differ between windows, one naturally wonders whether this difference is due to changes in nonlinear structure. A direct comparison of  $m_2$ , the weighted contribution of second cumulants to the measure, on the two windows may suggest identical linear structure, at least within the limits of the measure. However, we cannot immediately conclude that differences in  $m_3$  or  $m_4$  are the result of changes in nonlinear structure. These differences may simply result from statistical fluctuation of the measure. The question then is how to reject the possibility that the

same results could be obtained on data sharing the same linear structure but which retains none of the purported nonlinear structure of the original data. In other words, how can we assess the significance of fluctuations in higher order moments?

It is impossible to obtain an analytical expression for the distribution of this measure on most data sets. This is generally true of nonlinear measures. Consequently, one of the pitfalls of nonlinear time series analysis is to see evidence of nonlinearity and in particular deterministic chaos in results that can in fact be explained by a coincidence of autocorrelation in the data. To avoid this trap, analysts employ the method of surrogate data [82].

Artificial time series preserving certain characteristics of the original data but in which all other structure is lost are generated. These are the surrogates. The simplest example is that of “scrambled surrogates”, where the raw data are reordered randomly, thus preserving the marginal distribution of the original TS but destroying any statistical temporal correlations. With the surrogates and an appropriate discriminating statistic, one can test the null hypothesis of independence, i.e. that these data were produced by sampling a fixed probability distribution. This is, of course, not a particularly useful test.

In this chapter, rather, we will use *phase* scrambled surrogates generated using the amplitude adjusted Fourier transform algorithm [82] with iterative refinement described by Schreiber and Schmitz in [75] and implemented as the `surrogates` routine in the TISEAN package [38]. This method produces surrogates with the same distribution and Fourier amplitudes as the original data, which allow one to test the null hypothesis that the data was produced by a linear Gaussian process rescaled by the action of an invertible, static measurement process. Any nonlinearity that might be present in the data would thus be introduced by the latter. We limit the number of refinement iterations to 100. The routine stops if the surrogate series converges before that, i.e. if the change in the distribution or power spectrum between iterations is less than machine precision. We have accepted the default option of exact

duplication of the distribution (that is the stopping criterion) over exact duplication of the power spectrum.

There are two ways to compare the value obtained for a given discriminating statistic, which we will denote  $\gamma$ , on the raw or test data and those obtained on the surrogates. If  $\gamma[i]$  is the value of  $\gamma$  computed on the  $i^{\text{th}}$  surrogate, one can consider the mean  $\mu_s = \langle \gamma[i] \rangle$  (angle brackets indicate averages over all surrogates) and standard deviation

$$\sigma_s = \sqrt{\langle (\gamma[i] - \mu_s)^2 \rangle} \quad (4.20)$$

of the surrogate results. Having chosen an acceptable level of significance, Student's  $t$  test can then be used to ascertain whether the null hypothesis can be rejected. Use of statistical tests like Student's  $t$  require certain assumptions to be made about the distribution of the measure in question  $\gamma$  on the raw and surrogate data. If the validity of these assumptions is in question, it may be preferable to use rank-order statistics. In this case, for a one-sided test, if  $N$  surrogates are produced and the value of  $\gamma$  on the raw data is greater than on all the surrogates, there is a  $1/(N + 1)$  chance of *falsely* rejecting the null hypothesis.

## 4.4 Experiments on Artificial Data

The purpose of the experiments on the artificial data generated using the Bondarenko model was to assess the influence of window length and sampling rate on the IIF measure's sensitivity to different degrees of chaos in the TSs. Some tests were also done using additive noise, the results of which can be readily summarized by saying that for noise levels up to as much as 20% of the rms amplitude of the signal, there was no appreciable impact on the discriminating power of the measure. The IIF value decreased on all TSs with increasing noise as the statistical dependencies between past and future were gradually eroded by the latter.



### 4.4.1 Methods

The same procedure was followed for both the  $c$  series and the  $e$  series. Hereafter, we will just say “the control parameter”. For each value of this parameter and each sampling interval  $\Delta t \in \{1.0, 10.0\}$ , IIF was computed on one window of length  $\ell_w$  samples from each of the 25 TSs in the data set for that combination of control parameter and  $\Delta t$ . Two window lengths,  $\ell_w = 1024$  and 8192, were considered. Measure parameters used in the computation were  $n = 10$  and  $r_{\max} = 1$  (single-step prediction).

With the exception of the results in Fig. 4.3 where a lag of  $\tau = 10$  was introduced, consecutive samples were used in forming the blocks of  $n + 1$  points (i.e.  $\tau = 1$ ). The experiment with the longer lag allows us to compare the influence of sampling rate to that of using non-consecutive points in forming the blocks on the value of the measure. On over-sampled data, we would want to be able to introduce an appropriate lag to recover sensitivity to higher order correlations; this test explores that possibility.

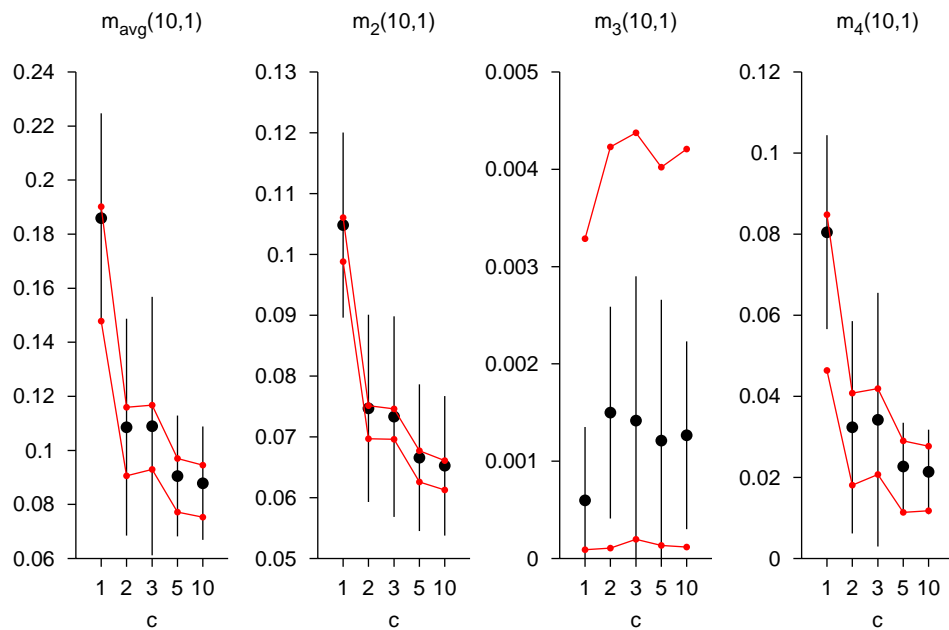
Furthermore, for *each* of the 25 windows, 49 surrogates were generated. The measure was computed on these surrogates using the same measure parameters. Our use of surrogates will, in fact, not be so formal; they serve more as a “sanity check” for observations we will make about the information flow measure’s sensitivity to nonlinear structure we know to be present in the artificial data. On the EEG data, we have in the end not used surrogates systematically. As we shall see, practically all of the changes in the measure which might be used to anticipate a seizure stem from spectral changes.

## 4.4.2 Results and Discussion

### 4.4.2.1 The $c$ Series

Using a window length of  $\ell_w = 1024$  on the TS sampled at  $\Delta t = 1.0$  (Fig. 4.1), considerable overlap in the results for each value of  $c$  obscures the difference between

Figure 4.1: IIF results on the  $c$  series with  $\ell_w = 1024$  and  $\Delta t = 1.0$ . The mean value obtained for  $m_{\text{avg}}(10, 1)$  and the weighted contribution of each order ( $m_2$ ,  $m_3$  and  $m_4$ ) on the 25 windows for a given value of  $c$  is indicated by a black dot and the SD by the black bars extending from that dot. The maximum and minimum values of  $m_{\text{avg}}$ ,  $m_2$ ,  $m_3$  and  $m_4$  obtained on the surrogates were recorded for each window. The red dots, connected with red lines to improve readability of the graph, indicate the averages of these 25 maxima and 25 minima for each value of  $c$ .



the different TS. Nonetheless, the average values of  $m_{\text{avg}}$ ,  $m_2$  and  $m_4$  already show the anticipated decrease in information flow with increasing Lyapunov exponent, which increases more or less monotonically  $c$  on our data (cf. Fig. 2.5). As the degree of chaos increases, information about the past is lost more rapidly and so less information flows from the block of  $n$  samples to the future. With longer windows (Fig. 4.4) and the same sampling rate, the SDs decrease sufficiently to allow us to distinguish three categories of TS:  $c \in \{1\}$ ,  $c \in \{2, 3\}$  and  $c \in \{5, 10\}$ . Here the surrogates appear to reproduce second-order results quite well and still overlap much of the range of  $m_4$ .

The third-order results are about two orders of magnitude smaller than the second and fourth order results and there is no apparent trend. Furthermore the surrogates fully contain even the error bars on these results. Increasing window length by a

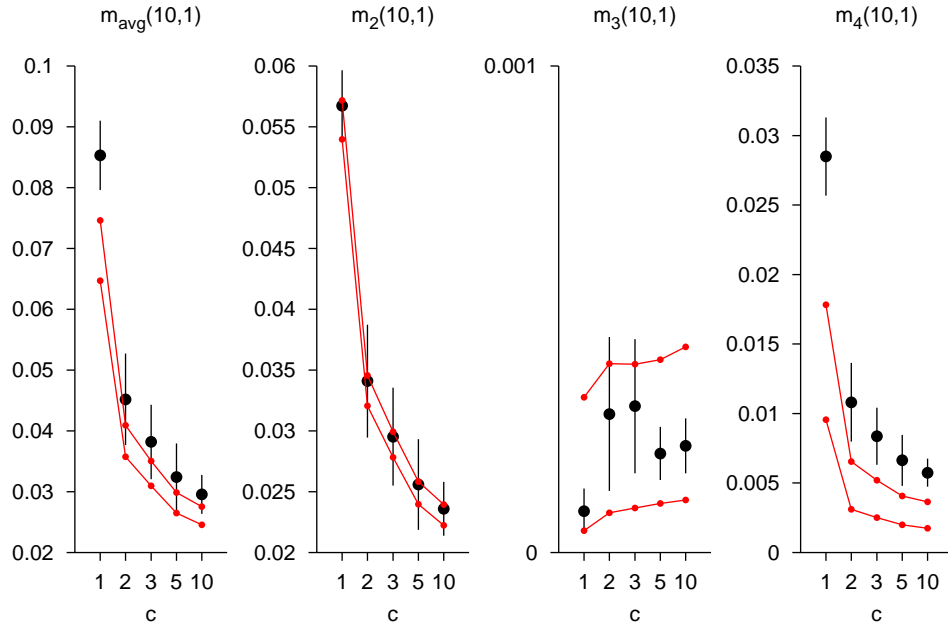
factor of eight (Fig. 4.4) decreases the  $m_3$  values by another order of magnitude while producing little effect on the other orders. Third-order cumulants thus appear to deviate from zero only because of statistical fluctuations in the measure.

We were initially concerned that this result stemmed from a implementation error. However, the same code had been used by F. Beaulieu in [4]. In his thesis, IIF was computed on an artificial time series consisting of  $N = 16384$  random deviates on  $[0, 1]$  followed by  $N$  iterates of the tent map and finally  $N$  iterates of the binary shift map (Bernoulli map). It was precisely the differences in  $m_3$  between the segments that proved necessary to distinguish the tent map segment from the others.

While we cannot prove that  $m_3$  tends to zero on data produced by the Bondarenko model, or for that matter on EEG data, we can make some observations about third moments in general and the distribution of these data. The third moment of a 1D distribution characterizes the degree of asymmetry about the distribution's mean [66]. It should be fairly evident from the sample model output in Figs. 2.2 and 2.3 that this signal is symmetrical about zero. To check this, we concatenated the 25 time series for  $c = 2$  and computed histograms from the resulting 204 800 data for each of the 10 outputs (neurons). These histograms (not shown here) were indeed symmetrical about zero, although by no means normal; this distribution is preserved in the phase-shuffled surrogates we have used. It is also known that “the third-order moment of a normal process with zero mean is identically zero...[because] the joint density of three jointly normal random variables with zero mean is symmetrical with respect to the origin.” [62]. In light of symmetry considerations such as these, the apparent tendency of  $m_3$  towards zero on these data is less alarming.

For IIF computed on data sampled at a lower frequency or by introducing a delay, we see the average fourth-order results rise clearly above the average maximum of surrogates. In Fig. 4.2, windows of  $\ell_w = 1024$  data sampled at intervals of  $\Delta t = 10.0$  were used with no delay, whereas in Fig. 4.3 a delay of  $\tau = 10$  was introduced in windows of  $\ell_w = 8192$  data sampled at  $\Delta t = 1.0$  intervals. The results in both cases

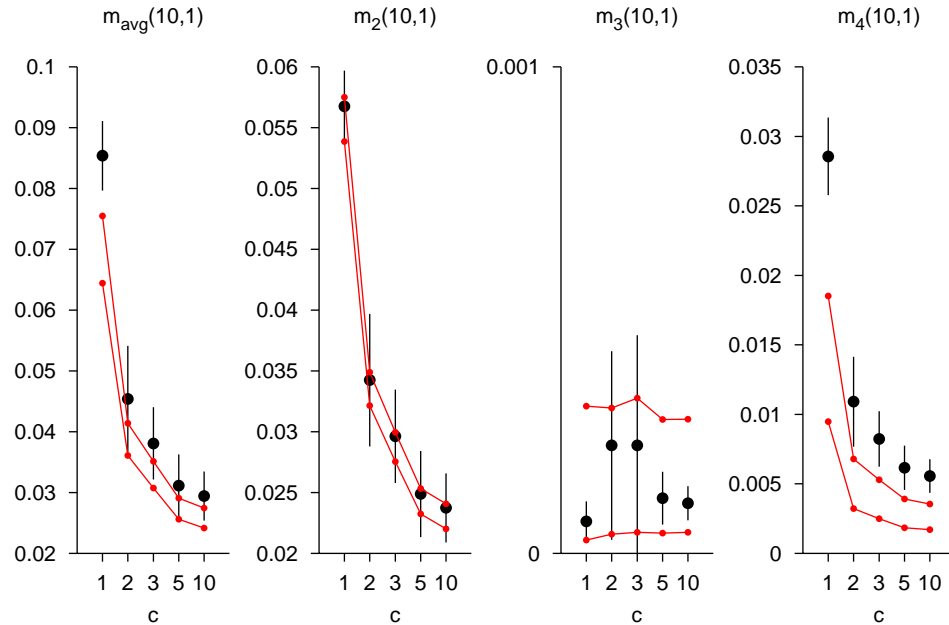
Figure 4.2: IIF results on the  $c$  series with  $\ell_w = 1024$  and  $\Delta t = 10.0$ . Further details in the caption to Fig. 4.1.



are very similar except for  $m_3$ , which is still apparently lost in the noise floor. This similarity is important because it means that use of a delay on oversampled data is sufficient to access the sensitivity to higher order correlations of the process that a more appropriate sampling rate would have produced.

Computing IIF on data sampled at intervals much shorter than the characteristic time scales of the dynamical processes underlying the TS makes the measure oversensitive to short-range correlations; a lot of information flows from the  $n$  points in the past to the future point because the system has hardly changed. This explains why the average values obtained without using a lag or on data sampled at a lower rate are about twice those obtained using samples more distant in time. This is also reflected in the fact that with a long window of  $\Delta t = 1.0$  data, second-order results on the surrogates match the average result on the raw data closely (Fig. 4.4) but this is lost when a delay is introduced (Fig. 4.5). In the former case, the IIF measure is dominated by linear correlations which, by construction, are all that matter in the

Figure 4.3: IIF results on the  $c$  series with  $\ell_w = 8192$  and  $\Delta t = 1.0$  where **a lag of  $\tau = 10$  has been introduced in computing the measure**. Further details in the caption to Fig. 4.1.

