



DIPLOMARBEIT

Localizing the focus of epileptic seizures using modern measures from multivariate time series analysis

ausgeführt am
Institut für Wirtschaftsmathematik
der Technischen Universität Wien

unter der Leitung von
o.Univ.Prof. Dr. Manfred Deistler

in Kooperation mit
AIT Austrian Institute of Technology GmbH

von

THOMAS SCHUSTER
Friedhofgasse 6
2540 Bad Vöslau

und

ULRIKE KALLIAUER
Loiskandlzeile 4
3550 Langenlois

Wien, Oktober 2009

Statutory Declaration

We hereby certify that all work presented in this diploma thesis is our own, no other than the sources and aids referred to were used and that all parts which have been adopted either literally or in a general manner from other sources have been indicated accordingly.

Furthermore we permit the Austrian Institute of Technology - AIT to use any of our work included in this diploma thesis for their purposes.

Vienna, October 2009

Thomas Schuster and Ulrike Kalliauer

Abstract

In this diploma thesis, we present different methods to realize the visualization of the spatio-temporal evolution of an epileptic seizure's focus based on multi-channel ECoG data. Furthermore, our methods allow for a precise localization of the seizure's initial focus, which is possible, as that focus exists because all our data stems from patients suffering from temporal lobe epilepsy.

Two main methods are presented, starting with a frequency domain approach. Based on the coefficients of a fitted AR-model, modern measures like partial directed coherence (PDC) are derived and discussed. Exhaustive analysis of PDC's problems further leads to a generalized and far better performing version of PDC.

In the second part, a Recursive Least Squares (RLS) Algorithm is performed instead of the ordinary least squares approach in frequency domain. This RLS-Algorithm helps us to cope with the instationarities of the ECoG signals far better.

Based on the time-dependent AR-model, Granger causality is used to indicate interactions between channels of the multivariate signals. Starting with Granger's basic idea and the analysis of coupling effects between two signals, stepwise partialization leads to improved results.

Generalized PDC as well as the Granger causality finally lead to a visualization of the seizure's evolution. The results presented are in excellent accordance with descriptions from clinical experts.

This diploma thesis is the work of two authors, where

- **Chapter 2**, which is the frequency domain approach, was **written by Thomas Schuster**, as well as the Medical Introduction in Section 1.1, Conclusion and Outlook,
- **Chapter 3**, which shows the Granger causality approach, Section 1.2 and the Appendix were **written by Ulrike Kalliauer**,
- the other chapters of the work have been elaborated cooperatively.

Summary

In this diploma thesis, we present different methods for the epileptic seizure propagation analysis. They allow to track the temporal evolution of the seizure focus. Especially the initial focus can be localized, which is essential if an epilepsy surgical intervention is necessary.

We will start in this work with a medical introduction in Chapter 1, where the reader will be made familiar with all the medical and mathematical terminology. Temporal lobe epilepsy is explained in more detail, which is one of the most common types of epilepsy, and furthermore one of the most promising for mathematical analysis. Consequently, all data used in our experiments stem from patients suffering from temporal lobe epilepsy.

Further, connections between mathematical and medical analysis are presented. The differences between EEG and ECoG data are explained, as well as the characteristic coupling effects between those signals during epileptic seizures, which will be identified by mathematical measures.

Basic information about our data and the doctors estimate about the evolution of the epileptic seizures are given. Figures will be presented, which will clarify the resulting plots from all our analysis.

In Chapter 2 (the frequency domain approach) classical measures like ordinary coherence are used as well as more recent ones, which are based on the coefficients of fitted AR-models. Partial directed coherence will be derived from the parametric representation of ordinary coherence. The problems of both measures are explained exhaustively and possible corrections are presented.

Next, generalized partial directed coherence is developed and proven to be an adequate measure for seizure propagation analysis. Results derived from that measure are visualized and will turn out to coincide quite good with the opinion of clinical experts.

In Chapter 3 a Recursive Least Squares Algorithm (RLS) will be used, which allows us to cope with the instationarity of the given biosignals even better. Based on the resulting time-dependent AR-model coefficients, we will recognize that Granger causality is an appropriate approach to track epileptic seizures.

Analysis will start with a simple bivariate Granger causality approach, whereas interactions between channels are analyzed pairwise without any respect to all the other channels available. Stepwise partialization then will lead to conditional and further partial Granger causality.

Again, results of all three measures will be visualized. Especially results from partial Granger causality will turn out to be in perfect accordance with the clinical epileptic seizure descriptions. Due to RLS and the partialization, the evolution of the seizure focus could be visualized even better than when using the frequency domain approach.

Finally in Chapter 4 the main achievements are summarized and results from Chapters 2 and 3 are compared. Critical assumptions made are emphasised and combined with ideas for improvements. Additionally a list of possible approaches for further research is given.

Acknowledgment

We hope, that this diploma thesis will arouse interest at the reader, the medical as well as the mathematical expert. Further we hope that this work will help to treat patients suffering from temporal lobe epilepsy, even in not too far future.

We like to thank for everyone's support during our work.

First, our special thanks are addressed to Andreas Graef, who himself has written a diploma thesis on a similar topic less than one year before our work. He provided us with all his experience, not only on mathematical or medical issues, but on a lot of relevant issues that appeared during our work on that diploma thesis. He arranged meetings and patiently answered emails, even in the evening or on weekends.

Our sincere thanks like to go to our supervising professor of this diploma thesis: o.Univ.Prof. Dr. Manfred Deistler, professor of economics at the "Institute for Mathematical Methods in Economics" and head of the research group "Econometrics and System Theory" at the Vienna University of Technology. He constantly helped us with inputs and critics based on his unbelievable experience. We are extremely grateful for his ability to make complex mathematical problems much clearer.

Furthermore, we want to thank all our colleagues at the Austrian Institute of Technology - AIT, especially Dr. Tilman Kluge, the head of the research group. He provided all data as well as a lot of advice and explained us the implications of epileptic seizures onto the ECoG data graphs. He is the one, who made us to understand, why mathematical research should be able to track the evolution of an epileptic seizure's focus.

Our special thanks also go to Manfred Hartmann from AIT, who especially helped us with mathematical and programming-related questions, and to Andreas Flamm and Georg Görg, both students of Vienna University of Technology, too. Their critics and explanations have made some aspects much clearer for us.

Last but not least, we want to thank our parents for their support throughout our studies. Without them, we impossibly would have proceeded that far and would not even have thought about writing a diploma thesis.

Thomas Schuster and Ulrike Kalliauer
Vienna, October 2009

Contents

1	Introduction	1
1.1	Medical background	1
1.1.1	Basic facts	1
1.1.2	Definition of epilepsy and epileptic seizures	2
1.1.3	Classifying epileptic seizures	2
1.1.4	Temporal lobe epilepsy	2
1.2	The link between medical and mathematical analysis	3
1.2.1	Data recording: EEG and ECoG	3
1.2.2	Data basis	5
1.2.3	Finding appropriate measures	6
1.3	Notation and Terminology	7
1.3.1	Terminology	7
1.3.2	Conventions for Notation	7
2	Localizing the focus of epileptic seizures using measures in Frequency Domain	8
2.1	Frequency domain basics	8
2.2	Measures derived from nonparametric spectrum estimators	10
2.2.1	Nonparametric estimation of the spectrum	10
2.2.1.1	Periodogram	10
2.2.1.2	Smoothed spectral estimates	14
2.2.1.3	Welch's Method	16
2.2.2	Coherence	19
2.2.3	Partial (spectral) coherence (PSC)	21
2.2.4	A simple signal model	24
2.2.5	Computational formula for the partial coherence	25
2.3	Measures derived from parametric spectrum estimators	30
2.3.1	Parametric estimation of the spectrum	30
2.3.1.1	Setting up a parametric model	30
2.3.1.2	Transformation into frequency domain	31
2.3.2	Ordinary coherence	32
2.3.3	The partial coherence function	33
2.3.4	The partial directed coherence factor (PDCF)	34
2.3.4.1	PDCF and factorization of the partial coherence function	34
2.3.4.2	Instantaneous causality	35
2.3.5	Partial directed coherence (PDC)	38
2.3.5.1	Characteristics and interpretation of PDC	38

	2.3.5.2	Link to Granger causality	40
	2.3.5.3	Alternative normalizations	41
	2.3.5.4	Statistical tests and confidence intervals	42
	2.3.5.5	Problems and disadvantages of PDC and its construction	44
	2.3.6	Generalized partial directed coherence (GPDC)	47
	2.3.7	Comparing PDC and GPDC using signal models	47
2.4	Implementation		51
	2.4.1	The Fast Fourier transform (FFT) algorithm	51
	2.4.1.1	Deriving the algorithm	51
	2.4.1.2	Computational time savings	53
	2.4.1.3	Using FFT to compute the autocovariance function and the Periodogram	53
	2.4.2	Estimating the parameters of an AR-model	56
	2.4.3	Channel selection	56
	2.4.4	Windowing and the signal instationarity	57
	2.4.5	Derivation of a frequency independent measure and the choice of a suitable threshold	57
2.5	Results		59
	2.5.1	Partial coherence (nonparametric)	59
	2.5.1.1	Patient 1, seizure 1	59
	2.5.1.2	Patient 2, seizure 3	60
	2.5.2	Partial directed coherence (parametric)	62
	2.5.2.1	Patient 4, seizure 1	62
	2.5.2.2	Patient 2, seizure 3	64
	2.5.2.3	Patient 1, seizure 1	65
	2.5.3	GPDC results for patient 2	66
	2.5.3.1	Seizure 1	66
	2.5.3.2	Seizure 3	67
	2.5.3.3	Seizure 4	68
	2.5.3.4	Seizure 2	69
	2.5.3.5	Conclusion	70
	2.5.4	GPDC results for patient 4	70
	2.5.4.1	Seizure 1	70
	2.5.4.2	Seizure 2	71
	2.5.4.3	Seizure 3	72
	2.5.4.4	Conclusion	72
	2.5.5	GPDC results for patient 1	73
	2.5.5.1	Seizure 1	73
	2.5.5.2	Seizure 2	74
	2.5.5.3	Seizure 3	75
	2.5.5.4	Conclusion	76
3	Localizing the focus of epileptic seizures using Granger causality		77
3.1	Bivariate Granger causality		77
	3.1.1	Definition	77
	3.1.1.1	Definition by C. W. Granger	77
	3.1.1.2	Bivariate Granger causality	79

3.1.1.3	Bivariate Granger causality for a partitioned process	80
3.1.2	Characterizing Granger causality with AR coefficients . . .	80
3.2	Conditional Granger causality	86
3.2.1	Definition	86
3.2.2	Comparison between bivariate and conditional Granger causality	89
3.3	Partial Granger causality	94
3.3.1	Definition	95
3.3.1.1	Linear model	95
3.3.2	Comparison between conditional and partial Granger causality	97
3.4	Implementation	100
3.4.1	Recursive Least Square algorithm	100
3.4.1.1	Preliminary remarks	100
3.4.1.2	Derivation of the RLS	102
3.4.1.3	Summary of the RLS algorithm	105
3.4.2	Channel selection	105
3.4.3	Selection of the forgetting factor λ	109
3.4.4	Order selection	111
3.5	Results	112
3.5.1	Bivariate Granger causality	112
3.5.1.1	Why Common Average data are necessary	112
3.5.1.2	Patient 2, seizure 3	115
3.5.1.3	Patient 1, seizure 1	117
3.5.1.4	Patient 1, seizure 3	118
3.5.2	Conditional Granger causality	119
3.5.2.1	Channel selection	119
3.5.2.2	Patient 2, seizure 3	120
3.5.2.3	Patient 1, seizure 1	121
3.5.2.4	Patient 1, seizure 3	123
3.5.3	Partial Granger causality	124
3.5.3.1	Patient 2, seizure 1	125
3.5.3.2	Patient 2, seizure 2	126
3.5.3.3	Patient 2, seizure 3	127
3.5.3.4	Patient 2, seizure 4	128
3.5.3.5	Patient 1, Seizure 1	129
3.5.3.6	Patient 1, seizure2	131
3.5.3.7	Patient 1, seizure 3	132
3.5.3.8	Patient 4, seizure 1	133
3.5.3.9	Patient 4, seizure 2	134
3.5.3.10	Patient 4, seizure 3	135
3.5.3.11	Conclusion	136
4	Conclusion and Outlook	138
4.1	Conclusion	138
4.2	Outlook	140

A	Clinical description	141
A.1	Patient 1	141
A.2	Patient 2	142
A.3	Patient 4	144

List of Figures

1.1	The 4 lobes of the cortex	3
1.2	ECoG data during an epileptic seizure	5
1.3	Channel numbers of ECoG electrodes of patient 1:	6
2.1	Illustration of data segmentation	16
2.2	Coherence: direct vs. indirect influences	21
2.3	The signal model to demonstrate PSC	25
2.4	The signal model and ordinary coherences	26
2.5	The signal model and spectral coherences	26
2.6	Normalization of the PDC	39
2.7	Alternative Normalization of the PDC	42
2.8	Scale-invariance of PDC	48
2.9	Analyzing the signal model using GPDC	50
2.10	Analyzing the signal model using PDC	51
2.11	FFT vs. DFT - computational time	54
2.12	Periodogram - Definition vs. FFT calculation	55
2.13	Typical graphs for frequency-dependent measures	57
2.14	Brainplots for partial coherence: data from patient 1, seizure 1	60
2.15	Brainplots for partial coherence: data from patient 2, seizure 3	61
2.16	Channel pairs with high PSC values, but no epileptic activity	62
2.17	Brainplots and residual error variances for patient 4, seizure 1	63
2.18	Brainplots and residual error variances for patient 2, seizure 3	64
2.19	Brainplots and residual error variances for patient 1, seizure 1	65
2.20	GPDC results for patient 2, seizure 1	66
2.21	GPDC results for patient 2, seizure 3	67
2.22	GPDC results for patient 2, seizure 4	68
2.23	GPDC results for patient 2, seizure 2	69
2.24	GPDC results for patient 4, seizure 1	70
2.25	GPDC results for patient 4, seizure 2	71
2.26	GPDC results for patient 4, seizure 3	72
2.27	GPDC results for patient 1, seizure 1	73
2.28	GPDC results for patient 1, seizure 2	74
2.29	GPDC results for patient 1, seizure 3	75
3.1	Example 1: bivariate vs. conditional: Structure of the Simulation	90
3.2	Example 1: bivariate vs. conditional: Pair-wise bivariate Granger causality index	90
3.3	Example 1: bivariate vs. conditional: The result of the bivariate analysis	91

3.4	Example 1: bivariate vs. conditional: conditional Granger causality index	91
3.5	Example 1: bivariate vs. conditional: conditional Granger causality	92
3.6	Example indirect cause: structure	92
3.7	Example: bivariate vs. conditional: results: indirect influence	92
3.8	Example: bivariate vs. conditional: causality without connection	93
3.9	Example: bivariate vs. conditional: causality without connection: results	93
3.10	Example: Inversion of a causal relationship when using bivariate methods	93
3.11	Results of example: inversion of causal relationship in bivariate case	94
3.12	Structure	98
3.13	Analysis without latent variables	99
3.14	Analysis with latent variables	100
3.15	Analysis with latent variables having different influences	101
3.16	Example: Channel selection - structure	107
3.17	Example Channel selection: Conditional and bivariate Granger causality index: whole system	107
3.18	Example Channel selection: Conditional Granger causality index: partitioned system	108
3.19	Example Channelization: bivariate Granger causality (partition system)	109
3.20	Channels used to find an optimal lambda	110
3.21	Forgetting factor lambda vs. variance of the one-step prediction error	110
3.22	AIC criterion for the ECoG data:	111
3.23	Bivariate analysis of seizure 3, patient 2 - without common average	113
3.24	Graph of data measured with reference to a channel or to the common average	114
3.25	Variances of data with reference to a channel and to the common average	114
3.26	Percentage of variance reduction	115
3.27	Bivariate Granger causality index of seizure 3, patient 2	116
3.28	Bivariate analysis for seizure 3, patient 2	117
3.29	Bivariate analysis of seizure 1, patient 1	118
3.30	Bivariate Analysis of seizure 3, Patient 1	119
3.31	Conditional Granger causality index: Patient2 seizure3	120
3.32	Conditional analysis of seizure 3, patient 2	121
3.33	Conditional analysis of seizure 1 from patient 1	122
3.34	Conditional analysis of seizure 3, patient 1	123
3.35	Partial analysis: Patient 2, seizure 1	125
3.36	Partial analysis: Patient 2, seizure 2	126
3.37	Partial analysis of seizure 3, Patient 2	127
3.38	Partial analysis of seizure 4, patient 2	129
3.39	Partial analysis of seizure 1, patient 1	130
3.40	Channels 27 to 29 of seizure1, Patient 1	131
3.41	Partial analysis of seizure 3, patient 1	131
3.42	Partial analysis of seizure 3, patient 1	132
3.43	Partial analysis of seizure 1, patient 4	134

3.44	Partial analysis of seizure 2, patient 4	135
3.45	Partial analysis of seizure 3, patient 4	136
A.1	Numeration of the electrodes of the ECoG from patient 1	141
A.2	Numeration of the electrodes of the ECoG from patient 4	142
A.3	Numeration of the electrodes of the ECoG from patient 2	144

List of Tables

3.1	Dimensions of the block matrices of the partitioned polynomial coefficient matrix $\Phi(z)$	81
3.2	Connection: forgetting factor λ to the number M of non-negligible samples and the memory time constant (τ_λ)	111
A.1	Clinical description of seizure 1 from patient 1	141
A.2	Clinical description of seizure 2 from patient 1	142
A.3	Clinical description of seizure 3 from patient 1	142
A.4	Clinical description of seizure 1 from patient 2	143
A.5	Clinical description of seizure 2 from patient 2	143
A.6	Clinical description of seizure 3, patient 2	143
A.7	Clinical description of seizure 4, patient 2	143
A.8	Clinical description of seizure 1, patient 4	144
A.9	Clinical description of seizure 2, patient 4	144
A.10	Clinical description of seizure 3, patient 4	145

Chapter 1

Introduction

We hope, that our work will give a lot of suggestions to medics and to mathematicians. Thus we want to use the first chapter to clarify the whole terminology: the medical as well as the mathematical one.

If the reader is not familiar with some common mathematical expressions, we recommend [5]. For any medical questions, a standard medical dictionary like [1] would be helpful.

1.1 Medical background

As the reader might have a mathematical or at least a technical background, but not necessarily any medical knowledge, we want to start this diploma thesis with a medical introduction.

We want to give a short overview, which contains the information necessary to understand the basic idea behind the given task: localizing the focus of epileptic seizures. This overview is based on [19] and [4], which are recommended to the interested reader.

1.1.1 Basic facts

In industrial countries, year by year, some kind of epilepsy is diagnosed by between 24 and 53 out of 10,000 inhabitants. Thus it is one of the most common serious neurological disorders, nearly comparable with diabetes.

For most people affected, symptoms start either during their first year of life, or after the age of 70. Based on statistics, the risk of falling ill is minimal in adulthood. Furthermore, it is slightly smaller for women than for men.

As for many diseases the relationship holds, that the better the socioeconomic status, the lower the risk of getting infected. Consequently, the number of incidences lies approximately three times higher in developing countries, compared with the numbers mentioned for industrial countries. In fact, that might

also be the reason, why blacks tend to suffer from epilepsy more often than whites.

1.1.2 Definition of epilepsy and epileptic seizures

Although those terms are mixed up quite often, medics clearly differentiate between epilepsy and epileptic seizures

Definition 1.1 (Epileptic seizure). An *epileptic seizure* is the clinical manifestation of excessive hyper-synchronous discharges of neurons in the cerebral cortex.

Symptoms mostly depend on the function of the affected brain area. Epileptic seizures can lead, for example, to the disorder of advanced brain functions, unconsciousness, abnormal sensory perceptions or spasms.

Epileptic seizures may also lead to severe complications. First of all, they might cause a lot of damage to affected brain regions, which is probably based on a high concentration of calcium within the nerve cells. On the other hand, the inability to control the extremities within a seizure may cause car accidents as well as downfalls from ladders or even drowning.

Definition 1.2 (Epilepsy). Epilepsy is a chronic disease, a heterogeneous group of affections with various syndromes and different causes. All of them have one characteristic in common: recurrent, unprovoked seizures.

In other words, medics speak of epilepsy, if seizures appear (without any exogenous reasons) repeatedly.

1.1.3 Classifying epileptic seizures

At first, a basic classification of epileptic seizures was presented by the Commission on Classification and Terminology of the International League against Epilepsy nearly thirty years ago. It is still in use and quite simple to understand.

1. **Focal (partial) seizures:** The initial focus of the epileptic seizure is localized only in one part of one cerebral hemisphere. It does not matter, whether the epileptic activity later is proceeding to different regions of the brain, or not.
2. **Generalized seizures:** They initially involve significant parts of both cerebral hemispheres. It is impossible to localize an initial focus.
3. **Not-classifiable seizures:** They do not fit into one of the first two points due to missing data.

1.1.4 Temporal lobe epilepsy

Temporal lobe epilepsy is one of the most common forms of focal epilepsy. Consequently all data used in this work stem from patients suffering from this form of focal epilepsy. As we will see in the result sections, some of the seizures may generalize, i.e. both hemispheres may be affected as epileptic activity expands.

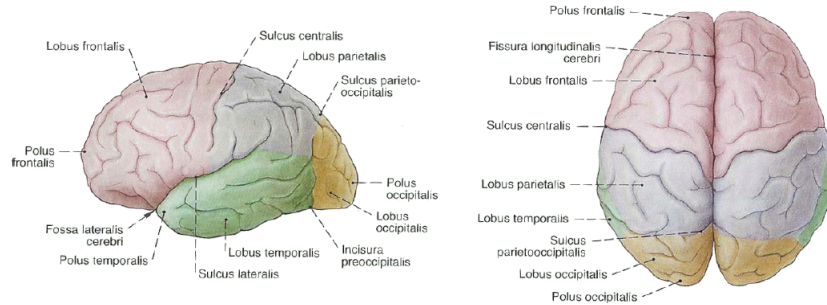


Figure 1.1: **The 4 lobes of the cortex.** Pictures represent the brain as seen from the right side (left picture) or from above (right picture). The temporal lobe (*Lobus temporalis*), we are mainly interested in, is indicated in green.

But, as classified before, it is in fact possible to find an initial focus for all of them.

The temporal lobe (*Lobus temporalis*) is located on both sides of the brain. Its position is painted with green color in Figure 1.1. The temporal lobe contains important brain regions, e.g. the auditory cortex and structures important for the brain's memory function.

Patients suffering from temporal lobe epilepsy mostly report the so called *Aurea* at the beginning of the seizure, which is a special, hard to describe, kind of feeling. Other typical symptoms are (which of course coincides with the brain functions located in the affected area) a limitation of consciousness, movements of the whole body, confusion or speech disorder.

For every patient suffering from temporal lobe epilepsy, medicamentous therapy is what doctors try first and helps for a majority of them. On the other hand, if the administration of drugs could not help to control seizures, an epilepsy surgical intervention (if possible) might be necessary. Obviously, for that intervention it is necessary to know the position of the initial focus as precise as possible. Thus the localization of the focus is exactly what we want to realize using measures from multivariate time series analysis.

Furthermore, not every surgical intervention is successful. Mathematical analysis might help to increase the number of operations which go well.

1.2 The link between medical and mathematical analysis

1.2.1 Data recording: EEG and ECoG

Before a surgical intervention can take place, the patient has to make long-lasting presurgical tests (lasting on average one week), as, obviously, a precise localization of the initial focus is needed. The surgical intervention is not re-

versible and may leave long lasting consequences (for example if the speech region is affected).

The first part of this long-lasting process is to analyze the epileptic activity with the help of the ***Electroencephalography (EEG)*** to receive an idea, where the initial focus (the source of the seizure) might be located. The EEG records electric cortical activity (caused by neurons within the brain). Usually, about 20 electrodes are placed along the scalp.

Medics analyze the EEG data and try to find abnormal activities and oscillation (like typical spikes or synchronous oscillations of neurons) between the variety of frequencies, which appear in the data.

Afterwards, to localize the focus in a more precise way, the ***electrocorticography (ECoG)*** is used. Therefore, a surgeon has to remove a part of the skullcap to expose the brain surface, where the electrodes will be directly implanted. The previous examination with the EEG is needed to constrain the exposure of the brain and to guide the placement of the electrodes.

Contrary to the EEG signal, we see low voltage and high frequency components more clearly in the ECoG data, because in the EEG data the measured signals are attenuated due to the bone's low conductivity. In fact, this is exactly the reason why only ECoG data will be used in our work.

Furthermore, the distance between two electrodes is less than one centimeter and therefore the medics obtain a precise localization of the focus. This is important because the difficulty of the surgery intervention is to avoid neurological deficits caused by the operation.

For about 30 to 50% of the patients, the epileptic region cannot be localized and the electrodes are removed without any surgery. The problem of analyzing the ECoG data is, that medics with a high experience are necessary to identify the epileptic activity. This is done through a visual interpretation of the ECoG data and problems will arise especially if there is a rapid seizure generalization.

Although signals of the affected electrodes show simultaneous oscillations or typical patterns (like spikes), for someone who is not specialized on analyzing this sort of data, there seems to be no chance of finding epileptic activity. To have an idea how difficult this is, we want to present in figure the ECoG data of one hemisphere, where channels 7, 8 and 9 show epileptic actions after five seconds in figure 1.2.

Contrary to the other channels, could be the same rhythm in the graph of channel 7, 8 and 9. Furthermore, they seem to have regular and simultaneous, high frequented oscillations compared to the other graphs. One effect which occurs often with channels showing epileptic activity is the tendency to have a higher amplitude, but this phenomenon is not significant for these channels. (This data belongs to the first seizure of patient 2.)

The difficulty of finding the initial focus just based on visualized ECoG signals combined with the fact, that even experts themselves often are not sure, where the focus exactly is located, has led to our diploma thesis. An automatic tool for the localization of the patient's epileptic "epicenter" would be extremely

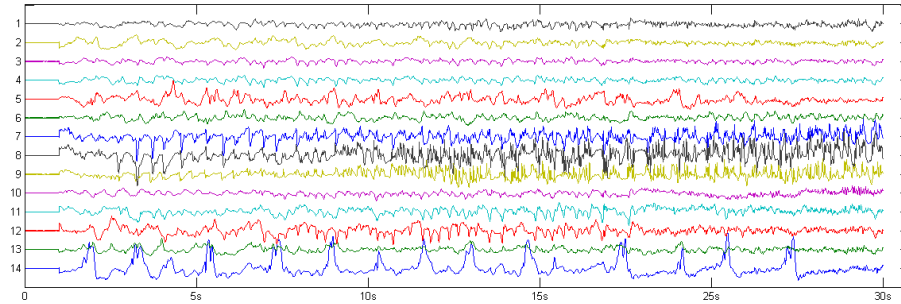


Figure 1.2: *ECoG data during an epileptic seizure: After 5 seconds an epileptic activity at channels 7, 8 and 9 is identified. (simultaneous oscillation).*

helpful.

1.2.2 Data basis

The ECoG recordings used in this diploma thesis, stem from three patients¹ (patient 1, patient 2 and patient 4) suffering from temporal lobe epilepsy. For patients 1 and 4, a number of 32 electrodes was used to find the focus, whereas for patient 2 only 28 electrodes were implanted.

ECoG signal recordings were made using a frequency of 256Hz and after filtering the line interference (50 Hz), the data was down-sampled to 128 Hz. As a reference value a not affected channel was chosen (far away from the focus). Every data point describes the potential difference between the chosen and the reference channel. For our analysis we use slightly adjusted time series, whereas the potential difference to the average data signal is measured for each data point².

Exemplary we want to give a detailed description of seizure 3 of patient 2, because this one is used often in the later analysis. Figure 1.3 gives a median view of the brain's surface, with circles indicating electrodes mounted on bands and marked with different colors.

The information provided by clinical experts (shown in Figure 1.3) might sound a bit cryptic to the reader. In fact, it means nothing else than

1. Seizure onset takes place at channels 26 and 27 at 12:31:41.
2. At 12:31:52 (11 seconds after the beginning) the epileptic activity propagates to the other half of the hemisphere and infects channels 10 and 11.
3. After the infection of channels 10 and 11, channel 12 shows epileptic activity, too.
4. At 12:32:39 (nearly one minute after the beginning) the seizure ends.

¹treated at Vienna General Hospital, Department of Neurology

²for a detailed discussion of common average data see Section 3.5.1.1

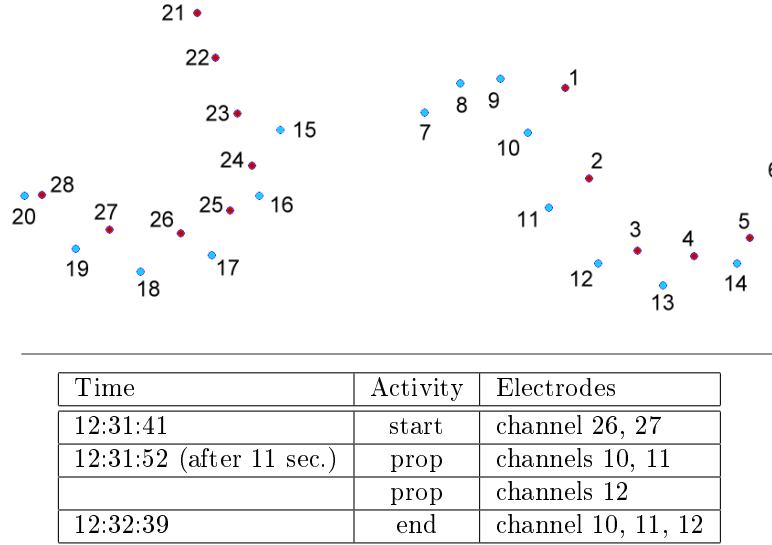


Figure 1.3: *Channel numbers of ECoG electrodes of patient 2: Upper plot visualizes the ECoG electrodes as mounted on the brain's surface. The left half of the plot corresponds with the right hemisphere and vice versa. Channel numbers used in the analysis are situated next to the electrode nodes. Furthermore the table under the plot shows all the information provided from clinical experts.*

Similar figures and recordings for all other seizures analyzed can be found in the Appendix.

1.2.3 Finding appropriate measures

The whole work is mainly based on the fact, that synchronization mechanisms play a major role during long-lasting epileptic activities. (Synaptic) effects have a positive feedback and lead to synchronization. Depending on the location of the temporal focus, further regions of the cortex might be affected. Thus obviously the localization of the initial focus and the tracking of its evolution are two coinciding activities.

In a more mathematical context those synchronization effects suggest the use of measures indicating causality. C.W. Granger introduced an appropriate measure, a lot of extensions have been applied and representations in time as well as in frequency domain have been derived.

1.3 Notation and Terminology

Before starting with any mathematical analysis, we have to clarify some expressions, that will be used in our work quite often.

We will always deal with discrete real-valued time series of length T consisting of K components. Thus, we may write for short

$$x_t \in \mathbb{R}^{T \times K} \text{ with } t = 1, \dots, T.$$

As most publications related to are established in the context of neurological signal treatment, the notation used here does not always coincide with that mathematicians are used to. To avoid confusion we will present a list of the most common terms.

1.3.1 Terminology

The following terminology will be used in our work repeatedly:

- We call any discrete real-valued time-series a **signal**.
- Each value of any given signal at a certain time, $x[n = n_0]$ is called a **sample**. Consequently, the index of the time axis is termed the **sample index** n .
- The frequency defined by

$$f_s = \frac{\text{number of samples}}{\text{length of the signal in seconds}}$$

is called the **sampling frequency**.

- Every component of any signal is called a **channel**, and its index is called **channel index**.

1.3.2 Conventions for Notation

In contrast to the last section, the reader with mathematical background will be quite familiar with the following conventions:

- **Sample indices:** when describing the k -th channel at time $n = n_0$, we will shortly write $x_k[n_0]$ or (to avoid confusion) even x_k .
- **Stochastic symbols:** As usual, the expected value is denoted by \mathbb{E} , the variance by \mathbb{V} and the variance-covariance-matrix by Σ . Estimated values will further be marked by $\hat{\cdot}$.
- $\overline{x_N}$ will denote the sample mean: $\overline{x_N} = \frac{1}{N} \sum_{i=1}^N x_i$
- Based on the different authors and the different approaches used it seemed more appropriate to use in each chapter a slightly different notation. In Chapter 3, matrices and vectors are denoted by bold letters, whereas in Chapter 2 it was decided not to use bold letters. We hope, that will not confuse the reader.