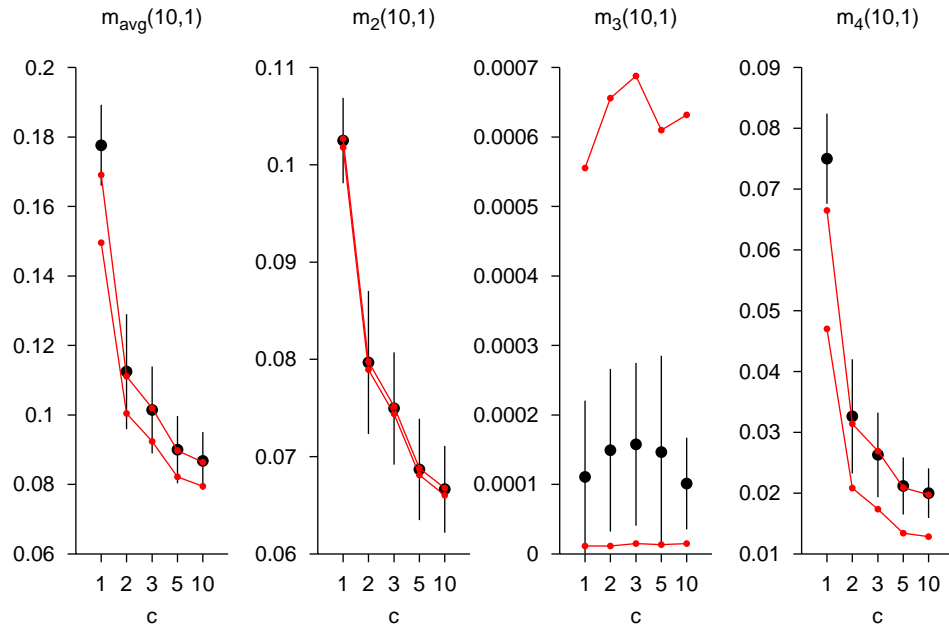


surrogates. This is of course related to the problem of choosing a delay for delay coordinate embedding in dynamical reconstruction.

The SDs in these results are similar to those obtained with the longer windows on data sampled at intervals of  $\Delta t = 1.0$  with no delay (Fig. 4.4). This suggests that total window *duration* has the most influence on variations in IIF values from window to window for a given value of the control parameter and not just the number of samples in the window; the windows in Fig. 4.4 and Fig. 4.2 are about 10 times longer in duration than those in Fig. 4.1. Windows longer in duration “see” more of the overall dynamics of the system and so yield more consistent values for IIF on stationary TS.

Finally, the IIF results for different values of  $c$  using long windows ( $\ell_w = 8192$  samples) of data sampled at intervals of  $\Delta t = 10.0$  (Fig. 4.5) are significantly distinct, even the error bars now clear the average surrogate maxima for  $m_4$ , the  $m_2$  results

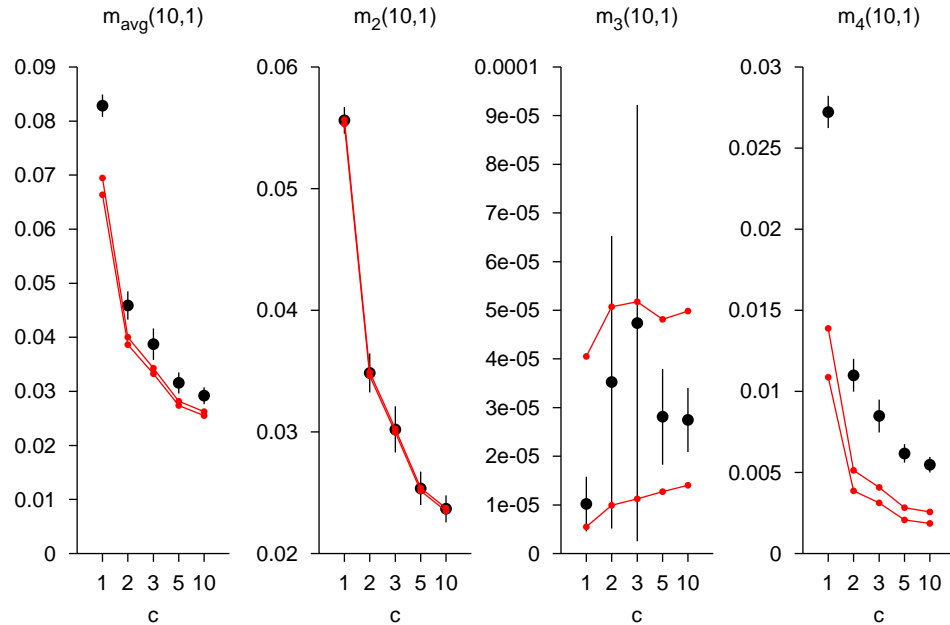
Figure 4.4: IIF results on the  $c$  series with  $\ell_w = 8192$  and  $\Delta t = 1.0$ . Further details in the caption to Fig. 4.1.



for raw and surrogate data are once again very similar and the overall decrease in information flow with increasing  $c$  is maintained. One must keep in mind that these results were obtained using eight times as many data points as those in Fig. 4.1, on windows spanning an interval 80 times longer. So while it is satisfying to see such clear discrimination, it is not that surprising.

In principle, linear correlations in the random process do not influence higher order cumulants. For example, we see a correction for these correlations in the expression for fourth-order cumulants (B.12), where combinations of second moments are subtracted from the fourth moments to yield the fourth cumulants. Third and higher order cumulants in *perfect* phase-shuffled surrogates—in which the linear correlations and distribution of values are perfectly preserved and all other correlations destroyed—would equal zero in the  $n \rightarrow \infty$  limit. Another way of saying this is that in our surrogates, in contrast to the raw, chaotic data, no information is conveyed through higher order correlations. When we can only estimate the cumulants of a process from

Figure 4.5: IIF results on the  $c$  series with  $\ell_w = 8192$  and  $\Delta t = 10.0$ . Further details in the caption to Fig. 4.1.



finite amounts of data, as is always the case in practice, and must limit the block length  $n$  to a manageable length, this correction must be imperfect. This explains in part why the values of  $m_4$  on the surrogates are clearly not just fluctuations above zero; they preserve the trend observed in  $m_2$ . As the number of data in and duration of the windows used in estimating the cumulants increases, the value of  $m_4$  on the raw data appears to stabilize (compare Fig. 4.4 and Fig. 4.5) whereas the average maxima on the surrogates continues to decrease. On the other hand, the average minima on the surrogates seem to have stabilized and still show the same trend. This would be a consequence of limiting  $n$  to ten.

#### 4.4.2.2 The $e$ Series

In these TS, it is the frequency and duration of bursts of periodicity that increase with increasing values of  $e$ . Figures 4.6 through 4.9 show the results obtained with different window lengths and sampling intervals. All of the remarks from the previous section concerning window length and duration, the behaviour of  $m_3$ , the values of  $m_4$  for the surrogates and the influence of sampling rate apply to these results as well. The main difference we note is the persistence of large error bars in  $m_4$  for the higher values of  $e$  even when  $\ell_w = 8192$  and  $\Delta t = 10.0$  (conditions under which the error bars had all but vanished in Fig. 4.5). As more and more of the TS becomes periodic, we see a corresponding increase in information flow; the future of a periodic signal is entirely dependent on its past.



Figure 4.6: IIF results on the  $e$  series with  $\ell_w = 1024$  and  $\Delta t = 1.0$ . Further details in the caption to Fig. 4.1.

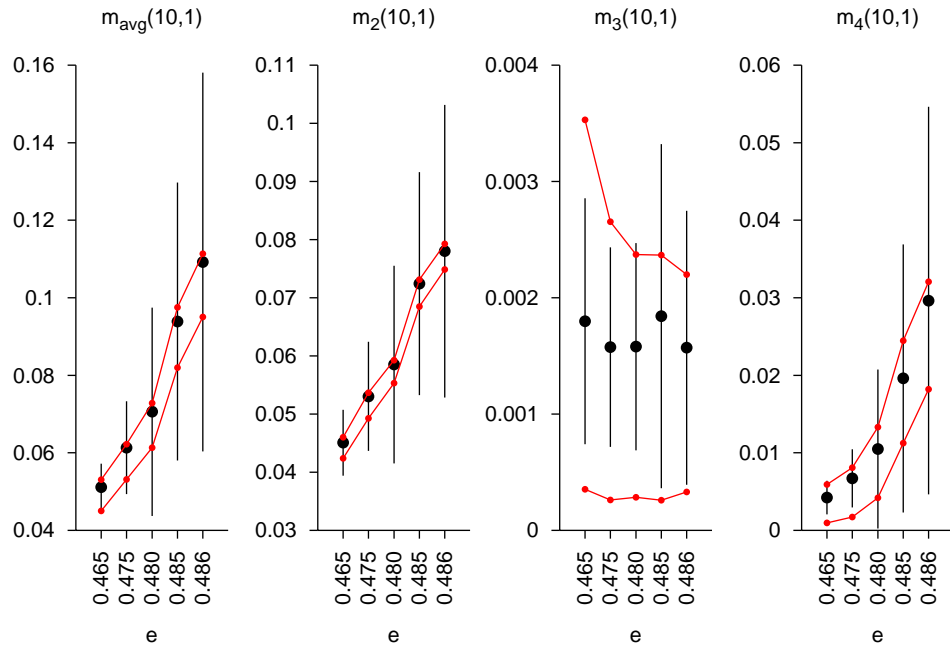


Figure 4.7: IIF results on the  $e$  series with  $\ell_w = 8192$  and  $\Delta t = 1.0$ . Further details in the caption to Fig. 4.1.

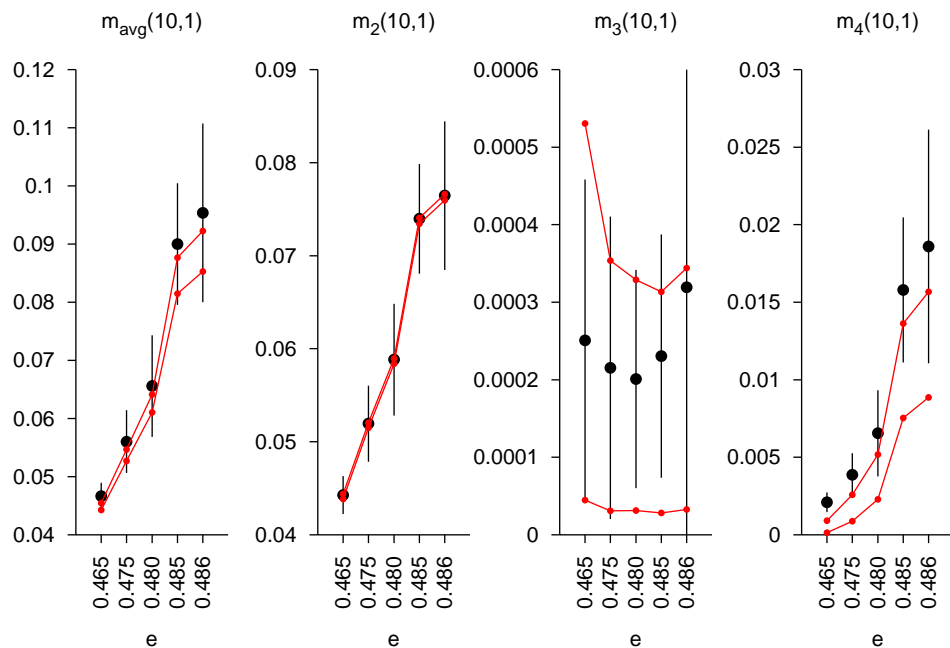


Figure 4.8: IIF results on the  $e$  series with  $\ell_w = 1024$  and  $\Delta t = 10.0$ . Further details in the caption to Fig. 4.1.

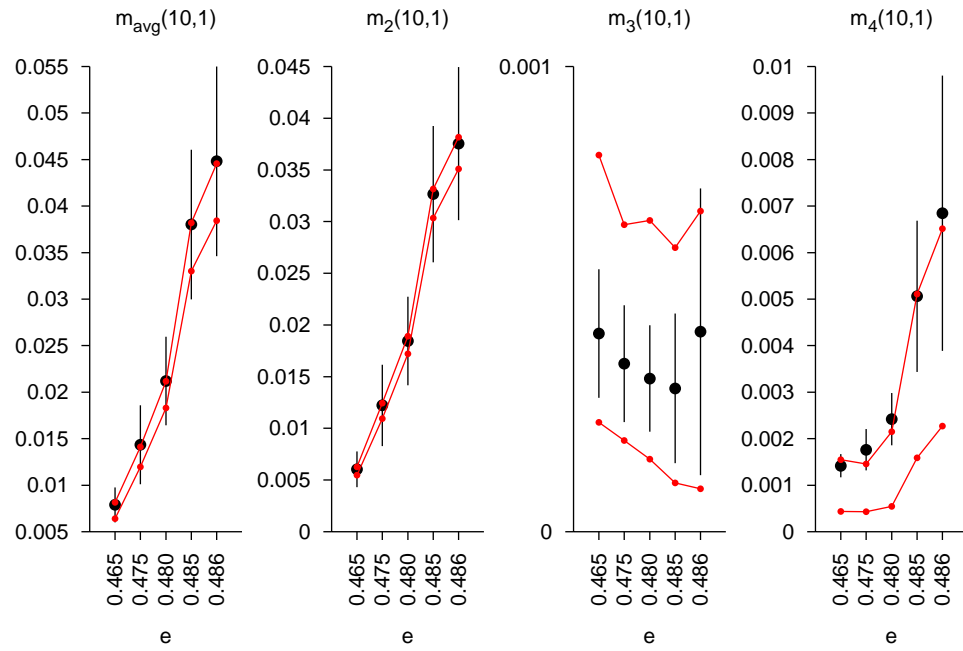
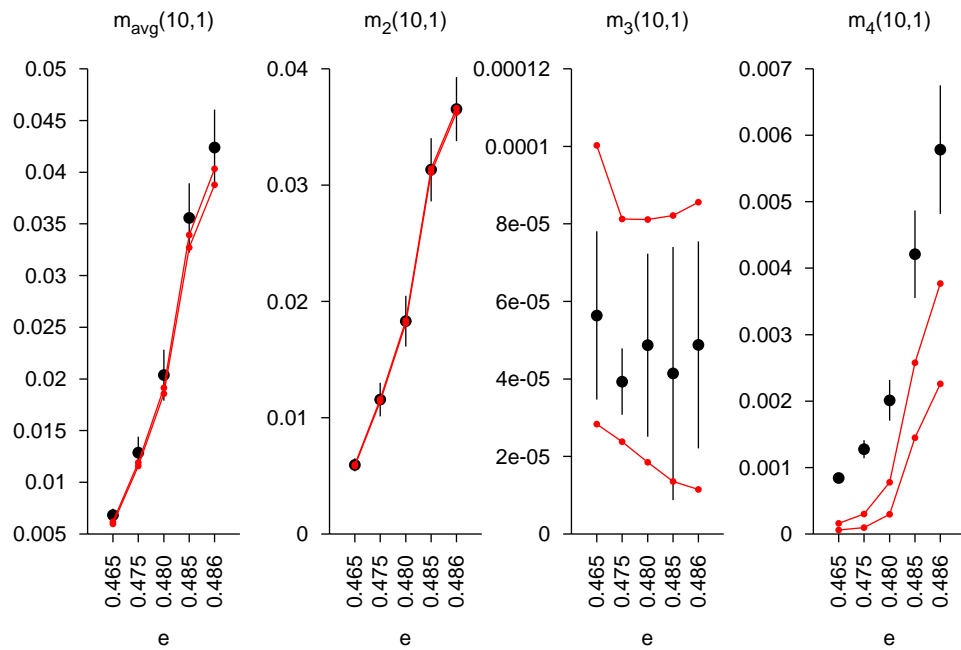


Figure 4.9: IIF results on the  $e$  series with  $\ell_w = 8192$  and  $\Delta t = 10.0$ . Further details in the caption to Fig. 4.1.



## 4.5 Experiments on EEG Recordings

In previous studies, the cumulant-based measure had been computed on a single, short TS of EEG data ending just before a clinical seizure [18] and two TS of manually corrected intervals between interictal spikes (interspike interval (ISI)) [4], which are characteristic of epileptiform brain activity, also ending just prior to a seizure. Results of these experiments were encouraging but inconclusive. This is the first time the measure has been computed on a large set of EEG recordings from epileptic brains. We will be investigating the measure's overall behaviour, its strengths and weakness and performance in anticipating epileptic seizures.

### 4.5.1 Methods

We have used the results of the previous section as a guide in choosing window lengths. Fair to excellent discrimination of our artificial time series, which are characterized by varying degrees of chaoticity or intermittency and qualitative similarity to EEG signals, was obtained using the cumulant-based measure of information flow on windows spanning an interval of about  $10^4$  time units. We had identified several strong peaks in the spectra of these data, the strongest having a frequency of approximately  $0.02$  [time units] $^{-1}$  or a period of 50 time units (Fig. 2.4). Thus, on the order of 200 cycles of this oscillation would be contained in a typical window. If we associate this dominant oscillatory component with the approximately 10 Hz alpha rhythm [1, 8], this would be equivalent to 20 seconds of EEG data or  $4 \times 10^3$  samples of data digitized at 200 Hz. We settled on window lengths of  $\ell_w = 8192$  samples or 42 sec and  $\ell_w = 2048$  sample (the FFT algorithm used as part of the surrogate generation routine requires that  $\ell_w = 2^n, n \in \mathbb{Z}^+$ ).

The cumulant-based information flow measure (4.14) was computed on nonoverlapping windows of these lengths spanning the entirety of every signal in every recording in the data set using various combinations of values of the measure parameters: block lengths of either  $n = 5$  or  $n = 10$ , maximum look-aheads  $r_{\max}$  of 1 or 5 and

delays  $\tau$  of 5 or 10 samples. A quick survey of autocorrelation functions computed using the first 8192 samples after the starting point designated for reference window 1 (see Fig. 2.8), i.e. about 5 min into the recording, on the first intracranial signal of several recordings revealed first zero crossings of this function at between 20 and 40 samples. Our choice of  $\tau$  values was guided by this observation. Although the delay is necessary to reduce the influence of short range correlations, results should not depend strongly on the precise value used and indeed do not appear to do so.

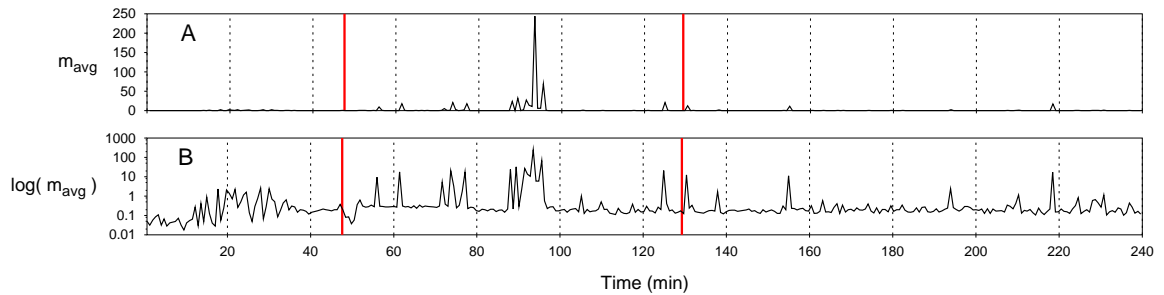
## 4.5.2 Results and Discussion

For the same values of  $n$ ,  $r_{\max}$  and  $\tau$ , results for the two different window lengths were quite similar, with greater statistical fluctuation on the shorter windows as one would expect. Interestingly, values obtained with  $\ell_w = 2048$  seldom dipped below those obtained on the longer windows. The exception to this is on sharp peaks, which generally extend to higher values on the longer windows. We will concentrate on the  $\ell_w = 8192$  results.

One of the main difficulties with this measure is immediately apparent in Figure 4.10, which shows its variation over the *LA1* channel of *amysez01*. As the measure is bounded only from below, regions of the TS with strong periodic components—such as frequently occur during seizures—will yield high values of the measure. By considering the full range of values obtained on a given series, less pronounced changes that might occur in the putative preictal period are obscured. Use of a logarithmic scale can help overcome this effect, but more interesting features are seen by “zooming in” on the low end of the scale, as we have done in Fig. 4.11 and subsequent figures.

There is a rise in  $m_{\text{avg}}$  before the seizure but it may be surprising not to see a very strong response in the measure *during* the seizure. There *is* a strong rise in the weighted value of  $m_2$  during the seizure, which then gradually returns to the levels seen at the start of the recording. Since this component is only a third of the total  $m_{\text{avg}}$  signal and  $m_3$  and  $m_4$  do not rise during the seizure, the fluctuation in  $m_2$  is

Figure 4.10: Full range (A) of  $m_{\text{avg}}(10, 1)$  obtained using a delay of  $\tau = 10$  and  $\ell_w = 8192$  on channel *LA1* (the first intracranial signal) of *amysez01*. A logarithmic scale (B) reveals a possible preictal trend. Red vertical lines mark onset of clinical seizures.

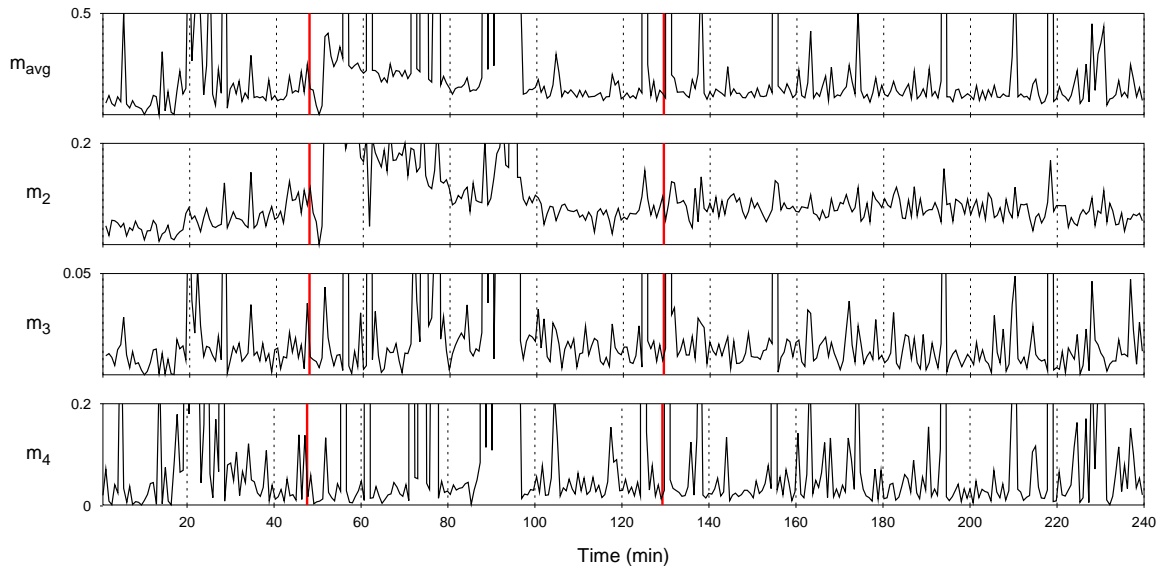


not easy to pick out in the average measure. A strong periodic component will have this influence on  $m_2$  rather than the high orders. If we consider only the aggregate measure,  $m_{\text{avg}}$ , it could be said that we have traded sensitivity to spectral changes for sensitivity to high order correlations; this trade-off may not prove fruitful. We should recall that this measure is not sensitive to changes in signal amplitude that are slow relative to the window length.

In Fig. 4.11, we see a rise in  $m_2$  which echoes the changes observed in the amplitude-frequency analysis of the same data (cf. Table 2.2). The same trend is even more pronounced in, for example, the *LA3* channel data from the same recording (Fig. 4.12). In the latter results, we also clearly see a “bump” in the  $m_3$  and  $m_4$  values before the seizure sufficient large to persists in  $m_{\text{avg}}$ . It is unclear to us why preictal fluctuations in the higher order components do not persist and increase during the seizure.

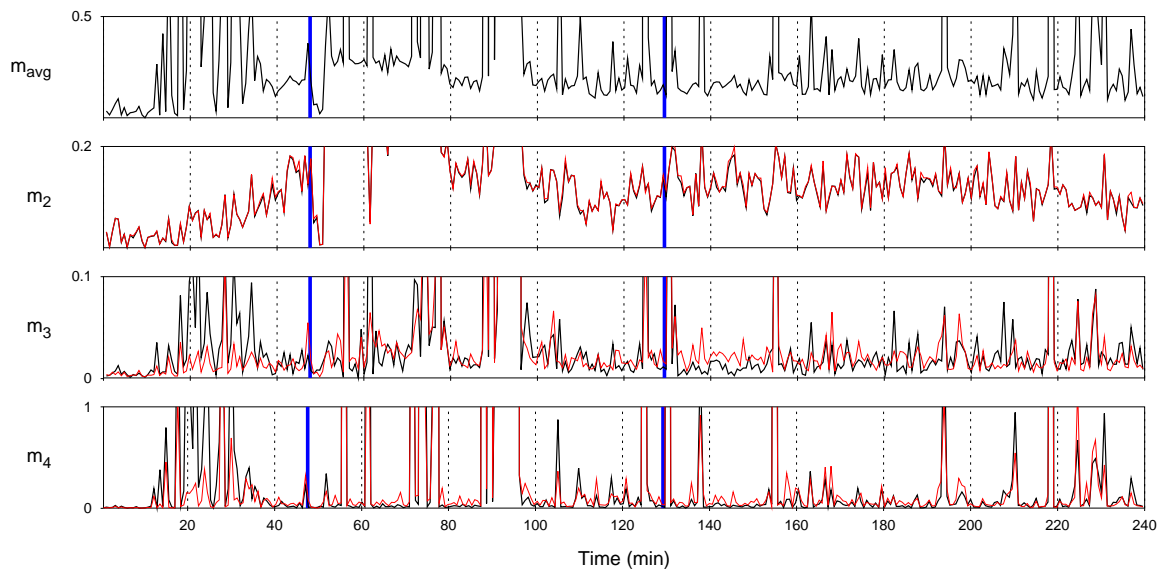
One would like to attribute these fluctuations in the higher order cumulants to preictal changes in brain dynamics, to a gradual build-up of synchrony and decrease in complexity and an attendant rise in information flow. The method of surrogate data described earlier allows us at the very least to avoid making such a claim if the fluctuations can in fact probably be caused by linear correlation of the data. We have not systematically computed surrogates on all of the EEG data. One reason for this

Figure 4.11: Values of  $m_{\text{avg}}(10, 1)$  with breakdown into weighted second, third and fourth cumulant contributions— $m_2$ ,  $m_3$  and  $m_4$ —obtained using a delay of  $\tau = 10$  and  $\ell_w = 8192$  on channel *LA1* of *amysez01*. The vertical scale is adjusted in each graph to highlight variations in the measure floor. Red lines mark onset time of clinical seizures.



is that, to produce and run the computation on a reasonable number of surrogates, it would have taken months to complete even a single run of the experiment, given and length and number of TS involved.. Furthermore, the influence of interictal spikes on the phase-shuffled surrogates was a concern (a point raised in discussions with M. Le Van Quyen and J. Martinerie). Isolated large amplitude spikes in the raw data can produce multiple large spikes in the surrogates, a sort of ringing artifact of the Fourier transforms, and these will likely influence the value of measures on these surrogates [75]. On the basis of the difference between values on raw and surrogate TS, one might conclude that nonlinearity is present when in fact, the surrogates are clearly, qualitatively different from the original TS. The cost of systematic use of surrogates seemed to outweigh the insight they might bring to the analysis. It was also apparent that on many of the recordings, there was either no appreciable preictal trend in the higher order cumulants or that any trend in  $m_{\text{avg}}$  was due primarily to changes in  $m_2$ , i.e. autocorrelation in the data.

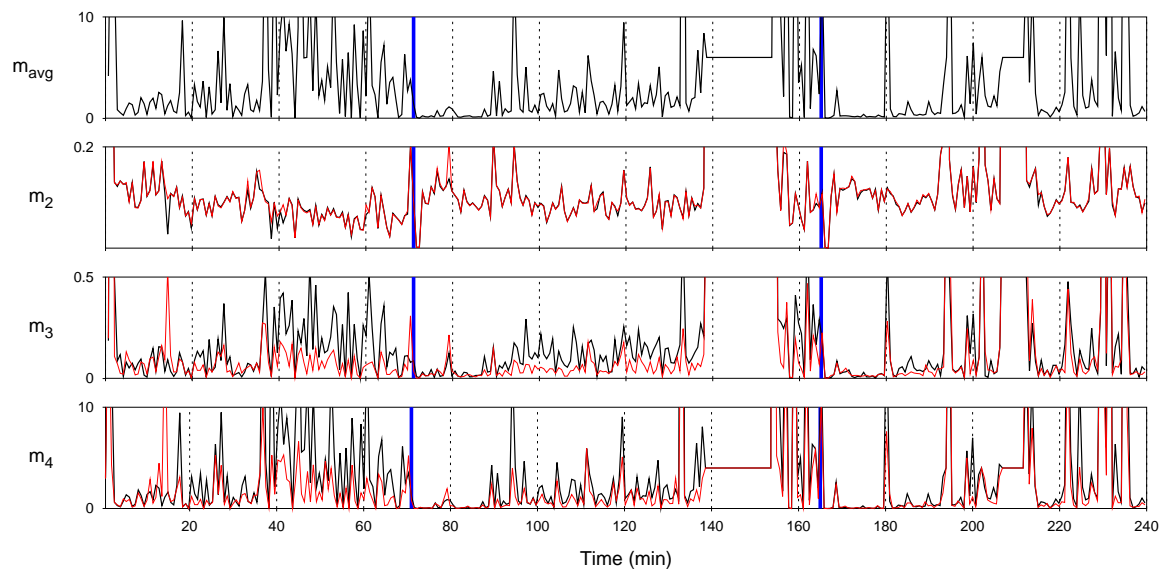
Figure 4.12: Values of  $m_{\text{avg}}(10, 1)$  with breakdown into weighted second, third and fourth cumulant contributions— $m_2$ ,  $m_3$  and  $m_4$ —obtained using a delay of  $\tau = 10$  and  $\ell_w = 8192$  on channel *LA3* of *amysez01* and surrogates. 19 surrogates were generated per window; the *maximum* value of the measure and its components on the surrogates is indicated by the thin red line. The vertical scale is adjusted in each graph to highlight variations in the measure floor. Red lines mark onset time of clinical seizures.



Nonetheless, we have included two examples of TS on which  $N_s = 19$  surrogates were computed on each window, signal *LA3* from *amysez01* (Fig. 4.12) and signal *LA1* from *bobsez01* (Fig. 4.13). The maximum surrogate value is shown. Using one-sided rank order statistics, a value for the raw data above this maximum would indicate that the null hypothesis (i.e. data generated by a linear Gaussian process filtered by a possibly non-linear measurement function) could be rejected at the 95% confidence level. In Fig. 4.13, we see a slight drop in  $m_2$  which we would have anticipated from the amplitude-frequency analysis of this recording presented in Chapter 2.2. That does not explain, however, the higher values and overall rise in  $m_3$  and  $m_4$  in the  $\sim 30$  min before the seizure. Surrogate data fails to reach similar levels, so it would be reasonable to think that these fluctuations were the result of changes in the nonlinear structure underlying the TS.

We do observe preictal increases in the measure on some of the recordings, but

Figure 4.13: Values of  $m_{\text{avg}}(10, 1)$  with breakdown into weighted second, third and fourth cumulant contributions— $m_2$ ,  $m_3$  and  $m_4$ —obtained using a delay of  $\tau = 10$  and  $\ell_w = 8192$  on channel LA3 of *bobsez01* and surrogates. 19 surrogates were generated per window; the *maximum* value of the measure and its components on the surrogates is indicated by the thin red line. The vertical scale is adjusted in each graph to highlight variations in the measure floor. Vertical red lines mark onset time of clinical seizures.





at the end of the day, it is the performance of the measure in anticipating epileptic seizures on these recordings which we wish to evaluate. Can these fluctuations be detected in a timely, *automated* manner? That will be our concern for the remainder of the section. A second problem arises here, in direct opposition to what should be one of the strengths of this measure. As mentioned at the head of this chapter, the cumulant-based IIF is an absolute rather than a relative measure. The increase in neuronal synchrony believed to precede a seizure should produce an increase in the absolute value of this information flow measure. It is computed on each window independently of the others. In light of this, one might imagine that some absolute *threshold* value in  $m_{\text{avg}}$  might be determined for a given patient or group of patients. When the measure rose above this threshold and remained there for a certain time, the system would raise a flag in anticipation of a seizure.