Chapter 2

Localizing the focus of epileptic seizures using measures in Frequency Domain

2.1 Frequency domain basics

Definition 2.1 (process with orthogonal increments). A stochastic process $(z(\lambda) \mid \lambda \in [-\pi, \pi])$ with random variables $z(\lambda) : \Omega \rightarrow \mathbb{C}^n$ is called a *process with orthogonal increments*, if¹

- (i) $z(-\pi) = 0, z(\pi) = x_o$ a.e.
- (ii) $\lim_{\epsilon \downarrow 0} z(\lambda + \epsilon) = z(\lambda)$ for $\lambda \in [-\pi, \pi)$
- (iii) $\mathbb{E}z(\lambda)^* z(\lambda) < \infty$ for all $\lambda \in [-\pi, \pi)$
- (iv) $\mathbb{E}(z(\lambda_4) - z(\lambda_3))(z(\lambda_2) - z(\lambda_1)) = 0$ for all $\lambda_1 < \lambda_2 \leq \lambda_3 < \lambda_4$.

Theorem 2.2 (Spectral Representation Theorem). *For every stationary process $(x[n])$, there exists a process with orthogonal increments $(z(\lambda) \mid \lambda \in [-\pi, \pi])$ such that²*

$$x[n] = \int_{-\pi}^{\pi} e^{i\lambda n} dz(\lambda) \quad \text{a.e.} \quad (2.1)$$

holds, where $i = \sqrt{-1}$ is the imaginary unit. The process $(z(\lambda))$ is a.e. uniquely determined from $(x[n])$.

Definition 2.3 (spectral distribution function). The function defined by

$$F(\lambda) = \mathbb{E}z(\lambda)z(\lambda)^*,$$

whereas $F : [-\pi, \pi] \rightarrow \mathbb{C}$, is called the *spectral distribution function*.

¹the symbol *l.i.m* denotes convergence in the mean squares sense

²“a.e.” means “almost everywhere”

Theorem 2.4. *For every covariance function $\gamma(n)$ of a stationary process $(x[n])$, there exists a spectral distribution function $F(\lambda)$ such that*

$$\gamma(n) = \int_{-\pi}^{\pi} e^{i\lambda n} dF(\lambda),$$

where $F(\lambda)$ is uniquely determined from $\gamma(n)$.

Proof. Without any loss of generality, assume that $\mathbb{E}x[n] = 0$. Thus, according to Theorem 2.2

$$\begin{aligned} \gamma(n) &= \mathbb{E}x[n]x[0]^* \\ &= \mathbb{E} \left(\int_{-\pi}^{\pi} e^{i\lambda n} dz(\lambda) \right) \left(\int_{-\pi}^{\pi} e^{i\lambda 0} dz(\lambda) \right)^* \\ &= \int_{-\pi}^{\pi} e^{i\lambda n} d(\mathbb{E}z(\lambda)z(\lambda)^*) \\ &= \int_{-\pi}^{\pi} e^{i\lambda n} dF(\lambda). \end{aligned}$$

A proof for the uniqueness can be found e.g. in [5]. □

Definition 2.5 (spectral density function). If a function $f : [-\pi, \pi] \rightarrow \mathbb{C}^{n \times n}$, such that

$$F(z) = \int_{-\pi}^z f(\lambda) d\lambda$$

exists at least almost everywhere, it is called the *spectral density function* of the stationary process $(x[n])$ ³.

Proposition 2.6. *If $(x[n])$ is a stationary process and $(z(\lambda))$ denotes the corresponding process with orthogonal increments, we have $(\lambda_1 < \lambda_2)$*

$$\begin{aligned} \int_{\lambda_1}^{\lambda_2} f(\lambda) d\lambda &= F(\lambda_2) - F(\lambda_1) \\ &= \mathbb{E} (z(\lambda_2) - z(\lambda_1)) (z(\lambda_2) - z(\lambda_1))^* \\ &= \mathbb{E} |z(\lambda_2) - z(\lambda_1)|^2. \end{aligned}$$

Thus, $\int_{\lambda_1}^{\lambda_2} f(\lambda) d\lambda$ measures the expected amplitude for the spectrum of frequencies between λ_1 and λ_2 .

Theorem 2.7. *If the stationary process $(x[n])$ has a spectral density⁴, then it is given by $(-\pi \leq \lambda \leq \pi)$*

$$f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-i\lambda n} \gamma(n)$$

³Obviously, not every stationary process has a spectral density function

⁴For example, according to [5], a stationary process with absolutely additive autocovariance has a spectral density

and the autocovariance function can be expressed by

$$\gamma(n) = \int_{-\pi}^{\pi} e^{i\lambda n} f(\lambda) d\lambda.$$

2.2 Measures derived from nonparametric spectrum estimators

2.2.1 Nonparametric estimation of the spectrum

2.2.1.1 Periodogram

The Periodogram, although introduced by A. Schuster [30] more than 100 years ago, is still one of the most common and established spectrum estimators in time series analysis. Clearly, modifications have been developed and will be discussed during this sections, but the main idea is still the same.

In this section, let $x[n]$ denote a scalar time series, with $n = 1, \dots, N$.

Definition 2.8. The Periodogram $I_N(\lambda)$ at one specified frequency λ is defined by

$$I_N(\lambda) = \frac{1}{N} \left| \sum_{n=1}^N x[n] e^{-i\lambda n} \right|^2, \quad (2.2)$$

where $i = \sqrt{-1}$ is the imaginary unit. In the following, we will mainly consider the so called Fourier frequencies

$$\lambda_j = \frac{2\pi j}{N}, \quad j = -\left\lfloor \frac{N-1}{2} \right\rfloor, \dots, 0, \dots, \left\lfloor \frac{N}{2} \right\rfloor.$$

Here, $\lfloor x \rfloor$ denotes the largest integer that is less than or equal to x . Therefore, for example, $\lfloor \frac{N}{2} \rfloor$ is either $\frac{N}{2}$ if N is even or $\frac{N-1}{2}$ for odd N .

The use of Fourier frequencies instead of arbitrarily chosen frequencies relieves some of the analysis⁵:

Theorem 2.9. *The complex N -vectors*

$$e_j = \frac{1}{\sqrt{N}} \begin{pmatrix} e^{-i\lambda_j 1} \\ e^{-i\lambda_j 2} \\ \vdots \\ e^{-i\lambda_j N} \end{pmatrix}, \quad j = -\left\lfloor \frac{N-1}{2} \right\rfloor, \dots, 0, \dots, \left\lfloor \frac{N}{2} \right\rfloor \quad (2.3)$$

form an orthonormal basis for \mathbb{C}^N .

⁵As a frequency's property of being a Fourier frequency depends only on the value of N , any fixed frequency can be approximated arbitrarily close by a Fourier frequency by simply increasing N . Thus, when talking about asymptotic ($N \rightarrow \infty$) properties, each frequency λ might be called a Fourier frequency.

Proof. Let us take a look at the scalar product:

$$\langle e_j, e_k \rangle = \frac{1}{N} \sum_{n=1}^N e^{-i\lambda_j n} e^{i\lambda_k n} = \frac{1}{N} \sum_{n=1}^N e^{i(\lambda_k - \lambda_j)n} \quad (2.4)$$

and therefore

$$\langle e_j, e_k \rangle = 1 \text{ if } j = k.$$

Furthermore,

$$\begin{aligned} \sum_{n=1}^N e^{i(\lambda_k - \lambda_j)n} &= \sum_{n=1}^N \left(e^{i(\lambda_k - \lambda_j)} \right)^n = e^{i(\lambda_k - \lambda_j)} \sum_{n=0}^{N-1} \left(e^{i(\lambda_k - \lambda_j)} \right)^n \\ &= e^{i(\lambda_k - \lambda_j)} \frac{1 - \left(e^{i(\lambda_k - \lambda_j)} \right)^N}{1 - e^{i(\lambda_k - \lambda_j)}} \\ &= 0 \end{aligned}$$

because $\left(e^{i(\lambda_k - \lambda_j)} \right)^N = e^{i(\lambda_k - \lambda_j)N} = e^{i2\pi(k-j)} = 1$. Thus, $\langle e_j, e_k \rangle = 0$ if $j \neq k$, which completes the proof. \square

Clearly, $I_N(\lambda_j) = |\langle e_j, X_N \rangle|^2$ holds for every Fourier frequency, where $X_N = (x[1], \dots, x[N])'$. Using the property of orthogonality proven above especially for $k = 0$,

$$\langle e_j, e_0 \rangle = 0 \quad j \neq 0,$$

this leads further to

$$I_N(\lambda_j) = |\langle e_j, X_N - c e_0 \rangle|^2, \quad (2.5)$$

for any $c \in \mathbb{C}$. In fact, all the entries of e_0 are the same: $e_0 = \frac{1}{\sqrt{N}} \mathbf{1}$, with $\mathbf{1} = (1, \dots, 1)'$. Thus, when choosing $c = \sqrt{N} \overline{x_N}$:

$$I_N(\lambda_j) = \frac{1}{N} \left| \sum_{n=1}^N e^{-i\lambda_j n} (x[n] - \overline{x_N}) \right|^2. \quad (2.6)$$

Based on (2.6), which is even used as the Periodogram's definition in some publications, it is quite easy to find an interpretation for the whole concept: the Periodogram is nothing else than a frequency dependent squared complex covariance (between x and an harmonic oscillation).

(2.6) yields

$$\begin{aligned} I_N(\lambda_j) &= \frac{1}{N} \left(\sum_{n=1}^N e^{i\lambda_j n} (x[n] - \overline{x_N}) \right) \left(\sum_{m=1}^N e^{-i\lambda_j m} (x[m] - \overline{x_N}) \right) \\ &= \frac{1}{N} \sum_{m,n=1}^N e^{-i\lambda_j(m-n)} (x[n] - \overline{x_N})(x[m] - \overline{x_N}). \end{aligned}$$

As, obviously, $x[m] = x[n + (m - n)]$, each additive part of the expression on the right hand side does not depend on the current value of m , but only on the

difference $(m - n)$ between m and n . Therefore, by substituting $u = m - n$ and rearranging the elements of the sum

$$\begin{aligned} I_N(\lambda_j) &= \frac{1}{N} \sum_{u=-N+1}^{N-1} \sum_{n=1}^{N-u} e^{-i\lambda_j u} (x[n] - \overline{x_N})(x[n+u] - \overline{x_N}) \\ &= \sum_{|u| < N} \hat{\gamma}(u) e^{-i\lambda_j u}. \end{aligned} \quad (2.7)$$

Thus the Periodogram $I_N(\lambda_j)$ for any Fourier frequency $\lambda_j \neq 0$ is the Fourier transform of the sample autocovariance function $\hat{\gamma}$. Because of the representation of the spectral density from theorem 2.7,

$$f(\lambda) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \gamma(s) e^{-i\lambda s},$$

the Periodogram (2.7) seems to be a suitable estimator for $2\pi f(\lambda_j)$. In fact, it is even an asymptotically unbiased estimator for the spectral density (for frequencies unequal to zero)⁶. Unfortunately, as the following example shows, the Periodogram is not a consistent estimator of the spectral density, i.e. its variance does not converge to zero.

Example 2.10. For simplicity let $x[n]$ be Gaussian white noise, i.e. each finite vector $x_c[n] = (x[n-c], \dots, x[n])'$ is normally distributed:

$$x_c[n] \sim N(0, \sigma^2 I). \quad (2.8)$$

The N vectors defined by

$$c_j = \sqrt{\frac{2}{N}} \begin{pmatrix} \cos \lambda_j 1 \\ \vdots \\ \cos \lambda_j N \end{pmatrix} \text{ and } s_j = \sqrt{\frac{2}{N}} \begin{pmatrix} \sin \lambda_j 1 \\ \vdots \\ \sin \lambda_j N \end{pmatrix}$$

for $1 \leq j \leq \frac{N}{2}$ and

$$c_k = \sqrt{\frac{1}{N}} \begin{pmatrix} \cos \lambda_k 1 \\ \vdots \\ \cos \lambda_k N \end{pmatrix}$$

for $k = 0$ and $k = \frac{N}{2}$ if $\frac{N}{2}$ is integer form an orthonormal basis for \mathbb{R}^N similar to the one defined for \mathbb{C}^N in (2.3). Thus,

$$\alpha_j = \langle c_j, X_N \rangle$$

and

$$\beta_j = \langle s_j, X_N \rangle$$

are independent random variables with zero mean and variance σ^2 .

⁶A proof can be found in Deistler [7] for example

Let λ_j be any Fourier frequency, different from 0 and π , in the following context.⁷ Furthermore frequencies larger than π or lower than zero need not be treated separately because of the spectral density's symmetry.

The Periodogram's definition (2.6) may now be expressed in terms of the scalar products α_j and β_j :

$$\begin{aligned}
\frac{\alpha_j^2 + \beta_j^2}{2} &= \frac{1}{2} \left(|\langle c_j, X_N \rangle|^2 + |\langle s_j, X_N \rangle|^2 \right) \\
&= \frac{1}{2} \left(\left| \sum_{n=1}^N \sqrt{\frac{2}{N}} x[n] \cos \lambda_j n \right|^2 + \left| \sum_{n=1}^N \sqrt{\frac{2}{N}} x[n] \sin \lambda_j n \right|^2 \right) \\
&= \frac{1}{N} \left| \sum_{n=1}^N x[n] \underbrace{(\cos \lambda_j n - i \sin \lambda_j n)}_{=e^{-i\lambda_j n}} \right|^2 \\
&= I_N(\lambda_j).
\end{aligned}$$

Thus, as $\frac{1}{\sigma} \alpha_j \sim N(0, 1)$ and therefore $\frac{1}{\sigma^2} \alpha_j^2 \sim \chi_1^2$ (and $\frac{1}{\sigma^2} \beta_j^2 \sim \chi_1^2$), it follows that the process's Periodogram is Chi-squared distributed with two degrees of freedom,

$$\frac{2}{\sigma^2} I_N(\lambda_j) \sim \chi_2^2.$$

Note, that, because the α_j and β_k are independent, $I_N(\lambda_j)$ and $I_N(\lambda_k)$ are independent too, i.e. $Cov(I_N(\lambda_j), I_N(\lambda_k)) = 0$ for $j \neq k$.

The expectation and the variance⁸ of the Periodogram can now be calculated easily⁹:

$$\mathbb{E} I_N(\lambda_j) = \frac{\sigma^2}{2} 2 = \sigma^2 \quad (= 2\pi f(\lambda_j))$$

and

$$\mathbb{V} I_N(\lambda_j) = \frac{\sigma^4}{4} 4 = \sigma^4 \quad (= 2\pi f(\lambda_j))^2.$$

Like we wanted to show, the Periodogram is an unbiased, but not a consistent estimator for the spectral density, because its variance does not converge to zero as $N \rightarrow \infty$. As the property of consistency is quite important in statistics, modifications have been made to the upper approach and will be presented during the following sections.

⁷Our example would apply for both frequencies too, but we would have to deal with them separately. For the sake of clarity, we will concentrate on Fourier frequencies unequal to 0 and π .

⁸if X is Chi-squared distributed with n degrees of freedom, its expectation is equal to n and its variance is equal to $2n$.

⁹Note, that the spectral density of white noise is constant and equal to $\frac{\sigma^2}{2\pi}$

2.2.1.2 Smoothed spectral estimates

Let $(x[n])$ be a scalar stationary process of the form

$$x[n] = \sum_{j=-\infty}^{\infty} b_j \varepsilon[n-j], \quad (2.9)$$

where

$$(\varepsilon[n]) \sim IID(0, \sigma^2) \text{ and } \sum_{j=-\infty}^{\infty} |b_j| \sqrt{|j|} < \infty. \quad (2.10)$$

Let $I_N(\lambda)$ denote the Periodogram of $(x[1], \dots, x[N])'$. If the spectral density $f(\lambda)$ of $(x[n])$ is greater than zero for every $\lambda \in [-\pi, \pi]$ and $0 < \lambda_1 < \dots < \lambda_m < \pi$, according to Brockwell & Davis [5], *the random vector $(I_N(\lambda_1), \dots, I_N(\lambda_m))'$ converges in distribution to a vector of independent and exponentially distributed random variables whose i -th component has mean $2\pi f(\lambda_i)$, $i = 1, \dots, m$.*

Using this result, we may write the Periodogram in the form

$$I_N(\lambda_j) = 2\pi f(\lambda_j) + u[j], \quad (2.11)$$

where the sequence $(u[j])$ is asymptotically white noise $WN(0, 1)$. The spectral density estimation problem therefore might be interpreted as a regression problem in the frequency domain. $f(\lambda_j)$ is nothing else than the trend, which we want to estimate. Hence, representation (2.11) suggests that we might reduce the Periodogram's variance by averaging over frequencies, i.e. smoothed Periodograms are used as an estimate for $f(\lambda)$.

Definition 2.11 (direct spectral estimate). A smoothed Periodogram

$$\hat{f}(\lambda_j) = \frac{1}{2\pi} \sum_{|u| \leq m_N} w_{u,N} I_N \left(\underbrace{\lambda_j - \frac{2\pi u}{N}}_{=: \lambda_{j-u}} \right) \quad (2.12)$$

is named a *direct spectral estimate*, if the so called *filter weights* $w_{u,N}$ are symmetric, i.e. $w_{u,N} = w_{-u,N}$, and non-negative. $I_N(\lambda)$ is the periodic extension of the Periodogram (2.2) from the interval $(-\pi, \pi]$ to \mathbb{R} .¹⁰

First of all, as

$$\mathbb{E} \hat{f}(\lambda_j) = \sum_{|u| \leq m_N} w_{u,N} \underbrace{\frac{1}{2\pi} \mathbb{E} I_N(\lambda_{j-u})}_{\xrightarrow{N \rightarrow \infty} f(\lambda_{j-u}) \approx f(\lambda_j)} \xrightarrow{N \rightarrow \infty} f(\lambda_j) \sum_{|u| \leq m_N} w_{u,N},$$

the direct spectral estimate in (2.12) is an asymptotically unbiased estimate, if

$$\sum_{|u| \leq m_N} w_{u,N} = 1 \quad (2.13)$$

¹⁰Therefore, $I_N \left(\frac{2\pi(T+1)}{T} \right) = I_N \left(\frac{2\pi}{T} \right)$ for example

and

$$\frac{m_N}{N} \xrightarrow{N \rightarrow \infty} 0 \quad (2.14)$$

because the spectral density is a continuous function (Together with (2.14) this assumptions guarantees, that $f(\lambda_{j-u}) = f(\lambda_j)$ holds asymptotically.). The infinite sum exists, because of (2.10).

Furthermore, if (2.10) holds and $\mathbb{E}\epsilon[n]^4 < \infty$, an asymptotic result for the covariance can be derived¹¹ ($\lambda_j, \lambda_k \geq 0$):

$$\text{Cov}(I_N(\lambda_j), I_N(\lambda_k)) \xrightarrow{N \rightarrow \infty} \begin{cases} 2(2\pi)^2 f(\lambda_j)^2 & \text{if } \lambda_j = \lambda_k \in \{0, \pi\} \\ (2\pi)^2 f(\lambda_j)^2 & \text{if } 0 < \lambda_j = \lambda_k < \pi \\ 0 & \text{if } \lambda_j \neq \lambda_k. \end{cases}$$

Thus, when using (2.20),

$$\begin{aligned} \mathbb{V}\hat{f}(\lambda_j) &= \sum_{|u| \leq m_N} w_{u,N}^2 \underbrace{\mathbb{V}\left(\frac{1}{2\pi} I_N(\lambda_{j-u})\right)}_{\rightarrow f^2(\lambda_j)} + \\ &\quad \sum_{|u|, |v| \leq m_N, u \neq v} w_{u,N} w_{v,N} \underbrace{\text{Cov}\left(\frac{1}{2\pi} I_N(\lambda_{j-u}), \frac{1}{2\pi} I_N(\lambda_{j-v})\right)}_{\rightarrow 0} \\ &\xrightarrow{N \rightarrow \infty} f^2(\lambda_j) \sum_{|u| \leq m_N} w_{u,N}^2, \end{aligned}$$

obviously the direct spectral estimate is a consistent estimate for the spectral density, if

$$\sum_{|u| \leq m_N} w_{u,N}^2 \rightarrow 0. \quad (2.15)$$

Let us state our results in a formal theorem:

Theorem 2.12 (asymptotic properties of smoothed spectral estimates). *Let $(x[n])$ be a stationary process of the form $x[n] = \sum_{j=-\infty}^{\infty} b_j \epsilon[n-j]$, where $(\epsilon[n])$ is white noise with variance σ^2 , $\sum_{j=-\infty}^{\infty} |b_j| \sqrt{|j|} < \infty$ holds¹² and $\mathbb{E}\epsilon[n]^4 < \infty$ is fulfilled. If $\hat{f}(\lambda)$ is a direct spectral estimate with $\frac{m_N}{N} \xrightarrow{N \rightarrow \infty} 0$, $\sum_{|u| \leq m_N} w_{u,N} = 1$ and $\sum_{|u| \leq m_N} w_{u,N}^2 \rightarrow 0$ then*

(i) \hat{f}_N is asymptotically unbiased, i.e. $\mathbb{E}\hat{f}_N(\lambda) = f(\lambda)$

$$(ii) \lim_{N \rightarrow \infty} \frac{\text{Cov}(\hat{f}_N(\lambda), \hat{f}_N(\omega))}{\sum_{|u| \leq m_N} w_{u,N}^2} = \begin{cases} 2f(\lambda)^2 & \text{if } \lambda = \omega \in \{0, \pi\} \\ f(\lambda)^2 & \text{if } 0 < \lambda = \omega < \pi \\ 0 & \text{else.} \end{cases}$$

Proof. A detailed proof of this theorem can be found in [5]. □

¹¹for a proof see e.g. [5]

¹²this condition holds especially for every ARMA process.

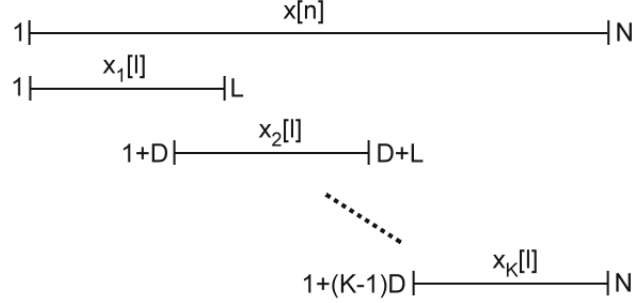


Figure 2.1: **Illustration of data segmentation.** The whole time series is splitted into overlapping segments of length L .

Nevertheless the presented concept here might be a bit abstract. Thus we want to consider the following example for demonstration:

Example 2.13. The simplest example possible for a direct spectral estimate is the Daniell estimate, which is in fact just a moving average of the Periodogram:

$$w_{u,N} = \begin{cases} \frac{1}{2m_N+1} & \text{for } |u| \leq m_N \\ 0 & \text{else.} \end{cases}$$

Clearly, the Daniell estimate is a consistent spectral density estimate according to Theorem 2.12 if $m_N \rightarrow \infty$ and $\frac{m_N}{N} \rightarrow 0$.

2.2.1.3 Welch's Method

The idea behind Welch's Method [33] is similar to that of smoothing in the previous section, but adds some kind of segmentation of the stochastic process $x[n]$, $n = 1, \dots, N$ to the basic concept.

For simplicity assume, that $x[n]$ is a stationary process with $\mathbb{E}x[n] = 0$ and spectral density $f(\lambda)$. Let λ_j again denote the Fourier frequencies.

Now take K segments of length L , that might overlap, but cover the whole record. Assume, that the difference between two starting points of a segment is always equal to D (Figure 2.1 illustrates the whole segmentation concept). If $x_k[l]$, $l = 1, \dots, L$; $k = 1, \dots, K$ describes the part of the stochastic process in the i -th segment, we get for any $1 \leq l \leq L$

$$\begin{aligned} x_1[l] &= x[l] \\ x_2[l] &= x[l + D] \\ &\vdots \\ x_K[l] &= x[l + (K - 1)D]. \end{aligned}$$

Next let us select a data window $W[l]$, $l = 1, \dots, L$ similar to the filter weights in Section 2.2.1.2, form sequences $x_k[l]W[l]$ for every $k = 1, \dots, K$ and transfer them into frequency domain. The resulting Fourier transforms $A_k(\lambda_n)$ are equal to

$$A_k(\lambda_n) = \frac{1}{L} \sum_{l=1}^L x_k[l]W[l]e^{-ikl\lambda_n}, \quad (2.16)$$

where $\lambda_n = \frac{2\pi n}{L}$, $n = 0, \dots, \frac{L}{2}$ in this context. Welch's K modified Periodograms are further defined by

$$I_K(\lambda_n) = \frac{L^2}{\sum_{l=1}^L W[l]^2} |A_k(\lambda_n)|^2 \quad (2.17)$$

and lead directly to

Definition 2.14 (*Welch's spectral estimate*). The average of Welch's modified Periodograms,

$$\hat{f}(\lambda_n) = \frac{1}{K} \sum_{k=1}^K I_K(\lambda_n), \quad (2.18)$$

is called *Welch's spectral estimate*.

As shown in [33], Welch's spectral estimate is just a special direct spectral estimate and therefore asymptotically unbiased and consistent.

For practical use, not only asymptotic properties are important - finite sample considerations are crucial too. Thus we want to analyze the variance of Welch's spectral estimate in detail now.

If we define¹³

$$d(j) = \text{Cov}(I_k(\lambda_n), I_{k+j}(\lambda_n)), \quad (2.19)$$

and consider the common formula for the variance of a sum of random variables X_i (a_i is an arbitrarily chosen real number),

$$\mathbb{V}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \mathbb{V}X_i + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_i a_j \text{Cov}(X_i, X_j), \quad (2.20)$$

which is simplified to

$$\mathbb{V}\left(\sum_{i=1}^n X_i\right) = n\gamma(0) + 2 \sum_{i=1}^{n-1} (n-i)\gamma(i) \quad (2.21)$$

if X is a stationary process, γ denotes its covariance function and $a_i = 1$. Then

¹³note, that the exact value of k does not matter as $(x[n])$ is a stationary process

an expression for the estimator's variance can be derived:

$$\begin{aligned}
\mathbb{V}\hat{f}(\lambda_n) &= \mathbb{V}\left(\frac{1}{K}\sum_{k=1}^K I_k(\lambda_n)\right) \\
&= \frac{1}{K^2}\left(Kd(0) + 2\sum_{j=1}^{K-1}(K-j)d(j)\right) \\
&= \frac{1}{K}\left(d(0) + 2\sum_{j=1}^{K-1}\frac{K-j}{K}d(j)\right).
\end{aligned}$$

After defining the correlation coefficient $\rho(j) = \frac{d(j)}{d(0)}$ we may further write

$$\mathbb{V}\hat{f}(\lambda_n) = \frac{d(0)}{K}\left(1 + 2\sum_{j=1}^{K-1}\frac{K-j}{K}\rho(j)\right). \quad (2.22)$$

Clearly, correlation is equal to zero if and only if the segments of $x[n]$ do not overlap, i.e. $D \geq L$. In that case, $\mathbb{V}\hat{f}(\lambda_n) = \frac{d(0)}{K} = \frac{Ld(0)}{N}$ and the variance converges to zero for $N \rightarrow \infty$. Therefore, non-overlapping segments are optimal, if N can be made sufficiently large.

Concluding, the best we can achieve by the use of segments which do not overlap, is a reduction of the variance by the factor $\frac{1}{K}$. In practice, the total number of points K cannot be made arbitrarily large, therefore an even greater reduction might be possible, if segments do in fact overlap. Welch suggests that they might overlap by exactly one half of their length, i.e. $D = \frac{L}{2}$.

Further analysis depends on the chosen data window $W[l]$. One of the reasonable choices¹⁴ is the shape $1 - t^2$, $-1 \leq t \leq 1$. When using that window, we get $\rho(1) \approx \frac{1}{9}$ and $\rho(j) = 0$ for any $j > 1$. Inserting all that assumptions into equation (2.22) yields

$$\mathbb{V}\hat{f}(\lambda_n) = \frac{d(0)}{K}\left(1 + 2\underbrace{\frac{K-1}{K}}_{=1-\frac{1}{K}}\frac{1}{9}\right) \approx \frac{11}{9}\frac{d(0)}{K},$$

because the second term $\frac{2d(0)}{9K^2}$ usually is insignificantly small. At first sight, the factor $\frac{11}{9}$ inflates the variance compared to the non-overlapping segments approach. However, an overall reduction in variance is achieved, because for non-overlapping segments,

$$K = \frac{N}{L/2} - 1 \approx \frac{2N}{L}$$

¹⁴another reasonable choice might be the Parzen spectral window, which has the shape $1 - |t|$, $-1 \leq t \leq 1$

instead of $K = \frac{N}{L}$ if segments do not overlap. Therefore,

$$\mathbb{V}\hat{f}(\lambda_n) = \frac{11}{18} \frac{d(0)L}{N}.$$

Thus, for fixed N and L , by overlapping a reduction of the variance by the factor $\frac{11}{18}$ was achieved.

2.2.2 Coherence

Coherence is a frequency dependent measure based on the concept of cross spectra between the components of multivariate time series. In this section, we shall consider only stationary bivariate processes $x[n] = (x_1[n], x_2[n])'$ with mean zero and covariances

$$\gamma_{i,j}(h) = \mathbb{E}x_i[n+h]x_j[n] \quad (2.23)$$

satisfying

$$\sum_{h=-\infty}^{\infty} |\gamma_{i,j}(h)| < \infty \quad \forall i, j = 1, 2. \quad (2.24)$$

As mentioned earlier, condition (2.24) guarantees the existence of the spectral densities f_{11} and f_{22} for the component processes $x_1[n]$ and $x_2[n]$. The matrix $\Gamma(h) = (\gamma_{i,j})_{i,j=1,2}$ is called the *covariance matrix function* of the time series $(x[n])$.

Definition 2.15 (cross spectrum). If $(x[n])$ is a stationary process like discussed above, the function

$$f_{12}(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma_{1,2}(h)$$

is called the *cross spectrum* of $x_1[n]$ and $x_2[n]$. The matrix $f(\lambda) = (f_{i,j}(\lambda))_{i,j=1,2}$ is called the *spectral density matrix* or the *spectrum* of the process $(x[n])$. Clearly,

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \Gamma(h).$$

Note, that the covariances $\gamma_{i,j}$ if $i \neq j$ (which is different to the case $i = j$) are not in general symmetric around zero. Therefore, the cross spectrum $f_{i,j}(\lambda)$ is typically a complex-valued function.

Let $(z_i(\lambda) \mid \lambda \in [-\pi, \pi])$ denote the process with orthogonal increments corresponding to the univariate component process $(x_i[n])$, where $i = 1, 2$, as discussed in the Spectral Representation Theorem 2.2. Thus,

$$x_i[n] = \int_{-\pi}^{\pi} e^{i\lambda n} dz_i(\lambda).$$

Furthermore, as shown in Proposition 2.6,

$$\int_{\lambda_1}^{\lambda_2} f_{i,i}(\lambda) d\lambda = \mathbb{E} |z_i(\lambda_2) - z_i(\lambda_1)|^2, \quad (2.25)$$

whereas $-\pi \leq \lambda_1 \leq \lambda_2 \leq \pi$. The cross spectrum $f_{i,j}(\lambda)$ has a similar representation, namely

$$\int_{\lambda_1}^{\lambda_2} f_{i,j}(\lambda) d\lambda = \mathbb{E} (z_i(\lambda_2) - z_i(\lambda_1)) \overline{(z_j(\lambda_2) - z_j(\lambda_1))'}. \quad (2.26)$$

>> From (2.26) we can see, that

$$f_{i,j}(\lambda) = \overline{f_{j,i}(\lambda)'} \text{ a.e.} \quad (2.27)$$

This implies, that the spectral density matrix $f(\lambda)$ is in fact a Hermitian matrix, i.e. $f(\lambda) = f(\lambda)^*$.

If the expression $dz_i(\lambda)$ denotes $z_i(\lambda_2) - z_i(\lambda_1)$, where $\lambda_1 < \lambda < \lambda_2$, the correlation between $dz_i(\lambda)$ and $dz_j(\lambda)$ is equal to

$$\begin{aligned} \text{Cor}(dz_i(\lambda), dz_j(\lambda)) &= \frac{\text{Cov}(dz_i(\lambda), dz_j(\lambda))}{\sqrt{\mathbb{E} dz_i(\lambda) \mathbb{E} dz_j(\lambda)}} \\ &= \frac{\mathbb{E} dz_i(\lambda) \overline{dz_j(\lambda)'}}{\sqrt{\mathbb{E} |dz_i(\lambda)|^2 \mathbb{E} |dz_j(\lambda)|^2}}. \end{aligned}$$

Making the difference $\lambda_2 - \lambda_1$ infinitesimally small while using representations (2.25) and Proposition 2.6 yields

$$\text{Cor}(dz_i(\lambda), dz_j(\lambda)) = \frac{f_{i,j}(\lambda)}{\sqrt{f_{i,i}(\lambda) f_{j,j}(\lambda)}},$$

which results in the following

Definition 2.16 (ordinary coherence). The *squared (ordinary) coherence* between two components $x_i[n]$ and $x_j[n]$ of a multivariate stationary process $(x[n])$ at one fixed frequency $\lambda \in [-\pi, \pi]$ is defined by

$$\chi_{i,j}^2(\lambda) = \frac{|f_{i,j}(\lambda)|^2}{\sqrt{f_{i,i}(\lambda) f_{j,j}(\lambda)}}. \quad (2.28)$$

According to (2.27), $\chi_{i,j}^2(\lambda) = \chi_{j,i}^2(\lambda)$ a.e..