To get an idea of how this might work, we proceeded as follows. A straight line was fit to the  $m_{\text{avg}}$  results for every signal in the recordings with seizures on an interval *selected by hand* for each recording. In choosing the interval, we generally tried keep as much of the pre-seizure data as possible while avoiding gaps and in some cases favouring results closer to the seizure. If the slope  $m$  of the fit was positive, it was recorded. We denote the number of signals which yield a positive slope on each recording as  $N_+$ . Let  $\bar{m}_+$  denote the average of all the positive slopes obtained for a given recording. As a rough indicator of the possible rise in the measure floor over the interval on this recording, we use  $\gamma = \bar{m}_+(t_{\text{end}} - t_{\text{start}})/2$  which is the rise of the line over half the interval. Lastly the average value of the measure on the interval  $\bar{m}_{\text{avg}}$  was also recorded. These values are shown in Table 4.1.

On the recordings with low  $N_+$  we already anticipate problems with the threshold approach. Based on the values of  $\gamma$  for each patient's recordings and favouring those recordings with a high  $N_+$ , we chose a threshold to try for that set of recordings. Using a time threshold of  $\tau_{\text{th}} = 4$  windows and no grace period along with these threshold values, we find the hot zones depicted in Fig. 4.14 (the caption to Fig. 3.11

Table 4.1: Positive slope statistics on recordings with seizures. See text for details.

Recording	Fit Interval		$N_+$	$\gamma$	$\bar{m}_{\text{avg}}$	Threshold
	Start	End				
<i>amysez01</i>	0	2500	59	0.12	0.25	0.1
<i>amysez02</i>	2000	7780	54	0.097	0.14	
<i>amysez03</i>	8000	10000	47	0.052	0.11	
<i>bobsez01</i>	500	4000	14	1.54	2.29	0.3
<i>bobsez02</i>	0	7000	12	0.23	1.18	
<i>bobsez03</i>	0	10000	31	0.51	1.45	
<i>carlsez01</i>	8000	10000	32	0.22	0.35	0.15
<i>carlsez02</i>	2000	8000	28	0.13	0.17	
<i>carlsez03</i>	0	10000	35	0.052	0.17	
<i>dansez01</i>	0	12500	49	0.12	0.42	0.15
<i>dansez02</i>	0	15000	43	0.076	0.20	
<i>dansez03</i>	0	8000	15	1.00	1.03	
<i>evesez01</i>	7000	15500	59	0.35	0.37	0.2
<i>evesez02</i>	0	2560	50	1.22	0.92	
<i>evesez03</i>	2250	5000	19	0.027	0.25	

$N_+$  is the number of signals which yield a positive slope

$\gamma$  is the rise over half of fit interval

$\bar{m}_{\text{avg}}$  is the average value of  $m_{\text{avg}}$  on the fit interval

on p. 66 explains in detail how to read these figures). Things start out well with the *amy* recordings, as usual. After that, however, the results are less encouraging. We see a tremendous number of free hot zones in all of the recordings, including the *awake* series. There is an overall increase in number and duration of hot zones in the lead up to the seizures in *carlsez01* and *carlsez02*.

Absolute thresholds would of course be more meaningful if the measure could be normalized or at least an upper limit established, but this problem is directly related to that of finding the distribution of the measure on data from a given process. When a good model of the governing dynamics of a system is available—which is not the case here—synthetic data can be generated and the distribution estimated numerically. It would be futile to try to “optimize” the thresholds for the few recordings we have for each patient.<sup>4</sup> Better to try a different approach.

With the exception of the cross-recording experiments, the EEG results of the previous chapter were obtained using a unique threshold for *each* channel on *each*

---

<sup>4</sup>I say this having, of course, spent far too much time in the quest for the “magic threshold” that would make the preictal period leap off the graph. It doesn’t appear to exist.

---

Figure 4.14 (*following page*): Intervals on which  $m_{\text{avg}}$  remains *above* the threshold selected for that patient’s recordings,  $m_{\text{th}}$ , for at least  $\tau_{\text{th}} = 4$  windows are defined as “hot zones”. These zones are indicated in the figure for each signal from every recording. If  $m_{\text{avg}}$  dips below  $m_{\text{th}}$ , the hot zone ends (i.e. no grace period). Results for  $n = 10$ ,  $r_{\text{max}} = 1$ ,  $\tau = 10$  and  $\ell_{\text{w}} = 8192$ . Results from intracranial EEG channels are in black, surface EEG is in purple. Red lines mark the start of clinical seizures. Blue bars along the bottom show the location and extent of gaps in the recording. Black tic marks on the bottom edge delimit the length of the recording, the full width of the box corresponding to 420 min. Black tic marks on the left-hand border indicate the extent of the signals for that recording; a box can accommodate a maximum of 64 signals. Generally, channels in the top half of this range were in the right hemisphere and those in the bottom the left. Ineligible channels are marked with a green tic on the left (this mark overwrites the black mark in some cases). Those on which the IIF algorithm failed for any reason are marked with a green tic on the right.

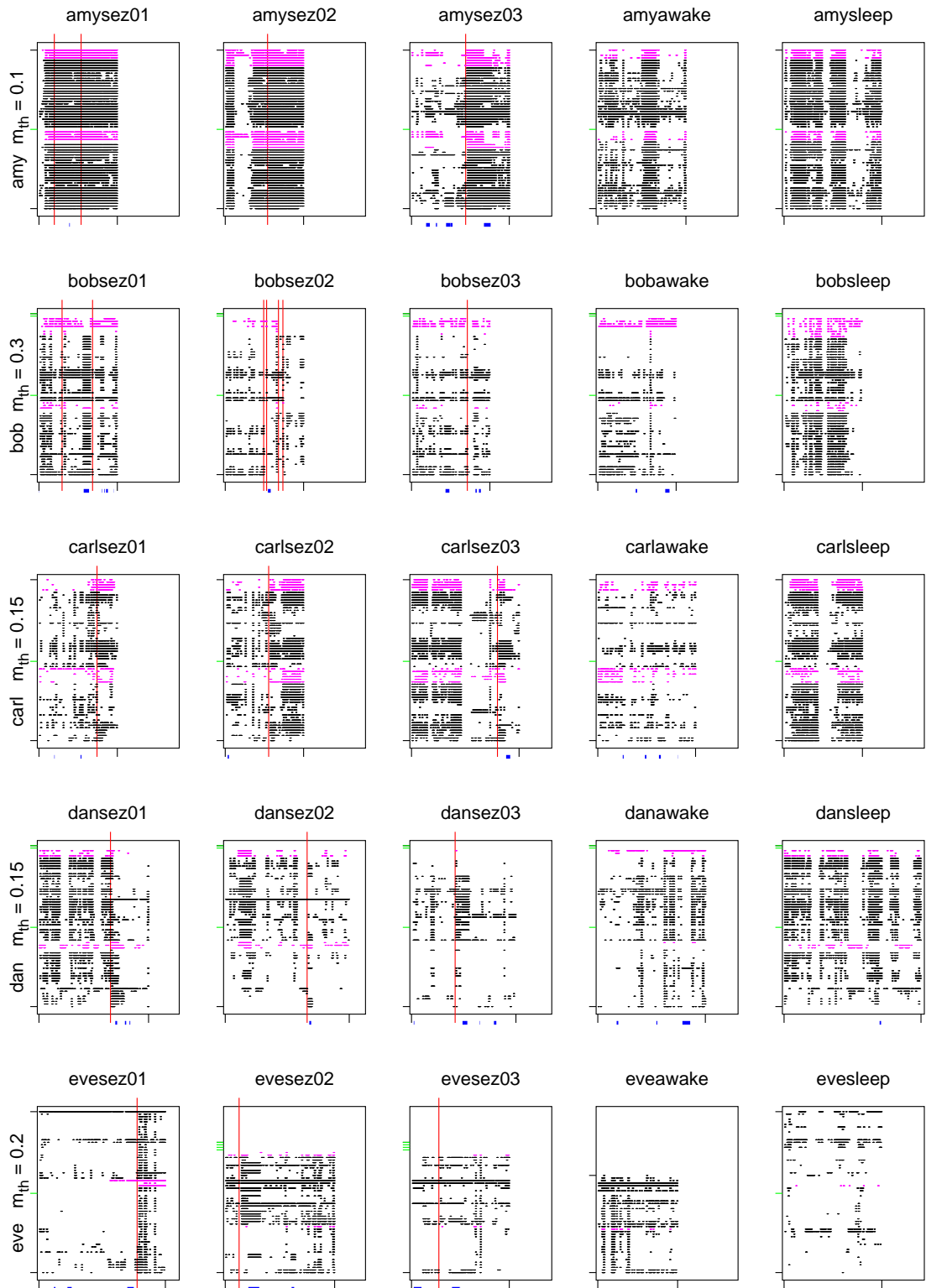
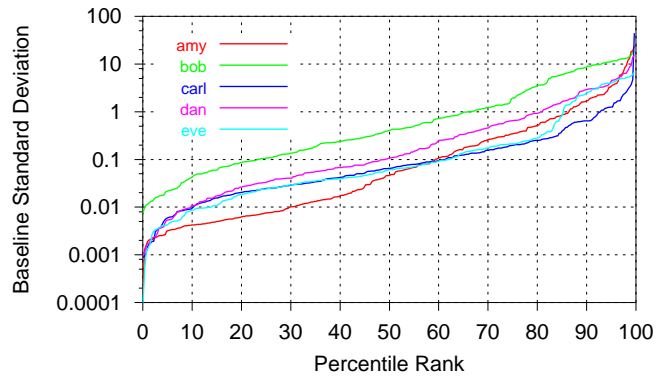


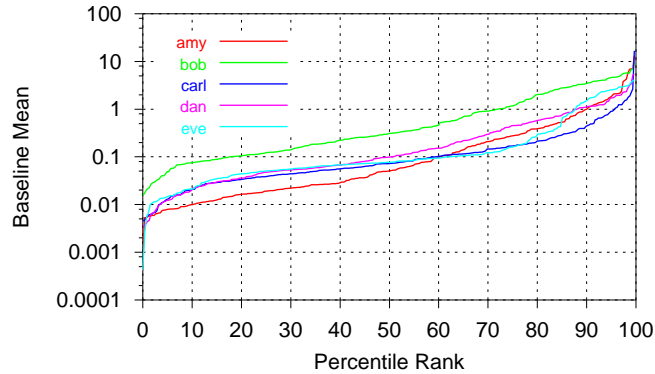
Figure 4.15: Percentile rank of baseline standard deviation values  $\sigma$  for  $m_{\text{avg}}$  for each patient's recordings.



recording. As we pointed out in our discussion there, the mean and variance of the similarity values on the reference windows were generally very similar across all the channels in a recording. Indeed, the cross-recording experiments demonstrated that a good reference window—and hence its baseline statistics—was generally good across all recordings for the patient in a similar state of arousal. We should give the information flow measure a fair chance by putting aside its “absolute” nature and using some sort of baseline statistics. Taking the first 8 windows after the start of the first reference window (see Fig. 2.8) as our baseline results for each channel, we compute the average  $\mu$  and standard deviation  $\sigma$  of, to begin with,  $m_{\text{avg}}$ .

Contrary to what we observed with the similarity measure, here the values of both statistics vary enormously from signal to signal within a recording. Figure 4.15 shows the percentile rank of different values of  $\sigma$  for each of the patient’s recordings (i.e. its rank in the approximately 300 baseline values from the signals in those recordings). Even on the “normal” percentile range of 25-75,  $\sigma$  covers two orders of magnitude! We notice that the values chosen for the threshold approach to anticipation correspond generally to the 50 percentile level. The values for  $\mu$  (Fig. 4.16) are similarly distributed as is to be expected when  $m_{\text{avg}}$  fluctuates widely on the baseline; whatever strong spikes there are in those 8 windows will dominate the mean. The variance is

Figure 4.16: Percentile rank of baseline average values  $\mu$  for  $m_{\text{avg}}$  for each patient's recordings.



then just related to the difference between those spikes and the remaining, small values. Because of this characteristic of the baseline statistics, it is pointless to use the “ $n$  sigma’s from the baseline mean” approach to identifying hot zones. Since the baseline mean is generally inflated by spikes, it is unlikely that the measure will rise significantly above it for a sustained period of time. We tried scaling down  $\mu$  and  $\sigma$  in various ways as well as simply using  $\mu$  times a constant factor as the threshold. For  $m_{\text{avg}}$ , the results were always *much* worse than those obtained with fixed patient thresholds; signals seemed to be all hot or all cold with clumps of each alternating in most recordings. On the other hand, using  $\sigma_{\text{th}} = 2$  and  $\tau_{\text{th}} = 4$  windows and indeed the usual “sigma’s from mu” approach on the  $m_2$  results produced considerably more satisfactory results as shown in Fig. 4.17. When a seizure in this data set is clearly announced by changes in IIF, it would appear to be the linear terms in the measure doing the talking.

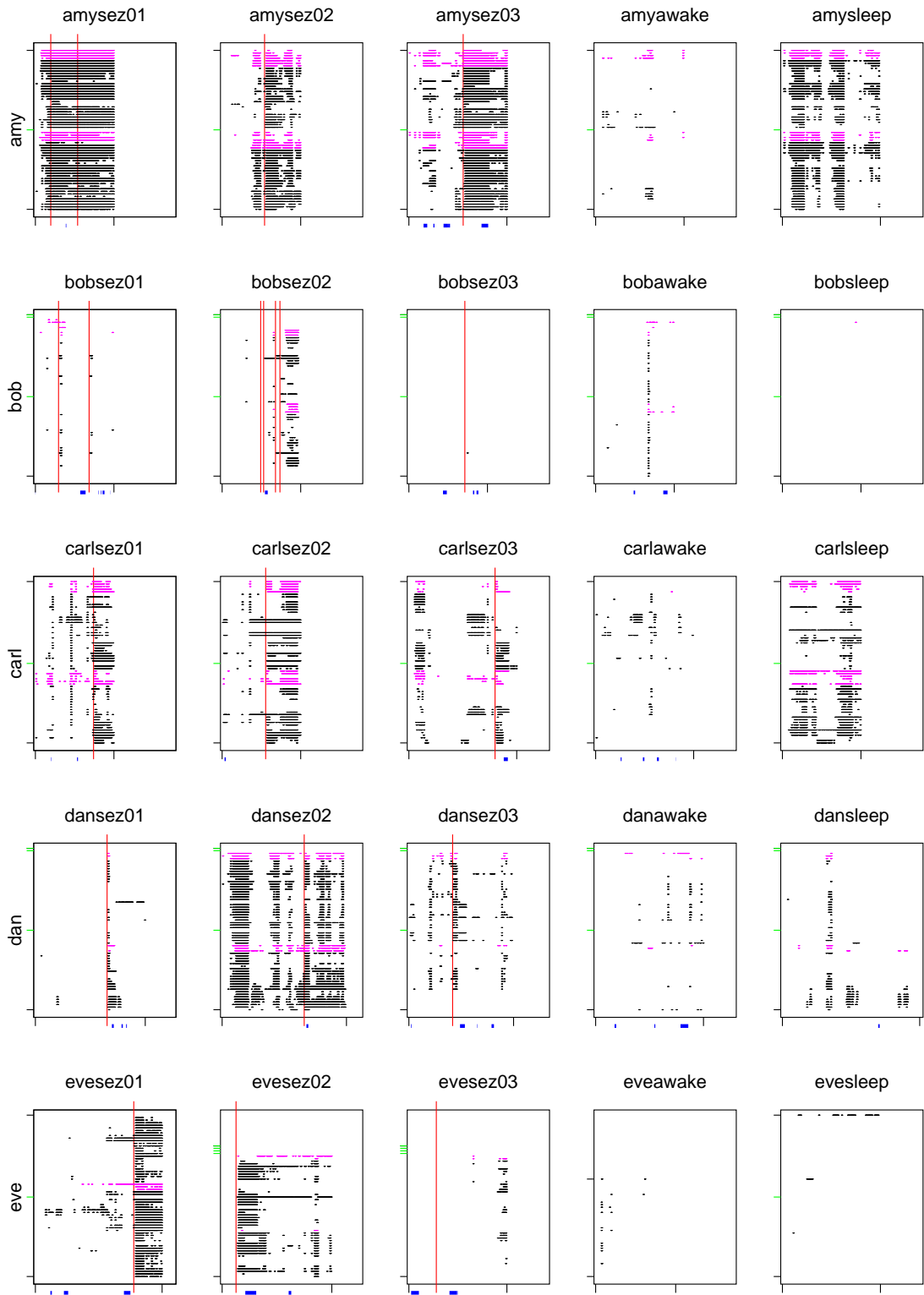
The same exercise with the third order cumulant term  $m_3$  produces basically no hot zones when the same threshold parameters are used. Various absolute thresholds produced what seemed like random bands of hot zones and a peppering of isolated hot zones. Results were similar with  $m_4$ , with echoes of the  $m_2$  results. We have found the

main factors decreasing the visibility of overall trends in this measure to be the large fluctuations in the terms involving higher order cumulants and the presence of large spikes in the measure. Although the statistical fluctuations of the higher order terms could be reduced by using larger windows, one must not lose sight of the fundamental assumption of stationarity on any given window. Perhaps we could “smooth out” the fluctuations with a moving average filter or some interpolation scheme such as fitting a Bezier curve to a moving window of results? Even if only causal filters are used (this is supposed to be a predictive measure, after all), these methods add to the “inertia” of the measure, making it less responsive. We nonetheless did examine moving averages of the result; there was no improvement in performance.