The Cauchy-Schwarz equality ensures, that $0 \leq \chi_{i,j}^2(\lambda) \leq 1$, where a value of $\chi_{i,j}^2(\lambda)$ near 1 indicates a strong linear relationship between $dz_i(\lambda)$ and $dz_j(\lambda)$. Obviously, due to its symmetric definition, the ordinary coherence cannot be a directed measure.

Example 2.17. Let $x_1[n]$ be the single input and $x_2[n]$ the single output of a linear system. If the system has constant parameters only, the coherence $\chi_{1,2}^2(\lambda)$ will be equal to one. If $x_1[n]$ and $x_2[n]$ are completely unrelated, the coherence will be equal to zero. If $\chi_{1,2}^2(\lambda)$ is greater than zero, but less than one, either noise is entering the measurements, the function relating input and output is not really linear, or in fact more inputs are affecting $x_2[n]$.

In other words: the coherence of a linear system is nothing else than the fractional part of the output power that is affected by the input at one specified frequency.

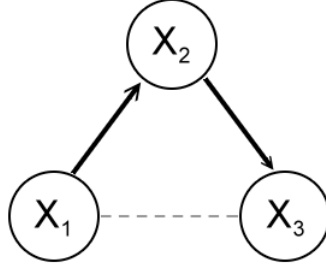


Figure 2.2: **Coherence and its inability to differ between direct and indirect influences.** When simulating any model, where channel 1 (represented by X_1) causes channel 2 and channel 2 further causes channel 3, according to the coherence's definition (2.28), ordinary coherence will always indicate coupling effects between channel 1 and 3, too.

2.2.3 Partial (spectral) coherence (PSC)

The ordinary coherence concept, which we presented during the previous section, is one of the least complicated concepts when analyzing coupling effects between components of multivariate time series. Unfortunately it simply does not work in many cases. One weakness of coherence is its inability to distinguish between direct and indirect influences (Figure 2.2 illustrates this problem). Dahlhaus' approach [6] to avoid the indication of indirect influences is partialization of time series.

In the following context, $x[n] = (x_1[n], \dots, x_k[n])'$ is a multivariate stationary time series. Our goal is to achieve only direct influences between two of the time series's components $x_i[n]$ and $x_j[n]$. The remaining components are denoted by

$$y_{i,j}[n] = \{x_k[n] \mid k \neq i, j\}.$$

First of all, we want to remove the linear effects of $y_{i,j}$ from the component $x_i[n]$. Thus, we set up a linear model, namely

$$x_i[n] = \sum_{r=-\infty}^{\infty} d_i[r] y_{i,j}[n-r] + \epsilon_i[n], \quad (2.29)$$

where $\epsilon_i[n]$ denotes the remaining (i.e. nonlinear) parts of $x_i[n]$. Obviously, our problem (i.e. withdrawing the linear effects optimally) is now equivalent to minimizing

$$\mathbb{E} \left(x_i[n] - \sum_{r=-\infty}^{\infty} d_i[r] y_{i,j}[n-r] \right)^2, \quad (2.30)$$

where optimal values of the filter $(d_i[r])$ are to be determined.

We will discuss two different cases to find a solution of the minimization problem (2.30). First, assume that our linear model is not an infinite one, i.e. only a finite number of values appears in equation (2.29). This assumption simplifies calculations and might even help to understand the proof of the second

(infinite) case, because methods are similar. Furthermore, Theorem 2.18 helps to understand Partial Granger causality in Section 3.3.

Theorem 2.18. *Let X (r -dimensional) and Y (s -dimensional) denote random variables with mean zero¹⁵ and covariance matrix*

$$\mathbb{E} \begin{pmatrix} X \\ Y \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}' = \begin{pmatrix} \mathbb{E}XX' & \mathbb{E}XY' \\ \mathbb{E}YX' & \mathbb{E}YY' \end{pmatrix} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma'_{xy} & \Sigma_{yy} \end{pmatrix}.$$

Then the optimal value for the $(r + s) \times s$ -matrix d , that is solving the minimization problem

$$\min_d \mathbb{E} (X - dY) (X - dY)' \quad (2.31)$$

is given by $d = \Sigma_{xy} \Sigma_{yy}^{-1}$.

Proof. After rearranging, we can find a different representation for our minimization problem (2.31):

$$\begin{aligned} \mathbb{E} (X - dY) (X - dY)' &= \\ &= \mathbb{E} (XX' - XY'd' - dYX' + dYY'd') \\ &= \Sigma_{xx} - \Sigma_{xy}d' - d\Sigma'_{xy} + d\Sigma_{yy}d' \\ &= \Sigma_{xx} - \Sigma_{xy}d' - \underbrace{d\Sigma_{yy}\Sigma_{yy}^{-1}\Sigma'_{xy}}_{=I_s} + d\Sigma_{yy}d' + \underbrace{\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma'_{xy} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma'_{xy}}_{=0} \\ &= \Sigma_{xx} + (d\Sigma_{yy} - \Sigma_{xy})d' - (d\Sigma_{yy} - \Sigma_{xy})\Sigma_{yy}^{-1}\Sigma'_{xy} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma'_{xy} \\ &= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma'_{xy} + (d\Sigma_{yy} - \Sigma_{xy})(d' - \Sigma_{yy}^{-1}\Sigma'_{xy}) \\ &= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma'_{xy} + (d\Sigma_{yy} - \Sigma_{xy})\Sigma_{yy}^{-1}(d\Sigma_{yy} - \Sigma_{xy})'. \end{aligned}$$

Thus obviously

$$\mathbb{E} (X - dY) (X - dY)' \geq \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma'_{xy},$$

where equality holds, if $d\Sigma_{yy} = \Sigma_{xy}$ and therefore $d = \Sigma_{xy}\Sigma_{yy}^{-1}$, which completes the proof. \square

We can now find the optimal solution for the complete optimization problem (2.30): The result of Theorem 2.19 is called the *Wiener filter formula*.

Theorem 2.19. *Let $(x[n])$ and $(y[n])$ denote stationary processes with mean zero and covariance functions $\gamma_{i,j}(h) = \mathbb{E}i[n+h]j[h]'$, where $i \in \{x, y\}$ and $j \in \{x, y\}$. Assume further, that all the covariance functions are absolutely additive. The optimal filter $(d_i[r])$ solving minimization problem (2.30) is then given by*

$$d_i[r] = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_{xy}(\lambda) f_{yy}(\lambda)^{-1} e^{ir\lambda} d\lambda,$$

where $f_{ij}(\lambda)$ are cross spectra, i.e. Fourier transforms of the corresponding covariance functions $\gamma_{i,j}$.

¹⁵if $\mathbb{E}X = \mu_x > 0$ and/or $\mathbb{E}Y = \mu_y > 0$, the theorem can be applied similarly to centered variables $X - \mu_x$ and $Y - \mu_y$

Proof. Again, we want to find a different representation for minimization problem (2.30). For simplicity, we will write $d[r]$ instead of $d_i[r]$ and neglect the expression ' (λ) ' for all frequency dependent terms:

$$\begin{aligned}
& \mathbb{E} \left(x[n] - \sum_{j=-\infty}^{\infty} d[j]y[n-j] \right) \left(x[n] - \sum_{k=-\infty}^{\infty} d[k]y[n-k] \right)' = \\
&= \gamma_{x,x}(0) - \sum_{j=-\infty}^{\infty} \gamma_{x,y}(j)d'_j - \sum_{j=-\infty}^{\infty} d_j \gamma_{y,x}(-j) + \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_j \gamma_{y,y}(k-j)d'_k \\
&= \int_{-\pi}^{\pi} f_{xx} - f_{xy} \sum_{j=-\infty}^{\infty} e^{i\lambda j} d'_j - \underbrace{\sum_{j=-\infty}^{\infty} d_j e^{-i\lambda j} f_{yx}}_{=:A} + \sum_{j,k=-\infty}^{\infty} d_j e^{i\lambda(k-j)} f_{yy} d'_k d\lambda \\
&= \int_{-\pi}^{\pi} f_{xx} - f_{xy} \bar{A} - A f_{yx} + A f_{yy} \bar{A} - \underbrace{f_{xy} f_{yy}^{-1} f_{yx} + f_{xy} f_{yy}^{-1} f_{yx}}_{=0} d\lambda \\
&= \int_{-\pi}^{\pi} f_{xx} - f_{xy} f_{yy}^{-1} f_{yx} - f_{xy} (\bar{A} - f_{yy}^{-1} f_{yx}) - A f_{yy} (f_{yy}^{-1} f_{yx} - \bar{A}) d\lambda \\
&= \int_{-\pi}^{\pi} f_{xx} - f_{xy} f_{yy}^{-1} f_{yx} + (A f_{yy} - f_{xy}) (\bar{A} - f_{yy}^{-1} f_{yx}) d\lambda \\
&= \int_{-\pi}^{\pi} f_{xx} - f_{xy} f_{yy}^{-1} f_{yx} + \underbrace{(A f_{yy} - f_{xy}) f_{yy}^{-1} (A f_{yy} - f_{xy})'}_{\geq 0} d\lambda.
\end{aligned}$$

Thus, an optimum is given, if the underbraced expression on the right side is equal to 0, i.e. $A f_{yy} - f_{xy} = 0$ and therefore

$$A(\lambda) = f_{xy}(\lambda) f_{yy}(\lambda)^{-1}. \quad (2.32)$$

As $A(\lambda)$ was defined by $\sum_{j=-\infty}^{\infty} d[j]e^{-i\lambda j}$, the spectral representation theorem (2.7) yields the optimal values for the linear filter:

$$d[j] = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda) e^{ij\lambda} d\lambda. \quad (2.33)$$

□

>> From now on, $\epsilon_i[n]$ denotes the remainder of the linear model, i.e.

$$\epsilon_i[n] = x_i[n] - \sum_{r=-\infty}^{\infty} d_i[r] y_{i,j}[n-r] \quad (2.34)$$

(clearly, $\epsilon_j[n]$ is defined in the same way), where values for the filter $d_i[r]$ are chosen optimally, like discussed in Theorem 2.18. $\epsilon_i[n]$ contains all the information from the time series component $x_i[n]$, but influences from other components (i.e. from $y_{i,j}[n]$) have been removed, which makes the following definition obvious.

Definition 2.20 (partial covariance and partial correlation). The *partial covariance* between two components $x_i[n]$ and $x_j[n]$ of a stationary time series is defined by the ordinary covariance between $x_i[n]$ and $x_j[n]$ after all the influences from the remaining components $y_{i,j}[n]$ have been removed, i.e.

$$\text{Cov}(x_i[n], x_j[n] \mid y_{i,j}[n]) \triangleq \text{Cov}(\epsilon_i[n], \epsilon_j[n]). \quad (2.35)$$

As correlation is just a normalized covariance, the *partial correlation* between $x_i[n]$ and $x_j[n]$ is therefore equal to

$$r_{i,j|y}^2 \triangleq \frac{\text{Cov}(\epsilon_i[n], \epsilon_j[n])}{\sqrt{\text{Var}(\epsilon_i[n])\text{Var}(\epsilon_j[n])}}. \quad (2.36)$$

The coherence definition (2.28) was directly based on the concept of cross spectra, which are Fourier transforms of corresponding covariance function. Partial coherence will be defined similarly. Let first denote

$$\gamma_{i,j|y}(h) = \text{Cov}(\epsilon_i[n+h], \epsilon_j[n])$$

the partial covariance function between the two components $x_i[n]$ and $x_j[n]$ of our stationary time series. If $\gamma_{i,j|y}$ is absolutely additive, i.e. $\sum_{h=-\infty}^{\infty} |\gamma_{i,j|y}(h)| < \infty \forall i, j$, the so called *partial cross spectrum* between $x_i[n]$ and $x_j[n]$ exists and is defined by

$$f_{i,j|y}(\lambda) \triangleq \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{i,j|y}(h) e^{-i\lambda h} \quad (2.37)$$

like discussed in Theorem 2.7. Clearly, this is exactly the cross spectrum between the two residuals $\epsilon_i[n]$ and $\epsilon_j[n]$ from equation (2.34). Rescaling leads to the partial coherence.

Definition 2.21 (Partial coherence). The (*squared*) *partial (spectral) coherence (PSC)* between two components $x_i[n]$ and $x_j[n]$ of a stationary time series is defined by the normalization of the partial spectrum (2.37), i.e.

$$R_{i,j|y}^2(\lambda) \triangleq \frac{|f_{i,j|y}^2(\lambda)|}{f_{i,i|y}(\lambda)f_{j,j|y}(\lambda)} \quad (2.38)$$

for any frequency $\lambda \in [-\pi, \pi]$. Again, partial coherence is a symmetric and therefore undirected measure for coupling effects between components of multivariate time series. Like ordinary coherence it is bounded by 0 and 1.

2.2.4 A simple signal model

Before discussing further improvements for the calculation of partial coherences, we want to look at a simple signal model. It will demonstrate some of the coherence's problems during the previous sections and will show, that partial coherences in fact might be a solution.

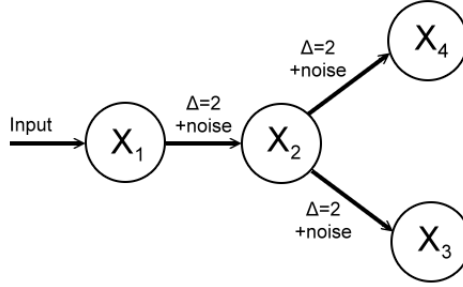


Figure 2.3: The signal model.

Our signal model is shown in Figure 2.3. We arbitrarily picked the first 2000 observations of one of the ECoG signals from the results section and defined it the input signal X_1 of the signal model. Furthermore, every arrow describes a shift of the data vector to the right for two times. Thus, for example

$$\tilde{X}_2[1] = \tilde{X}_2[2] = 0$$

and

$$\tilde{X}_2[i] = X_1[i - 2] \text{ for } i > 2.$$

After shifting the signal, white noise of approximately half of the signal's amplitude is added to \tilde{X}_2 . The remaining time series is called X_2 . Executing the same procedure with X_2 as input signal yields X_3 and X_4 .

Of course, we cannot expect arrows to be a result from any analysis using PSC or ordinary coherence. They both are not directed measures and therefore can indicate only interactions between two channels.

As expected and shown in Figure 2.4, ordinary coherence cannot distinguish between direct and indirect influences. Thus, it indicates interactions between channels 1 and 3 and between channels 1 and 4. Unfortunately, it also suggests that there is interaction between channels 3 and 4, which is in contrast to the model's definition. Of course, both signals are similar in the way that they both are derived from channel 2 by shifting and adding noise, which leads to a high correlation between them. However, it is not satisfiable that this similarity yields our measure to indicate interactions between the two channels.

Fortunately, partial spectral coherence meets all the expectations. As shown in Figure 2.5, it only indicates direct interactions between channels that in fact do interact. PSC therefore is a definitive improvement compared to ordinary coherence.

2.2.5 Computational formula for the partial coherence

Without any doubt, partial coherence is a useful tool in numerous different applications. On the other hand, it is quite computation-intensive to perform all the linear optimization problems and Fourier transformations necessary to obtain partial cross spectra as defined above. Therefore, Dahlhaus [6] developed

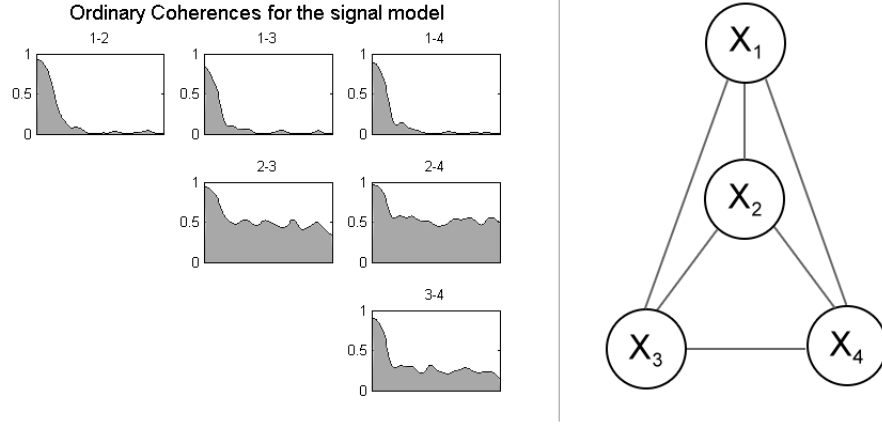


Figure 2.4: *The signal model and ordinary coherences.* Analyzing pairwise interactions between different channels of our signal model. Ordinary coherences indicate coupling effects between any pair of channels, which is in contrast to the model's definition.

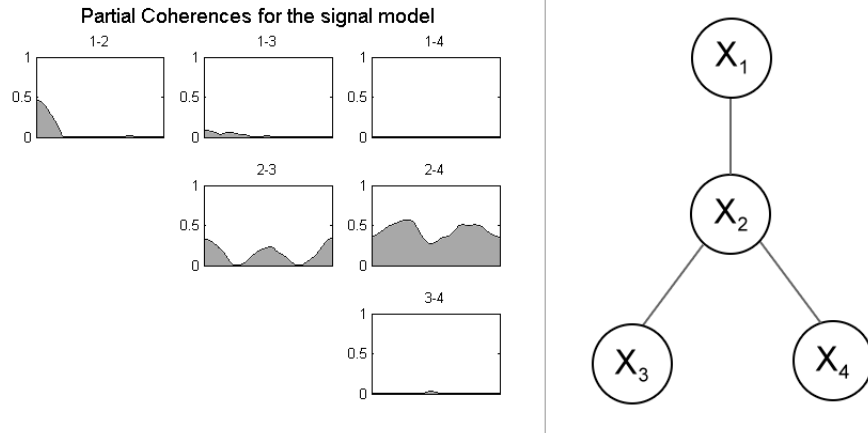


Figure 2.5: *The signal model and partial spectral coherences.* Analyzing pairwise interactions between different channels of our signal model. Partial spectral coherence indicates only coupling effects between pairs of channel if they are in accordance with the signal model's definition.

a simple computational formula to calculate partial coherences.

First of all, we need a different representation for the partial cross spectral density $f_{ij|y}(\lambda)$.

Theorem 2.22. *The partial cross spectral density (2.37) can be obtained by calculating*

$$f_{ij|y}(\lambda) = f_{ij}(\lambda) - f_{i,y}(\lambda)f_{yy}(\lambda)^{-1}f_{y,j}(\lambda),$$

where $f_{i,y}$ is the cross spectral density between $x_i[n]$ and $y_{i,j}[n]$.

Proof. The proof is quite similar to the one shown for the Wiener filter formula in Theorem 2.19. Again, for simplicity we will omit the expression ' (λ) '. We have

$$\begin{aligned} \gamma_{i,j|y}(h) &= \mathbb{E}\epsilon_i[h]\epsilon_j[0]' \\ &= \mathbb{E}\left(x_i[h] - \sum_{r=-\infty}^{\infty} d_i[r]y_{i,j}[h-r]\right)\left(x_j[0] - \sum_{s=-\infty}^{\infty} d_j[s]y_{i,j}[-s]\right)' \\ &= \underbrace{\mathbb{E}x_i[h]x_j[0]'}_{=\gamma_{i,j}(h)} - \sum_{s=-\infty}^{\infty} \underbrace{\mathbb{E}x_i[h]y_{i,j}[-s]'}_{=\gamma_{i,y}(h+s)} d_j[s]' - \sum_{r=-\infty}^{\infty} d_i[r] \underbrace{\mathbb{E}y_{i,j}[h-r]x_j[0]'}_{=\gamma_{y,j}(h-r)} \\ &\quad + \sum_{r,s=-\infty}^{\infty} d_i[r] \underbrace{\mathbb{E}y_{i,j}[h-r]y_{i,j}[-s]'}_{=\gamma_{yy}(h-r+s)} d_j[s]' \\ &= \int_{-\pi}^{\pi} e^{i\lambda h} \left(f_{ij} - f_{i,y} \sum_{s=-\infty}^{\infty} e^{i\lambda s} d_j[s]' - \sum_{r=-\infty}^{\infty} e^{-i\lambda r} d_i[r] f_{y,j} \right. \\ &\quad \left. + \left(\sum_{r=-\infty}^{\infty} d_i[r] e^{-i\lambda r} \right) f_{yy} \left(\sum_{s=-\infty}^{\infty} e^{i\lambda s} d_j[s] \right)' \right) \end{aligned}$$

According to the Wiener filter formula proven before,

$$\sum_{r=-\infty}^{\infty} d_i[r] e^{-i\lambda r} = f_{i,y} f_{yy}^{-1}$$

and

$$\sum_{s=-\infty}^{\infty} d_j[s] e^{-i\lambda s} = f_{j,y} f_{yy}^{-1} \triangleq A_j.$$

Thus,

$$\begin{aligned} \gamma_{i,j|y}(h) &= \int_{-\pi}^{\pi} e^{i\lambda h} \left(f_{ij} - f_{i,y} \overline{A_j'} - f_{i,y} f_{yy}^{-1} f_{y,j} + f_{i,y} f_{yy}^{-1} f_{yy} \overline{A_j'} \right) d\lambda \\ &= \int_{-\pi}^{\pi} e^{i\lambda h} (f_{ij} - f_{i,y} f_{yy}^{-1} f_{y,j}) d\lambda, \end{aligned}$$

which finishes our proof, because $\gamma_{i,j|y}(h) = \int_{-\pi}^{\pi} e^{i\lambda h} f_{i,j|y} d\lambda$ and therefore $f_{i,j|y} = f_{ij} - f_{i,y} f_{yy}^{-1} f_{y,j}$ a.e.

□

Using this result will finally help to prove the computational formula. It shows, that partial coherences are in fact just negative values of the rescaled inverse spectrum.

Theorem 2.23 (computational formula for partial coherences). *Let $(x[n])$ denote a k -dimensional stationary process with existing cross spectra $f_{ij}(\lambda)$, $i, j = 1 \dots k$. The matrix $f_{xx}(\lambda)$ is defined by $(f_{ij}(\lambda))_{i,j=1 \dots k}$. Assume, that f_{xx} has full rank, then $g(\lambda) = f_{xx}^{-1}(\lambda)$ is its inverse. Let further denote*

$$d(\lambda) = \begin{pmatrix} g_{11}(\lambda)^{-\frac{1}{2}} & & 0 \\ & \ddots & \\ 0 & & g_{kk}(\lambda)^{-\frac{1}{2}} \end{pmatrix} g(\lambda) \begin{pmatrix} g_{11}(\lambda)^{-\frac{1}{2}} & & 0 \\ & \ddots & \\ 0 & & g_{kk}(\lambda)^{-\frac{1}{2}} \end{pmatrix}.$$

The partial coherence $R_{i,j|y}(\lambda)$ can then be calculated by $R_{i,j|y}(\lambda) = -d_{ij}(\lambda)$ for any $i \neq j$.

Proof. Without loss of generality assume, that $i = 1$ and $j = 2$. The stationary process $x[n]$ might be written as $(x_1[n], x_2[n], z[n])$ ¹⁶. This notation yields¹⁷

$$f_{xx} = \begin{pmatrix} \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix} & f_{z,(2,1)} \\ f_{z,(1,2)} & f_{zz} \end{pmatrix}, \quad (2.39)$$

where $f_{z,(1,2)} = \begin{pmatrix} f_{z1} \\ f_{z2} \end{pmatrix}$ is the cross spectrum between x_1 and the rest. $f_{(1,2),z}$ is defined in the same way. As the inverse of any regular matrix of the form

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

is given by

$$B^{-1} = \begin{pmatrix} E^{-1} & -E^{-1}B_{12}B_{22}^{-1} \\ -B_{22}^{-1}B_{21}E^{-1} & B_{22}^{-1} + B_{22}^{-1}B_{21}E^{-1}B_{12}B_{22}^{-1} \end{pmatrix}$$

with $E = (B_{11} - B_{12}B_{22}^{-1}B_{21})^{-1}$, we can find a representation for the inverse of the spectral matrix f_{xx} :

$$f_{xx}^{-1} = \begin{pmatrix} E^{-1} & -E^{-1}f_{(1,2),z}f_{zz}^{-1} \\ -f_{zz}^{-1}f_{z,(1,2)}E^{-1} & f_{zz}^{-1} + f_{zz}^{-1}f_{z,(1,2)}E^{-1}f_{(1,2),z}f_{zz}^{-1} \end{pmatrix}, \quad (2.40)$$

where

$$\begin{aligned} E &= \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix} - f_{(1,2),z}f_{zz}^{-1}f_{z,(1,2)} \\ &= \begin{pmatrix} f_{11} - f_{1z}f_{zz}^{-1}f_{z1} & f_{12} - f_{1z}f_{zz}^{-1}f_{z2} \\ f_{21} - f_{2z}f_{zz}^{-1}f_{z1} & f_{22} - f_{2z}f_{zz}^{-1}f_{z2} \end{pmatrix} \end{aligned} \quad (2.41)$$

and, according to Cramer's rule, the inverse of matrix E is given by

$$E^{-1} = \frac{1}{E_{11}E_{22} - E_{12}E_{21}} \begin{pmatrix} E_{22} & -E_{12} \\ -E_{21} & E_{11} \end{pmatrix}. \quad (2.42)$$

¹⁶During this proof, we are only interested in the first two time series components, thus z simply denotes the rest

¹⁷we will abandon from writing (λ) again

We have now finished all the necessary preparations and can finally prove the statement we wanted to prove:

$$\begin{aligned}
d &= \begin{pmatrix} (f_{xx})_{11}^{-\frac{1}{2}} & 0 \\ & (f_{xx})_{22}^{-\frac{1}{2}} \\ 0 & \ddots \end{pmatrix} f_{xx}^{-1} \begin{pmatrix} (f_{xx})_{11}^{-\frac{1}{2}} & 0 \\ & (f_{xx})_{22}^{-\frac{1}{2}} \\ 0 & \ddots \end{pmatrix} \\
&= \begin{pmatrix} 1 & \frac{(f_{xx}^{-1})_{12}}{\sqrt{(f_{xx})_{11}(f_{xx})_{22}}} & \times \\ \frac{(f_{xx}^{-1})_{21}}{\sqrt{(f_{xx})_{11}(f_{xx})_{22}}} & 1 & \times \\ \times & \times & \times \end{pmatrix} \\
&= \begin{pmatrix} 1 & \frac{(f_{xx}^{-1})_{12}}{\sqrt{(f_{xx})_{11}(f_{xx})_{22}}} & \times \\ \frac{(f_{xx}^{-1})_{21}}{\sqrt{(f_{xx})_{11}(f_{xx})_{22}}} & 1 & \times \\ \times & \times & \times \end{pmatrix} \\
&= \begin{pmatrix} 1 & \frac{-E_{12}}{\sqrt{E_{11}E_{22}}} & \times \\ \frac{-E_{21}}{\sqrt{E_{11}E_{22}}} & 1 & \times \\ \times & \times & \times \end{pmatrix}, \tag{2.43}
\end{aligned}$$

where the symbol ' \times ' characterizes matrix entries we are not interested in during this proof. Combining (2.43) with the expression for E (2.41) and the results from Theorem 2.22 finally yields the desired results:

$$d_{12} = \frac{-E_{21}}{\sqrt{E_{11}E_{22}}} = -\frac{f_{12} - f_{1z}f_{zz}^{-1}f_{z2}}{\sqrt{(f_{11} - f_{1z}f_{zz}^{-1}f_{z1})(f_{22} - f_{2z}f_{zz}^{-1}f_{z2})}} = -R_{12|z}$$

and analog for d_{21} . □

2.3 Measures derived from parametric spectrum estimators

Deriving measures from parametric spectra is a quite common approach in literature and can be found in numerous publications (e.g. [20, 21, 26]). Its main idea is to fit an autoregressive (AR) model to preselected EEG/ECOG data, to transform the estimated AR-coefficients into frequency domain and then to calculate dependency measures, which are based on these coefficients.

2.3.1 Parametric estimation of the spectrum

Two major steps are necessary to compute parametric spectrum estimators:

2.3.1.1 Setting up a parametric model

In this section, we will always consider a linear approach, i.e. an AR model with order p . We therefore assume, that the value of the signal x at moment n depends only on the p previous values and a random component, i.e. it especially does not depend on the signal's present or future values.

We further assume, that K channels are left after channel preselection (channel selection will be discussed in Section 2.4.3). At each moment n , we will denote those channels by a K -dimensional vector $x[n] = (x_1[n], \dots, x_K[n])'$, such that the fitted AR(p) model can then be expressed as

$$x[n] = \sum_{j=1}^p A_j x[n-j] + \epsilon[n], \quad (2.44)$$

where the $K \times K$ -matrices A_j are the AR model's coefficients and the K -dimensional vector $\epsilon[n]$ represents white noise with mean zero and covariance matrix Σ_ϵ . Note, that the AR coefficient A_0 is set to identity in the upper definition.

As result this first step delivers the model's p estimated coefficient matrices. They will be represented by \hat{A}_j in the following context.

Despite its simplicity this approach is widely and successfully used in many different applications and especially in the analysis of EEG/ECOG signals (we will ourselves use a similar approach again in Chapter 3, when dealing with Granger causalities).

Unfortunately, using an autoregressive model implies the assumption of stationarity and therefore a constant variance. At our application, obviously, the signal's variance at least changes when the seizure starts and ends. Furthermore, Graef showed [12], that the assumption of stationarity does not even hold in short windows. Nevertheless, we will use this approach because numerous publications (for example [10, 21]) detected, that it might be sufficient for the approximation of EEG/ECOG signals and works quite well in practice.

2.3.1.2 Transformation into frequency domain

We may write the definition of our AR model (2.44) in a different, equivalent way, by using the backshift operator z , which is defined as $zx[n] = x[n-1]$:

$$x[n] = \sum_{j=1}^p A_j z^j x[n] + \epsilon[n].$$

After moving $x[n]$ completely to the equation's left hand side, we further receive

$$\left(I - \sum_{j=1}^p A_j z^j \right) x[n] = \epsilon[n]. \quad (2.45)$$

When denoting the expression between the brackets by $A(z)$, we receive the standard *Z-transformation* of our AR model, which holds in fact for every $z \in \mathbb{C}$, such that

$$\det A(z) \neq 0 \quad \forall |z| \leq 1. \quad (2.46)$$

According to Deistler [7] or Brockwell and Davis [5], this assumption guarantees stationarity of our model (furthermore it also assures the invertibility of $A(z)$, which will be used in the following context).

Transformation from Z-domain to frequency domain can now be managed by setting $z = e^{-i\lambda}$, where $\lambda \in [-\pi, \pi]$ represents the (angular) frequencies¹⁸ and i is the imaginary unit. Like in Z-domain, we will denote

$$A(\lambda) = I - \sum_{j=1}^p A_j e^{-i\lambda j}. \quad (2.47)$$

Therefore, we obtain the frequency domain representation of equation (2.45):¹⁹

$$A(\lambda)X(\lambda) = E(\lambda),$$

and further

$$X(\lambda) = H(\lambda)E(\lambda),$$

where the complex-valued matrix $H(\lambda) = A(\lambda)^{-1}$ is called the *transfer function* of our autoregressive model.

We can now easily calculate the *power spectrum* $S(\lambda)$ (a proof can be found in Brockwell and Davis [5] for example):

$$S(\lambda) = H(\lambda)\Sigma_\epsilon H^*(\lambda), \quad (2.48)$$

where $*$ denotes the transposed of the complex conjugate. Note, that the variance of noise, Σ_ϵ , does not depend on the frequency in our model.

¹⁸In literature, frequencies are often represented by the expression $2\pi f$, $f \in [0, 1]$. Obviously, both representations are equivalent.

¹⁹Fourier transformed vectors are set in capital letters like matrices. However, differences should be quite obvious.

Finally, we obtain the desired *estimated power spectrum* $\hat{S}(\lambda)$, by inserting the estimated model coefficients \hat{A}_j from the first step into equation (2.48). When talking about measures derived from nonparametric spectra in this section, we always talk about measures that are derived either from $\hat{S}(\lambda)$, from the estimated transfer function or directly from the estimated model coefficients \hat{A}_j .