In light of results obtained with this measure and the comments above, it would be pointless to compute and present further “performance indicators” such as average anticipation times, number of active channels, etc. as was done in the previous chapter (Tables 3.1 through 3.3). There, such statistics were in part necessary to allow comparison with previously published results; there are no such results for the cumulant-based IIF. We have ample material to allow a performance comparison of the two measures in the concluding chapter. With the exception perhaps of *amy-sez03*, the information flow measure would appear to be sensitive to the same preictal changes as the dynamical similarity measure. We have convincing evidence that these

---

Figure 4.17 (*following page*): Hot zones using  $m_2$  only ( $\sigma_{\text{th}} = 2$ ,  $\tau_{\text{th}} = 4$ ,  $\tau_{\text{g}} = 0$ , only deviations above the baseline mean are considered). Results for  $n = 10$ ,  $r_{\text{max}} = 1$ ,  $\tau = 10$  and  $\ell_{\text{w}} = 8192$ . Results from intracranial EEG channels are in black, surface EEG is in purple. Red lines mark the start of clinical seizures. Blue bars along the bottom show the location and extent of gaps in the recording. Black tic marks on the bottom edge delimit the length of the recording, the full width of the box corresponding to 420 min. Black tic marks on the left-hand border indicate the extent of the signals for that recording; a box can accommodate a maximum of 64 signals. Generally, channels in the top half of this range were in the right hemisphere and those in the bottom the left. Ineligible channels are marked with a green tic on the left (this mark overwrites the black mark in some cases). Those on which the IIF algorithm failed for any reason are marked with a green tic on the right.





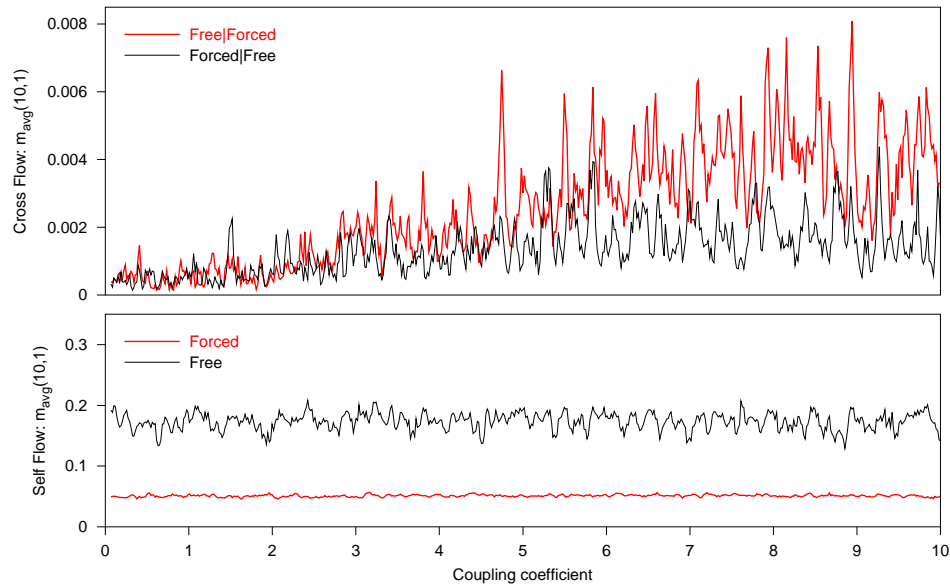
are changes in the *linear* structure of the data.

## 4.6 Cross-Channel Information Flow

Finally, we would like to briefly discuss the results of experiments done with cross-channel integrated information flow (XIIF). Rather than taking blocks of  $n + 1$  points (with a possible delay between points) from the same TS, the final, future point is taken from the same time in another TS. The idea behind the experiment was to detect synchronization across spatially distant brain regions in the preictal period. In our experiments thus far, this spatial dimension has been entirely neglected in favour of temporal correlations. Increased synchronization would result in increased statistical dependence between two distant signals.

Initial tests with toy systems like the coupled Rössler oscillator demonstrated that XIIF could track the change from independent oscillation to phase locked oscillation produced with increased coupling strength and otherwise detected by means of the Hilbert transform [71]. We then generated two TSs using the neuron 1 output from each of two mean-field coupled Bondarenko brains ( $m = 10$  neurons each,  $\tau = 10$ ,  $c = 5.0$  in brain 1 and  $c = 1.0$  in brain 2,  $e = 0.0$  in both). In other words, brain 2 is the “less chaotic” brain. Brain 1 was coupled unidirectionally to brain 2:  $T_{21} = 0$  and  $T_{12}$  was increased from 0 to 10 in steps of 0.1 (cf. equation 2.7). At each coupling strength, 12000 output steps were recorded ( $\Delta t = 1.0$  with integrations step size  $h = 0.01$ ). The same matrix of coupling coefficients and initial conditions were used for both brains; we verified that the outputs were indeed chaotic under these conditions. On each TS separately,  $m_{\text{avg}}(10, 1)$  was computed on windows of  $\ell_w = 8192$  samples, a shift of  $\ell_s = 2048$  and a delay of 1 (not the best choice in retrospect). Using the same measure parameters, the XIIF was then computed two ways: (1) taking the future point from the forced brain and (2) taking the future point from the free brain. We denote these two cases as Free|Forced (i.e. forced brain—brain 1—conditioning free brain) and Forced|Free. The results of these computations are

Figure 4.18: Normal IIF and cross-channel information flow computed on artificial time series generated with coupled Bondarenko models, “brains”, of varying degrees of chaoticity. See text for details.



shown in Fig. 4.18. While there is no appreciable change in the (self) flow, we see an increase in cross flow, especially Free|Forced, with increased coupling strength. The rms amplitude of the free, less chaotic signal is about one-fifth that of the forced signal. Consequently, even with a mean-field coupling constant of  $T_{12} = 10$ , the influence on the driven oscillator is still quite weak. The cross-information nevertheless detects the change.

Our experiments did not go much beyond that although we did compute the cross-channel  $m_{\text{avg}}(10, 1)$  on every channel pair on every recording in our EEG data set. We used windows of  $\ell_w = 8000$  with a shift of  $\ell_s = 6000$ , no surrogates and a delay of  $\tau = 100$  (another questionable choice, in retrospect). The result: for the most part, the XIIF values resemble the same-channel results. In particular, far from a seizure the signals never appear to be “independent” (i.e.  $m_{\text{avg}} \approx 0$ ). Synchronization of distant brain regions is vital for normal brain function; an epileptic seizure is synchronous abnormal discharging of large groups neurons. What exactly

brings about this synchronization, what drives a the brain toward seizure is still not fully understood. Recent work on synchrony in the brain is shedding light on the mechanisms of epilepsy, as well as offering a very promising new approach to seizure anticipation [45]. Cross-channel information flow, however, does not appear to have a great deal more to tell us here than conventional IIF.

## Chapter 5

# Closing Discussion and Conclusion

We have studied two measures of dynamical change, each motivated by different theoretical considerations and involving different approximations to decrease sensitivity to measurement noise or computational cost. The behaviour of each measure was first examined using artificial time series as control data. Epileptic seizure anticipation is a problem which has received renewed attention with the introduction of methods of nonlinear time series analysis. We looked for changes in each measure on EEG recordings prior to clinical seizures and on seizureless recordings from the same patients several hours in duration.

The first measure we considered was based solidly in dynamical systems theory. The similarity in the dynamics underlying nonoverlapping segments of a time series was gauged by the overlap in the distribution of states in phase spaces reconstructed from these segments [42]. The use of intercrossing intervals (ICIs) to form the embedding vectors and projection onto the eigenvectors corresponding to the largest singular values essentially serve as nonlinear filters, increasing the measure's sensitivity to changes in the system's most significant degrees of freedom and—as a bonus—considerably reducing the computational burden of the measure. We call this a *relative* measure, because it explicitly involves comparison with a reference dynamics. It is strictly positive and normalized to a maximum value of unity.

We then examined measures of information flow, both the formal and the practical. It is the extent to which the future depends on information about the past

which is being measured. A formal measure of information flow on dynamical systems, the Kolmogorov-Sinai (KS) entropy, was defined and the problems associated with its estimation discussed. The cumulant-based integrated information flow measure [4, 18] proceeds from the assumption that windows of data on the time series are “sufficiently” stationary and can be represented as a vector of random variables, i.e. a random process. If the last variable in the process, occurring  $r$  steps/samples after the previous one, is statistically independent from all the previous random variables, all cumulants simultaneously involving both this variable and any of the others will equal zero. This was the basis of the IIF measure, although as a result of approximations introduced to make its computation feasible, statistical independence is in practice only the *most probable* reason for a result of zero.

Insofar as it does not involve direct comparison between segments of the time series and hence does not require *a priori* identification of a reference dynamics, this measure is *absolute*. Unlike KS entropy, it does not measure information flow in bits but rather on a strictly positive dimensionless scale. An upper limit for this measure has not been found analytically. This and the strong statistical fluctuation observed even with relatively long windows of data make it difficult to establish any useful absolute thresholds for this absolute measure.

## 5.1 Experiments with Artificial Data

The asymmetric neural network with delay studied by Bondarenko served as a source of artificial time series for comparison of the two measures on stationary control data. For parameter values in the chaotic regime, this model produces output qualitatively similar to EEG signals. We verified that the “degree of chaos”, reflected in maximal Lyapunov exponent estimates, increased with increasing global coupling strength,  $c$ . Although whether a given realization of the network (random coupling matrix) will produce chaotic output for values of  $c$  in the ostensibly “chaotic regime” seems to depend on the initial conditions, the data set retained for subsequent experiments

did consist of chaotic solutions. We also examined the influence on the measures' behaviour of intermittent bursts of quasiperiodicity, increasing in duration and frequency with network asymmetry  $e$ .

The discriminating power of the information flow measure on these artificial data was greater overall than that of the similarity measure for the same window length. Using the longer sampling interval of  $\Delta t = 10.0$  on the  $c$  series, what little difference there was between  $c$  values on the similarity results in fact vanished; we said the data had been “undersampled”. However, the very same sampling interval or an equivalent delay actually improved IIF results considerably and increased the significance of nonlinear contributions, as revealed by comparison with phase shuffled surrogates.

It is equally important that the measures remain fairly stable on stationary data. In that respect, similarity can be said to have performed better than IIF. If we consider the  $\ell_w = 8192$  results for  $\Delta t = 1.0$ , error bars on the former for  $c_{\text{ref}} = c_t$  were easily less than 1% of full scale, whereas those on the IIF results were over 10% of the mean result for all values of  $c$ . We attribute these two differences—greater discrimination between levels of chaoticity for IIF and higher stability on stationary data for similarity—in large part to the nonlinear filtering steps in the latter. In addition, for a constant window length, statistical fluctuations of cumulant estimates increase with the cumulant order being estimated. Since the contributions of each order are weighted equally, it is not surprising that the IIF measure fluctuates more. Differences such as these are precisely what we speculated would lead to improved anticipation of epileptic seizures by combining the measures.

The results on the  $e$  series illustrate the importance of using windows of sufficiently *long duration*, since the *apparent* stationarity of a system is a question of time scale. To capture both the chaotic and quasiperiodic phases of the intermittent behaviour found in these TSs requires adequate sampling over a sufficient period of time. Both measures were remarkably robust in regards to additive or “measurement” noise. At levels less than 20% of the unperturbed signal's rms amplitude, the discriminating

power of neither IIF nor dynamical similarity (such as it was) was affected.

## 5.2 Seizure Anticipation from EEG Recordings

Before applying the nonlinear methods under investigation in this paper to our data set of 25 EEG recordings, we examined a representative signal from an intracranial electrode for spectral and variance fluctuations on windows approximately 10 sec in length. On several of the recordings, significant changes in the power spectrum were observed prior to the first clinical seizure captured in the recording. When these changes appeared to persist up to the time of seizure onset and were unique in the interval before this time, we called the time from the change to the seizure the “anticipation time”. As expected, overall changes in periodogram intensity, i.e. total power, were found to correspond to increases in signal amplitude, reflected in an increase in variance. Seizureless recordings made while the patient was asleep were generally highly nonstationary, showing rapid changes in overall power. Recordings made during sleep were relatively quiescent.

Both measures were computed on the entire set of EEG recordings using values for the measure parameters cited by the method’s developers [44] in the case of dynamical similarity and several combinations of values for IIF. Furthermore, the similarity was calculated using each of two reference windows chosen on each recording (the same segment for all channels) and, finally, crossed recordings and reference windows for the same patient. Various thresholding schemes were then considered in identifying “hot zones” on each TS.

For the similarity measure, various statistics concerning these hot zones, in particular longest and average anticipation times and numbers of active channels, were calculated. Five of the 15 initial/isolated seizures were preceded by bound hot zones in over 15% of iEEG channels. In all, eight of the seizures were preceded by bound hot zones in no less than 4 channels. *Awake* recordings were found to contain the greatest density of free hot zones.

Comparing results obtained with each of the two RWs, we find that the similarity measure is not symmetrical. That is, if RW1 produces a hot zone overlapping RW2, RW2 will not necessarily yield a hot zone overlapping RW1. The projection onto different bases explains this in part but also the variation in number of threshold crossings in the RW. Nevertheless, the asymmetry is often weak. Furthermore, the cross-recording results show that when a reference window on one recording from a patient reveals preictal changes on that recording, it will often do so on other recordings with seizures. This supports the claim that the RW in some way captures the patient's interictal dynamics.

No anticipation statistics were computed on the IIF results: the only convincing preictal changes were on the *amy* recordings and there were *many* free hot zones everywhere. Such nonspecificity in the threshold results is a consequence of the large statistical fluctuation of the IIF measure; it seems impossible to detect a rise in information flow robustly without violating causality, i.e. using a noncausal smoothing filter. Use of baseline statistics nonetheless resulted in fewer free hot zones than absolute thresholds, with the *awake* recordings remaining relatively hot. The anticipation results for spectral, similarity and information flow measures are summarized in Table 5.1. The aim of this table is not to compare the *performance* of each measure in anticipation, but rather to reflect the tendency of all the measures to anticipate the same seizures to varying degrees on our data set.

The main conclusion to be drawn from these results is that the nonlinear measures of dynamical change were sensitive to changes in the *linear* structure of the signals, in the *amysez* and *carlsez01* recordings especially, and that seizures that were not preceded by such changes were not anticipated.