2.3.2 Ordinary coherence

Definition of coherence is similar to the one we already discussed exhaustively in Section 2.2.2. A parametric estimator for cross-spectral densities has already been derived in the last section, because the non-diagonal elements of the estimated power spectrum are in fact nothing else than cross-spectral densities of the signals $x_i(n)$, $i = 1, \dots, K$.²⁰

$$S(\lambda) = \begin{pmatrix} S_{11}(\lambda) & S_{12}(\lambda) & \cdots & S_{1K}(\lambda) \\ S_{21}(\lambda) & S_{22}(\lambda) & \cdots & S_{2K}(\lambda) \\ \vdots & \vdots & \ddots & \vdots \\ S_{K1}(\lambda) & S_{K2}(\lambda) & \cdots & S_{KK}(\lambda) \end{pmatrix}.$$

So (squared) ordinary coherence between two channels i and j is defined by

$$C_{ij}(\lambda) = \frac{|S_{ij}(\lambda)|^2}{S_{ii}(\lambda)S_{jj}(\lambda)}.$$

For frequencies λ , where $C_{ij}(\lambda)$ is high, interactions between the two signals X_i and X_j can be modeled using a linear operator T_{ij} , so that

$$X_i(\lambda) = T_{ij}X_j(\lambda),$$

and further (because $T_{ji} = T_{ij}^{-1}$)

$$X_j(\lambda) = T_{ij}^{-1}X_i(\lambda).$$

Thus, obviously, ordinary coherence is an undirected measure again.

Unfortunately, high coherence does not necessarily explain the development of epileptic seizures. Two neighboring electrodes might show similar behavior over time, because distance between them is small and not because of epileptic activity.

Surprisingly, although ordinary coherence has been a quite common approach in literature for years, only little attention has been given to its evolution. Baccalá and Sameshima determined a new concept in neural structure determination in 2001 [3], which is based on the decomposition of the interactions between two signals into a “feedback” and a “feedforward” effect and finally leads to a measure called partial directed coherence. We will discuss the necessary steps in detail during the following sections.

²⁰For the sake of clarity, we will denote estimated values of power spectra without using the symbol $\hat{\cdot}$.

2.3.3 The partial coherence function

Before starting with the desired decomposition of interactions between two signals, we first have to establish a definition similar to partial spectral coherence from Section 2.2.3. According to the *Computational formula for the partial coherence* (2.23), we have already proven before that the partial spectral coherences can be obtained as the negative rescaled inverse of the power spectrum matrix. Thus,

$$R_{X_i X_j | Y_{i,j}}^2 = \left| \frac{g_{ij}(\lambda)}{\sqrt{g_{ii}(\lambda)g_{jj}(\lambda)}} \right|^2, \quad i \neq j, \quad (2.49)$$

where $g(\lambda) = f_{xx}(\lambda)^{-1}$ represents the nonparametric estimation of the inverted power spectrum.

Using the parametric approach, the inverse spectrum can easily be derived from (2.48):

$$\begin{aligned} S(\lambda)^{-1} &= (H(\lambda)\Sigma_\epsilon H^*(\lambda))^{-1} \\ &= H^*(\lambda)^{-1}\Sigma_\epsilon^{-1}H(\lambda)^{-1} \\ &= A^*(\lambda)\Sigma_\epsilon^{-1}A(\lambda). \end{aligned}$$

Therefore element (i, j) of the inverse spectrum is

$$[S(\lambda)^{-1}]_{i,j} = a_{\bullet i}^*(\lambda)\Sigma_\epsilon^{-1}a_{\bullet j}(\lambda), \quad (2.50)$$

where $a_{\bullet j}(\lambda)$ denotes the j -th column of the matrix $A(\lambda)$.

Inserting the resulting expression for the elements of the inverse spectrum (2.50) into the computational formula for partial coherence (2.49) leads to the following representation of partial spectral coherence:

$$\left| \frac{a_{\bullet i}^*(\lambda)\Sigma_\epsilon^{-1}a_{\bullet j}(\lambda)}{\sqrt{a_{\bullet i}^*(\lambda)\Sigma_\epsilon^{-1}a_{\bullet i}(\lambda)}\sqrt{a_{\bullet j}^*(\lambda)\Sigma_\epsilon^{-1}a_{\bullet j}(\lambda)}} \right|^2.$$

Now, we can summarize this derivation with the following

Definition 2.24 (partial coherence function). The *partial coherence function* is given by²¹ $\kappa_{i,j}(\lambda) = \frac{a_{\bullet i}^*(\lambda)\Sigma_\epsilon^{-1}a_{\bullet j}(\lambda)}{\sqrt{a_{\bullet i}^*(\lambda)\Sigma_\epsilon^{-1}a_{\bullet i}(\lambda)}\sqrt{a_{\bullet j}^*(\lambda)\Sigma_\epsilon^{-1}a_{\bullet j}(\lambda)}}$ for any $i \neq j$.²² As proven above, the squared partial coherence function is exactly the parametric approach representation of partial spectral coherence (2.34).

²¹Note that Bacalá and Sameshima use a slightly different notation in their definition [3] of $\kappa_{i,j}(\lambda)$. In fact, their matrix $A_{ij}(\lambda)$ is exactly the same as our matrix $A_{ij}(\lambda)$.

²²we always study interactions between two different channels, therefore the case $i = j$ is not an interesting one.

Unfortunately, the partial coherence function is still an undirected measure. It describes the direct influence between X_i and X_j at one specific frequency, i.e. influences from all the other $K - 2$ channels are eliminated during calculations, but it still does not provide any information about the direction of those influences.

Although partialization is a key result on our way to partial directed coherence, the provided result is still not what we wanted to derive. The desired factorization will finally take place in the next section.

2.3.4 The partial directed coherence factor (PDCF)

2.3.4.1 PDCF and factorization of the partial coherence function

Definition 2.25 (partial directed coherence factor). The *partial directed coherence factor (PDCF)* from j to i is given by

$$\pi_{i,j}(\lambda) \triangleq \frac{A_{ij}(\lambda)}{\sqrt{a_{\bullet j}^*(\lambda) \Sigma_\epsilon^{-1} a_{\bullet j}(\lambda)}}, \quad (2.51)$$

where $a_{\bullet j}(\lambda)$ denotes the j th column of matrix $A(\lambda)$ like defined in (2.47) and $A_{ij}(\lambda)$ is element (i, j) of $A(\lambda)$. The *partial directed coherence factor matrix* is then represented by $\Pi \triangleq (\pi_{i,j})_{i=1 \dots K, j=1 \dots K}$.

The partial directed coherence factor ensures the desired factorization as we will see in the following

Theorem 2.26. *The partial coherence function can be expressed in terms of partial directed coherence factors by*

$$\kappa_{i,j}(\lambda) = \pi_{\bullet i}^* \Sigma_\epsilon^{-1} \pi_{\bullet j},$$

where, in analogy to the previous notation, $\pi_{\bullet i}$ denotes the i th column of the partial directed coherence factor matrix Π .

Proof. For every i, j with $i \neq j$ we have

$$\begin{aligned} \pi_{\bullet i}^* \Sigma_\epsilon^{-1} \pi_{\bullet j} &= \\ &= \left(\begin{array}{c} \frac{A_{1i}(\lambda)}{\sqrt{a_{\bullet i}^*(\lambda) \Sigma_\epsilon^{-1} a_{\bullet i}(\lambda)}} \\ \vdots \\ \frac{A_{Ki}(\lambda)}{\sqrt{a_{\bullet i}^*(\lambda) \Sigma_\epsilon^{-1} a_{\bullet i}(\lambda)}} \end{array} \right)^* \Sigma_\epsilon^{-1} \left(\begin{array}{c} \frac{A_{1j}(\lambda)}{\sqrt{a_{\bullet j}^*(\lambda) \Sigma_\epsilon^{-1} a_{\bullet j}(\lambda)}} \\ \vdots \\ \frac{A_{Kj}(\lambda)}{\sqrt{a_{\bullet j}^*(\lambda) \Sigma_\epsilon^{-1} a_{\bullet j}(\lambda)}} \end{array} \right) \\ &= \frac{1}{\sqrt{a_{\bullet i}^*(\lambda) \Sigma_\epsilon^{-1} a_{\bullet i}(\lambda)} \sqrt{a_{\bullet j}^*(\lambda) \Sigma_\epsilon^{-1} a_{\bullet j}(\lambda)}} (A_{1i}^*(\lambda) \ \dots \ A_{Ki}^*(\lambda)) \Sigma_\epsilon^{-1} \begin{pmatrix} A_{1j}(\lambda) \\ \vdots \\ A_{Kj}(\lambda) \end{pmatrix} \\ &= \frac{1}{\sqrt{a_{\bullet i}^*(\lambda) \Sigma_\epsilon^{-1} a_{\bullet i}(\lambda)} \sqrt{a_{\bullet j}^*(\lambda) \Sigma_\epsilon^{-1} a_{\bullet j}(\lambda)}} a_{\bullet i}^*(\lambda) \Sigma_\epsilon^{-1} a_{\bullet j}(\lambda) \\ &= \kappa_{i,j}(\lambda), \end{aligned}$$

which is exactly what we wanted to show. \square

Obviously, the partial directed coherence factor is an asymmetric measure as long as $A(\lambda)$ is not a symmetric matrix. In practice, $A(\lambda)$ is derived from the AR model's estimated coefficients and therefore, without much doubt, not symmetric.

Unfortunately, the appearance of Σ_ϵ in the denominator mixes the desired effects with so called instantaneous causality. As we want to investigate the development of epileptic seizures over time, we need to completely remove those instantaneous effects to derive a meaningful measure. In the following section, we will therefore exhaustively discuss the concept of instantaneous causality and find a characterization that helps us to remove all the instantaneous effects from the PDCF.

2.3.4.2 Instantaneous causality

Suppose, $x[n]$ and $z[n]$ are two stochastic processes and Ω_n is a set containing all the relevant information available up to period n . Furthermore, we denote by $\hat{z}[n|\Omega_n]$ the optimal 1-step prediction error of the process $z[n]$ given the information in Ω_n , i.e. the 1-step predictor that minimizes MSE²³. The corresponding forecast MSE itself will be denoted by $\hat{\Sigma}_z[n|\Omega_n]$ in the following context.

Definition 2.27 (instantaneous causality). We say that there is *instantaneous causality* between $x[n]$ and $z[n]$ if, in period n , adding the future value $x[n+1]$ to the information set improves the forecast of $z[n+1]$, i.e. ²⁴

$$\hat{\Sigma}_z[n|\Omega_n \cup \{x[n+1]\}] < \hat{\Sigma}_z[n|\Omega_n]. \quad (2.52)$$

According to Lütkepohl [23] this concept of causality is a symmetric one, i.e. whenever there is instantaneous causality between $x[n]$ and $z[n]$, there is also instantaneous causality between $z[n]$ and $x[n]$. So, of course, the notion “instantaneous causality from $x[n]$ to $z[n]$ ” is not used in our definition. This is in contrary to the definition of Granger causality in section 3.1.

Our goal (according to Lütkepohl [23] again) is to study the concept of instantaneous causality in the framework of a K -dimensional AR process of order p , represented by

$$y[n] = \sum_{j=1}^p B_j y[n-j] + u[n], \quad (2.53)$$

where the $K \times K$ -matrices B_j are the model coefficients and the K -dimensional vector $u[n]$ denotes a white noise process. We suppose, that the AR process (2.53) has the canonical MA representation

$$y[n] = \mu + \sum_{j=0}^{\infty} \Phi_j u[n-j] = \mu + \Phi(z)u[n]. \quad (2.54)$$

²³MSE=mean squared error. If θ denotes the estimated parameter and $\hat{\theta}$ denotes its estimator, the mean squared error is given by $MSE(\theta, \hat{\theta}) = \mathbb{E}(\theta - \hat{\theta})^2$

²⁴For two matrices A and B , the notation $A < B$ means, that $B - A$ is a positive definite matrix.

Again, the $K \times K$ -matrices Φ_j are the model coefficients, with Φ_0 set to identity I_K . Furthermore, $\Phi(z) = \sum_{j=0}^{\infty} \Phi_j z^j$ where z is the backshift operator. The (nonsingular) covariance matrix of the white noise process $u[n]$ will be denoted by Σ_u in the following context.

For further investigation, we need a MA representation with orthogonal residuals (i.e. noise errors have uncorrelated components). As Σ_u is a covariance matrix and therefore positive (semi-)definite, it can be written as

$$\Sigma_u = PP',$$

where P is a nonsingular lower triangle matrix.²⁵ Obviously, representation (2.54) is equivalent to

$$y[n] = \mu + \sum_{j=0}^{\infty} \Phi_j PP^{-1}u[n-j] = \mu + \sum_{j=0}^{\infty} \Theta_j w[n-j], \quad (2.55)$$

with $\Theta_j = \Phi_j P$ and $w[n] = P^{-1}u[n]$. The process's residuals are now orthogonal because (Σ_w denotes the covariance matrix of $w[n]$)

$$\begin{aligned} \Sigma_w &= \mathbb{E}[w[n]w[n]'] = \mathbb{E}[P^{-1}u[n]u[n]'(P^{-1})'] \\ &= P^{-1}\Sigma_u(P^{-1})' = P^{-1}PP'(P^{-1})' \\ &= I_K. \end{aligned}$$

As a final assumption, we claim that the K -dimensional process $y[n]$ consists of the M -dimensional process $z[n]$ (with $M < K$) and the $(K - M)$ -dimensional process $x[n]$, i.e. $y[n] = \begin{pmatrix} z[n] \\ x[n] \end{pmatrix}$. Applying this partitioning to the MA representation (2.55) of the AR model therefore leads to

$$y[n] = \begin{pmatrix} z[n] \\ x[n] \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \Theta_{11}(z) & \Theta_{12}(z) \\ \Theta_{21}(z) & \Theta_{22}(z) \end{pmatrix} \begin{pmatrix} w_1[n] \\ w_2[n] \end{pmatrix}. \quad (2.56)$$

Clearly, as we supposed $\Phi_0 = I_K$, the matrix Θ_0 is equal to the triangle matrix P which we have used to decompose Σ_u . Hence, representation (2.56) is equal to

$$\begin{pmatrix} z[n] \\ x[n] \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \Theta_{11,0} & 0 \\ \Theta_{21,0} & \Theta_{22,0} \end{pmatrix} \begin{pmatrix} w_1[n] \\ w_2[n] \end{pmatrix} + \begin{pmatrix} \Theta_{11,1} & \Theta_{12,1} \\ \Theta_{21,1} & \Theta_{22,1} \end{pmatrix} \begin{pmatrix} w_1[n-1] \\ w_2[n-1] \end{pmatrix} + \dots$$

This representation of the AR model holds for every n , so

$$z[n+1] = \mu_1 + \Theta_{11,0}w_1[n+1] + \Theta_{11,1}w_1[n] + \Theta_{12,1}w_2[n] + \dots$$

and

$$x[n+1] = \mu_2 + \Theta_{21,0}w_1[n+1] + \Theta_{22,0}w_2[n+1] + \Theta_{21,1}w_1[n] + \Theta_{22,1}w_2[n] + \dots$$

²⁵A prove can be found in [3] for example.

Using this result, we now can derive the one-step prediction error for $x[n]$:

$$\begin{aligned}\hat{x}[n|\{y[s]|s \leq n\} \cup \{z[n+1]\}] &= \hat{x}[n|\{w[s]|s \leq n\} \cup \{w_1[n+1]\}] \\ &= \Theta_{21,0}w_1[n+1] + \underbrace{\mu_2 + \Theta_{21,1}w_1[n] + \Theta_{22,1}w_2[n] + \dots}_{=\hat{x}[n|\{y[s]|s \leq n\}]}\end{aligned}$$

The latter equality holds because of $\mathbb{E}[w_2[n+1]] = 0$. Apparently, the instantaneous causality condition

$$\hat{x}[n|\{y[s]|s \leq n\} \cup \{z[n+1]\}] = \hat{x}[n|\{y[s]|s \leq n\}] \quad (2.57)$$

holds, if and only if

$$\Theta_{21,0} = 0$$

and consequently

$$P = \Theta_0 = \begin{pmatrix} \Theta_{11,0} & 0 \\ \Theta_{21,0} & \Theta_{22,0} \end{pmatrix} = \begin{pmatrix} \Theta_{11,0} & 0 \\ 0 & \Theta_{22,0} \end{pmatrix}.$$

Hence, the covariance matrix

$$\Sigma_u = PP' = \begin{pmatrix} \Theta_{11,0}'\Theta_{11,0} & 0 \\ 0 & \Theta_{22,0}'\Theta_{22,0} \end{pmatrix}$$

has to be a block-diagonal matrix, too.

We can conclude this derivation with the following

Proposition 2.28 (characterization of instantaneous causality in VAR models). *Let a stochastic process $y[n]$ be defined as in (2.53) and (2.56). Then there is no instantaneous causality between $x[n]$ and $z[n]$ if and only if the corresponding residuals $u_1[n]$ and $u_2[n]$ are uncorrelated, i.e.²⁶*

$$\mathbb{E}[u_1[n]u_2[n]'] = 0.$$

Generalization of Proposition 2.28 finally leads to

Proposition 2.29 (a further characterization of instantaneous causality in VAR models). *Let a K -dimensional stochastic process $y[n]$ be defined like in (2.53). Assume, that $y[n]$ consists of K one-dimensional components $y_i[n]$, $i = 1 \dots K$, i.e. $y[n] = (y_1[n], y_2[n], \dots, y_K[n])'$. Let Σ_u again be the covariance matrix of the white noise process $u[n] = (u_1[n], u_2[n], \dots, u_K[n])'$. Then there is no instantaneous causality between any pair of components $y_i[n]$ and $y_j[n]$ with $i \neq j$ if and only if Σ_u is a diagonal matrix, i.e.*

$$\mathbb{E}[u_i[n]u_j[n]'] = 0 \quad \forall i \neq j.$$

The latter proposition will help to explain the crucial step from the PDCF to partial directed coherence in the following section.

²⁶ $u[n] = \begin{pmatrix} u_1[n] \\ u_2[n] \end{pmatrix}$ is partitioned in the same way as $w[n] = \begin{pmatrix} w_1[n] \\ w_2[n] \end{pmatrix}$ was earlier in this section

2.3.5 Partial directed coherence (PDC)

As we have seen in the previous section, removing all the instantaneous effects from the partial directed coherence factor can be obtained by simply removing all the non-diagonal elements from our covariance matrix Σ_ϵ .

Beyond that, Baccalá and Sameshima [3] even made one extra step and defined partial directed coherence by simply setting the covariance matrix equal to identity, i.e. $\Sigma_\epsilon = I_K$, which leads to the following

Definition 2.30 (partial directed coherence). The *partial directed coherence from j to i* is defined by

$$\pi_{i,j}^2(\lambda) \triangleq \frac{|A_{ij}(\lambda)|^2}{\sum_{k=1}^K |A_{kj}(\lambda)|^2}, \quad (2.58)$$

where $A_{ij}(\lambda)$ is element (i, j) of the matrix $A(\lambda)$ defined in (2.47).

2.3.5.1 Characteristics and interpretation of PDC

We will start our analysis of PDC with the derivation and interpretation of some of its basic characteristics:

- First, we want to show that *partial directed coherence is well-defined*, i.e. the denominator $\sum_{k=1}^K |A_{kj}(\lambda)|^2$ is greater than zero, if the stability condition (2.46) holds for the AR-model (2.44).

By definition of the stability condition, $\det A(z) \neq 0 \forall |z| \leq 1$, where $A(z) = I_K - A_1 z - A_2 z^2 - \dots - A_p z^p$, $z \in \mathbb{C}$ is a polynomial matrix. As the stability condition assures, that $A(z)$ is not a singular matrix, all of its eigenvalues are different from zero. Replacing z with $e^{-i\lambda}$ proves, that all the eigenvalues of $A(\lambda)$ are different from zero, too.

Obviously, $A(\lambda)^* A(\lambda)$ is a positive semi-definite matrix as $x^* A(\lambda)^* A(\lambda) x = (A(\lambda)x)^* (A(\lambda)x) \geq 0$ for every vector x . Furthermore,

$$x^* A(\lambda)^* A(\lambda) x = (A(\lambda)x)^* (A(\lambda)x) = 0$$

for an arbitrarily chosen $x \neq 0$ if and only if $A(\lambda)x = 0$, i.e. if one of the eigenvalues of $A(\lambda)$ is equal to zero. So, if the stability condition (2.46) holds, $A(\lambda)^* A(\lambda)$ is a strictly positive definite matrix, i.e.

$$A(\lambda)^* A(\lambda) > 0 \quad (2.59)$$

Now, let $e_j = (0, \dots, 0, 1, 0, \dots, 0)'$ be the j th base vector consisting of zeros except for a single 1 on the j th position. Let $\|\cdot\|$ further denote the Euclidean norm. Using this notation, the denominator

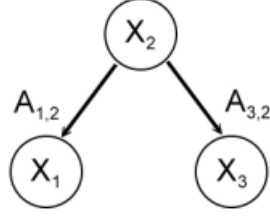


Figure 2.6: **Normalization of the PDC.** Exemplary 3-dimensional system with source channel $x_2[n]$ and two target channels $x_1[n]$ and $x_3[n]$. Partial directed coherences $\pi_{1,2}^2(\lambda)$ and $\pi_{3,2}^2(\lambda)$ are both normalized with respect to the sum of all outflows from the source channel, i.e. $|A_{1,2}(\lambda)|^2 + |A_{2,2}(\lambda)|^2 + |A_{3,2}(\lambda)|^2$.

of partial directed coherence is equal to

$$\begin{aligned}
 \sum_{k=1}^K |A_{kj}(\lambda)|^2 &= \|A(\lambda)e_j\|^2 \\
 &= (A(\lambda)e_j)^*(A(\lambda)e_j) \\
 &= e_j' \underbrace{A(\lambda)^* A(\lambda)}_{>0} e_j > 0.
 \end{aligned}$$

- As the denominator cannot be zero, clearly, $\pi_{i,j}^2(\lambda)$ is bounded by 0 and 1, i.e.

$$0 \leq \pi_{i,j}^2(\lambda) \leq 1 \quad (2.60)$$

for every λ, i, j .

- Partial directed coherence is a normalized measure, therefore, if one channel j is chosen, all the partial directed coherences from j to i , $i = 1 \dots K$ sum to 1, i.e.

$$\sum_{i=1}^K \pi_{i,j}^2(\lambda) = \sum_{i=1}^K \frac{|A_{ij}(\lambda)|^2}{\sum_{k=1}^K |A_{kj}(\lambda)|^2} = 1. \quad (2.61)$$

- According to Baccalá [3], because of the normalization condition (2.61), the partial directed coherence $\pi_{i,j}^2(\lambda)$ represents *the relative coupling strength of the interactions between a given signal source channel j and another channel i compared to j 's interactions to all other channels*.

In other words, $\pi_{i,j}^2(\lambda)$ is the ratio of the outflow from source channel j to channel i normalized to all the outflows from the source channel. Figure 2.3.5.1 illustrates this concept using a 3-dimensional system (i.e. the number of channels is equal to 3).

- Furthermore, when studying the normalization conditions (2.60) and (2.61), another interpretation of PDC is obvious:
PDC ranks the relative interactions between two channels with respect to one given signal source. So, values of $\pi_{i,j}^2(\lambda)$ near one indicate, that source

channel j has high influence onto channel i . On the other hand, values near²⁷ zero indicate, that there might not be any influence from j onto i ²⁸.

- Clearly, there could be high influence from channel j onto i for one frequency λ_1 , i.e.

$$\pi_{i,j}^2(\lambda_1) \approx 1$$

but nearly no influence for some different frequency $\lambda_2 \neq \lambda_1$, i.e.

$$\pi_{i,j}^2(\lambda_2) \approx 0.$$

In practice, we will need a general, frequency independent measure for the interactions between two channels. Mostly, this problem is solved by summation or integration. We will go into even more detail when discussing the implementation of PDC in Section 2.4.

- Interpretation of PDC for $i = j$ is slightly different from the interpretations mentioned before at $i \neq j$. $\pi_{i,i}^2(\lambda)$ describes the fraction of outflow from channel i that is not explicable by the other measured time series. In other words, $\pi_{i,i}^2(\lambda)$ represents how much of the i th channel's current state is explained by it's own past.
- To obtain PDC, there is no need to calculate the matrix $H(\lambda)$, i.e. it's not necessary to invert $A(\lambda)$. As $A(\lambda)$ could be close to singular in practice, this property provides a huge computational advantage of PDC compared to different measures that might be derived from parametric spectral estimators.

2.3.5.2 Link to Granger causality

Up to now, we always considered PDC as the factorization of partial spectral coherence. Actually, Baccalá and Sameshima's idea behind the definition of PDC [3] was completely different - the initial intention was to create a frequency domain representation of Granger causality, which is a fundamental tool in the analysis of multivariate time series. Casually spoken, an observed time series $x[n]$ *Granger causes* another time series $y[n]$ if knowledge of past values of $x[n]$ significantly improves the prediction of $y[n]$ ²⁹.

Let's take a look at the definition (2.58) of PDC, again:

$$\pi_{i,j}^2(\lambda) \triangleq \frac{|A_{ij}(\lambda)|^2}{\sum_{k=1}^K |A_{kj}(\lambda)|^2},$$

where the matrix $A(\lambda)$ was defined by $A(\lambda) = I_K - \sum_{j=1}^p A_j e^{-i\lambda j}$. As

$$A_{ij}(\lambda) = \begin{cases} 1 - \sum_{r=1}^p (A_r)_{ij} e^{-i\lambda r} & j = k \\ - \sum_{r=1}^p (A_r)_{ij} e^{-i\lambda r} & j \neq k \end{cases},$$

²⁷In practice, as we already discussed, partial directed coherences are derived from estimated AR coefficients. Therefore, it's rather unrealistic, that $\pi_{i,j}^2(\lambda)$ is exactly equal to zero for any chosen frequency λ .

²⁸Of course, as PDC is not a symmetric measure, there could be high influence from channel j onto channel i .

²⁹We will define and discuss the concept of Granger causality more accurate in Chapter 3.

partial directed coherence depends mostly on the elements $(A_r)_{ij}$ of the estimated AR model coefficients A_r , which are, in fact, $K \times K$ matrices. In our autoregressive model (2.44), $x[n] = \sum_{r=1}^p A_r x[n-r] + \epsilon[n]$, the coefficients $(A_r)_{ij}$ describe, how the present values of one channel x_i depend linearly on the past values of all the other channels x_j .

As discussed and partially proven in Section 3.1.2, zeros in the coefficient matrices $A_r(\lambda)$ indicate, that there is no Granger causality between the corresponding channels, i.e. $(A_r)_{ij} = 0 \ \forall r$ is equivalent to the statement “the time series $x_j[n]$ does not *Granger cause* $x_i[n]$, given all the other components $x_k[n]$, $k = \{1 \dots K\} \setminus \{i, j\}$ ”.

Putting the pieces together leads to the following, very important

Theorem 2.31 (Link between partial directed coherence and conditional Granger causality). *The partial directed coherence (2.58) indicates conditional Granger causality in the multivariate AR model (2.44):*

$$\pi_{i,j}^2(\lambda) \equiv 0 \ \forall \lambda \Leftrightarrow x_j[n] \stackrel{G}{\nrightarrow} x_i[n],$$

where, $i \neq j$ ³⁰ and the symbol $\stackrel{G}{\nrightarrow}$ denotes, that there is no Granger causality.

2.3.5.3 Alternative normalizations

Although PDC has been derived carefully from partial spectral coherence and, hopefully, all steps in this derivation sound plausible to the reader, one might suggest different normalizations in the PDC definition (2.58).

As discussed in Section 2.3.5.1, according to definition (2.58) $\pi_{i,j}^2(\lambda)$ is the normalized outflow from source channel j to channel i . On the other hand it is expected, that the influence of a source channel j on another channel i decreases, if the number of channels, that affect i increases. Therefore, partial directed coherence could also be defined using a different normalization term:

$$\tilde{\pi}_{i,j}^2(\lambda) \triangleq \frac{|A_{ij}(\lambda)|^2}{\sum_{k=1}^K |A_{ik}(\lambda)|^2}. \quad (2.62)$$

This alternative definition of PDC represents the flow from source channel j to another channel i normalized to all the other inflows at channel i . It could be derived from partial coherence similarly, using adequate factorization terms. In some applications, definition (2.62) might in fact be more suitable than the mostly used genuine definition (2.58). Figure 2.3.5.3 illustrates this alternative normalization concept.

Furthermore, it might not be clear, why all the interactions between two channels are weighted equally. As discussed in Section 2.3.5, Baccalá and Sameshima simply set $\Sigma_\epsilon = I_K$ to remove all instantaneous effects and therefore received the PDC definition (2.58). Using Theorem 2.29, we can indeed prove,

³⁰as already mentioned earlier, the case $i = j$ is not an interesting one, because we want to analyze interactions between different ECoG channels.

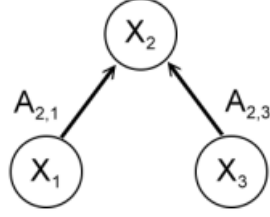


Figure 2.7: **Alternative Normalization of the PDC.** Exemplary 3-dimensional system with target channel $x_2[n]$ and two source channels $x_1[n]$ and $x_3[n]$. Partial directed coherences $\pi_{2,1}^2(\lambda)$ and $\pi_{2,3}^2(\lambda)$ are both normalized with respect to the sum of all inflows into the target channel, i.e. $|A_{2,1}(\lambda)|^2 + |A_{2,2}(\lambda)|^2 + |A_{2,3}(\lambda)|^2$.

that all the non-diagonal elements of the model's covariance matrix should be set to zero to remove instantaneous effects, but it cannot explain, why all the diagonal elements have to be equal (to one).

In fact, as we will see in Section 2.3.5.5, this simplification leads to a severe problem which might even make PDC useless in some applications. Correction of the PDC definition, using the diagonal elements of Σ_ϵ for normalization will lead to Generalized PDC in Section 2.3.6.

2.3.5.4 Statistical tests and confidence intervals

Typically, statistical tests and confidence intervals for modern measures of multivariate time series analysis can only be determined by simulation. However, for PDC, as it depends mostly on estimated AR model coefficients, it is possible to derive the necessary significance levels (according to Schelter et al. [26], the AR-model coefficients are usually asymptotically normally distributed and unbiased). We will omit some details here as a few parts of the prove are extremely complex, but want to present the concept behind this derivation here.

First, we need a special kind of covariance matrix³¹ for our AR process $x[n]$ defined in (2.44):

$$R(k, l) = \begin{pmatrix} R_{11}(k, l) & \cdots & R_{1K}(k, l) \\ \vdots & \ddots & \vdots \\ R_{K1}(k, l) & \cdots & R_{KK}(k, l) \end{pmatrix} \quad (2.63)$$

with entries

$$R_{ij}(k, l) = \text{cov}(x_i[n - k], x_j[n - l]) \quad (2.64)$$

for every $i, j = 1 \dots K$ and $k, l = 1 \dots p$. Let further denote $H(k, l) = R^{-1}(k, l)$ the inverse of covariance matrix (2.63).

For the sake of clarity, the diagonal elements of the AR process's covariance matrix, $(\Sigma_\epsilon)_{ii} = \mathbb{V}\epsilon_i[n]$ are represented simply by σ_i^2 in the following context.

³¹this representation traces back to Lütkepohl [23]

Schelter et al. proved in their paper [26], that under the null hypothesis $H_0 : |A_{ij}(\lambda)|^2 = 0$ (and therefore $\pi_{ij}^2(\lambda) = 0$) the expression

$$\frac{K \left| \hat{A}_{ij}(\lambda)^2 \right|}{\hat{C}_{ij}(\lambda)} \sim \chi^2(1) \quad (2.65)$$

is asymptotically Chi-squared distributed with one degree of freedom, where

$$C_{ij}(\lambda) = \sigma_i^2 \left[\sum_{k,l=1}^p H_{jj}(k, l) (\cos(k\lambda) \cos(l\lambda) + \sin(k\lambda) \sin(l\lambda)) \right] \quad (2.66)$$

and the symbol $\hat{\cdot}$ in the nominator and denominator denotes, that the values of the AR model coefficients $(A_k)_{ij}$ have been replaced by estimated ones.

If α is the significance level and $\chi_{1,1-\alpha}^2$ represents the $(1 - \alpha)$ -quantile of a Chi-squared distribution with one degree of freedom,

$$\mathbb{P} \left(\frac{K \left| \hat{A}_{ij}(\lambda)^2 \right|}{\hat{C}_{ij}(\lambda)} \leq \chi_{1,1-\alpha}^2 \right) = 1 - \alpha \quad (2.67)$$

holds. According to definition (2.58),

$$\pi_{i,j}^2(\lambda) \triangleq \frac{|A_{ij}(\lambda)|^2}{\sum_{k=1}^K |A_{kj}(\lambda)|^2}.$$

Simple usage of Taylor's Theorem shows³², that under the null hypothesis,

$$\hat{\pi}_{i,j}^2 = \pi_{i,j}^2(\lambda) + R, \quad (2.68)$$

where the remainder R is negligible compared with the main term (i.e. the real value of PDC). Putting (2.67), (2.58) and (2.68) together leads to

$$\mathbb{P} \left(\frac{K \pi_{i,j}^2(\lambda) \sum_{k=1}^K \left| \hat{A}_{kj}(\lambda) \right|^2}{\hat{C}_{ij}(\lambda)} \leq \chi_{1,1-\alpha}^2 \right) = 1 - \alpha \quad (2.69)$$

and further

$$\mathbb{P} \left(\pi_{i,j}^2(\lambda) \leq \frac{\hat{C}_{ij}(\lambda) \chi_{1,1-\alpha}^2}{K \sum_{k=1}^K \left| \hat{A}_{kj}(\lambda) \right|^2} \right) = 1 - \alpha. \quad (2.70)$$

Hence, PDC is significantly (at a $\alpha\%$ probability of failure) different from zero, if

$$\pi_{i,j}^2(\lambda) > \frac{\hat{C}_{ij}(\lambda) \chi_{1,1-\alpha}^2}{K \sum_{k=1}^K \left| \hat{A}_{kj}(\lambda) \right|^2} \quad (2.71)$$

³²Details can be found in Schelter et al. [26], again

for every given frequency $\lambda \in [-\pi, \pi]$. The right hand side of this equation is called the α -*significance level for PDC*.

Unfortunately, as the significance level depends on the current frequency λ , PDC mostly will be significant for some frequencies and non-significant for the remaining frequencies. Beyond that, the largest value of $\{\pi_{i,j}^2(\lambda), \lambda \in [-\pi, \pi]\}$ could be non-significant, while at other frequencies smaller values of $\pi_{i,j}^2(\lambda)$ are significant.

Applied to our problem, we need a frequency independent measure derived from PDC, i.e. we need to know if coupling effects between a source channel j and a target channel i are significantly different from zero. As it is not quite clear how to derive a suitable measure from PDC combined with the significance level as long as significance is only given for a short range of frequencies (of course, this range can even differ if two different channels are chosen), we will not use the information that is provided by the significance level in the implementation Section 2.4.

Nevertheless further research should be based on the idea behind the construction of a significance level in (2.71). For example, Schelter et al. even defined a completely different kind of PDC using normalization conditions derived from (2.65) [25].

2.3.5.5 Problems and disadvantages of PDC and its construction

During the last subsections, we carefully have explained and derived the concept of PDC. Up to now, we mostly talked about pros and ignored cons. In fact, PDC is used in numerous publications. It works rather good in a lot of applications, so it does in its generalized form in the localization of the focus of epileptic seizures (as we will see in Section 2.5.3). Nevertheless, the construction gives several problems, which we want to discuss now:³³

- (i) PDC measures the strength of influences between time series *relative* to one given signal source. Thus, comparison of PDC values for different source processes might be difficult, as well as comparison for different frequencies, as the denominator in PDC definition (2.58) depends on the chosen frequency λ .

In addition, the strength of coupling effects between two channels i and j depends on the number of channels involved in our analysis, i.e. the value of PDC decreases if the number of channels K increases even if the added channels do not affect both, i and j . This property of PDC might cause severe problems in practice.

- (ii) We showed, that the partial directed coherence function (2.51) and partial coherence (2.34) are exactly the same. Unfortunately, it is not quite clear,

³³a detailed discussion of PDC problems can also be found in [14]

why the property of partialization still holds after factorization in Section 2.3.4.

- (iii) It is obvious, that PDC is an asymmetric measure, but is it also a directed measure? In our context, we assumed that asymmetric and directed measures are in fact the same. If, in contrast, the definition of a directed measure would somehow be based on causal aspects, asymmetry might not be a sufficient condition.
- (iv) PDC is not scale-invariant, i.e. it depends on the units of measurement of the processes involved. Even worse, changing the scale can lead to values of partial directed coherence arbitrarily near to zero. We want to illustrate that problem using a simple example:

Assume a 2-dimensional (i.e. $K = 2$) autoregressive model of order $p = 1$, where element $(2, 1)$ of the AR model coefficient A_1 is set to zero. Furthermore denote $x[n] = \begin{pmatrix} x_n \\ y_n \end{pmatrix}$ during this example. The AR(1) model can therefore be written as

$$\underbrace{\begin{pmatrix} x_n \\ y_n \end{pmatrix}}_{=x[n]} = \underbrace{\begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix}}_{=A_1} \underbrace{\begin{pmatrix} x_{n-1} \\ y_{n-1} \end{pmatrix}}_{=x[n-1]} + \underbrace{\begin{pmatrix} \varepsilon_{x,n} \\ \varepsilon_{y,n} \end{pmatrix}}_{=\epsilon[n]}, \quad (2.72)$$

where $\epsilon[n]$ denotes a stochastic component with zero mean³⁴.

Assume further, that the second row of equation (2.72) is scaled by a factor $\alpha \in \mathbb{R}$, i.e.

$$\alpha y_n = a_{22}\alpha y_{n-1} + \alpha \epsilon_{y,n}. \quad (2.73)$$

After defining

$$u_n \triangleq \alpha y_n, \quad (2.74)$$

we receive a different AR model in x and u with a different coefficient matrix \tilde{A}_1 :

$$x_n = a_{11}x_{n-1} + \frac{a_{12}}{\alpha}u_{n-1} + \epsilon_{x,t} \quad (2.75)$$

$$u_n = a_{22}u_{n-1} + \underbrace{\alpha \epsilon_{y,t}}_{=\epsilon_{u,t}}. \quad (2.76)$$

Normally, if x Granger causes y , it should also Granger cause u , independent from the value of α . Therefore, we would expect $\pi_{xy}^2(\lambda)$ to be equal to $\pi_{xu}^2(\lambda)$ for any chosen frequency λ . We will now show, that, unfortunately, this expectation does not hold for PDC.

By definition, the elements of the matrix $A(\lambda)$ we need to calculate PDC are equal to

$$A_{ij}(\lambda) = \begin{cases} 1 - (A_1)_{ij}e^{-i\lambda} & i = j \\ -(A_1)_{ij}e^{-i\lambda} & i \neq j \end{cases}$$

³⁴Note, that $\epsilon[n]$ need not necessarily be white noise in this context.

Hence,

$$|A_{ij}(\lambda)| = \begin{cases} |1 - (A_1)_{ij}e^{-i\lambda}| & i = j \\ |(A_1)_{ij}| & i \neq j \end{cases} \quad (2.77)$$

and therefore

$$\begin{aligned} \pi_{xu}^2(\lambda) &= \frac{\left| \left(\tilde{A}_1 \right)_{12}(\lambda) \right|^2}{\left| 1 - \left(\tilde{A}_1 \right)_{22}(\lambda)e^{-i\lambda} \right|^2 + \left| \left(\tilde{A}_1 \right)_{12}(\lambda) \right|^2} \\ &= \frac{\left| \frac{a_{12}}{\alpha} \right|^2}{\left| 1 - a_{22}(\lambda)e^{-i\lambda} \right|^2 + \left| \frac{a_{12}}{\alpha} \right|^2} \xrightarrow{\alpha \rightarrow \infty} 0. \end{aligned}$$

As shown in this example, simple scaling of channel y results in an arbitrarily small value of PDC. In practice, this problem becomes important especially when variables are not measured on common scale.

To put things right, we may define an alternative version of PDC:

$$\pi_{xu}^2(\lambda) = \frac{\frac{\sigma_y^2}{\sigma_x^2} \left| \frac{a_{12}}{\alpha} \right|^2}{\left| 1 - a_{22}(\lambda)e^{-i\lambda} \right|^2 + \frac{\sigma_y^2}{\sigma_x^2} \left| \frac{a_{12}}{\alpha} \right|^2}, \quad (2.78)$$

where $\sigma_u^2 = \mathbb{V}\epsilon_{u,t}$ and $\sigma_x^2 = \mathbb{V}\epsilon_{x,t}$ are the variance's of the AR model's stochastic components. As

$$\sigma_u^2 = \alpha^2 \sigma_y^2, \quad (2.79)$$

α cancels out in the nominator and in the denominator. Therefore, $\pi_{xu}^2(\lambda)$ remains unchanged even if the scaling factor α changes.

Finally, (2.78) is equivalent to

$$\pi_{xy}^2(\lambda) = \frac{\frac{1}{\sigma_x^2} |a_{12}|^2}{\frac{1}{\sigma_y^2} |1 - a_{22}(\lambda)e^{-i\lambda}|^2 + \frac{1}{\sigma_x^2} |a_{12}|^2}, \quad (2.80)$$

which is exactly the generalized partial directed coherence from y to x we will define in Section 2.3.6.

2.3.6 Generalized partial directed coherence (GPDC)

Generalization of equation (2.80) leads to

Definition 2.32 (generalized partial directed coherence). The *generalized partial directed coherence* from j to i is defined by

$$\pi_{i,j}^2(\lambda) \triangleq \frac{\frac{1}{\sigma_i^2} |A_{ij}(\lambda)|^2}{\sum_{k=1}^K \frac{1}{\sigma_k^2} |A_{kj}(\lambda)|^2}, \quad (2.81)$$

where $A_{ij}(\lambda)$ stands for element (i, j) of the matrix $A(\lambda)$ defined in (2.47) and σ_i^2 denotes the i th diagonal element of the AR model's covariance matrix Σ_ϵ as discussed in (2.44).³⁵

GPDC can be obtained from the partial directed coherence factor (2.51), if instantaneous effects are removed by setting all non-diagonal elements of the covariance matrix Σ_ϵ equal to zero. This approach is in perfect accordance with Proposition 2.29 and does therefore sound more plausible than the approach used to obtain PDC definition (2.58).

Hence in GPDC definition, all elements are weighted by corresponding precisions $\frac{1}{\sigma_i^2}$ from the AR model's definition (2.44). Obviously, as GPDC in fact does not differ from PDC except for different weights in the denominator's sum, PDC characteristics and interpretations discussed in Section 2.3.5 remain the same, as well as the critics do.

2.3.7 Comparing PDC and GPDC using signal models

We already discussed some problems of the partial directed coherence in Section 2.3.5.5 and proved, that PDC is not invariant to simple scaling of time series components. Nevertheless, the scenario might sound unrealistic to the reader. Therefore we want to discuss the impacts of differently scaled components using simple signal models.

Example 2.33. We want to start the analysis using a simulated 2-dimensional AR(1)-process $x[n] = (x_1[n], x_2[n])' = Ax[n-1] + \epsilon[n]$, where the coefficient matrix A is defined by

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} 0.5 & 0.5 \\ 0 & 0.5 \end{pmatrix} \quad (2.82)$$

and $\epsilon[n]$ denotes white noise with mean zero and variance one. An initial value $x[0] = 0$ was chosen for the simulation of this process in Matlab. As $a_{21} = 0$, we await, that $x_2[n]$ is only affected by its own past and not by the past of the first channel, whilst $x_1[n]$ itself is influenced by the past of both channels, as a_{11} and a_{12} are both greater than zero.

Clearly, PDC and GPDC indicate this system, which is extremely simple, correctly. Unfortunately, that changes (for PDC) if the second channel is scaled by a positive factor. We will discuss two different scenarios:

³⁵Baccalá presented the idea of GPDC in [2]

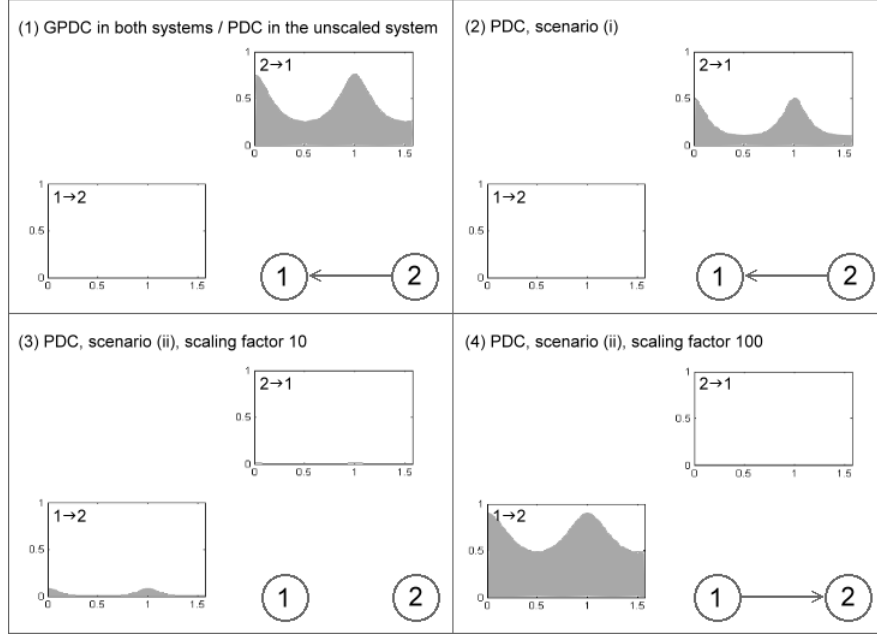


Figure 2.8: **Scale-invariance of PDC.** If the unscaled system is analyzed by using either PDC or GPDC or if GPDC is used in the scaled system, interaction from channel 2 to channel 1 **is** identified correctly (1). If scaling is added using scenario (i), values of PDC decrease rapidly, but the direction is still correct (2). Implementing scenario (ii) lets PDC indicate either no interaction between 1 and 2 if the second component of $x[n]$ is scaled by a factor 10 (3) or interaction even in the wrong direction, if it is scaled by a factor 100 (4).

- (i) The coefficient matrix A is held constant, whilst only the simulated values of $x_2[n]$ are scaled. This is exactly the situation discussed in Section 2.3.5.5 and leads to a sharp decrease in the values of PDC in practice, although we did not manage to reach values of zero as proven in 2.3.5.5. Plot (2) of Figure 2.8 shows the results of our simulation.
- (ii) The coefficient matrix A is estimated separately for the simulated and for the scaled system. Of course, none of the estimated coefficients is equal to those defined in (2.82) - Matlab returns an estimated coefficient matrix

$$\hat{A} = \begin{pmatrix} 0.5024 & 0.5188 \\ 0.0146 & 0.5039 \end{pmatrix} \quad (2.83)$$

for the unscaled system.

All the values of channel 2 are now scaled by a factor 10 and a new AR-model is fitted. Obviously, a_{11} and a_{22} will not differ for the scaled system, as the optimal derivation of $x[n]$ from its own past does not change, and $y[n]$ is obtained from its past value in the same way as $10y[n]$ is obtained from $10y[n-1]$. Unfortunately, those are the coefficients of matrix A that do not really matter when calculating the PDC.

On the other hand, the values of the off-diagonal elements (the more important ones for PDC calculation) do change. In the old system, a_{12} described the optimal explanation of $x[n]$ from the past value $y[n-1]$. As the values of $x[n]$ are not affected by scaling, it is obvious that a reduction of a_{12} by a factor of 10 must equalize the increase in $y[n-1]$. For the same reason, a_{21} increases by a factor 10, too.

Consequently, Matlab returns estimated coefficients

$$\hat{A}_s = \begin{pmatrix} 0.5024 & 0.0519 \\ 0.1458 & 0.5039 \end{pmatrix}$$

for the scaled system. As PDC is based on those coefficients only, it will indicate (depending on the chosen threshold) no interactions between channels 1 and 2, or³⁶ even interaction from channel 2 to channel 1 instead of the opposite direction.

Fortunately, GPDC is able to resolve this problem because the coefficients are weighted by the variances of the corresponding residuals. Clearly, the variance of $10y[n]$ is 100-times larger than the variance of $y[n]$, thus the variance of the residual error of channel 2 will increase by a factor 100 too, while the variances of $x[n]$ and the residuals of channel 1 are not affected. (3) and (4) in Figure 2.8 illustrate that circumstances using a Matlab simulation.

Although this first example already illustrated the PDC's scale invariance problems, we will now show an even more impressive one:

Example 2.34. A 4-dimensional AR(1) model $x[n] = (x_1[n], x_2[n], x_3[n], x_4[n])' = Ax[n-1] + \epsilon[n]$ is chosen, with coefficient matrix

$$A = \begin{pmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0.5 \end{pmatrix}.$$

Again, $\epsilon[n]$ denotes white noise with mean zero and variance one. The model was simulated using Matlab, again, with initial value $x[0] = (0, 0, 0, 0)'$.

Obviously, we await interactions from channel 1 to channel 3 and further from channel 2 and channel 3 to channel 4. The whole model was identified correctly by both, PDC and GPDC. Clearly, we did not expect them to fail in any simple AR(1) model.

Like in the previous example, we will scale the second channel of our time series by a factor 100. Again, the scaled model is identified correctly when using generalized PDC, as shown in Figure 2.9, which is in perfect accordance with the results from Example 2.33 and the theoretical considerations from Section 2.3.5.5.

³⁶interaction in the wrong direction is obtained at least if the second channel is scaled by an even higher factor

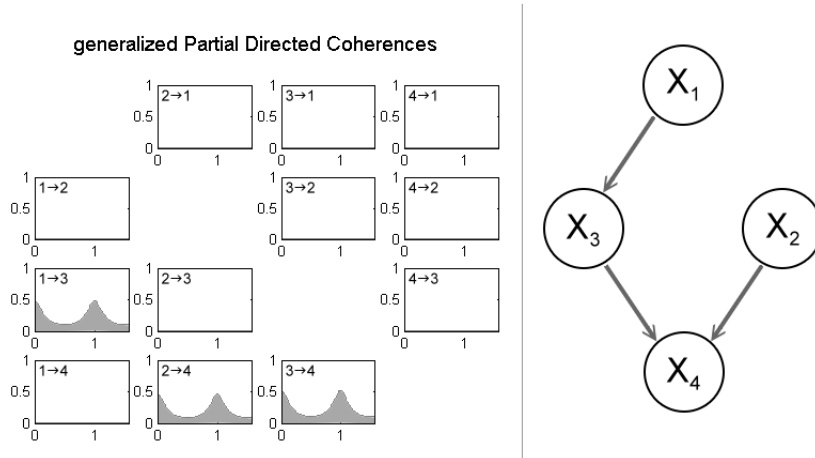


Figure 2.9: *Analyzing the signal model using GPDC. All interactions are identified correctly.*

PDC is not able to identify the scaled model correctly, but the results, that are presented in Figure 2.10, are worse than expected. None of the interactions was identified correctly (without leaving any doubt, if PDC values are not too low according to the chosen threshold³⁷).

First of all, like in the two-dimensional model from example 2.33, coupling effects from all the other channels to the scaled channel are indicated. Explanation is still the same - scaling leads to AR model coefficients which are scaled themselves, either by a factor 100 or by its inverse. Unfortunately, zeros in the coefficient matrix A will be estimated by a small number, which is different from zero. If this number is multiplied by the factor 100, completely wrong arrows will appear.

Secondly, PDC values for the interactions between channels 1 and 3 and channels 3 and 4 are small. Fortunately, the system here is quite simple, therefore it is still possible to recognize the difference compared to channel pairs without any interactions to be noticed. In practice systems are much more complicated with a lot of noise and unobservable variables. Thus it is necessary to suppress those disturbing factors by choosing a sufficiently high threshold. Under this circumstances it is obvious, that the interactions between 1, 3 and 4 would in fact not be identified.

³⁷In practice, we need a frequency-independent measure. Therefore, PDC values are added (or integrated) over the whole spectrum. An interaction is then indicated if the resulting value is greater than the chosen threshold.

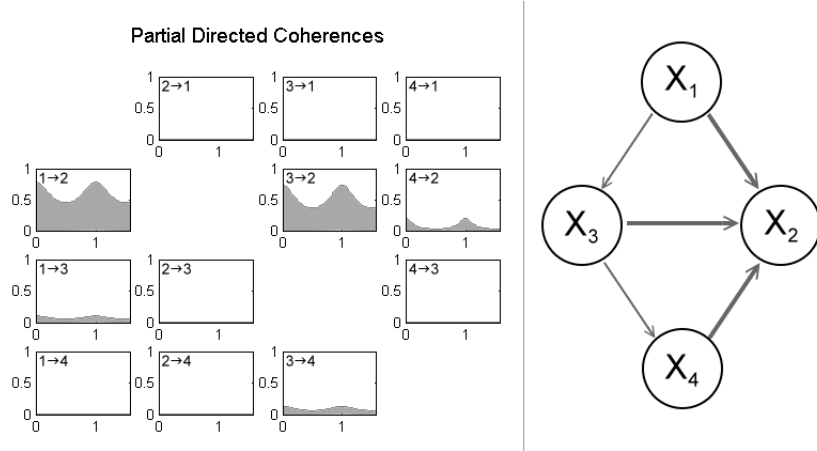


Figure 2.10: **Analyzing the signal model using PDC.** Interactions from 1 to 3 and from 3 to 4 are identified correctly, but PDC values are small. On the other hand it is completely wrong, that channel 2 is influenced by any of the other channels. The arrow from X_2 to X_4 from the correctly identified Figure 2.9 even changed its direction.

2.4 Implementation

2.4.1 The Fast Fourier transform (FFT) algorithm

When taking a close look onto the discrete Fourier transform's definition (2.84), it is obvious that the calculation of a “long”³⁸ vector's Fourier transform is really computation intensive. Although, nowadays most computers are quite powerful, computational time still is an important factor when implementing mathematical methods.

Thus, the Fast Fourier transform algorithm was one key factor in the rapid development of spectral analysis during the last decades. The algorithm itself was developed by Cooley and Tukey in 1965. Obviously, a lot of modifications and extensions exist, but we want to present a quite simple version that makes clear the idea behind FFT.

2.4.1.1 Deriving the algorithm

Assume, that the number of observations n is of the form

$$n = 2m,$$

where m is a natural number, too. The *discrete Fourier transform (DFT)* of a complex vector $x = (x_0, \dots, x_{n-1})' \in \mathbb{C}^n$ is given by

$$a_k = \sum_{j=0}^{n-1} x_j e^{-\frac{2\pi i j k}{n}} = \sum_{j=0}^{n-1} x_j w^{kj}, \quad (2.84)$$

³⁸By “long”, we mean a vector containing at least more than thousand elements. Figure 2.11 shows, how long those computations might need in practice.

if we substitute $w = e^{-\frac{2\pi i}{n}}$ to simplify the notation.

The idea behind the Fast Fourier transform is the reduction of the number of elements in the summation in equation (2.84) from n to $\frac{n}{2}$. Thus, we need to find suitable representations for even numbers n :

$$\begin{aligned} a_{2l} &= \sum_{j=0}^{n-1} x_j w^{2lj} = \sum_{j=0}^{\frac{n}{2}-1} \left(x_j w^{2lj} + x_{j+\frac{n}{2}} w^{2l(j+\frac{n}{2})} \right) \\ &= \sum_{j=0}^{\frac{n}{2}-1} w^{2lj} \left(x_j + x_{j+\frac{n}{2}} \underbrace{w^{ln}}_{=1} \right) \end{aligned}$$

and for odd numbers n :

$$\begin{aligned} a_{2l+1} &= \sum_{j=0}^{n-1} x_j w^{(2l+1)j} = \sum_{j=0}^{\frac{n}{2}-1} \left(x_j w^{(2l+1)j} + x_{j+\frac{n}{2}} w^{(2l+1)(j+\frac{n}{2})} \right) \\ &= \sum_{j=0}^{\frac{n}{2}-1} w^{2lj} \left(x_j w^j + x_{j+\frac{n}{2}} w^j \underbrace{w^{\frac{n}{2}}}_{=-1} \underbrace{w^{ln}}_{=1} \right) \\ &= \sum_{j=0}^{\frac{n}{2}-1} (x_j - x_{j+\frac{n}{2}}) w^j w^{2lj}, \end{aligned}$$

whereas $l = 0, \dots, m-1$. Clearly, if the number of observations n is of the form $n = 2^p$ with $p \in \mathbb{N}$, the idea from above can be applied recursively for p times. The whole concept will be described in more detail in algorithm 2.1.

Algorithm 2.1 The Fast Fourier transform (FFT) algorithm

Inputs: number of observations $n = 2^p$, $p \in \mathbb{N}$, vector $x = (x_0, \dots, x_{n-1})' \in \mathbb{C}^n$

Output: vector $a = (a_0, \dots, a_{n-1})$, which is the discrete Fourier transform of x

function $FFT(n, x)$

if $n = 1$

$a_0 = x_0$

else

$w = e^{-\frac{2\pi i}{n}}$

$m = \frac{n}{2}$

$g_j = (x_j + x_{j+m}) \quad \forall j = 0, \dots, m-1$

$h_j = (x_j - x_{j+m}) w^j \quad \forall j = 0, \dots, m-1$

$(a_0, a_2, \dots, a_{n-2})' = FFT(m, (g_0, \dots, g_{m-1})')$

$(a_1, a_3, \dots, a_{n-1})' = FFT(m, (h_0, \dots, h_{m-1})')$

end

return a

end

2.4.1.2 Computational time savings

We want to discuss the amount of computational time, that might be saved by the use of FFT algorithm compared to the calculation of DFT values as defined in (2.84). Therefore let $C(n)$ denote the number of operations necessary to perform a FFT for a vector of length n . If k denotes a constant, whose exact value does not matter in this context, algorithm 2.1 yields ($n = 2^p$)

$$\begin{aligned} C(2^p) &\leq 2C(2^{p-1}) + k2^p \\ &\leq 2(2C(2^{p-2}) + k2^{p-1}) + k2^p = 2^2C(2^{p-2}) + 2k2^p \\ &\leq 2^3C(2^{p-3}) + 3k2^p \\ &\vdots \\ &\leq 2^pC(2^0) + pk2^p \end{aligned}$$

and further, if k_2 denotes another constant,

$$C(n) \leq nk_2 + \log_2(n) kn.$$

Therefore $\mathcal{O}(n \log n)$ operations are necessary to calculate a FFT for any n -dimensional vector x . Clearly this is a relevant reduction in computational time, as $\mathcal{O}(n^2)$ operations were required to compute the DFT as shown in (2.84)³⁹. Figure 2.11 gives an illustrative example of concrete time savings during a Matlab simulation.

2.4.1.3 Using FFT to compute the autocovariance function and the Periodogram

We already know, that the Periodogram is simply the Fourier transform of the autocovariance function. Thus, it is obvious, that FFT helps to save computational time whenever the Periodogram is needed. On the other hand, the autocovariance function itself might be interpreted as the Fourier transform of some expression, too. We want to derive that expression during this section.

Let $\gamma(k)$, $|k| < n$ denote the sample autocovariance function for any given time series $x = (x_1, \dots, x_n)$. As n need not be of the form 2^p in general, we want to expand this time series. To simplify calculations we further use a centered version of x :

$$y_i = \begin{cases} x_i - \bar{x} & i \leq n \\ 0 & n < i \leq 2n-1 \end{cases},$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ as usual. Obviously, $\bar{y} = \sum_{i=1}^{2n-1} y_i = 0$.

According to definition (2.84), the discrete Fourier transform of y is given by

$$a_k = \sum_{j=1}^{2n-1} x_j e^{-ij\lambda_k},$$

³⁹for any of the n components of the Fourier transformed vector a , $(n-1)$ summations and n multiplications are needed.

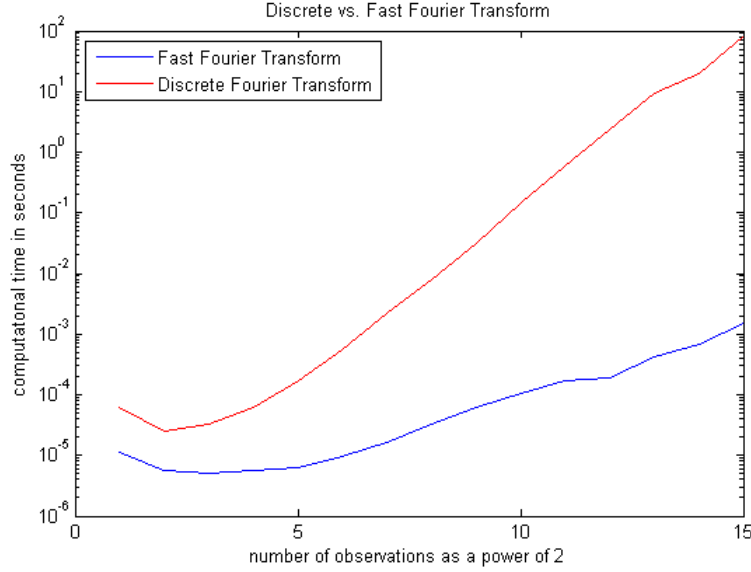


Figure 2.11: **Comparing DFT and FFT.** Time needed to compute the Fourier transform of a complex vector of length 2^p , if either DFT according to its definition (upper, red line), or the FFT (lower, blue line) algorithm is used. If $2^{12} = 4096$ observations are available, ordinary DFT needs approximately one second to be computed, whereas FFT is finished after 0.1 milliseconds.

with $\lambda_k = \frac{2\pi k}{2n-1}$, where $k \in F_{2n-1} \triangleq \{j \in \mathbb{Z} : -\pi < \lambda_j \leq \pi\}$. The Periodograms definition (2.2) yields

$$I_{2n-1}(\lambda_k) = \left| \sum_{j=1}^{2n-1} x_j e^{-ij\lambda_k} \right|^2 = |a_k|^2, \quad (2.85)$$

and, because the Periodogram is just the Fourier transformation of the autocovariance function γ_y (of the vector y), we may write

$$I_{2n-1}(\lambda_k) = \sum_{|j| < 2n-1} \gamma_y(j) e^{-i\lambda_j k} = \frac{n}{2n-1} \sum_{|j| < n} \gamma(j) e^{-i\lambda_j k} \quad (2.86)$$

We know from Theorem 2.9, that

$$\sum_{t=1}^T e^{i(\lambda_j - \lambda_k)t} = \begin{cases} T & \text{if } j = k \\ 0 & \text{else.} \end{cases}$$

Thus,

$$\sum_{j \in F_n} e^{i\lambda_j s} e^{-i\lambda_j t} = \begin{cases} 2\pi & \text{if } s = t \\ 0 & \text{else.} \end{cases} \quad (2.87)$$

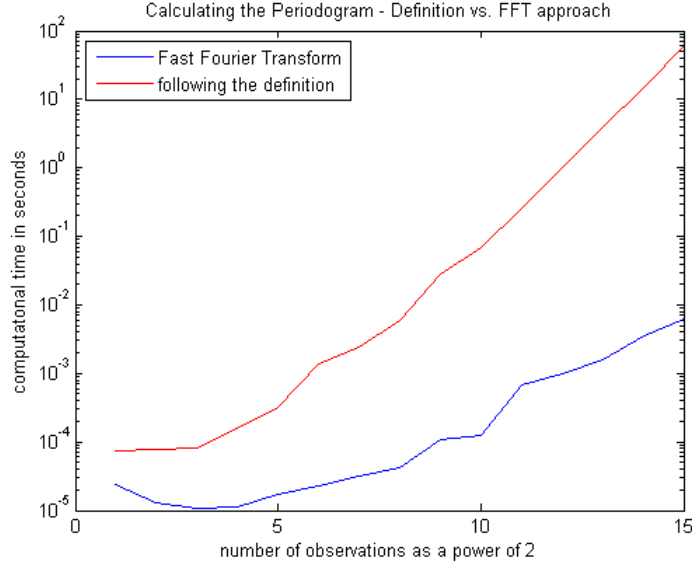


Figure 2.12: **Using FFT to compute the Periodogram.** Computational time used, when calculating the Periodogram either by following exactly the definition (upper red line) or by Fast Fourier transformation of the data vector. If $2^{12} = 4096$ observations are available, calculating the Periodogram according to its definition (2.2) needs approximately one second, whereas calculation is finished after less than one millisecond if FFT is used.

Combining (2.85), (2.86) and (2.87) finally leads to

$$\begin{aligned}
 \sum_{k \in F_{2n-1}} |a_k|^2 e^{im\lambda_k} &= \sum_{k \in F_{2n-1}} \frac{n}{2n-1} \sum_{|j| < n} \gamma(j) e^{-i\lambda_k j} e^{i\lambda_k m} \\
 &= \sum_{k \in F_{2n-1}} \frac{n}{2n-1} 2\pi \gamma(m) \\
 &= 2\pi n \gamma(m).
 \end{aligned}$$

We now have proved that the autocovariance function is the rescaled discrete Fourier transform of the vector $a = (a_1, \dots, a_n)'$, which is itself the discrete Fourier transform of our data vector x . When calculating the Periodogram, this might therefore be realized by applying Fourier transformation for three times.

It is obvious, that FFT would save a lot of computational time compared to the realization of the DFT definition (2.84). In fact, it is even much faster than the calculation of the Periodogram based on its definition (2.2) if n is not too small. Brockwell [5] suggests a minimum number of available observations of approximately 200 to be necessary. Our Matlab simulation in Figure 2.12 shows even more impressive results.

2.4.2 Estimating the parameters of an AR-model

For the estimation of all the parameters of our AR-models, the Matlab module *ARFIT* is used, which is based on a stepwise least squares algorithm, presented by Schneider and Neumaier in [27] and [28]. Given the time series, a lower bound p_{min} and an upper bound p_{max} for the model order, *ARFIT* automatically computes the optimal coefficient matrices and the optimal model order.