We find that the specificity of the similarity measure to preictal changes is higher than that of IIF, because free hot zones are shorter and fewer in results obtained with the former. Broadly speaking, the two measures show similar sensitivity to these changes. In short, we conclude that there is no complementarity in the re-



Table 5.1: Summary of seizure anticipation results. Spectral times are for the first iEEG signal; formal criteria were not used to obtain these times, but rather they reflect our general impression of the periodogram of that signal. On the contrary, the similarity results are the earliest anticipation times from Table 3.1 using the formal definition of bound hot zones presented in that chapter. The number of active channels is in parentheses. Asterisks in the IIF column indicate bound hot zones in over  $\sim 20\%$  of the channels preceded by a relative absence of free hot zones.

Recording	Spectral min	Similarity min (channels)	IIF
amysez01	30	23 (47)	*
amysez02	40?	3 (7)	*
amysez03	20	7 (8)	*
bobsez01	20	—	—
bobsez02	—	—	—
bobsez03	5	—	—
carlsez01	15	26 (18)	*
carlsez02	—	7 (4)	—
carlsez03	100	113 (15)	—
dansez01	—	10 (4)	—
dansez02	20	9 (4)	—
dansez03	—	—	—
evesez01	—	—	—
evesez02	—	—	—
evesez03	—	—	—

sults for the two measures which would allow us to improve anticipation performance (sensitivity/specificity to preictal changes) by combining them.

### 5.3 Conclusion

In this thesis, two measures of dynamical change, one grounded in dynamical systems theory and the other in information theory, were presented. We discussed the value of “knowing one’s tools” before undertaking any form of analysis. This we have endeavoured to do by first examining the response of these measures to inputs with known properties and stationarity, inputs which were also qualitatively similar to the real time series we ultimately wished to analyze. Significant differences in sensitivity to difference degrees of chaos and variability on stationary data were found. These observations guide us in our use of the measure on real data from human brains.

There is hardly any doubt that nonlinear processes play an important part in the dynamics of the brain [47, 55]; this is not to say that nonlinearity will necessarily be detectable in EEG recordings. As for the role or even presence of certain specific nonlinear phenomena in creating mind, particularly deterministic chaos, it would appear that the jury is still out. Various methods of nonlinear time series analysis have now been applied to the problem of detecting and characterizing changes in brain dynamics prior to epileptic seizures, with several groups reporting evidence that there is a discernible preictal period beginning at much as 20 minutes prior to a seizure [45]. Measures of *synchronization* are yielding still longer anticipation times as well as insight into the as yet poorly understood *mechanisms* underlying seizure precursors.

With such encouraging findings in the literature, why do our results seem primarily to reiterate what linear measures can already tell us about these EEG recordings? Why does the specificity of these changes to preictal periods appear to be so low? One simple reason is that seizures preceded by changes in linear measures have not been actively excluded from previous studies and results obtained with linear measures

are not typically reported alongside the nonlinear results. As for specificity, it has rarely been addressed anywhere but in the discussion section of papers, and there only in passing. Our experiments were designed to probe these issues. The question of correlation between results from linear and nonlinear measures is relevant and an active subject of debate [53, 56].

Recent results from Navarro et al. [59] (a collaboration between the group at Pitié-Salpêtrière and J. Gotman of the MNI) are a departure from earlier studies in that the *causes* of observed preictal fluctuations in dynamical similarity are investigated. Out of 51 seizures in that study, 31 were preceded by preictal changes in the measure (bound hot zones) but of these only 7 were *not* also preceded by visually detectable changes in the EEG. These results echo our own findings of about 50% detection involving mostly seizures preceded by, for example, a change in signal strength. While our investigation did not extend to visual examination of the EEG signals, we do feel it is important to consider not just one but a well chosen battery of measures in analysis of this sort, including linear measures. [68]

Those who suffer from severe epilepsy—approximately 1% of the population is affected by this common neurological disease—must live in fear of their next seizure. In the last 5–10 years, dynamicists have given us a glimpse of the real potential of novel measures of dynamical change to anticipate epileptic seizures and to alleviate this fear. Much work remains to be done before these methods will be useful in a clinical setting, such as studies on longer and more varied data sets including data from nonepileptic brains [49]. We have also seen that the behaviour of potential measures on suitable artificial data can provide insight into the relevance and usage of a given measure in a given setting. Lastly, increasing collaboration between research groups will help to identify the relative strengths and weaknesses of potential methods for seizure anticipation more rapidly and effectively.

# Bibliography

- [1] *The New Encyclopaedia Britannica*. Encyclopaedia Britannica, Chicago, 15th edition, 1992.
- [2] K. Aihara, T. Takabe, and M. Toyoda. Chaotic neural networks. *Phys. Lett. A*, 144(6/7):333–340, 1990.
- [3] A. Babloyantz and A. Destexhe. Low-dimensional chaos in an instance of epilepsy. *Proc. Nat. Acad. Sci. USA*, 83(10):3513–3517, 1986.
- [4] F. Beaulieu. Réduction non-linéaire du bruit et détection de non-stationnarité. Master’s thesis, Laval University, April 2000.
- [5] L. H. Billington. Humanizing the information revolution. *Proc. Am. Phil. Soc.*, 145(4):579–586, 2001.
- [6] M. H. Boisot. *Knowledge Assets: Securing Competitive Advantage in the Information Economy*. Oxford University Press, Oxford, 1999.
- [7] V. E. Bondarenko. A simple neural network model produces chaos similar to the human EEG. *Phys. Lett. A*, 196(3/4):195–200, 1994.
- [8] V. E. Bondarenko. Analog neural network model produces chaos similar to the human EEG. *Int. J. Bifurcation. and Chaos.*, 7(5):1133–1140, 1997.
- [9] V. E. Bondarenko. High-dimensional chaotic neural network under external sinusoidal force. *Phys. Lett. A*, 236:513–19, 1997.

- [10] V. E. Bondarenko. Self-organization processes in chaotic neural networks under external periodic force. *Int. J. Bifucation and Chaos*, 7(8):1887–1895, 1997.
- [11] V. E. Bondarenko. Control and ‘anticontrol’ of chaos in an analog neural network with time delay. *Chaos, Solitons and Fractals*, 13:139–54, 2002.
- [12] V. E. Bondarenko. Discussion of implementation details and methods. Private communication, November 2002.
- [13] L. Brillouin. *La Science et la théorie de l’information*. Éditions Jacques Gabay, 1988. This is reprint of the original 1959 edition published by Masson et Cie. (Paris).
- [14] T. Chen and S. I. Amari. Stability of asymmetric Hopfield networks. *IEEE Trans. Neural Networks*, 12:159, 2001.
- [15] T. M. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons, Inc., 1991.
- [16] P. K. Das II, W. C. Schieve, and Z.-J. Zeng. Chaos in an effective four-neuron neural network. *Phys. Lett. A*, 161(1):60–66, 1991.
- [17] G. Deco and D. Obradovic. *An Information-Theoretic Approach to Neural Computing*. Perspectives in Neural Computing. Springer, 1996.
- [18] G. Deco, C. Schittenkopf, and B. Schürmann. Dynamical analysis of time series by statistical tests. *Int. J. Bifucation and Chaos*, 7:2629–2652, 1997.
- [19] H. M. Deitel and P. J. Deitel. *C: How to Program*. Prentice Hall, New Jersey, 3rd edition, 2001.
- [20] A. Destexhe. Oscillations, complex spatiotemporal behavior, and information transport in networks of excitatory and inhibitory neurons. *Phys. Lett. A*, 132(2/3):101–106, 1993.

- [21] P. F. Drucker. Beyond the Information Revolution. *The Atlantic*, 284(4):47–57, October 1999.
- [22] J.-P. Eckmann and D. Ruelle. Fundamental limitations for estimating dimensions and lyapunov exponents in dynamical systems. *Physica D*, 56:185–187, 1992.
- [23] P. Ehrenfest and T. Ehrenfest. *The Conceptual Foundations of the Statistical Approach in Mechanics*. Dover Publications, Inc., New York, 1990.
- [24] W. Feller. *An Introduction to Probability Theory and Its Applications*. Wiley, New York, 3rd edition, 1968. 2 v.
- [25] D. Fish and M. Walker. *Epilepsy: Comprehensive Review and Case Discussions*, chapter 2. Martin Dunitz, 2001.
- [26] M. Fogarty and C. Bahls. Information overload: Feel the pressure? *The Scientist*, 16(16):16–18, 2002.
- [27] C. W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer Series in Synergetics. Springer, 2nd edition, 1985.
- [28] M. Gell-Mann and S. Lloyd. Information measures, effective complexity and total information. *Complexity*, 2(1):44–52, 1996.
- [29] P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica D*, 9:189–208, 1983.
- [30] J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer-Verlag, New York, 1983.
- [31] L. M. Hively, P. C. Gailey, and V. A. Protopopescu. Detecting dynamical change in nonlinear time series. *Phys. Lett. A*, 258:103–114, 1999.
- [32] L. M. Hively and V. A. Protopopescu. Timely detection of dynamical change in scalp EEG signals. *Chaos*, 10(4):864–875, December 2000.

- [33] J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Nat. Acad. Sci. USA*, 81:3088–3092, 1984.
- [34] J. J. Hopfield. Brain, neural networks and computation. *Rev. Mod. Phys.*, 71:S431–437, 1999.
- [35] N. Janson et al. Reconstruction of dynamical and geometrical properties of chaotic attractors from threshold-crossing interspike intervals. *Phys. Rev. E*, 58:R4–7, 1998.
- [36] N. Kanou et al. A current-mode circuit of a chaotic neuron model. *IEICE Trans. Fundamentals E*, 76-A(4):642–644, 1993.
- [37] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*, volume 7 of *Cambridge Nonlinear Science Series*. Cambridge University Press, Cambridge, 1997.
- [38] H. R. Kantz and T. Schreiber. Practical implementation of nonlinear time series methods: the TISEAN package. *Chaos*, 9(413), 1999.
- [39] J. Karl. *Introduction to Digital Signal Processing*. Academic Press, 1989.
- [40] A. N. Kolmogorov. A new invariant of transitive dynamical systems. *Dokl. Akad. Nauk. SSSR*, 119:861, 1958.
- [41] K. E. Kürten and J. W. Clark. Chaos in neural systems. *Phys. Lett. A*, 114:413–418, 1986.
- [42] M. Le Van Quyen et al. Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings. *NeuroReport*, 10:2149–2155, 1999.
- [43] M. Le Van Quyen et al. Spatio-temporal characterizations of non-linear changes in intracranial activities prior to human temporal lobe seizures. *Eur. J. Neurosci.*, 12:2124–2134, 2000.

- [44] M. Le Van Quyen et al. Anticipation of epileptic seizures from standard EEG recordings. *The Lancet*, 357:183–188, 20 January 2001.
- [45] M. Le Van Quyen et al. Characterizing neurodynamic changes before seizures. *J. Clinical Neurophysiology*, 18(3):191–208, 2001.
- [46] K. Lehnertz and C. E. Elger. Can epileptic seizures be predicted? Evidence from nonlinear time series analysis of brain electrical activity. *Phys. Rev. Lett.*, 80:5019–5022, 1998.
- [47] K. Lehnertz et al., editors. *Workshop on Chaos in Brain?* World Scientific Pub. Co., Singapore, 2000.
- [48] B. Litt et al. Epileptic seizures may begin hours in advance of clinical onset: A report of five patients. *Neuron*, 30:51–64, April 2001.
- [49] B. Litt and K. Lehnertz. Seizure prediction and the preseizure period. *Current Opinion in Neurology*, 15(2):173–177, 2002.
- [50] R. Manuca and R. Savit. Stationarity and nonstationarity in time series analysis. *Physica D*, 99:134–161, 1996.
- [51] C. M. Marcus and R. M. Westervelt. Stability of analog neural networks with delay. *Phys. Rev. A*, 39:347, 1989.
- [52] J. Martinerie et al. Epileptic seizures can be anticipated by non-linear analysis. *Nature Med.*, 4:1173–1176, 1998.
- [53] J. Martinerie et al. Response to: Mcsharry, p. et al. “prediction of epileptic seizures: are nonlinear methods relevant?”, *Nature*, 9(3):241-42, 2003. *Nature Medecine*, 9(3):242, 2003.
- [54] J. T. Mathews. The information revolution. *Foreign Policy*, (119):63–66, Summer 2000.



- [55] T. M. McKenna, T. A. McMullen, and M. F. Shlesinger. The brain as a dynamical system. *Neuroscience*, 60(3):587–605, 1994.
- [56] P. McSharry, L. Smith, and L. Tarassenko. Prediction of epileptic seizures: are nonlinear methods relevant? *Nature Medecine*, 9(3):241–242, 2003.
- [57] M. Nakagawa and M. Okabe. On the chaos region of the modified Nagumo-Sato model. *J. Phys. Soc. Japan*, 61(4):1121–1124, 1992.
- [58] K. Nakamura and M. Nakagawa. On the associative model with parameter controlled chaos neurons. *J. Phys. Soc. Japan*, 62(8):2942–2955, 1993.
- [59] V. Navarro et al. Analysis of the preictal EEG changes detected by a nonlinear method in an unselected population of patients with focal epilepsy [abstract]. *Epilepsia*, 43, 2002.
- [60] E. Ott, T. Sauer, and J. A. Yorke. *Coping with Chaos: Analysis of Chaotic Data and the Exploitation of Chaotic Systems*. Wiley Series in Nonlinear Science. John Wiley & Sons, 1994.
- [61] K. Pakdaman, C. Grotta-Ragazzo, and C. P. Malta. Transient duration in continuous-time neural networks with delay. *Phys. Rev. E*, 58:3623, 1998.
- [62] A. Papoulis and S. U. Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw Hill, Boston, 4th edition, 2002.
- [63] A. N. Pavlov et al. Extracting dynamics from threshold-crossing interspike intervals: possibilities and limitations. *Phys. Rev. E*, 61(5):5033–5044, May 2000.
- [64] J. P. Pijn. Chaos or noise in EEG signals: dependence on state and brain site. *Electroenceph. and Clin. Neuro.*, 79(5):371–381, 1991.
- [65] A. Pikovsky et al. Phase synchronization of chaotic oscillators by external driving. *Physica D*, 104:219–230, 1997.

- [66] W. Press et al. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 1992.
- [67] V. A. Protopopescu, L. M. Hively, and P. C. Gailey. Epileptic event forewarning from scalp EEG. *J. Clin. Neuro.*, 18:223–245, 2001.
- [68] G. Rangarajan and M. Ding. Integrated approach to the assessment of long range correlation in time series data. *Phys. Rev. E*, 61:4991–5001, 2000.
- [69] F. Reif. *Fundamentals of statistical and thermal physics*. McGraw-Hill, Toronto, 1965.
- [70] U. Riedel, R. Kühn, and J. L. van Hemmen. Temporal sequences and chaos in neural nets. *Phys. Rev. A*, 38(2):1105–1108, 1988.
- [71] M. Roseblum, A. Pikovsky, and J. Kurths. Phase synchronization of chaotic oscillators. *Phys. Rev. Lett.*, 76(11):1804–1807, 11 March 1996.
- [72] T. Sauer. Reconstruction of dynamical systems from interspike intervals. *Phys. Rev. Lett.*, 72:3811–3814, 1994.
- [73] T. Sauer. Interspike interval embedding of chaotic signals. *Chaos*, 5:127–32, 1995.
- [74] T. Sauer, J. A. Yorke, and M. Casdagli. Embedology. *J. Stat. Phys.*, 65(3/4):579–616, 1991.
- [75] T. Schreiber and A. Schmitz. Surrogate time series. *Physica D*, 142(346), 2000.
- [76] C. Shannon and W. Weaver. *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana, Illinois, 1949.
- [77] K. Shimizu, M. Aihara, and M. Kotani. An electronic circuit model of chaotic neural networks. *Electron. Commun. Japan, Pt. 3*, 74(1):51–64, 1991.