As usual, the optimum order of an AR model is chosen by an order selection criterion. The default selection criterion of *ARFIT*, which was in fact used for all calculations in this diploma thesis, is Schwarz's Bayesian Criterion [31], as this has led, on average, to the smallest one-step prediction error of the fitted AR models, compared to several other criteria⁴⁰.

The stepwise least squares algorithm is computationally efficient, especially for higher-dimensional systems. It uses a normalized QR factorization for an AR model of order p_{max} . Optimal parameters are calculated by stepwise downdating of this factorization (i.e. the QR factorization for an AR model of order $p_{max} - 1$ is calculated recursively from the QR factorization of the model of order p_{max} , and so on), instead of calculating a new one for each model order between p_{min} and p_{max} , which would be the typical procedure implemented in AR-model estimation algorithms.

2.4.3 Channel selection

When talking about the analysis of ECoG signals, this always includes a necessary processing of multivariate time series with at least 28 channels. As already explained during the introduction, these channels correspond with electrodes, whereas 6 or 8 electrodes are mounted onto the same band, and a handful of bands is attached on the patient's brain surface. Obviously, as we want to find the epileptic seizure's focus as precise as possible, the distance between two neighboring electrodes is quite small, which leads to correlation between them.

On one hand, we appreciate those coupling effects, as they are exactly what we want to analyze during the diploma thesis. On the other hand, the large number of channels combined with significant correlation between many of them leads to variance-covariance-matrices which are scaled badly. To avoid falsely estimated parameters of the AR model, we have to restrict ourselves to a smaller amount of channels selected for the analysis.

During our work, the set of selected channels was always chosen manually, whereas not less than a third but not more than one half of the available channels have been omitted. Obviously, it is sometimes quite easy to surrender some channels, as the corresponding electrodes are located far distant from any epileptic activity. A quick look on the time series graphs helps to find an approximate choice, fine-tuning is then based on the overall results from the analysis of different patients, seizures and measures.

⁴⁰Lütkepohl presented that comparison in [22]

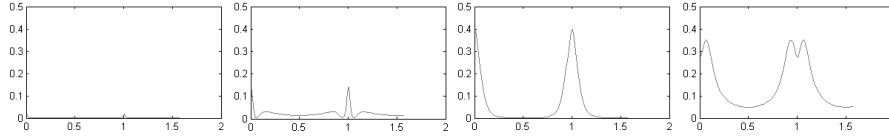


Figure 2.13: **Typical graphs for frequency-dependent measures.** Upper plots yield from GPDC calculations, indicating no or weak interactions in the first two plots and stronger interactions in the third and fourth plot. x -axis values are normalized frequencies, i.e. a value of 1 means 1π .

Of course, the perfect way of channel selection would be given if an automatic algorithm was used. In fact, we have tried a quite simple version of a channel selection algorithm, which is based on an idea of Hartman et al [17] and presented in more detail in Graef [12]. But, unfortunately, that approach seems to be too simple for our demands and leads to really bad results. Thus, none of those results are presented in this diploma thesis.

2.4.4 Windowing and the signal instationarity

As already mentioned in Section 2.3.1.1, every ECoG signal's variance rises during an epileptic seizure. Thus, obviously, time series cannot be stationary over the whole time frame, which leads to the idea to split them into shorter windows of just a few seconds.

Graef unfortunately showed in his diploma thesis [12] (which is based on the same ECoG data we use in our work), that ECoG signals are not even short time stationary, i.e. the assumption of stationary is wrong even within a time frame of just one second for example. On the other hand, results in Section 2.5 will justify the assumption as well as numerous publications⁴¹ do.

During our work with measures in frequency domain, we always have used time windows with three or four seconds length. At the given sampling frequency of 128 Hz, this is equal to 384 and 512 samples respectively.

2.4.5 Derivation of a frequency independent measure and the choice of a suitable threshold

All the measures presented during this section have one thing in common: they all depend on the chosen frequency. In principle, we only need a binary measure, indicating if there is interaction between two channels, or not (clearly, the direction of the interaction should be indicated too, if this is supported by the chosen measure). Thus, we have to “collect” all the information of the given range of frequencies in one measure. As the reader might have expected, that will be done by integration.

Figure 2.13 shows a typical graph of any of the frequency dependent measures, which we applied to ECoG data. Surprisingly, graphs are similar for

⁴¹see [10, 21] for example

PSC, PDC and GPDC. On the x-axis, we have normalized frequencies, whereas a range between 0 and $\frac{\pi}{2}$ has been chosen. A rise in PDC values always takes place at low frequencies (approximately in the interval $I_1 = [0, 0.2]$) and around the normalized frequency 1 (approximately within $I_2 = [0.8, 1.2]$).

Clearly, we have four plausible areas, where integrals could be calculated: I_1 , I_2 , a combination of both or the whole range from 0 to $\frac{\pi}{2}$. In fact, we tried all of those possibilities and could not find significant differences. Thus, we decided to use all the available information and will therefore always integrate from frequency 0 to $\frac{\pi}{2}$.

After the integral has been calculated, it is necessary to decide, whether its value is big enough to indicate interaction between the selected channels. Thus, a threshold has to be chosen. For simplicity reasons, the choice will be made manually during our analysis, although it is a bit heuristic. We always start with a very low threshold and raise it until most of the unwanted arrows have disappeared. Clearly, only the arrows remain, that coincide with the highest values of the chosen measure, which justifies the whole procedure in some way.

Obviously, for an even more exact analysis, the calculation of a frequency dependent confidence interval would be necessary for each graph and measure. One possible approach could be based on the idea of surrogate data, which is presented for example in [32]. As that calculation is quite complex, it has not been implemented within our analysis.

2.5 Results

We want to start with some remarks. During this section we will often use plots with small circles and crosses (we call them brainplots). Both represent electrodes, at which ECoG signals are measured, whereas circles stand for channels selected for calculations and crosses represent ignored channels.

The reader has to take care especially about one fact: At all brainplots we look onto the brain from above (“median view”). Thus, when talking about the right half of the brain, this is in fact the left half of our brainplot and vice versa.⁴²

2.5.1 Partial coherence (nonparametric)

Partial spectral coherence is the only undirected measure we used to analyze epileptic seizures. Furthermore, it is somehow a “classic measure” compared to PDC or Granger causality. As it indicates in fact not more than correlation in frequency domain, before calculating the presented Figures we were quite skeptic, if it really could help to track the propagation of epileptic seizures.

To anticipate the conclusion of the PSC analysis, some of the results were better than awaited. On the other hand, none of them was good enough to mark one single focus. We could not base our decision, where a patient would have to be treated for surgery, on PSC results only.

2.5.1.1 Patient 1, seizure 1

We want to start the presentation of partial spectral coherence results with patient 1, which is, as we will see later, in fact the only patient where useful results could be obtained using PSC.

According to medics, the seizure should start at channel 25 and proceed to the surrounding channels 26, 17 and 18 between seconds 25 and 30 after kick-off. We can see in Figure 2.14, that the area of the seizure’s focus is obtained correctly. In all six time windows presented, there is heavy action with channels 17, 18 and 25 involved. Thus, PSC would suggest the seizure’s focus to be somewhere between those channels.

On the other hand, PSC indicates several interactions that should not be taken into account because they do not result from epileptic activity. While coupling effects between channels 27 and 28 could be identified with all measures used in this diploma thesis, those on the left side of the brain have not. Especially the high values of PSC between channels 9, 10 and 11 starting in the third time window from Figure 2.14 cannot be based on epileptic activity, because the corresponding electrodes are located far away from the focus channels.

Our hypothesis now is, that PSC “prefers” neighboring channels, i.e. obtained values tend to be higher, if the distance between the electrodes is low. Results from patient 2 presented in the next section will prove that hypothesis

⁴²the interested reader might find an introduction to graphical modeling in [9]

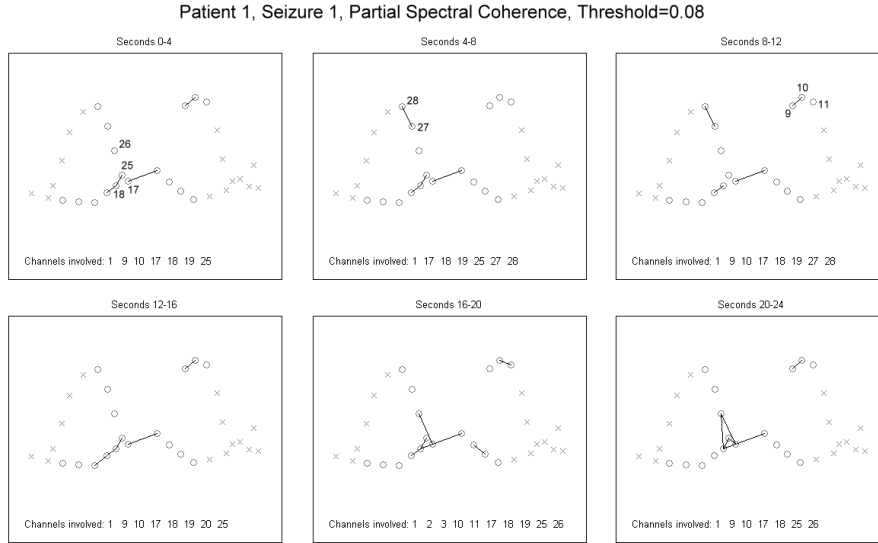


Figure 2.14: *Resulting brainplots for partial coherence using the first 24 seconds of data from patient 1, seizure 1. Lines indicate interactions between affected channels. Thus, partial spectral coherence might locate the focus near channels 17, 18 and 25 which is in accordance with the medics' opinion.*

in an impressive way.

Due to those doubts, the reader might take Figure 2.14 as a quite good result. We want to notice, that it is in fact not that hard to find an approximate location for the focus here. All the channels involved in epileptic activity (according to the medics) are located nearby each other. Comprising the fact that PSC does not give any information about the direction of interactions, results are not that powerful anymore. Some could also interpret channel 19 or channel 1 to be the seizure focus - one of them is located far distant from seizure onset, the other one is located even on the wrong side of the brain.

2.5.1.2 Patient 2, seizure 3

We will focus ourselves on the left half of the brain, as interactions on the other half are negligible, according to the medics opinion and our results from different measures. Furthermore, the resulting figures from this analysis show the problem of our measure derived from PSC perfectly.

According to Figure 2.15, we cannot find the seizure's focus based on PSC analysis. Seizure onset should take place at channels 7, 8 and 9. Clearly, we can identify interactions between those channels, but they are not significantly higher than the interactions between channels 1 and 2, and 3 and 4 respectively.

Epileptic activity should further proceed to channels 10 and 11 after 22 seconds. Actually, we could find high values of PSC between channels 10 and 11 after second 20, but those values were not higher than in the second or third

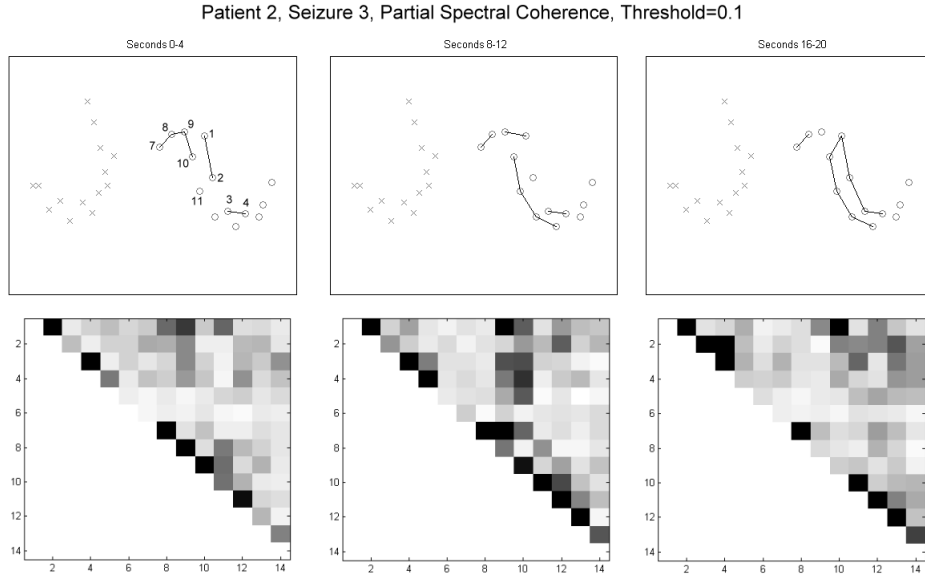


Figure 2.15: *Resulting brainplots for partial coherence using data from patient 2, seizure 3. Squares in plots on the lower half indicate the intensity of interactions between the channels involved, whereas darker values indicate higher values of PSC. All graphs show, that PSC values tend to be higher for neighboring channels.*

time window shown in figure 2.15 for example. And again, coherences between those two focus channels are not significantly higher than between a lot of other channels on the left brain half. In fact we can see nearly the whole electrode bands connected due to high PSC values.

As mentioned before, when analyzing the results from patient 1's data, we see, that PSC values between neighboring channels tend to be high, although there is no epileptic activity on the channels involved. To clarify that hypothesis, we present a second diagram for each time window in figure 2.15, where darker squares correspond with high values of PSC and lighter squares indicate, that there might be only weak interaction between the channels concerned. Obviously, most black squares are located on the line directly above the square's diagonal, i.e. they indicate interactions between neighboring electrodes. Furthermore, the dark squares not located on that line, e.g. between channels 1-10 or 2-11, mostly correspond to neighboring channels, too. The difference between the channel numbers is higher than 1, just because the electrodes are mounted on different bands (A detailed description of the channels and their numbers may be found in the appendix.).

Figure 2.16 tries to give an explanation for that behavior of PSC. We have drawn a few seconds of the time series graphs for channels 1 to 4. Obviously, the graphs for the chosen channel couples are nearly identical. As PSC is in fact just a correlation in frequency domain, high PSC values are not really surprising.

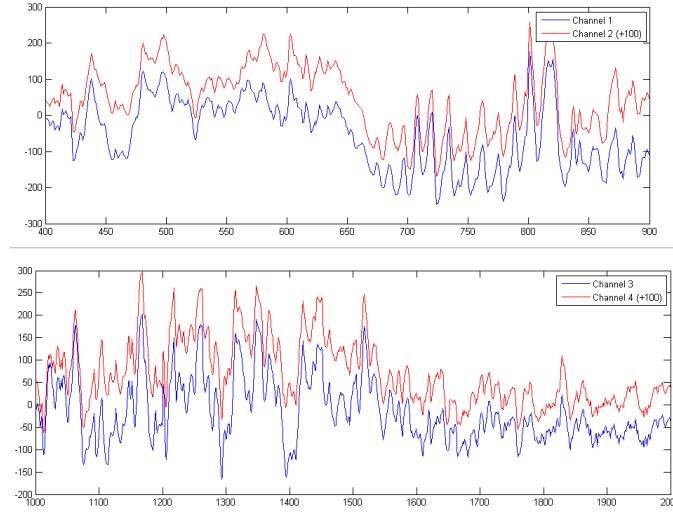


Figure 2.16: **Channel pairs with high PSC values, but no epileptic activity.** The graphs of channels 1 and 2, and 3 and 4 respectively, are nearly identical within the shown time frame. (128 samples (drawn on the x-axis) are equal to one second). As PSC is some kind of correlation, this leads to high PSC values.

Of course we cannot give a positive summary for all the PSC results presented. In fact, the Figure shown for patient 2, seizure 3 is even far the worst compared to all the other measures we have used during our analysis in this diploma thesis.

Mostly, interactions suggested by the medics could be confirmed by PSC. But on the other hand we were not able to find the focus channels - PSC values simply do not differ from the values obtained from a lot of different channel pairs. Of course, it is our goal to realize that direction. We want to present a mathematical method to track epileptic seizures and indicate channels as focus channels. Unfortunately, PSC cannot help to achieve this aim.

2.5.2 Partial directed coherence (parametric)

Although we already discussed several problems of the partial directed coherence, we want to apply that measure to our ECoG data. We choose three epileptic seizures from three different patients and present the resulting Matlab calculations.

Clearly, none of the calculations delivered great results, but some of them were much better than expected. They all fulfill at least one expectation: they justify generalized PDC's definition.

2.5.2.1 Patient 4, seizure 1

We want to start our evaluation of PDC results with patient 4's first recorded seizure. According to the medics opinion, that seizure has its focus on the right hemisphere at channels 27 and 28. Channels 21 and 22 should be affected about

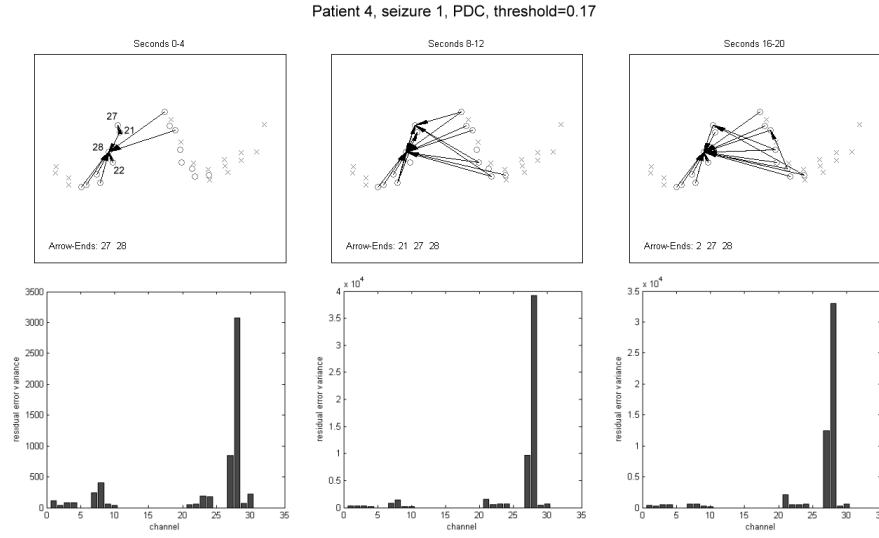


Figure 2.17: *Resulting Brainplots for PDC and residual error variances for patient 4, seizure 1. Residual variances are the highest for focus channels 27 and 28, therefore many arrows point towards the focus. Furthermore, as suggested by medics, an expansion of the seizure to the left hemisphere can be obtained as more and more arrows start on the brain's left side.*

one second later, and the whole left side of the brain is concerned 20 seconds after the beginning of the epileptic activity.

Results in Figure 2.17 show, that the residual error variance is the highest for the focus channels 27 and 28. Variances of other channel's residuals are significantly lower, at least 20 times and 10 times respectively. Thus we have lots of arrows (indicating interactions according to PDC) pointing towards the seizure's expected focus channels, which is in perfect accordance with examples 2.33 and 2.34.

That result seems to be plausible. The time series's instationarity increases if epileptic activity becomes "heavier", therefore its explanation by any AR-model gets worse, which results in a rise of the residuals and their variance. As PDC does only work (see example 2.33) if all the residual's variances are approximately equal, results are not correct here - with arrows pointing towards badly approximated channels.

When comparing the three time windows presented in Figure 2.17, we can see a rise in arrows starting on the left hemisphere in the second and third period, compared to the first four seconds observed, which might coincide with the seizure's expansion to the brain's left half. Unfortunately, PDC indicates nearly no interactions between neighboring channels on the left hemisphere, which makes those results implausible.

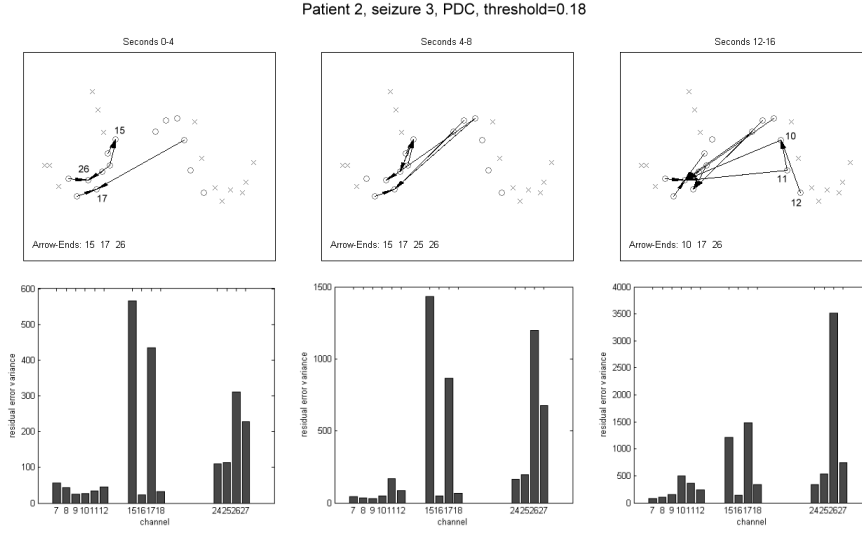


Figure 2.18: *Resulting Brainplots for PDC and residual error variances for patient 2, seizure 3. Interactions with channel 26 involved are identified correctly, but arrows pointing towards channel 17 cannot be explained. Although, channel 17 is located next to channel 26, none of our other measures indicated it to be a focus channel, neither did the medics.*

2.5.2.2 Patient 2, seizure 3

According to the medics, the seizure's focus should be located at channels 26 and 27. Expansion to channels 10 and 11 should take place after 12 seconds, with channel 12 affected a bit later.

Now residual error variances, as presented in Figure 2.18, are significantly higher for four channels compared to the rest. Two of them (26 and 27) are focus channels according to the medics opinion, and the third one (channel 15) seems to affect surrounding channels according to results from Granger causality and GPDC. Furthermore, channel 15 is a focus channel for the same patient's fourth observed seizure. But, unfortunately, channel 17, which is the fourth channel with higher residual error variances here, could not be identified as a focus channel during any of our analysis.

Thus, arrows indicating interactions on the right hemisphere around channels 26 and 15 are plausible, expansion of the epileptic activity towards the left hemisphere (channels 10, 11 and 12) is indicated perfectly in the third time window, but all the arrows pointing towards channel 17 are in principle not valid. Clearly, the electrode corresponding to channel 17 lies next to the electrode for channel 26 on the brain's surface and therefore could be affected as well. But, as GPDC and Granger causality will indicate better results (compared to the medics opinion), PDC seems to be a poor measure to locate the seizure's focus.

Concluding, analysis for patient 2, seizure 3 showed, that poor tracking of

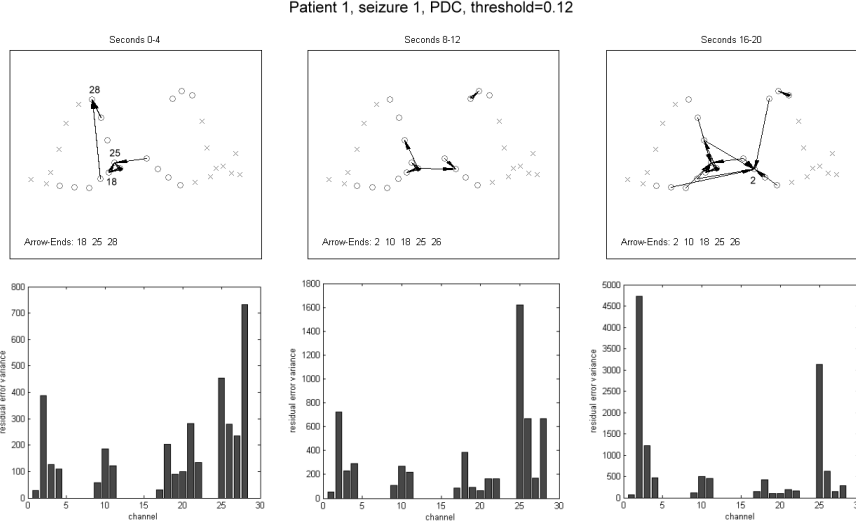


Figure 2.19: **Resulting Brainplots for PDC and residual error variances for patient 1, seizure 1.** Variances are similar during the first four seconds and brainplots confirm the medics opinion. Later, the residual error variance for channel 2 is the highest, leading to lots of arrows pointing towards channel 2, which cannot be confirmed by either the medics opinion or any of our other methods.

ECoG signals need not be based on epileptic activity. Clearly this result is no surprising one, as instationarities may be caused by an infinite number of reasons.

2.5.2.3 Patient 1, seizure 1

The first of the three time windows presented in Figure 2.19 seems to be plausible. Medics suggest channel 25 to be the focus channel of that seizure, and channel 18 to be a focus channel for different seizures of the same patient. Activations implicating channel 28 have been approved at least by Granger causality, but not by the medics.

Problems start in the second window. According to the medic's opinion, the epileptic seizure stays local around channels 17, 18, 25 and 26, whose corresponding electrodes are located next to each other on the brains surface. In contrast to that opinion (which has been approved by our analysis when different measures were used), arrows tend to start and end on the left hemisphere.

In the third time window, one would even suggest channel 2 to be the seizure's focus. Again, clearly, all the arrows pointing towards channel 2 are based on a very high variance of the corresponding residuals. And, as discussed for patient 2, seizure 3, the high variance is not based on epileptic activity.

While the first two presented results in this section did either suggest the correct focus, or at least a channel located next to the focus, we have a com-

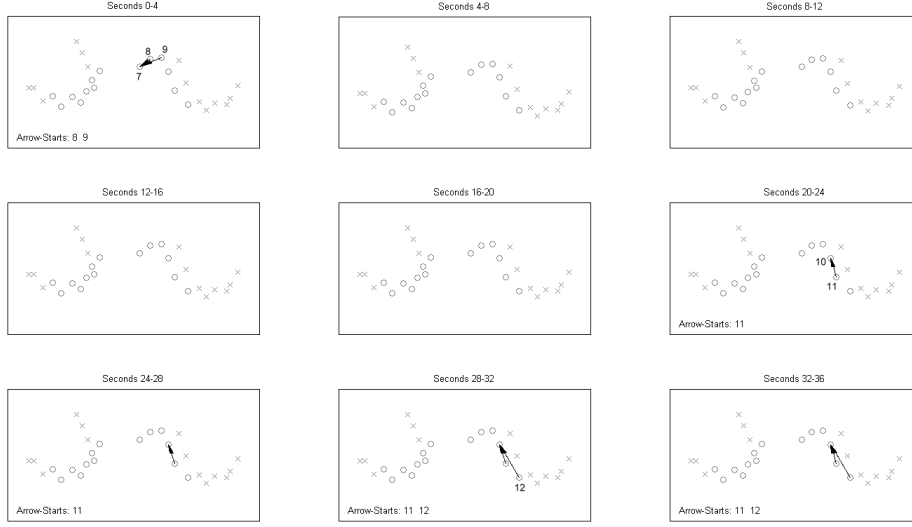


Figure 2.20: *GPDC results for patient 2, seizure 1. A threshold of 0.13 has been chosen. Upper plots represent the first 36 seconds after seizure onset.*

pletely wrong result here. Finally, it is clear, that PDC cannot be the right measure to track epileptic seizures.

2.5.3 GPDC results for patient 2

We want to conclude Chapter 2 with the results from generalized PDC. As awaited, they were far the best we could obtain from any measure in frequency domain we presented in this work. Only Granger causality will be able to deliver an even more helpful outcome.

At first the reader might be confused by the sequence of our result's presentation. In fact, it will simplify interpretation, as some of the indicated interactions between channels not suggested by the doctors might possibly be an important part of other seizures from the same patient, too. Thus, it is quite assumable that those interactions are in fact correct and the mathematical analysis is even superior to the medics opinion in some cases.

2.5.3.1 Seizure 1

We start the presentation of GPDC results with a seizure, that is quite out of normal behavior, as none of the other seizures from the same patient showed a similar result.

According to the medics, our results seem perfect. Seizure onset should take place at channels 7 to 9, which is in accordance with the first plot in Figure 2.20, where arrows from channels 9 to channels 8 and 7 were calculated. Doctors further see a diffusion of epileptic activity to channels 10 and 11 after approximately 22 seconds. Again, the sixth plot of Figure 2.20, which represents seconds 20-24 after seizure onset, approves that hypothesis, represented by an

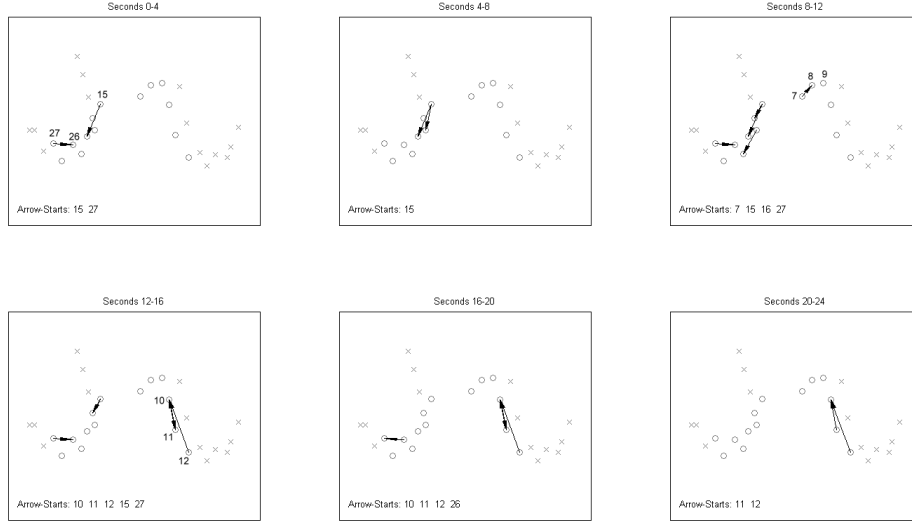


Figure 2.21: **GPDC results for patient 2, seizure 3.** A threshold of 0.17 has been chosen. Upper plots represent the first 24 seconds after seizure onset.

arrow pointing from channel 11 to 10.

Although GPDC results and the opinion of the doctors fit together perfectly, we are not completely happy with that plots. First of all, it is quite unrealistic that there is no epileptic activity between seconds 4 and 20 after seizure onset. And, second, all the other seizure's focuses from patient 1 will be located on the other half of the brain, either at channel 15 or around channel 26. Thus, the outcome of seizure 1's analysis is not very helpful if the exact location of the focus has to be located.

Furthermore, the chosen threshold (0.13) is far lower than for all other seizures presented for patient 2, because GPDC values are lower. Thus, one could argue that there is in fact no activity according to GPDC in the first five plots (i.e. 20 seconds) of Figure 2.20. On the other hand, that argumentation might even help us, as we do not really want to locate the focus on the left hemisphere, as explained before.

Anyway, our first results seem to be at least promising.

2.5.3.2 Seizure 3

Seizure 3 is somehow the “typical” seizure from patient 2, as three out of four seizures from that patient proceeded similarly.

The medics told us, that the seizure's focus should be located around channels 26 and 27. We can see in the first plot of Figure 2.21, that GPDC indicates interaction between those channels. Furthermore, channel 15 is causing surrounding channels during the first approximately 15 seconds. This action has not been identified by the medics for seizure 3. But in fact, it has been identified

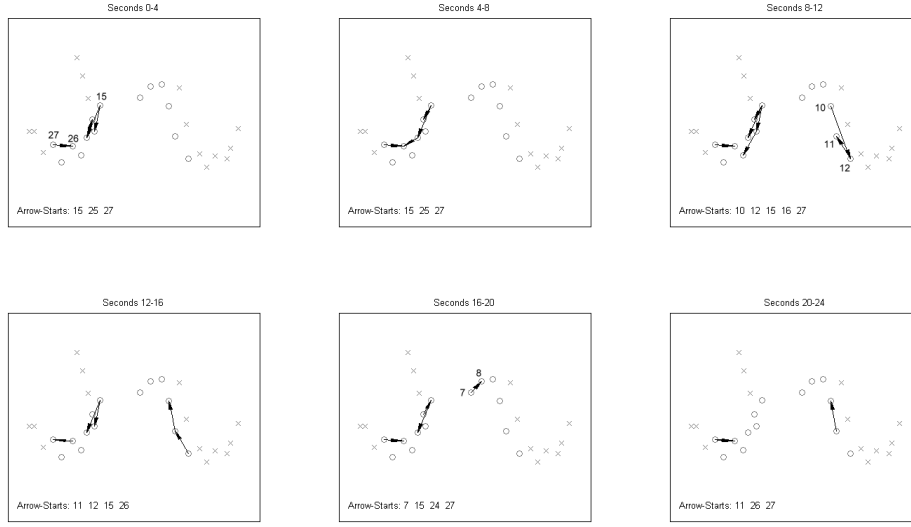


Figure 2.22: *GPDC results for patient 2, seizure 4. A threshold of 0.17 has been chosen. Upper plots represent the first 24 seconds after seizure onset.*

for the same patient's seizures 2 and 4. Thus, it is quite realistic, that epileptic action takes place at that channel, too.

After 12 seconds, we recognize an expansion of epileptic activity to channels 10, 11 and 12. Again, that coincides with the information received from the doctors. The only disturbing arrow drawn in Figure 2.21 might be the one from channel 7 to 8 in the third plot presented, although we already mentioned for seizure 1, that those channels might be involved in epileptic activity, too.