- [78] H. J. Sommers et al. Spectrum of large random asymmetric matrices. *Phys. Rev. Lett.*, 60:1897, 1988.
- [79] H. Sompolinsky, A. Crisanti, and H. J. Sommers. Chaos in random neural networks. *Phys. Rev. Lett.*, 61:259–62, 1988.
- [80] F. Takens. Detecting strange attractors in turbulence. In *Lecture Notes in Mathematics*, volume 898. Springer-Verlag, 1981.
- [81] J. Theiler. On the evidence for low-dimensional chaos in an epileptic electroencephalogram. *Phys. Lett. A*, 196:335–341, 1995.
- [82] J. Theiler et al. Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 58:77–94, 1992.
- [83] M. Zigmond et al., editors. *Fundamental Neuroscience*. Academic Press, San Diego, 1999.

# Appendix A

## Elements of Information Theory

### A.1 Introduction

Never before in history have people been as preoccupied with the role of information in their lives. In the popular and scholarly press, much has been made in recent years of the Information Revolution [5, 21, 54]. A common complaint of business people, consumers and academics everywhere today is information overload [26]. Yesterday’s software developer or computer salesperson is today’s “IT specialist”, where IT—information technology—runs from computers to cell phones to the Internet. These expressions speak to us because, among other reasons, we all have an intuitive grasp of what information is. We all need it and want it, but not too much of it. We understand that it can be transmitted and stored. We are increasingly adept at “filtering out” much, even most of the information sent our way. Some information, of course, has value and can even be bought and sold in our...information economy [6].

In light of all this, one might think that *information theory* was a product of the social sciences, meant to illuminate our modern condition. Ironically, one of the fundamental questions answered by this theory seems designed to aggravate the problem of “information overload”: what is the *maximum* transmission rate of a communication channel. It is of course a question of great practical importance. Shannon demonstrated that, contrary to what was commonly believed in the early 1940’s, increasing the transmission rate over a channel did not increase the probability of transmission

error as long as the rate was below “channel capacity”, which is strictly a function of the noise characteristics of the channel itself [76]. He also demonstrated that lossless data compression was ultimately limited by the *entropy* of that data. The entropy concept was a familiar one from thermodynamics and statistical mechanics [23, 69], the entropy of a system being roughly speaking a measure of the disorder of that system or the unpredictability of its state.

## A.2 Information and the Entropies

In this overview, we will use the notation found in Cover and Thomas’ excellent information theory primer [15]. Consider a data stream consisting of a string of symbols, each a realization of a discrete random variable  $X$ . Let  $\mathcal{X}$  denote the set of  $N$  possible outcomes of  $X$ , called its **alphabet**. In discussing random variables, we adopt the convention that upper case letters, e.g.  $X$ , refer to the variable itself whereas lower case letters, e.g.  $x$ , refer to a specific **realization** of that variable, thus  $x \in \mathcal{X}$ . We ask what is the degree of uncertainty in predicting the next symbol in the sequence irrespective of previous outcomes, since it is this level of uncertainty which limits compressibility.

Let  $p(x)$  denote the probability of the outcome  $x$ .<sup>1</sup> Shannon *defines* the information content inherent in each outcome  $x$  as

$$I(x) = \log p(x). \qquad \text{Shannon information (A.1)}$$

This definition forms the basis of information theory.<sup>2</sup> By defining information like

---

<sup>1</sup>It should be clear that the probability mass function (pmf) will generally be different for different random variables and that the notation  $p(x)$  implies the pmf of the random variable  $X$  just as  $p(y)$  would imply that of a different variable  $Y$ . Where there is a risk of confusion, we will use the explicit notation  $p_X(x)$ . We might also have used  $p(X = x)$ , another common notation.

<sup>2</sup>A distinction emphasized in [37] is that between *probabilistic* information theory, which will be our concern in this chapter, and algorithmic information theory. The latter deals with the *computational effort* required to obtain/compute a given piece of information.

this—with the logarithm—Shannon bestows certain properties on it, such as positivity and *additivity*, as Brillouin<sup>3</sup> [13] and others [28] are quick to point out. The joint probability  $p_{XY}(x; y)$  of obtaining outcomes  $x$  and  $y$  of two independent random variables  $X$  and  $Y$  is the *product* of their respective probabilities

$$p_{XY}(x; y) = p_X(x) \cdot p_Y(y). \quad (\text{statistical independence}) \quad (\text{A.2})$$

By virtue of their independence, we require furthermore that the information in the joint outcome be the *sum* of the information in each outcome separately. Thus we take the logarithm of  $p_{XY}(x; y)$ ,

$$\begin{aligned} I(x; y) \equiv \log p_{XY}(x; y) &= \log(p_X(x) \cdot p_Y(y)) \\ &= \log p_X(x) + \log p_Y(y) \\ &= I(x) + I(y), \end{aligned} \quad (\text{A.3})$$

which confirms that the Shannon information has this additive property. Any other measure of information content must have these properties [28].

Taking the weighted average of this information over all possible outcomes yields the Shannon entropy<sup>4</sup> of  $X$ ,

$$\begin{aligned} H(X) &= - \sum_{x \in \mathcal{X}} p(x) \log p(x) \\ &= -\langle \log p(x) \rangle = -\langle I(x) \rangle, \end{aligned} \quad \text{Shannon entropy} \quad (\text{A.4})$$

where  $\langle \cdot \rangle$  denotes the expectation value.  $H$  is strictly positive, with a minimum of zero when only one outcome  $x$  is possible, i.e.  $x$  occurs with probability 1.<sup>5</sup> In that case, there is no uncertainty in the outcome of the event. The entropy reaches a

---

<sup>3</sup>Brillouin's discussion of the connection between thermodynamics and information theory in this work is fascinating. In addition to its origins and continued importance in communications theory and intimate connection to physics, information theory is strongly related to mathematics, probability theory, economics and computer science. [15]

<sup>4</sup>The entropy is expressed in **bits** if  $\log_2$  is used, **nats** when  $\ln$  is used [15].

<sup>5</sup>We define  $p \log p \equiv 0$  for  $p = 0$ .

maximum when all outcomes are equiprobable, for then  $p(x) = 1/N$  for all  $x$  and  $H = \log N$ ; the degree of uncertainty in our prediction is also a maximum.

Consider now a block of  $n$  consecutive symbols and represent this as a series of  $n$  random variables  $X_n, \dots, X_1$ . The probability of a given outcome, i.e. a given sequence of  $n$  consecutive symbols, is written  $p(x_n, \dots, x_1)$ . The *block entropy* is then defined as

$$H(X_n, \dots, X_1) = - \sum_{x_n, \dots, x_1 \in \mathcal{X}} p(x_n, \dots, x_1) \log p(x_n, \dots, x_1).$$

**block entropy** (A.5)

The sum is taken over all  $N^n$  possible sequences, those which are not allowed or simply never occur being of measure zero. Lastly, if  $p(x_{n+r}|x_n, \dots, x_1)$  is the conditional probability of observing outcome  $x_{n+r}$  after having *already observed* the consecutive sequence  $x_1, \dots, x_n$  a number  $r$  symbols earlier in the stream, then the conditional entropy of the sequence of random variables  $X_{n+r}, X_n, \dots, X_1$  is

$$H(X_{n+r}|X_n, \dots, X_1) = - \sum_{x_{n+r}, x_n, \dots, x_1 \in \mathcal{X}} p(x_{n+r}|x_n, \dots, x_1) \log p(x_{n+r}|x_n, \dots, x_1).$$

**conditional entropy** (A.6)

It corresponds to the uncertainty in the combined outcome of a set of events given knowledge of the outcome of a subset of these events, a relation which is expressed as follows:

$$H(X_{n+r}|X_n, \dots, X_1) = H(X_{n+r}, X_n, \dots, X_1) - H(X_n, \dots, X_1). \quad (\text{A.7})$$

This is the level of uncertainty in our prediction of the symbol  $r$  steps in the future when we know the last  $n$  symbols. It is intuitively obvious that

$$0 \leq H(X_{n+r}|X_n, \dots, X_1) \leq H(X_{n+r}) = H(X), \quad (\text{A.8})$$

which is to say that *conditioning decreases entropy* (Theorem 2.6.5 in [15]). Our uncertainty in this case cannot be any greater than when we know nothing about the past.

# Appendix B

## Elements of Probability Theory

For a thorough introduction to probability theory with plenty of insightful and historically important examples, the reader may wish to consult the classic text by Feller [24]. Probability and random variables are also given a clear introductory treatment in Papoulis and Pillai [62].

### B.1 Descriptions of a Probability Distribution

Let  $\mathbf{X} = (X_1, \dots, X_n)$  be a vector of continuous **random variables** defined on  $\mathbb{R}^n$ . The distribution of  $\mathbf{X}$  over this space is described by the **probability density function (PDF)**  $p(\mathbf{X})$ ; that is,  $p(\mathbf{x}) d\mathbf{x}$  is the probability that a given **realization** of  $\mathbf{X}$  will fall in the infinitesimal volume element  $d\mathbf{x}$  centered on  $\mathbf{x}$ .

The expectation value of a scalar function of this random vector  $f(\mathbf{X})$  is the average value of that function over the space *weighted* by  $p(\mathbf{X})$  and is denoted by angle brackets:

$$\langle f(\mathbf{X}) \rangle = \int p(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}. \quad (\text{B.1})$$

The moments of the distribution are a particularly important example of expectation values. Their definition is as follows:

$$M_{q_1 \dots q_n} = \langle X_1^{q_1} \dots X_n^{q_n} \rangle, \quad (\text{B.2})$$



where the sum of the exponents

$$\sum q_i = m \tag{B.3}$$

is called the **order** of the moment. First-order moments ( $m = 1$ ) are the familiar weighted averages of each of the components of  $\mathbf{X}$ . For  $n = 2$ , these would be denoted  $M_{10} = \langle X_1^1 X_2^0 \rangle = \langle X_1 \rangle$  and  $M_{01} = \langle X_1^0 X_2^1 \rangle = \langle X_2 \rangle$ . Second-order moments ( $m = 2$ ) would be  $M_{20} = \langle X_1^2 \rangle$ ,  $M_{11} = \langle X_1 X_2 \rangle$  and  $M_{02} = \langle X_2^2 \rangle$ . The above notation for moments conveys their meaning with economy but it will be convenient to introduce an alternative notation at this point. Given a moment of order  $m$ , we observe that

$$\begin{aligned} M_{q_1 \dots q_n} &= \langle X_1^{q_1} \dots X_n^{q_n} \rangle \\ &= \langle \underbrace{X_1 \dots X_1}_{q_1 \text{ times}} \dots \underbrace{X_n \dots X_n}_{q_n \text{ times}} \rangle \\ &= \mathcal{M}_{\ell_1 \dots \ell_m}, \end{aligned} \tag{B.4}$$

such that  $q_1$  of the indices  $\ell_i$  equal 1,  $q_2$  equal 2, etc. for a total of  $m$  indices,  $\ell_i \in [1, n] \forall i$ . In this new notation, the first order moments (expectation values of individual components) of  $\mathbf{X}$  where  $n = 2$  would be  $M_{10} = \mathcal{M}_1 = \langle X_1 \rangle$  and  $M_{01} = \mathcal{M}_2 = \langle X_2 \rangle$ . Second-order moments then become  $M_{20} = \mathcal{M}_{11}$ ,  $M_{02} = \mathcal{M}_{22}$  and  $M_{11} = \mathcal{M}_{12} = \langle X_1 X_2 \rangle = \langle X_2 X_1 \rangle = \mathcal{M}_{21}$ . Note that in the new notation permutation of the indices does not change the value of the moment.

The Fourier space representation of the distribution is called its **characteristic function** with the definition

$$\Phi(\boldsymbol{\alpha}) = \langle \exp(i\boldsymbol{\alpha} \cdot \mathbf{X}) \rangle = \int d\mathbf{x} p(\mathbf{x}) \exp(i\boldsymbol{\alpha} \cdot \mathbf{X}), \tag{B.5}$$

where  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$ . In the power series expansion of  $\Phi(\boldsymbol{\alpha})$

$$\Phi(\alpha_1, \dots, \alpha_n) = \sum_{q_1, \dots, q_n=0}^{\infty} \frac{i^{(q_1 + \dots + q_n)}}{q_1! \dots q_n!} M_{q_1 \dots q_n} \alpha_1^{q_1} \dots \alpha_n^{q_n}, \tag{B.6}$$

we see that, to within a constant factor, the coefficients are the moments of the distribution. This shows that the distribution of  $\mathbf{X}$  on its space of possible outcomes is

completely described not only by  $p(\mathbf{X})$  but equally well by the characteristic function  $\Phi(\boldsymbol{\alpha})$  or the infinite set of moments.

Taking the natural logarithm of the characteristic function, we obtain the **cumulant characteristic function**,

$$\Psi(\boldsymbol{\alpha}) = \ln \Phi(\boldsymbol{\alpha}), \quad (\text{B.7})$$

the power series expansion of which can be written

$$\Psi(\boldsymbol{\alpha}) = \sum_{m=1}^{\infty} \frac{i^m}{m!} \sum_{\ell_1, \dots, \ell_m=1}^n \mathcal{K}_{\ell_1 \dots \ell_m} \alpha_{\ell_1} \cdots \alpha_{\ell_m}. \quad (\text{B.8})$$

Here it is not moments which appear as coefficients but rather **cumulants**. The same notation is used here as in (B.4)—order  $m$  cumulants have  $m$  indices each of which has a value from 1 to  $n$ , designating one of the random variables in  $\mathbf{X}$ —and the same property of invariance under permutation of the indices holds. By virtue of the defining relationship between the characteristic function and the cumulant characteristic function (B.7), the cumulants of order  $m$  can be expressed as a function of the moments of order  $m$  and lower. If the average of each random variable in the random process described by  $\mathbf{X}$  is zero,  $\langle X_i \rangle = 0 \forall i \in [1, n]$ , then the following expressions can be obtained for cumulants up to fourth order [4, 27]:

$$\mathcal{K}_{\ell_1} = \mathcal{M}_{\ell_1} = 0 \quad (\text{B.9})$$

$$\mathcal{K}_{\ell_1 \ell_2} = \mathcal{M}_{\ell_1 \ell_2} \quad (\text{B.10})$$

$$\mathcal{K}_{\ell_1 \ell_2 \ell_3} = \mathcal{M}_{\ell_1 \ell_2 \ell_3} \quad (\text{B.11})$$

$$\mathcal{K}_{\ell_1 \ell_2 \ell_3 \ell_4} = \mathcal{M}_{\ell_1 \ell_2 \ell_3 \ell_4} - \mathcal{M}_{\ell_1 \ell_2} \mathcal{M}_{\ell_3 \ell_4} - \mathcal{M}_{\ell_1 \ell_3} \mathcal{M}_{\ell_2 \ell_4} - \mathcal{M}_{\ell_1 \ell_4} \mathcal{M}_{\ell_2 \ell_3} \quad (\text{B.12})$$

## B.2 Expressions of Statistical Independence

The concept of *statistical independence* is fundamental to probability theory. The joint probability density  $p(\mathbf{X}, \mathbf{Y})$  of two random processes  $\mathbf{X}$  and  $\mathbf{Y}$  with probability

density functions  $p(\mathbf{X})$  and  $p(\mathbf{Y})$  respectively will satisfy the relation

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}) \cdot p(\mathbf{y}) \quad (\text{B.13})$$

only if the two processes are independent. From the definitions (B.5) and (B.7), it is easily demonstrated that if (B.13) is true then the relations between the characteristic functions

$$\Phi(\boldsymbol{\alpha}_X, \boldsymbol{\alpha}_Y) = \Phi(\boldsymbol{\alpha}_X) \cdot \Phi(\boldsymbol{\alpha}_Y) \quad (\text{B.14})$$

and between the cumulant characteristic functions

$$\Psi(\boldsymbol{\alpha}_X, \boldsymbol{\alpha}_Y) = \Psi(\boldsymbol{\alpha}_X) + \Psi(\boldsymbol{\alpha}_Y) \quad (\text{B.15})$$

are also satisfied. A consequence of this last relation is that all cumulants of the joint process simultaneously involving at least one random variable from both  $\mathbf{X}$  and  $\mathbf{Y}$  must *equal zero*. Since this consequence is the basis of the information flow measure defined in §4.2, we will reiterate it more formally. Let the integers 1 through  $n$  designate the  $n$  random variables in the vector  $\mathbf{X}$  and  $n + 1$  through  $n + n'$  the  $n'$  random variables in  $\mathbf{Y}$ . If the random processes described by  $\mathbf{X}$  and  $\mathbf{Y}$  are statistically independent, then the following condition will be satisfied by the cumulants of the joint process:

$$\begin{aligned} \mathcal{K}_{\ell_1 \dots \ell_m} = 0 \quad \text{if} \quad \exists \ell_i \in [1, n] \quad \text{and} \quad \exists \ell_j \in [n + 1, n + n'], \\ \text{where} \quad i, j \in [1, m] \quad \text{and} \quad i \neq j. \end{aligned} \quad (\text{B.16})$$

## Appendix C

# Coupling Matrix Studies of the Bondarenko Model

The asymmetric analogue neural network model with delays presented in § 2.1.1 was chosen to generate the synthetic time series for some of our experiments for two reasons. It was known to have chaotic solutions which were qualitatively similar to EEG signals. Secondly, the chaoticity or degree of chaos, as reflected in the maximal Lyapunov exponent, had been shown to be an increasing function of what we have called the global coupling strength  $c$ . Our expectations of the model were based on a series of papers by V. Bondarenko [7–11]. These same properties of the solutions prompted Hively, Gailey and Protopopescu to test a different set of measures for detection of dynamical change in scalp EEG signals on data from this model [32]. So when values of  $c$  that should have produced chaos gave fixed point or quasiperiodic solutions instead, we got worried.

Our main reference for Bondarenko's work was [8], not his earliest paper on the topic. When (at his suggestion) we took a look at this 1994 article, we found that a dependence of the type of solution on the *realization* of the coupling matrix had been observed. For given values of  $c$ ,  $M$  (number of neurons) and  $\tau$  (delay), the spectra and correlation dimension depend on the coupling matrix realization [12]. Finally, different initial conditions produced different solutions for the same matrix. In his experiments, Bondarenko ensured convergence [12] to the same attractor for all re-

alizations of the coupling matrix by choosing his initial conditions randomly from a uniform distribution on  $[-1 \times 10^{-50}, +1 \times 10^{-50}]$ , which he describes in [8] as “small random values”. Small indeed. We, on the other hand, used random deviates on  $[-2, +2]$  for both the matrix elements and initial conditions, as did Hively and his collaborators.

Once we were confident that the nonchaotic solutions were not the result of a bug in our program, we asked the question: what conditions on the coupling matrix, if any, assure us that the chaotic solution will be the stable one? Marcus and Westervelt had studied the stability of random *symmetric* analogue neural networks with delay [51]. Sompolinsky, Crisanti and Sommers had used mean field theory to analyze random asymmetric neural networks *without* delay [79]. We know of no analytical results, however, for asymmetrical random coupling with delay. What follows is a brief numerical exploration of different aspects of the problem.

For the symmetric case, Marcus and Westervelt obtain a delay-dependent expression for the “regions of stability” in the complex plane within which all the eigenvalues of the coupling matrix (the “connection eigenvalues”) must fall for the network not to oscillate. We undertook to examine the relationship between the connection eigenvalues of different matrices and the type of solution obtained with a given matrix using the Bondarenko model. The spectrum of large random fully asymmetric matrices is known to be uniform on a circle in the complex plane [78]. It was conceivable that on this circle, the eigenvalues of matrices producing chaos would fall in a discernible region.

To test this hypothesis, a routine was devised for automatically classifying the output of the network. 1688 separate realizations of the network were classified using this scheme. An  $M = 10$  neuron network ( $c = 3$ ,  $e = 0$ ,  $\tau = 10$ ) was used, always with the same initial conditions on the neurons but with a different coupling matrix each time. The matrix was output and its eigenvalues computed [66]. For each initialization of the network, the procedure for estimating the maximal Lyapunov

exponent described in §2.1.4.1 was followed except that: a) a transient of 3000 atu was skipped, b) the ratio  $\delta_n/\delta_0$  was stored only up to  $n = 1500$  and finally c) the value of  $\delta_n/\delta_0$  was averaged over three rather than 50 runs. Two straight lines were then fit to  $\ln(\delta_n/\delta_0)$  vs.  $n$ , one on  $n \in [200, 800]$  and the other on  $n \in [1000, 1500]$ .

Let  $m_1$  and  $m_2$  denote the slope of these two lines respectively,  $b_1$  and  $b_2$  their respective intercepts and  $\Delta m_1$ ,  $\Delta m_2$ ,  $\Delta b_1$  and  $\Delta b_2$  the errors in these quantities returned by the curve fitting routine. The following criteria were used, in the given order, to classify the results:

1. *Chaotic*, if

- (a)  $m_1 > 0.005$  and
- (b)  $\Delta m_1 < 0.0009$ .

A positive slope with small error implies linear scaling region in  $\ln(\delta_n/\delta_0)$  vs.  $n$  plot and thus exponential growth in distance between trajectories.

2. *Quasiperiodic*, if

- (a)  $m_1 < 0.005$  and
- (b)  $m_2 < 0.005$  and
- (c)  $|b_2 - b_1| < 2$  and
- (d)  $\Delta b_2 > 0.1$ .

For this type of solution, we expect no overall increase in the separation but some fluctuation (hence the non-zero  $\Delta b_2$ ).

3. *Fixed point*, if

- (a)  $|m_2| < 0.000001$  and
- (b)  $\Delta b_2 < 0.001$

A fixed point solution will yield both very flat lines *and* very little fluctuation.

4. otherwise *uncertain*.

The numerical values used in the various inequalities above were selected on the basis of visual inspection of the results for about 50-100 matrices.

Of the 1688 runs completed, 55% were identified as chaotic, 36% quasiperiodic, 5.5% fixed point and 3.6% uncertain. These results allow us to plot, in Fig. C.1, the connection eigenvalues of these matrices colour coded by type of solution obtained. A fairly uniform distribution on a circle is observed *for all solution types*; there are no discernible stable or chaotic regions. The deviations from a perfectly uniform distribution, in particular the high concentration of eigenvalues on the real axis, are a consequence of the finite dimensions of the matrix [51].

The next question we asked was whether a given realization of the coupling matrix would always yield the same type of solution for different sets of initial conditions. For each of 100 different realizations of the matrix, the long-term behaviour of the system was classified using the same scheme described above with 20 different sets of initial conditions, obtained by seeding the RNG for these random values differently. The same 20 sets of initial conditions were used for each coupling matrix (the same 20 seeds were used). Averaging over the results for all 100 matrices, we found that 55% of solutions were chaotic (SD 4%), 35% were quasiperiodic (SD 4%) and 6% were fixed point (SD 3%). 3% could not be classified (SD 2%). The type of solution obtained for each combination of coupling matrix and set of initial conditions is presented in Fig. C.2.

What is most striking about the results of these two experiments is the similarity in the proportion of each type of solution found. This suggests that for any given realization of the coupling matrix and set of initial conditions, there is about a 50/50 chance that the long-term behaviour of the network will be chaotic and that there is no way of determining on the basis of the coupling eigenvalues alone what that behaviour will be. We should point out that the matrix “Realization 1” was used to generate our artificial TSs. We see that for all 20 seed values, this particular matrix

Figure C.1: Connection eigenvalues of different realizations of Bondarenko model coupling matrix, colour coded by solution obtained: red for chaotic, blue for quasiperiodic, green for fixed point. Green points are more prominent on the real axis only because they were plotted last; the distribution of all solution types peaks strongly on this axis.

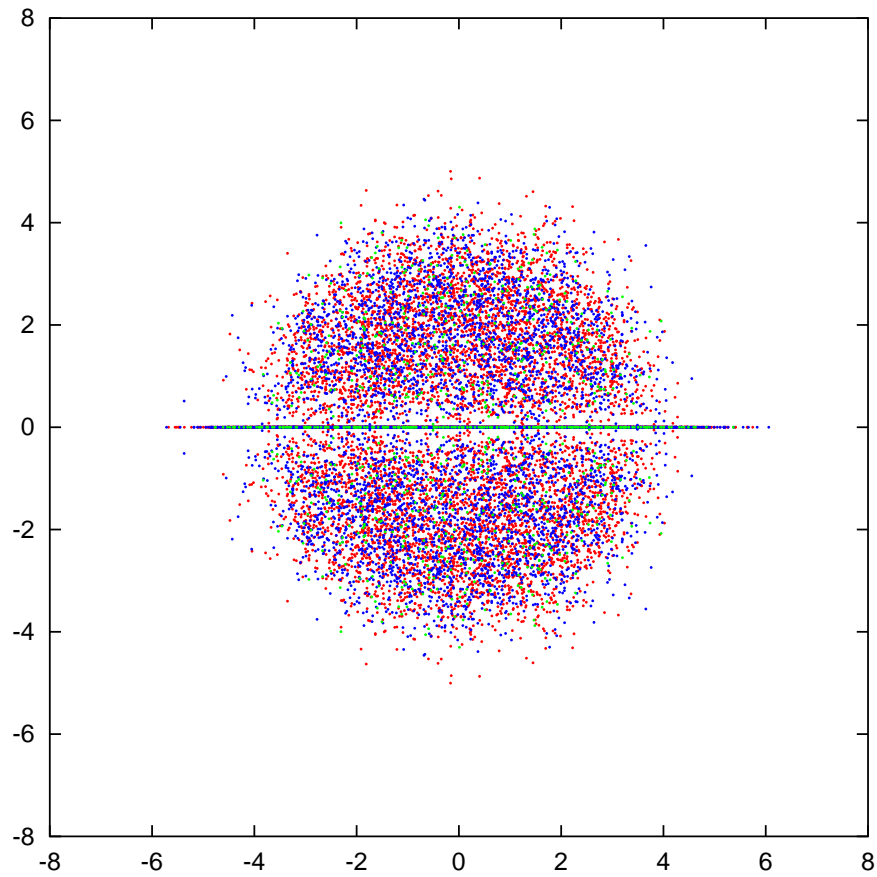
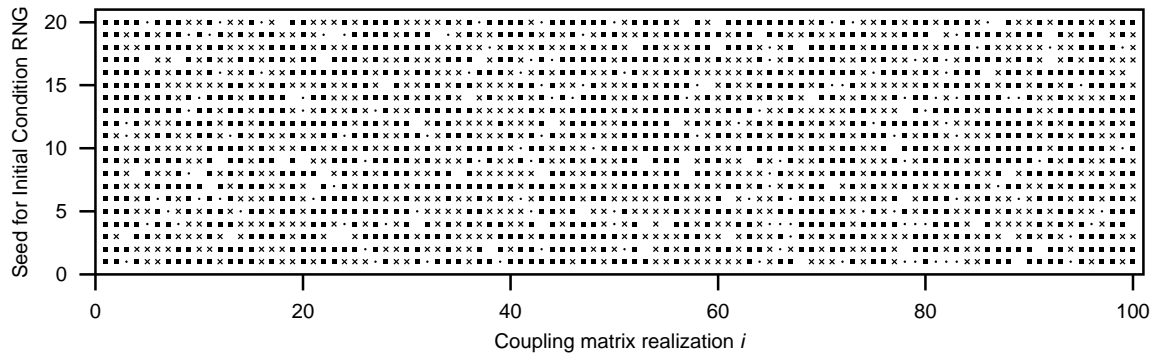




Figure C.2: For each combination of coupling matrix and seed for the initial conditions the type of solution obtained is indicated by a symbol: squares for chaos, crosses for quasiperiodicity, dots for fixed points. Where the behaviour could not be classified automatically, the space is left blank.



yields chaos. It happens to do so for seed values 21 through 25 as well, so that *all* the TSs in our data set for  $c = 3$  were chaotic. For other values of the measure parameters, the TSs were inspected to ensure that a nonchaotic solution had not slipped in.

# Appendix D

## Gap Cataloguing Algorithm

Since fluctuations of the measures studied in the preceding chapters are *ignored* on what we call “gaps” in the EEG recordings, the algorithm used for automatic detection of these gaps could have some bearing on our results. As it turns out, in our data set, gaps seldom occur prior to the first clinical seizure on recordings which capture such seizures, which is the interval of most concern in our analyses.

The sequence of “zeroes” which we call a gap does not always begin and end at *exactly* the same time in every channel of the recording. This algorithm is designed to catalogue the *earliest* start time of the gap on the first five channels and the *latest* end time. There are two main steps to the algorithm: gaps on each channel are first identified and then the start and end times of these gaps are compared to those found on previous signals in order to keep the earliest and latest times respectively.

Each of the first five channels of data from the recording is scanned for intervals satisfying three criteria: (1) consecutive samples on the interval differ by no more than  $\Delta x$ , (2) the absolute difference between the maximum and minimum values on the interval is no greater than  $\Delta x$  and (3) the interval is at least  $\Delta t$  samples long. This approach was used because the “zeroes” filling the gap were not identically equal to zero or even strictly constant, which is presumably a consequence of signal digitization. The lower limit on duration prevents identification as gaps of intervals on which the signal is simply varying slowly. We used values of  $\Delta x = 10.0$  mV and  $\Delta t = 200$  samples (i.e. 1 sec). The location of the beginning and end of every

interval is flagged with “start” and “end” markers on a “timeline” the same length as the recording,  $N$  samples.

After the candidate intervals have been logged for each signal, the timeline is scanned forward to the first “start” marker. From that point, we scan to the nearest “end” marker, erasing all intervening “start” markers. This procedure is repeated until only the first “start” in each unbroken sequence of “start” markers remains. Moving backwards along the timeline, the same procedure retains only the latest “end” in each unbroken series of “end” markers. After all five channels have been considered, the number of remaining “start” markers is logged as the number of gaps and the start and end of each gap recorded as the positions of consecutive pairs of “start” and “end” markers on the timeline.

A formal description of this algorithm (almost an implementation, really) is presented below. We note that it is similar in many respects to the algorithm used for logging hot zones with respect to the dynamical similarity measure.

**Gap cataloguing algorithm:**

```

for all  $i$  such that  $1 \leq i \leq N$  do
   $g_i \leftarrow 0$ 
for each of the first  $m$  signals in a multichannel recording do
  Read the signal into the vector  $\mathbf{x} = (x_1, \dots, x_N)$ 
   $q \leftarrow 0$ 
  for  $i = 2$  to  $N$  do
    if  $|x_i - x_{i-1}| < \Delta x$  then
      if  $q > 0$  then
        if  $x_i > x_{\max}$  then
           $x_{\max} \leftarrow x_i$ 
        if  $x_i < x_{\min}$  then
           $x_{\min} \leftarrow x_i$ 
      else
         $x_{\min} \leftarrow x_i$ 
         $x_{\max} \leftarrow x_i$ 
         $q \leftarrow i - 1$ 
    else
      if  $q > 0$  then
        if  $|x_{\max} - x_{\min}| < \Delta x$  and  $i - q \geq \Delta t$  then
           $g_q \leftarrow 1$ 

```

```

         $g_{i+1} \leftarrow -1$ 
         $q \leftarrow 0$ 
 $q \leftarrow 0$ 
for  $i = 1$  to  $N$  do
    if  $g_i = 1$  then
        if  $q = 1$  then
             $g_i \leftarrow 0$ 
        else
             $q \leftarrow 1$ 
    else
        if  $g_i = -1$  and  $q = 1$  then
             $q \leftarrow 0$ 
 $q \leftarrow 0$ 
for  $i = N$  to  $1$  do
    if  $g_i = -1$  then
        if  $q = 1$  then
             $g_i \leftarrow 0$ 
        else
             $q \leftarrow 1$ 
    else
        if  $g_i = 1$  and  $q = 1$  then
             $q \leftarrow 0$ 
Output number of  $g_i$  which satisfy  $g_i = 1$  (number of gaps)
 $i \leftarrow 1$ 
while  $i < N$  do
    if  $g_i = 1$  then
        Output  $i$  (gap start)
        while  $g_i \neq -1$  and  $i < N$  do
             $i \leftarrow i + 1$ 
        Output  $i$  (gap end)
     $i \leftarrow i + 1$ 

```