2.5.3.3 Seizure 4

Results presented for seizure 4 from patient 2 in Figure 2.22 look like a copy from seizure 3 on the first sight. Activity starts at channels 15 and 26 and proceeds to the left hemisphere to channels 10, 11 and 12 some time later.

In fact, there is a difference between seizures 3 and 4 in what the doctors told us. According to their opinion, seizure onset is at channel 15 only, while channels 24 to 27 are following after 5 seconds. Fortunately, we could reproduce that hypothesis by simply raising the chosen threshold from 0.17 to 0.2, but we decided to keep the same threshold for seizures 2, 3 and 4 for comparison reasons.

Two further observations have to be mentioned here. First of all, we see interaction between channels 7 and 8 (like for seizures 1 and 3) again, although the medics did not. And second, action on the left hemisphere around channels 11 and 12 is indicated earlier than awaited. Furthermore, it is disappearing in the fifth plot and appearing again one plot later. We think, that this early action is indicated wrongly, as it could not be accomplished by any of the Granger causality measures.

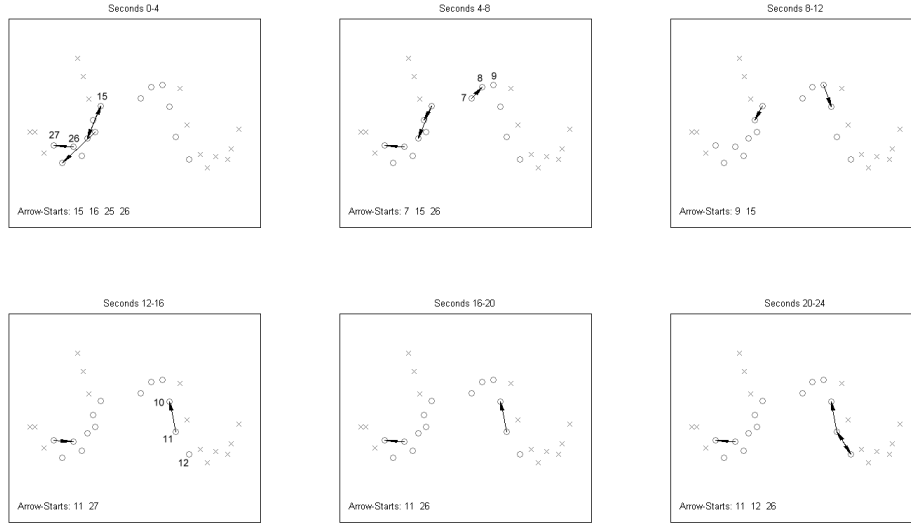


Figure 2.23: *GPDC results for patient 2, seizure 2. A threshold of 0.17 has been chosen. Upper plots represent the first 24 seconds after seizure onset.*

2.5.3.4 Seizure 2

Seizure 2 is in some way a special one. Medics assumed the seizure to start at channels 26 and 27, but found epileptic activity at the same time nearly on the brain's complete right side.

Kindly, results presented in Figure 2.23 are much clearer than that. They look quite similar to Figures 2.21 and 2.22 from the two seizures presented before. The only arrow we have not seen in one of the first plots of any figure, is the one pointing from channel 16 to 18. Fortunately, this arrow is “weaker” than all the others calculated for the first four seconds, i.e. it disappears if the threshold is raised from 0.17 to 0.21.

As we have already seen at the two seizures presented before, epileptic activity expands to the brain's left half. For seizure 2, that should happen after around 14 seconds. GPDC obviously indicates that action, but a quite interesting thing happens before that. First of all, we see the usual activity with channels 7 and 8 involved during seconds 4 to eight. Then, the seizure “moves downwards” the electrode band, to channel 9 in the third window and finally to channels 11 and 12 as suggested by the medics.

Concluding, the focus has its onset on the right hemisphere. After some seconds, it moves to the left half, and we find an activity at the electrode that is located nearest to the right half - channel 7. In fact, it sounds realistic, that the seizure does not “jump” directly from channel 15 or 26 to the far distant area between channels 10 and 11.

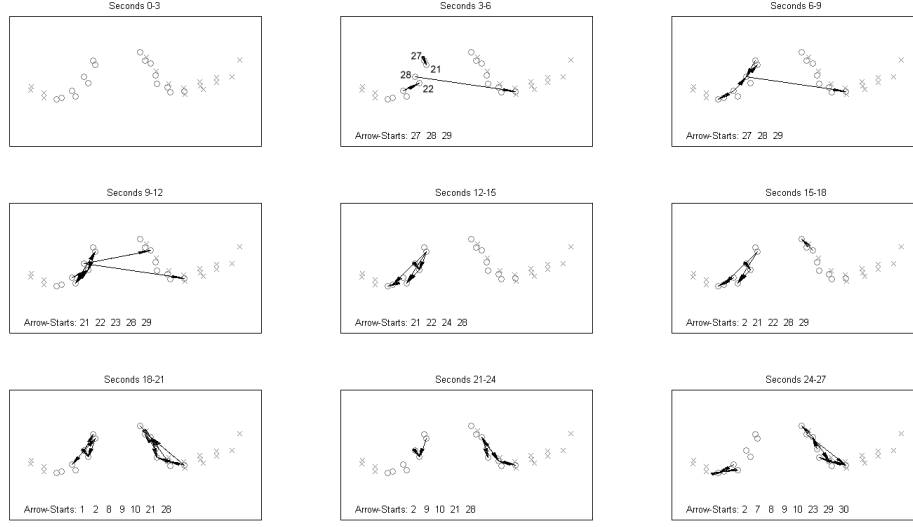


Figure 2.24: *GPDC results for patient 4, seizure 1. A threshold of 0.14 has been chosen. Upper plots represent the first 27 seconds after seizure onset.*

2.5.3.5 Conclusion

As presented, all of the GPDC results for patient 2's seizures were quite good. Only a few interactions might have been indicated falsely and none of them were found in the first few seconds of the seizure, which is in fact the time frame we are mostly interested in, if we want to find the seizure's initial focus.

Unfortunately, GPDC does not suggest an exact position of the patient's epileptic center. It could be located either at channel 15, which might be more likely when looking at the GPDC results, or at channel 26, which was slightly preferred to be the focus by the medics. As the two channels are located far distant, consequently, the patient has not been operated yet.

2.5.4 GPDC results for patient 4

2.5.4.1 Seizure 1

As the medics told us, seizure 1 should start at channels 27 and 28, proceed to 21 and 22 only one second later and expand to the whole left brain side another 13 seconds after that.

All channels indicated as focus channels are located nearby each other, which should help to find clear results within our analysis.

Surprisingly, GPDC indicated absolutely no action within the first time window (i.e. the first three seconds after seizure onset) in Figure 2.24. Of course, we tried to lower the chosen threshold, but it was impossible to find an appropriate value, i.e. a threshold that really indicates interaction between channels and not just noise.

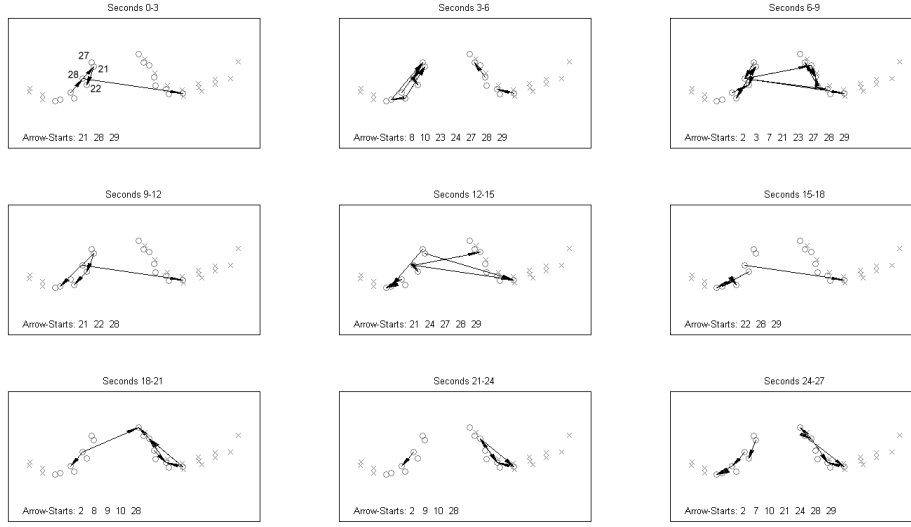


Figure 2.25: *GPDC results for patient 4, seizure 2. A threshold of 0.14 has been chosen. Upper plots represent the first 27 seconds after seizure onset.*

It could be, that the delay of the seizure's onset is based on our assumption of stationarity, which does of course not hold, as we discussed earlier. But following, a similar delay will be visible if partial Granger causality is used in Chapter 3, which is in fact not based on a simple AR-model.

During the remaining windows, tracking of the epileptic seizure works quite good. The arrows' starting points, which coincide with the focus of epileptic activity in our interpretation, are located around channels 21, 22, 27 and 28 as awaited. Furthermore, a lot of arrows point to the brain's left half, which sounds plausible, as the seizure should in fact proceed to the left side.

We only could criticize here, that the timing of action indicated by GPDC is not perfect. Interactions from the right to the left brain half are indicated immediately after brain start, whereas coupling effects between channels on the left half are noticeable after approximately 18 seconds. Obviously, none of that values coincides with the doctor's value of 14 seconds.

2.5.4.2 Seizure 2

Seizure 2 proceeds similarly to seizure 1. The doctors' descriptions for both seizures are identical, with the only difference, that expansion to the left side should take place 12 seconds after seizure onset, which is 2 seconds earlier than for the first seizure.

Again, the area of the focus is located around channels 21, 22, 27 and 28. Immediate interaction between the two brain halves is visible in Figure 2.25, with a lot of action on the whole left hemisphere some seconds later than expected.

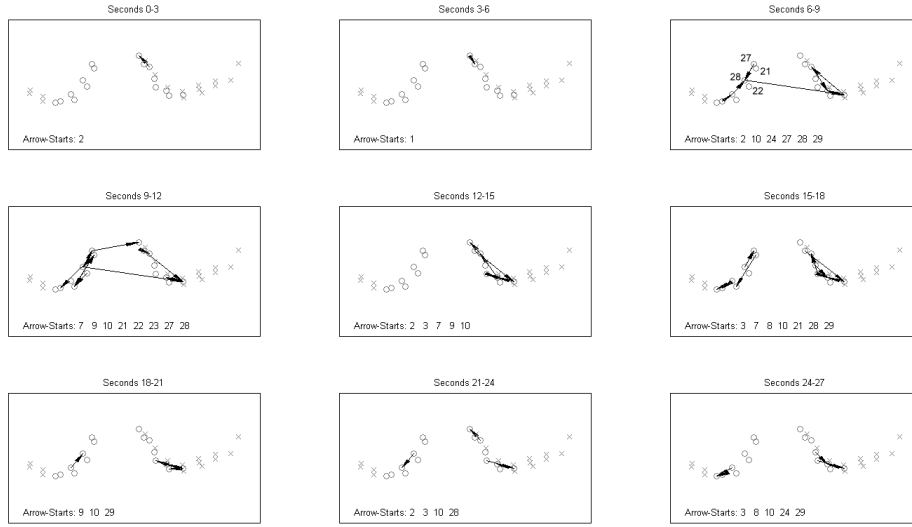


Figure 2.26: *GPDC results for patient 4, seizure 3. A threshold of 0.14 has been chosen. Upper plots represent the first 27 seconds after seizure onset.*

2.5.4.3 Seizure 3

Seizure 3 could be compared with the second seizure of patient 2. The doctors were not able not provide an exact diagnosis based on visualized ECoG data only. According to their opinion, all channels are affected immediately after seizure onset.

Based on that hypothesis, our result in Figure 2.26 is extremely surprising. Seizure's onset is delayed even more than for seizure 1. In fact, no activity was found by GPDC within at least the first five seconds. It could be, that this delay can be explained with a lot of interaction over the whole brain's surface, which would lead to instationarities and would further mean, that our AR-model is not suitable. On the other hand, the length of the delay seems to be too long for such a simple explanation.

Beyond that, results seem to be normal at the plots 3 to 9 of figure 2.26. Again, we have lots of arrows starting around the expected focus channels 21, 22, 27 and 28. Furthermore, there is immediate activity between the two halves and even on the left brain side. Unfortunately, we could therefore not base a clear interpretation only on that figure. But, as all the GPDC results presented for patient 4 look similar, it is quite obvious, that the same interpretation holds here, too.

2.5.4.4 Conclusion

Like for patient 2, tracking of all the seizures worked quite good, with a similar behavior noticed for all of them. We located the focus in the area between channels 21, 22, 27 and 28, which is in accordance with what the doctors told us. When taking a closer look at the three figures presented for patient 4, we

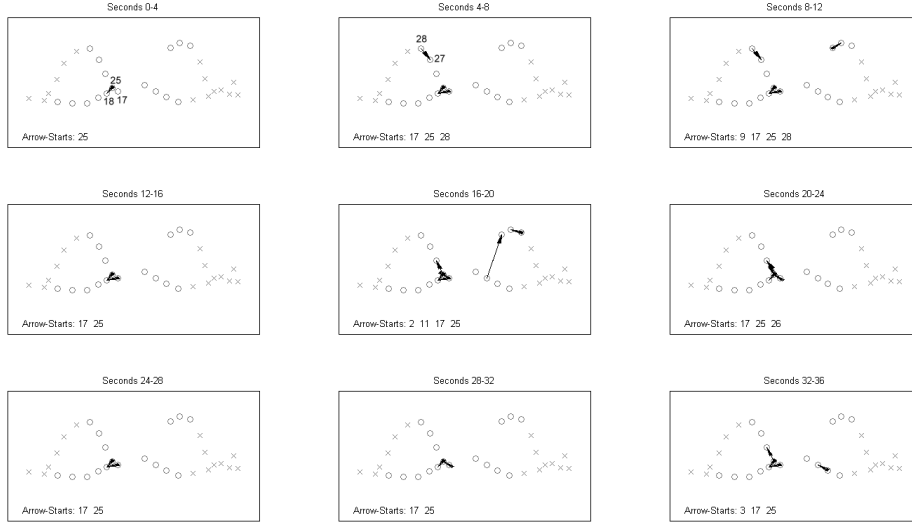


Figure 2.27: *GPDC results for patient 1, seizure 1. A threshold of 0.14 has been chosen. Upper plots represent the first 36 seconds after seizure onset.*

could even try to interpret channel 28 as the single focus: most of the arrows pointing from the right to the left brain half start at this channel. As channel 28 is in fact far distant from the other side of the brain, of course we have to assume, that there is heavy activity around this channel.

The only point we are not happy with is the timing of some activities. Seizure onset was delayed for two of three seizures analyzed, compared with what doctors told us. Furthermore, a lot of action on the left brain half was mentioned too late (for the first two seizures) while interaction between the two halves took place earlier than expected.

Anyway, we think that all the results presented here are quite pleasing, based on the simplifying assumption of stationarity.

2.5.5 GPDC results for patient 1

2.5.5.1 Seizure 1

Doctors believe, that epileptic activity starts at channel 25, proceeds to channel 26 after 25 seconds and further to channels 17 and 18 another five seconds later.

As presented in the first time window of Figure 2.27 (which represents the first four seconds after seizure onset), GPDC localizes the focus correctly. Arrows indicating high values of generalized partial directed coherence start at channel 25 in all presented plots.

Apart from that, results are not as pleasing as hoped for. Channel 17 is indicated to be an origin of epileptic activity not as recently as 30 seconds after seizure onset, but already in the second plot from Figure 2.14, which is more

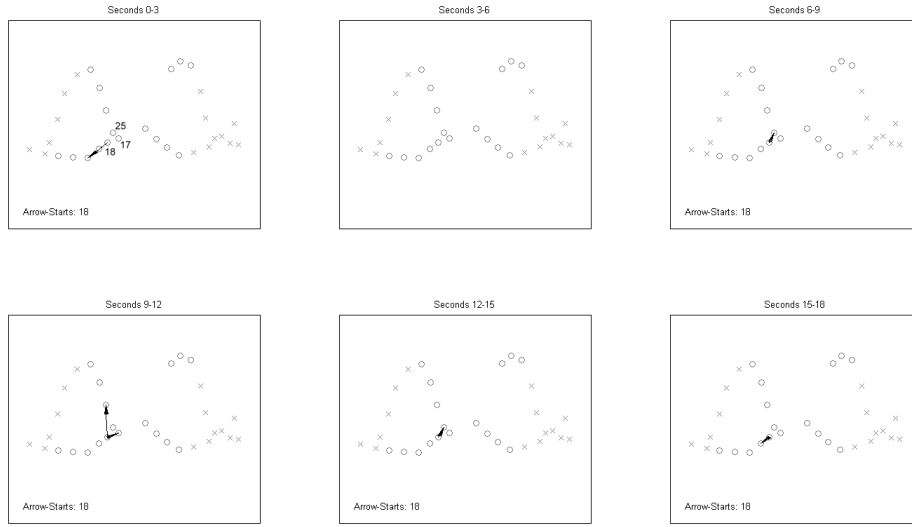


Figure 2.28: *GPDC results for patient 1, seizure 2. A threshold of 0.13 has been chosen. Upper plots represent the first 18 seconds after seizure onset.*

than 20 seconds too early. Moreover, no change in the behavior of channel 18 is noticeable due to GPDC after 30 seconds, which contradicts the doctor's opinion, too. On the other hand, it is quite likely, that epileptic activity proceeds from the seizure's focus to the neighboring electrode. Thus, GPDC visualizes effects, which the doctors could not see. In fact, our thesis will be accomplished by Granger causality and the fact, that 18 is a focus channel for seizure 2.

Furthermore, we see a lot of activity in the plots of Figure 2.27, that is not nice to see. First of all, GPDC shows channel 28 involved in epileptic activity quite often. It is not clear, why the activity should proceed to such a far distant electrode without affecting electrodes somewhere in the middle. Surprisingly, those arrows will appear again if they are based on some kind of Granger causality and have also appeared while PSC was used. We suggest, that there is in fact activity starting from channel 28, but not an epileptic one.

2.5.5.2 Seizure 2

According to the medics opinion, seizure 2 proceeds slightly different from seizure 1. Seizure onset should take place not only at channel 25, but on the neighboring channel 18, too. Doctors further noticed absolutely no epileptic activity between seconds 4 and 7, and a restart at channels 18, 25 and 26 after that break.

To visualize this hypothesis best, we decided to reduce the chosen window length compared to seizure 1 from four to three seconds. Thus, of course, we also had to reduce the chosen threshold slightly (from 0.14 to 0.13).

Results presented in Figure 2.28 help to track the doctors' description of the seizure quite good, including the expected break in time window 2. Surprisingly,

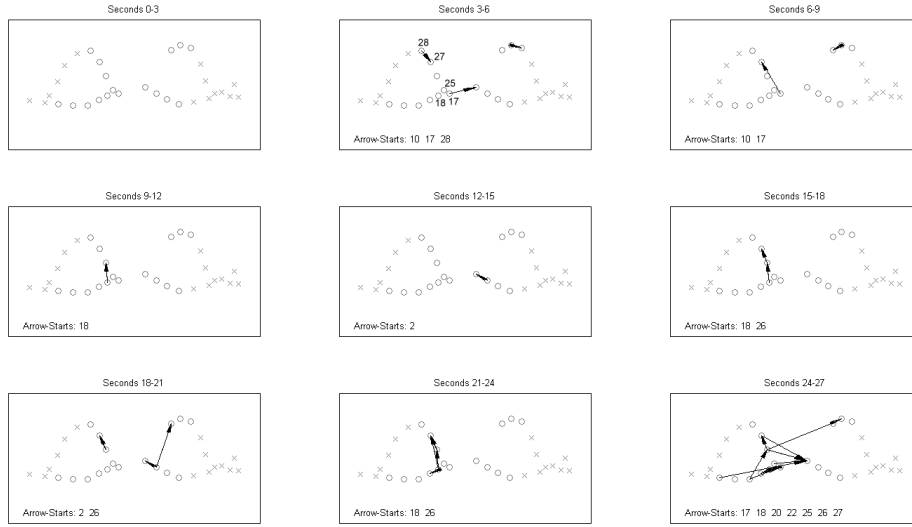


Figure 2.29: *GPDC results for patient 1, seizure 3. A threshold of 0.11 has been chosen. Upper plots represent the first 27 seconds after seizure onset.*

channel 18 is indicated by GPDC to be the initial focus, but arrows starting at that channel at least point towards channels 25 during the first three time frames after restart.

Obviously, there is extremely sparse activity in all the presented plots, which would in fact not change if the threshold was a little bit lowered. We want to mention here, that a lot of activity would be indicated around the focus channels 18, 25 and 26 after approximately 40 seconds, and looking to the time series graphs approves, that the activity is actually epileptic. But, as the doctors were not interested in that activity, we omit further details here.

2.5.5.3 Seizure 3

Seizure 3 should take place similarly to the first and second one presented for patient 1. Doctors localize the seizure's focus at channels 18 and 25, with channel 26 further involved after about 11 seconds.

Unfortunately, GPDC results are really bad in Figure 2.29. Seizure onset is far too late (in the second time window) and at a channel (17) that should not really be involved in epileptic activity. Furthermore, we have activity starting at channel 28 like for seizure 1, and even activity on the left brain side. Obviously, it is impossible to distinguish a focus channel based on GPDC plots here.

As for seizure 2, we noticed a lot of activity around the focus channels later in the seizure's progress. But again, doctors did not care about that fact.

2.5.5.4 Conclusion

Results for patient 1 were quite good for the first two seizures analyzed. The seizure's focus was localized correctly in the area between channels 17, 18 and 25, with 18 and 25 slightly favored. As this is a quite small area, and coincides with the doctors' opinion, we think that it would be possible to base an operation on that results.

On the other hand, tracking of epileptic activity apart from the localization of the focus did not work. We found interaction immediately after seizure onset, although it should take place half a minute later.

Furthermore results for seizure 3 were the worst during all the GPDC analysis. Neither the focus was located correctly, nor the tracking of the epileptic seizure did work in the following time windows.

Summarizing, based on the simplifying assumption of stationarity, we think that results are quite good - as, out of 10 analyzed seizures only one figure presented was useless.

Chapter 3

Localizing the focus of epileptic seizures using Granger causality

In this part we introduce another approach to analyze the ECoG-data, which is based on the definition of C. W. Granger (1969)[15], the so called Granger causality. The basic idea behind this concept is that the cause has to happen before the effect: If past values of the variable y help predicting the variable x , x is caused by y . With this concept some sort of directed flow between connected variables can be identified.

3.1 Bivariate Granger causality

In practice there exist three different concepts to deal with the abstract original definition of Granger. The first of them is called “bivariate Granger causality”.

3.1.1 Definition

Before introducing more appropriate Granger causality approaches, we are presenting the original definition by Granger.

3.1.1.1 Definition by C. W. Granger

Let $\bar{\mathbf{x}}_t$ represent the past values $\{\mathbf{x}_{t-i}, i = 1, \dots, \infty\}$ and let \mathbf{y}_t be another stationary random vector with mean zero. The predictor of \mathbf{x}_t only using \mathbf{U}_t ¹, the information up to time t , is denoted by $\mathbb{E}_{t-1}[\mathbf{x}_t|\bar{\mathbf{U}}_t]$. The *one step prediction error* of \mathbf{x}_t is defined as $\epsilon_t(\mathbf{x}_t|\bar{\mathbf{U}}_t) = \mathbf{x}_t - \mathbb{E}[\mathbf{x}_t|\bar{\mathbf{U}}_t]$ with a variance $\Sigma_X(1|\bar{\mathbf{U}})$. Analogously we define the variance of the *h step prediction error* by $\Sigma_X(h|\bar{\mathbf{U}})$.

Later in this work the prediction of \mathbf{x}_t , using only past values of \mathbf{x} , is needed. It is denoted by $\mathbb{E}_{t-h}[\mathbf{x}_t|\bar{\mathbf{X}}_t]$. The corresponding one step prediction error

¹In the definition by Granger in 1969, \mathbf{U}_t is defined to be all information in the universe. Because this is a completely unrealistic aspect he redefined \mathbf{U}_t to be the available information.

$\epsilon_t(\mathbf{x}_t|\bar{\mathbf{X}}_t)$ has the variance $\Sigma_X(1|\bar{\mathbf{X}})$. Analogously the predictor of \mathbf{x}_t , using past values of \mathbf{x} and \mathbf{y} , is denoted by $\mathbb{E}[\mathbf{x}_t|\bar{\mathbf{X}}_t, \bar{\mathbf{Y}}_t]$ with an error $\epsilon_t(\mathbf{x}_t|\bar{\mathbf{X}}_t, \bar{\mathbf{Y}}_t)$ and a variance $\Sigma_X(1|\bar{\mathbf{X}}, \bar{\mathbf{Y}})$.

To compare the quality of the prediction using the complete information set $\bar{\mathbf{U}}_t$ or using only a part of it, another notation is introduced: The term $\bar{\mathbf{U}}_t - \bar{\mathbf{Y}}_t$ is defined to be the whole information set, reduced by all information which result from \mathbf{y} .

Definition 3.1 (*Granger causality*). If and only if a better prediction of \mathbf{x}_t is obtained using all available information $\bar{\mathbf{U}}$ than if this information is reduced by that coming from past values of \mathbf{y}_t , \mathbf{y} is Granger causing \mathbf{x} .²

$$\Sigma_X(h|\bar{\mathbf{U}}) < \Sigma_X(h|\bar{\mathbf{U}} - \bar{\mathbf{Y}}) \quad (3.1)$$

It follows that we speak about Granger causality if the signal \mathbf{x} depends on past values of the signal \mathbf{y} . Hence directed relationship between these two signals is defined.

Remark. One thing to consider is that Granger in 1969 used in his work only stationary time series. Nevertheless he remarked that this definition may be used for non-stationary data, but in this case it has to be considered that a causal relationship might change from one moment to the other.

Further it should be taken in mind that this definition is universal and does not depend on any underlying process.

As in this work only autoregressive models are used, only linear Granger causality is considered.

Let $\mathbf{x}_t = (x_{1t}, \dots, x_{Kt})$ be a stationary random vector with mean zero. Furthermore \mathbf{x}_t is adjusted to a linear model, where the past and present values of the components of \mathbf{x}_t explain each other. The error of the adjustment is a white-noise vector ϵ_t .³

The following VAR(p)-model (with $p \in \mathbb{N} \cup \{\infty\}$) with coefficient matrices A_i , $i = 1, \dots, p$ represents the time series \mathbf{x}_t :

$$\mathbf{A}_0 \mathbf{x}_t = \sum_{i=1}^p \mathbf{A}_i \mathbf{x}_{t-i} + \epsilon_t$$

In this work we concentrate on simple models, where the instantaneous effects are neglected ($\mathbf{A}_0 = \mathbf{I}_{K \times K}$ ⁴).

Definition 3.2 (*Feedback*). *Feedback* between \mathbf{x} and \mathbf{y} is given, if \mathbf{x} Granger causes \mathbf{y} and \mathbf{y} Granger causes \mathbf{x} .

$$\Sigma_Y(h|\bar{\mathbf{U}}) < \Sigma_Y(h|\bar{\mathbf{U}} - \bar{\mathbf{X}}) \text{ and } \Sigma_X(h|\bar{\mathbf{U}}) < \Sigma_X(h|\bar{\mathbf{U}} - \bar{\mathbf{Y}}) \quad (3.2)$$

²If $\Sigma_X(h|\bar{\mathbf{U}})$ and $\Sigma_X(h|\bar{\mathbf{U}} - \bar{\mathbf{Y}})$ are matrices, it is said that $\Sigma_X(h|\bar{\mathbf{U}}) < \Sigma_X(h|\bar{\mathbf{U}} - \bar{\mathbf{Y}})$ if and only if $[\Sigma_X(h|\bar{\mathbf{U}} - \bar{\mathbf{Y}}) - \Sigma_X(h|\bar{\mathbf{U}})]$ is positive definite.

³A white-noise vector ϵ_t fulfills the following features: Its components are uncorrelated ($\mathbb{E}[\epsilon'_t \epsilon_s] = 0$ for $t \neq s$) and the variance of one component equals one ($\mathbb{E}[\epsilon'_t \epsilon_t] = 1$).

⁴ $\mathbf{I}_{K \times K}$ is the K -dimensional identity matrix.

3.1.1.2 Bivariate Granger causality

To analyze the ECoG data, represented by a K -dimensional signal $\mathbf{x}[n] = (x_1[n], \dots, x_K[n])$, a practical concept is introduced. To test for causal relationships of that data, each pair of channels is tested. The first way to deal with the abstract definition of Granger is the so called *bivariate Granger causality*.

Definition 3.3 (*Bivariate Granger causality*). Let $\mathbf{x}[n] = (x_1, \dots, x_K)$ be a K -dimensional signal. The component x_j is said to *bivariate Granger cause* the component x_i , if and only if knowledge of $x_j[n]$'s past significantly improves the prediction of $x_i[n]$:

The improvement is measured with the help of the variances. Thus the definition is equivalent to

$$\Sigma_{x_i}(h|\bar{\mathbf{X}}_i, \bar{\mathbf{X}}_j) < \Sigma_{x_i}(h|\bar{\mathbf{X}}_i). \quad (3.3)$$

Example 3.4 (*Bivariate (linear) Granger causality*). Based on the work of E. Möller et al.[24] an illustration of the definition is given. Let $\mathbf{x}[n] = (x_1[n], \dots, x_K[n])$ be a K -dimensional multivariate system (i.e. it consist of K channels).

Consider two of these channels: $x_i[n]$ and $x_j[n]$. To identify, if $x_j[n]$ Granger causes $x_i[n]$, an univariate and a bivariate AR(p)-model of these components are compared.

Univariate AR(p)-model:

$$\begin{aligned} x_i[n] &= \sum_{s=1}^p a_{1,s} x_i[n-s] + \epsilon_i[n] \\ x_j[n] &= \sum_{s=1}^p d_{2,s} x_j[n-s] + \epsilon_j[n] \end{aligned}$$

Bivariate AR(p)-model:

$$\begin{aligned} x_i[n] &= \sum_{s=1}^p a_{2,s} x_i[n-s] + \sum_{i=1}^p b_{2,s} x_j[n-s] + \tilde{\epsilon}_i[n] \\ x_j[n] &= \sum_{s=1}^p c_{2,s} x_i[n-s] + \sum_{s=1}^p d_{2,s} x_j[n-s] + \tilde{\epsilon}_j[n] \end{aligned}$$

The component $x_j[n]$ Granger causes $x_i[n]$ if and only if

$$\mathbb{V}[\tilde{\epsilon}_i[n]] < \mathbb{V}[\epsilon_i[n]]. \quad (3.4)$$

One problem that occurs with this definition is that the following relationship always holds:

$$\mathbb{V}[\tilde{\epsilon}_i[n]] \leq \mathbb{V}[\epsilon_i[n]]$$

Hence inequality (3.4) is almost surely fulfilled.

A sort of measure for Granger causality is necessary to identify a significant improvement of the prediction if information based on past values of the second component is added.

For stationary signals Geweke (1982)[11] introduced the following (non-negative) measure (the *bivariate Granger causality index*), which indicates if x_j Granger causes x_i :

$$bGCI_{x_j \rightarrow x_i} = \ln \left(\frac{\mathbb{V}[\epsilon_i[n]]}{\mathbb{V}[\tilde{\epsilon}_i[n]]} \right)$$

If this index receives zero, x_j is Granger not-causing x_i . Because dealing with non-stationary data, we refer the interested reader for detailed information to Geweke (1982) [11].

3.1.1.3 Bivariate Granger causality for a partitioned process

For the sake of completeness we extend the bivariate definition to the case of two sets of variables. This idea is found in A. Graef et al.[13].

Definition 3.5 (*Bivariate Granger causality for a partitioned process*). Let $\mathbf{x}_I[n]$ ($I = \{i_1, \dots, i_M\}$) be a M -dimensional and $\mathbf{x}_J[n]$ ($J = \{j_1, \dots, j_N\}$) be a N -dimensional subprocesses of the K -dimensional signal $\mathbf{x}[n]$. Furthermore the subprocesses are claimed to be disjoint and to fulfill $M + N = K$.

If and only if the past values of $\mathbf{x}_J[n]$ improve the prediction of $\mathbf{x}_I[n]$ significantly, $\mathbf{x}_J[n]$ is said to be Granger causing $\mathbf{x}_I[n]$.

$$\Sigma_{\mathbf{x}_I}(h|\overline{\mathbf{X}_I[n]}, \overline{\mathbf{X}_J[n]}) < \Sigma_{\mathbf{x}_I}(h|\overline{\mathbf{X}_I[n]}) \quad (3.5)$$

Remark. This definition does not imply anything about causal relationships between the individual components of both sets. For a better understanding we give a small example.

Let $\mathbf{x}[n] = (x_1[n], x_2[n], x_3[n])'$ be a three dimensional signal with $x_1[n]$ Granger causing the set $(x_2[n], x_3[n])'$. With this information we cannot decide if $x_1[n]$ Granger causes $x_2[n]$ or if $x_1[n]$ Granger causes $x_3[n]$.

3.1.2 Characterizing Granger causality with AR coefficients

In this section we concentrate on AR models and the characterization of Granger causality for this special case.

Following the work of Graef et al. [13] we will introduce a very simple form characterizing Granger causality, only using AR coefficients.

The first case considered here is the multivariate system, which consists of two disjoint variable sets. Therefore we speak about the bivariate Granger causality, characterized in Section 3.1.1.3.

Now a K -dimensional multivariate system, which is partitioned into a M -dimensional set of variables $\mathbf{x}_I[n]$ and a N -dimensional set of disjoint variables $\mathbf{x}_J[n]$, is considered. As before it is claimed that $\mathbf{x}[n] = \mathbf{x}_I[n] \cup \mathbf{x}_J[n]$ and $M + N = K$.

Lemma 3.6 (Characterization of Granger causality in MA⁵ processes). *Let $\mathbf{x}[n]$ be a MA process $\mathbf{x}[n] = \mu + \sum_{s=0}^{\infty} \Phi[s]\epsilon[n-s]$ with $\Phi[0] = \mathbf{I}_{K \times K}$, where $\epsilon[n]$ describes a white noise process with a non-singular covariance matrix Σ_{ϵ} . Granger causality is determined by the coefficient matrices $\Phi[s]$ only:*

If and only if $\mathbf{x}_J[n]$ Granger not causes $\mathbf{x}_I[n]$ ($\mathbf{x}_J[n] \not\stackrel{G}{\rightarrow} \mathbf{x}_I[n]$), the block matrix of the moving average coefficient $\Phi_{IJ}[s]$ is the zero matrix ($\Phi_{IJ}[s] = \mathbf{0}$).

Proof. Let $\mathbf{x}[n]$ be a K -dimensional moving average process, which is partitioned into two component processes $\mathbf{x}_I[n]$ (M -dimensional) and $\mathbf{x}_J[n]$ ($N = (K - M)$ -dimensional).

$$\mathbf{x}[n] = \begin{pmatrix} \mathbf{x}_I[n] \\ \mathbf{x}_J[n] \end{pmatrix} = \begin{pmatrix} \mu_I \\ \mu_J \end{pmatrix} + \begin{pmatrix} \Phi_{II}(z) & \Phi_{IJ}(z) \\ \Phi_{JI}(z) & \Phi_{JJ}(z) \end{pmatrix} \begin{pmatrix} \epsilon_I[n] \\ \epsilon_J[n] \end{pmatrix} \quad (3.6)$$

The matrix $\Phi(z)$ describes a polynomial matrix with the lag operator z . For a better overview we present table 3.1 with the dimensions of the block matrices of $\Phi(z)$.

Polynomial matrix	Dimension
$\Phi_{II}(z)$	$M \times M$ -dimensional
$\Phi_{IJ}(z)$	$M \times (K - M)$ -dimensional
$\Phi_{JI}(z)$	$(K - M) \times M$ -dimensional
$\Phi_{JJ}(z)$	$(K - M) \times (K - M)$ -dimensional

Table 3.1: *Dimensions of the block matrices of the partitioned polynomial coefficient matrix $\Phi(z)$. (The lag operator is denoted with z .)*

For the concept of Granger causality the one step prediction error is necessary. The h -step prediction formula for MA-processes is given by

$$\hat{\mathbf{x}}[n](h) = \mu + \sum_{s=h}^{\infty} \Phi[s]\epsilon[n+h-s]. \quad (3.7)$$

The first step will be to analyze the one step prediction error of $\mathbf{x}_I[n]$, using past values of $\mathbf{x}[n]$.

Following the one step prediction formula (equation (3.7), with $h = 1$), we obtain

$$\begin{aligned} \hat{\mathbf{x}}_I[n](1|\bar{\mathbf{x}}[n]) &= \begin{pmatrix} \mathbf{I}_{M \times M} & \mathbf{0}_{M \times (K-M)} \end{pmatrix} \hat{\mathbf{x}}[n](1) \\ &= \begin{pmatrix} \mathbf{I}_{M \times M} & \mathbf{0}_{M \times (K-M)} \end{pmatrix} \left(\mu + \sum_{s=1}^{\infty} \Phi[s]\epsilon[n+1-s] \right) \\ &= \begin{pmatrix} \mathbf{I}_{M \times M} & \mathbf{0}_{M \times (K-M)} \end{pmatrix} (\mu + \Phi(z)\epsilon[n]) \\ &= \begin{pmatrix} \mathbf{I}_{M \times M} & \mathbf{0}_{M \times (K-M)} \end{pmatrix} \left(\begin{pmatrix} \mu_I \\ \mu_J \end{pmatrix} + \begin{pmatrix} \Phi_{II}(z) & \Phi_{IJ}(z) \\ \Phi_{JI}(z) & \Phi_{JJ}(z) \end{pmatrix} \begin{pmatrix} \epsilon_I[n] \\ \epsilon_J[n] \end{pmatrix} \right) \\ &= \mu_I + \Phi_{II}(z)\epsilon_I[n] + \Phi_{IJ}(z)\epsilon_J[n]. \end{aligned}$$

⁵Moving Average

Using sum-notation we get

$$\begin{aligned}\hat{\mathbf{x}}_I[n](1|\bar{\mathbf{x}}[n]) &= \mu_I + \sum_{s=1}^{\infty} \Phi_{II}[s] \epsilon_I[n+1-s] \\ &\quad + \sum_{s=1}^{\infty} \Phi_{IJ}[s] \epsilon_J[n+1-s].\end{aligned}\quad (3.8)$$

To calculate the one step prediction error, the value $\mathbf{x}_I[n+1]$ has to be expressed with the help of the MA-representation (equation (3.6)).

$$\begin{aligned}\mathbf{x}_I[n+1] &= \mu_I + \Phi_{II}(z) \epsilon_I[n+1] + \Phi_{IJ}(z) \epsilon_J[n+1] \\ &= \mu_I + \sum_{s=0}^{\infty} \Phi_{II}[s] \epsilon_I[n+1-s] + \sum_{s=0}^{\infty} \Phi_{IJ}[s] \epsilon_J[n+1-s] \\ &= \mu_I + \Phi_{II}[0] \epsilon_I[n+1] + \sum_{s=1}^{\infty} \Phi_{II}[s] \epsilon_I[n+1-s] \\ &\quad + \Phi_{IJ}[0] \epsilon_J[n+1] + \sum_{s=1}^{\infty} \Phi_{IJ}[s] \epsilon_J[n+1-s]\end{aligned}$$

Restricting ourselves to the case, where $\Phi[0]$ is the identity matrix, the coefficient matrix $\Phi_{II}[0]$ is the M -dimensional identity matrix ($\Phi_{II}[0] = \mathbf{I}_{M \times M}$) and the coefficient matrix $\Phi_{IJ}[0]$ is a $M \times (K-M)$ -dimensional zero matrix ($\Phi_{IJ}[0] = \mathbf{0}_{M \times (K-M)}$).

Hence $\mathbf{x}_I[n+1]$ can be expressed as

$$\begin{aligned}\mathbf{x}_I[n+1] &= \mu_I + \epsilon_I[n+1] + \sum_{s=1}^{\infty} \Phi_{II}[s] \epsilon_I[n+1-s] \\ &\quad + \sum_{s=1}^{\infty} \Phi_{IJ}[s] \epsilon_J[n+1-s].\end{aligned}\quad (3.9)$$

The forecast error of $\mathbf{x}_I[n+1]$ is defined by

$$\Sigma_{\mathbf{x}_I}(1|\bar{\mathbf{x}}[n]) = \mathbf{x}_I[n+1] - \hat{\mathbf{x}}_I[n](1|\bar{\mathbf{x}}[n]).$$

Comparing the representations of the prediction (equation (3.8)) and the real value (equation (3.9)) of $\mathbf{x}_I[n+1]$, the one step prediction error can be calculated as

$$\Sigma_{\mathbf{x}_I}(1|\bar{\mathbf{x}}[n]) = \epsilon_I[n+1].\quad (3.10)$$

The projection of $\mathbf{x}_I[n+1]$ on its own past is defined as

$$\mathbf{x}_I[n+1] = \mu_I + \sum_{s=0}^{\infty} \mathbf{F}[s] \nu[n+1-s]\quad (3.11)$$

Similar to the projection on the space of \mathbf{x}_I and \mathbf{x}_J , we choose the representation with $\mathbf{F}[0] = \mathbf{I}_{M \times M}$.

Next we want to derive the one step prediction of $\mathbf{x}_I[n+1]$ considering only the past values of $\mathbf{x}_I[n]$. The prediction formula leads to

$$\hat{\mathbf{x}}_I[n](1|\bar{\mathbf{x}}_I[n]) = \mu_I + \sum_{s=1}^{\infty} \mathbf{F}[s] \nu[n+1-s].$$

The forecast error is equal to the difference between the real value $\mathbf{x}_I[n+1]$ and the estimated one:

$$\begin{aligned} \Sigma_{\mathbf{x}_I}(1|\bar{\mathbf{x}}_I[n]) &= \mathbf{x}_I[n+1] - \hat{\mathbf{x}}_I[n](1|\bar{\mathbf{x}}_I[n]) \\ &= \nu[n+1] \end{aligned} \quad (3.12)$$

Granger non-causality between \mathbf{x}_I and \mathbf{x}_J is given, if and only if both forecast errors in equations (3.10) and (3.12) are equal:

$$\begin{aligned} \Sigma_{\mathbf{x}_I}(1|\bar{\mathbf{x}}[n]) &= \Sigma_{\mathbf{x}_I}(1|\bar{\mathbf{x}}_I[n]) \\ \Leftrightarrow \epsilon_I[n+1] &= \nu[n+1] \quad \forall n. \end{aligned}$$

Under the assumption of equality, we rewrite the representation for $\mathbf{x}_I[n]$ projected on the past values of \mathbf{x}_I (equation (3.11)) and get

$$\begin{aligned} \mathbf{x}_I[n] &= \mu_I + \sum_{s=0}^{\infty} \mathbf{F}[s] \nu[n-s] \\ &= \mu_I + \sum_{s=0}^{\infty} \mathbf{F}[s] \epsilon_I[n-s] \\ &= \mu_I + \sum_{s=0}^{\infty} \begin{pmatrix} \mathbf{F}[s] & \mathbf{0}_{M \times (K-M)} \end{pmatrix} \epsilon[n-s]. \end{aligned}$$

Now we like to compare this result with the representation of $\mathbf{x}_I[n]$ based on past values of \mathbf{x} (equation (3.6)):

$$\mathbf{x}_I[n] = \mu_I + \sum_{s=0}^{\infty} \begin{pmatrix} \Phi_{II}[s] & \Phi_{IJ}[s] \end{pmatrix} \epsilon[n-s].$$

From the uniqueness of the moving-average representation, we obtain

$$\begin{aligned} \Phi_{II}[s] &= \mathbf{F}[s] \quad \forall s \\ \Phi_{IJ}[s] &= \mathbf{0}_{M \times (K-M)} \quad \forall s. \end{aligned}$$

Now we have shown that the corresponding polynomial coefficient matrix $\Phi_{IJ}(z) = \mathbf{0}_{M \times (K-M)}$ if the signal $\mathbf{x}_J[n]$ is Granger not-causal for $\mathbf{x}_I[n]$.

Because of equivalence of all transformations the inversion holds and the proof is completed. \square

Before we adapt this result to autoregressive processes, we need lemma for a block-wise inversion of a matrix.

Lemma 3.7 (Block-wise inversion of a matrix). *The inversion of the matrix*

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

where B_{ij} (for $i, j = 1, 2$) are matrices, is

$$B^{-1} = \begin{pmatrix} (B_{11} - B_{12}B_{22}^{-1}B_{21})^{-1} & -(B_{11} - B_{12}B_{22}^{-1}B_{21})^{-1}B_{12}B_{22}^{-1} \\ -B_{22}^{-1}B_{21}(B_{11} - B_{12}B_{22}^{-1}B_{21})^{-1} & B_{22}^{-1} + B_{22}^{-1}B_{21}(B_{11} - B_{12}B_{22}^{-1}B_{21})^{-1}B_{12}B_{22}^{-1} \end{pmatrix}$$

This statement is proofed by calculating $B B^{-1} = \mathbf{I}$.

Returning to our original problem - characterizing Granger causality in AR processes, we have the following theorem.

Theorem 3.8 (Characterization of Granger causality in AR processes). *Given an invertible⁶ autoregressive model*

$$\mathbf{x}[n] = \sum_{s=1}^{\infty} \mathbf{A}[s] \mathbf{x}[n-s] + \epsilon[n],$$

Granger causality can be determined by considering the coefficient matrices $\mathbf{A}[s]$:

If and only if the respective block matrices of the coefficient matrix are zero matrices, Granger non-causality is given.

$$\mathbf{A}_{IJ}[s] = \mathbf{0} \quad \forall s \quad \Leftrightarrow \quad \mathbf{x}_J[n] \not\stackrel{G}{\Rightarrow} \mathbf{x}_I[n]$$

Proof. Every stable invertible AR process can be transformed into a moving average process.

$$\begin{aligned} \mathbf{A}(z) \mathbf{x}[n] &= \epsilon[n] \\ \Leftrightarrow \mathbf{x}[n] &= \mathbf{A}^{-1}(z) \epsilon[n] \end{aligned}$$

The polynomial coefficient matrix (with lag operator z) is denoted by $\mathbf{A}(z)$. To use the same notation for the MA-process, $\mathbf{A}^{-1}(z)$ is redefined as $\Phi(z)$ with $\Phi[0] = \mathbf{I}$.

$$\mathbf{x}[n] = \sum_{s=0}^{\infty} \Phi[s] \epsilon[n-s] = \Phi(z) \epsilon[n]$$

We are still considering the case, where the K -dimensional signal $\mathbf{x}[n]$ consists of two sets of variables: $\mathbf{x}_I[n]$ (M -dimensional) and $\mathbf{x}_J[n]$ ($(K-M)$ -dimensional). We are interested if $\mathbf{x}_J[n]$ Granger causes $\mathbf{x}_I[n]$ or not.

Reminding lemma 3.6: We have shown that in the case of a canonical moving average process, a Granger non-causality relation ($\mathbf{x}_J[n] \not\stackrel{G}{\Rightarrow} \mathbf{x}_I[n]$) is obtained,

⁶i.e $\mathbf{A}[s]$ is a non-singular matrix for all $s = 0 \dots \infty$

if and only if the corresponding coefficient block matrices are zero matrices ($\Phi_{IJ}[s] = \mathbf{0}_{M \times (K-M)} \forall s$).

Due to the non-causal relation ($\mathbf{x}_J[n] \not\stackrel{G}{\Rightarrow} \mathbf{x}_I[n]$) the polynomial matrix $\Phi(z)$ becomes

$$\Phi(z) = \begin{pmatrix} \Phi_{II}(z) & \mathbf{0}_{M \times (K-M)} \\ \Phi_{JI}(z) & \Phi_{JJ}(z) \end{pmatrix}.$$

Reflecting this situation to the AR process, the polynomial matrix $\Phi(z)$ has to be inverted.

$$\begin{aligned} \mathbf{x}[n] &= \Phi(z)\epsilon[n] \\ \Phi^{-1}(z)\mathbf{x}[n] &= \epsilon[n] \end{aligned}$$

Because of the uniqueness of the representation of the AR process we get

$$\begin{aligned} \mathbf{A}(z) = \Phi^{-1}(z) &= \begin{pmatrix} \mathbf{A}_{II}(z) & -(\Phi_{II} - \Phi_{IJ}\Phi_{JJ}^{-1}\Phi_{JI})^{-1}\underbrace{\Phi_{IJ}\Phi_{JJ}^{-1}}_{=0} \\ \mathbf{A}_{JI}(z) & \mathbf{A}_{JJ}(z) \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}_{II}(z) & \mathbf{0}_{M \times (K-M)} \\ \mathbf{A}_{JI}(z) & \mathbf{A}_{JJ}(z) \end{pmatrix}. \end{aligned}$$

In case of Granger non-causality the corresponding block matrix of the coefficient matrix is zero.

Analogously $\Phi_{IJ}(z) = 0$ is received by inverting the polynomial AR-coefficient matrix $\mathbf{A}(z)$ with $\mathbf{A}_{IJ}(z) = \mathbf{0}$. Hence for the corresponding MA representation we obtain the following coefficient matrix

$$\Phi(z) = \begin{pmatrix} \Phi_{II}(z) & \mathbf{0}_{M \times (K-M)} \\ \Phi_{JI}(z) & \Phi_{JJ}(z) \end{pmatrix}.$$

This is (see lemma 3.6) equivalent to Granger non-causality, which completes the proof. \square

Example 3.9 (Identifying Granger causality via AR-coefficients in a 3-dimensional system). Let us consider a three dimensional AR(1) system

$$\begin{pmatrix} x_1[n] \\ x_2[n] \\ x_3[n] \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 3 \\ 0 & 8 & 4 \end{pmatrix}}_{=\mathbf{A}(1)} \begin{pmatrix} x_1[n-1] \\ x_2[n-1] \\ x_3[n-1] \end{pmatrix} + \begin{pmatrix} \epsilon_1[n] \\ \epsilon_2[n] \\ \epsilon_3[n] \end{pmatrix} :$$

We are splitting this system into two sets of variables: $x_1[n]$ and $(x_2[n], x_3[n])$. For analyzing Granger causality between $x_1[n]$ and $(x_2[n], x_3[n])$, theorem 3.8 tells us to consider the corresponding block matrices of the coefficient matrix $\mathbf{A}(1)$.

Considering the block matrix $\mathbf{A}_{1,(2,3)} = (0, 0)$, it follows that $(x_2[n], x_3[n])$ is Granger not-causing $x_1[n]$. Contrarily $x_1[n]$ Granger causes the set $(x_2[n], x_3[n])$, because some entries in the corresponding block matrix $\begin{pmatrix} 2 \\ 0 \end{pmatrix}$ differ from zero.

3.2 Conditional Granger causality

Up to now we have just considered two sets of components, which together amount to the whole information we are analyzing. Generally we want to use all available information. Furthermore causal relations should be analyzed between all variables (which do not have to amount to the whole information).

Separating two variables as done before and reducing the considered system to these two variables, we could get a causal relation, although they are only indirectly (through a third variable) connected. (For detailed information see Section 3.2.2.)

Because no uniform extension for the multivariate case exists, we are going to introduce two different methods, which are often used in a neuronal context. Both concepts are introduced in Guo et al.[16].

3.2.1 Definition

The first extension of the bivariate definition is called Conditional Granger causality.

Definition 3.10 (*Conditional Granger causality*). Let $\mathbf{x}[n]$ be a K -dimensional signal. A M -dimensional subprocess $\mathbf{x}_I[n]$, $I = \{i_1, \dots, i_M\}$ and a second N -dimensional, disjoint subprocess $\mathbf{x}_J[n]$, $J = \{j_1, \dots, j_N\}$ are observed. $\mathbf{x}_J[n]$ is Granger causing $\mathbf{x}_I[n]$, if and only if given the remaining components ($\mathbf{R} = \{x_1, \dots, x_K\} \setminus \{\mathbf{x}_I \cup \mathbf{x}_J\}$), the prediction of $\mathbf{x}_I[n]$ using past values of $\mathbf{x}_J[n]$ improves, compared to the one without this information.

$$\Sigma_{\mathbf{x}_I}(h | \overline{\mathbf{U}[n]}) < \Sigma_{\mathbf{x}_I}(h | \overline{\mathbf{U}[n] \setminus \mathbf{X}_J[n]}) \quad (3.13)$$

The term $\overline{\mathbf{U}[n]}$ denotes the whole information obtained by past values of the signal $\mathbf{x}[n]$ ($\overline{\mathbf{U}[n]} = \{x_1[s], \dots, x_K[s], s < n\} = \{\mathbf{x}_I[s] \cup \mathbf{x}_J[s] \cup \mathbf{R}[s]\}$).

This method is called Conditional Granger causality, because it indicates the relationship between two variables, with respect to the rest.

One advantage of this method is, that it detects whether a variable is directly caused by another one, or whether it is indirectly connected through a third observed variable.

In contradiction to the bivariate case with two sets of variables, it is not necessary to restrict the sets of variables to these spanning the whole space.

We are only interested in the identification of causal relations between two single variables. Hence we restrict ourselves to the simple case where $M = N = 1$ and to linear Granger causality.

Example 3.11 (*Linear Conditional Granger causality in the case of three variables*). Let x_1 , x_2 and x_3 be three time series. The task will be finding out if variable x_3 is causing variable x_1 or not. For the joint autoregressive system of

the first variables x_1 and x_2 , the following representation is obtained:

$$\begin{aligned} x_1[n] &= \sum_{i=1}^p a_{1,i}[n] x_1[n-i] + \sum_{i=1}^p b_{1,i}[n] x_2[n-i] + \epsilon_1[n] \\ x_2[n] &= \sum_{i=1}^p d_{1,i}[n] x_1[n-i] + \sum_{i=1}^p e_{1,i}[n] x_2[n-i] + \epsilon_2[n]. \end{aligned}$$

For this system the covariance matrix is

$$S_{12}[n] = \begin{pmatrix} \mathbb{V}(\epsilon_1[n]) & \text{Cov}(\epsilon_1[n], \epsilon_2[n]) \\ \text{Cov}(\epsilon_2[n], \epsilon_1[n]) & \mathbb{V}(\epsilon_2[n]) \end{pmatrix}.$$

To be able to test if the third variable x_3 is causing one of the other, we have to calculate the joint autoregressive representation of all three components:

$$\begin{aligned} x_1[n] &= \sum_{i=1}^p a_{2,i}[n] x_1[n-i] + \sum_{i=1}^p b_{2,i}[n] x_2[n-i] + \sum_{i=1}^p c_{2,i}[n] x_3[n-i] + \tilde{\epsilon}_1[n] \\ x_2[n] &= \sum_{i=1}^p d_{2,i}[n] x_1[n-i] + \sum_{i=1}^p e_{2,i}[n] x_2[n-i] + \sum_{i=1}^p f_{2,i}[n] x_3[n-i] + \tilde{\epsilon}_2[n] \\ x_3[n] &= \sum_{i=1}^p g_{2,i}[n] x_1[n-i] + \sum_{i=1}^p h_{2,i}[n] x_2[n-i] + \sum_{i=1}^p l_{2,i}[n] x_3[n-i] + \tilde{\epsilon}_3[n] \end{aligned}$$

The following covariance matrix is obtained:

$$\Sigma = \begin{pmatrix} \mathbb{V}(\tilde{\epsilon}_1[n]) & \text{Cov}(\tilde{\epsilon}_1[n], \tilde{\epsilon}_2[n]) & \text{Cov}(\tilde{\epsilon}_1[n], \tilde{\epsilon}_3[n]) \\ \text{Cov}(\tilde{\epsilon}_2[n], \tilde{\epsilon}_1[n]) & \mathbb{V}(\tilde{\epsilon}_2[n]) & \text{Cov}(\tilde{\epsilon}_2[n], \tilde{\epsilon}_3[n]) \\ \text{Cov}(\tilde{\epsilon}_3[n], \tilde{\epsilon}_1[n]) & \text{Cov}(\tilde{\epsilon}_3[n], \tilde{\epsilon}_2[n]) & \mathbb{V}(\tilde{\epsilon}_3[n]) \end{pmatrix}.$$

The component $x_3[n]$ improves the prediction of x_1 , if and only if the variance of the prediction error is reduced. With the help of the covariances the improvement of the prediction can be measured. x_3 Granger causes x_1 , if and only if $\mathbb{V}(\tilde{\epsilon}_1[n]) < \mathbb{V}(\epsilon_1[n])$. Analogous to the bivariate case the (non-negative) *conditional Granger causality index* is defined as

$$cGCI_{x_3 \rightarrow x_1} = \ln \left(\frac{\mathbb{V}(\epsilon_1[n])}{\mathbb{V}(\tilde{\epsilon}_1[n])} \right). \quad (3.14)$$

In the linear case the conditional Granger causality can be attributed to the bivariate case with the help of lemma 3.7:

Lemma 3.12 (Connection between linear bivariate and linear conditional Granger causality). *Eliminating the effects of the rest ($\{x_1[n], \dots, x_K[n]\} \setminus \{\mathbf{x}_I[n] \cup \mathbf{x}_J[n]\}$) from $\mathbf{x}_I[n]$ and $\mathbf{x}_J[n]$ and testing linear bivariate Granger causality between the two components, leads to the same result as considering linear conditional Granger causality between these two variables.*

Proof. For convenience reasons we redefine the M -dimensional variable $\mathbf{x}_I[n]$ as \mathbf{x} , the N -dimensional variable $\mathbf{x}_J[n]$ as \mathbf{y} and call the rest \mathbf{z} .

The matrices $\bar{\mathbf{X}} \in \mathbb{R}^{T \times N}$, $\bar{\mathbf{Y}} \in \mathbb{R}^{T \times M}$ and $\bar{\mathbf{Z}} \in \mathbb{R}^{T \times (K-N-M)}$ represent the partitioned data (the rows represent the T data points and the columns represent the different components).

First *linear Conditional Granger causality* between \mathbf{x} and \mathbf{y} (conditioned to the rest \mathbf{z}) is studied.

Let us consider the regression of \mathbf{x} projected on the space spanned by past values of \mathbf{x} and \mathbf{z} .

$$\mathbf{x} = (\bar{\mathbf{X}}, \bar{\mathbf{Z}})\beta + \mathbf{u}$$

The projector on the orthogonal space of $\bar{\mathbf{Z}}$ is denoted with $\mathbf{I} - \mathbf{P}_z = \mathbf{I} - (\mathbf{I} - \bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\bar{\mathbf{Z}})\bar{\mathbf{Z}}')$. The corresponding least squares estimator $\hat{\beta}$ is defined as

$$\hat{\beta} = \begin{pmatrix} \bar{\mathbf{X}}'\bar{\mathbf{X}} & \bar{\mathbf{X}}'\bar{\mathbf{Z}} \\ \bar{\mathbf{Z}}'\bar{\mathbf{X}} & \bar{\mathbf{Z}}'\bar{\mathbf{Z}} \end{pmatrix}^{-1} \begin{pmatrix} \bar{\mathbf{X}}'\mathbf{x} \\ \bar{\mathbf{Z}}'\mathbf{x} \end{pmatrix}.$$

With the matrix inversion lemma (lemma 3.7), the least squares estimator is

$$\begin{aligned} \hat{\beta} &= \begin{pmatrix} \overbrace{(\bar{\mathbf{X}}'(\mathbf{I} - \mathbf{P}_z)\bar{\mathbf{X}})^{-1}}^{\mathbf{A}} & -\mathbf{A}\bar{\mathbf{X}}\bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-1} \\ -(\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\bar{\mathbf{X}}\mathbf{A} & (\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-1} + (\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\bar{\mathbf{X}}\mathbf{A}\bar{\mathbf{X}}'\bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-1} \end{pmatrix} \begin{pmatrix} \bar{\mathbf{X}}'\mathbf{x} \\ \bar{\mathbf{Z}}'\mathbf{x} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}\bar{\mathbf{X}}'(\mathbf{I} - \mathbf{P}_z)\mathbf{x} \\ (\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'(-\bar{\mathbf{X}}\mathbf{A}\bar{\mathbf{X}}' + \mathbf{I} + \bar{\mathbf{X}}\mathbf{A}\bar{\mathbf{X}}'\mathbf{P}_z)\mathbf{x} \end{pmatrix}. \end{aligned}$$

Dealing with Granger causality we are only interested in the estimation errors. The one for the regression of \mathbf{x} on the past values of \mathbf{x} and \mathbf{z} is

$$\begin{aligned} \hat{\mathbf{u}} &= \mathbf{x} - (\bar{\mathbf{X}}, \bar{\mathbf{Z}})\hat{\beta} \\ &= (\mathbf{I} - \bar{\mathbf{X}}\mathbf{A}\bar{\mathbf{X}}'(\mathbf{I} - \mathbf{P}_z) + \mathbf{P}_z\bar{\mathbf{X}}\mathbf{A}\bar{\mathbf{X}}' - \mathbf{P}_z - \mathbf{P}_z\bar{\mathbf{X}}\mathbf{A}\bar{\mathbf{X}}'\mathbf{P}_z)\mathbf{x} \\ &= (\mathbf{I} - \mathbf{P}_z)(\mathbf{I} - \bar{\mathbf{X}}\mathbf{A}\bar{\mathbf{X}}'(\mathbf{I} - \mathbf{P}_z))\mathbf{x}. \end{aligned}$$

The second part of the proof deals with the modified time series \mathbf{x} , where all the influences from \mathbf{z} have been removed. Therefore the projected data $(\mathbf{I} - \mathbf{P}_z)\mathbf{x}$ is considered:

$$(\mathbf{I} - \mathbf{P}_z)\mathbf{x} = ((\mathbf{I} - \mathbf{P}_z)\bar{\mathbf{X}})'\beta + \mathbf{u}$$

Again our task is to find the prediction error. Hence the least squares estimation coefficient $\hat{\beta}$ is needed:

$$\hat{\beta} = (\bar{\mathbf{X}}'(\mathbf{I} - \mathbf{P}_z)'(\mathbf{I} - \mathbf{P}_z)\bar{\mathbf{X}})^{-1} \bar{\mathbf{X}}'(\mathbf{I} - \mathbf{P}_z)'(\mathbf{I} - \mathbf{P}_z)\mathbf{x}.$$

Reminding that we defined $\mathbf{A} = (\bar{\mathbf{X}}'(\mathbf{I} - \mathbf{P}_z)\bar{\mathbf{X}})^{-1}$ and that $(\mathbf{I} - \mathbf{P}_z)$ is a projector and therefore idempotent ($(\mathbf{I} - \mathbf{P}_z)'(\mathbf{I} - \mathbf{P}_z) = (\mathbf{I} - \mathbf{P}_z)$) we obtain

$$\begin{aligned} \hat{\beta} &= \underbrace{(\bar{\mathbf{X}}'(\mathbf{I} - \mathbf{P}_z)\bar{\mathbf{X}})^{-1}}_{\mathbf{A}} \bar{\mathbf{X}}'(\mathbf{I} - \mathbf{P}_z)\mathbf{x} \\ &= \mathbf{A}\bar{\mathbf{X}}'(\mathbf{I} - \mathbf{P}_z)\mathbf{x}. \end{aligned}$$

For a Granger causality analysis we have to determine the estimation error $\hat{\mathbf{u}}$ of the regressions (and its variance).

$$\begin{aligned}\hat{\mathbf{u}} &= (\mathbf{I} - \mathbf{P}_z) \mathbf{x} - (\mathbf{I} - \mathbf{P}_z) \bar{\mathbf{X}} \hat{\beta} \\ &= (\mathbf{I} - \mathbf{P}_z) \left(\mathbf{I} - \bar{\mathbf{X}} \mathbf{A} \bar{\mathbf{X}}' (\mathbf{I} - \mathbf{P}_z) \right) \mathbf{x}\end{aligned}$$

Comparing the estimation errors of both regressions, we see that they are identical.

Analogously the same result may be obtained for the regression of \mathbf{x} at the variables \mathbf{x} , \mathbf{y} and \mathbf{z}

$$\mathbf{x} = (\bar{\mathbf{X}}, \bar{\mathbf{Y}}, \bar{\mathbf{Z}}) \beta + \mathbf{u}.$$

Because both methods lead to the same errors, the same variances are obtained and the Granger causality index is the same in both cases. \square

Lemma 3.13 (Characterization of conditional Granger causality via AR-coefficients). *$\mathbf{x}_J[n]$ is Granger non-causal for $x_I[n]$, if and only if one of the two equivalent conditions holds:*

- (i) $|\mathbb{V}(\mathbf{x}_I[n] | \bar{\mathbf{x}}_{R \cup I}[n])| = |\mathbb{V}(\mathbf{x}_I[n] | \bar{\mathbf{x}}[n])|$
- (ii) $\mathbf{A}_{IJ}(s) = \mathbf{0} \quad \forall s \in \mathbb{N}$

3.2.2 Comparison between bivariate and conditional Granger causality

To justify the use of this multivariate extension, we are presenting some examples, where the pair-wise estimation fails identifying the true relations, while the analysis using the definition of conditional Granger causality reflects the correct system.

The following models is found in a modified way either in Kus et al.[21] or in Kaminski [20]. They show that only using the bivariate method can be very risky.

For the simulations we like to use any ECoG signal, between sample 1000 and 2000. In each step we delay the signal by one sample and add a white noise term, generated by a normal distribution with mean zero and variance 400.

Example 3.14 (Common problems using a bivariate method). Figure 3.1 illustrates the structure of the first example, which contains some problems we are going to analyze in more detail later.

Channel X_1 consists of the ECoG data mixed with a white noise process. By delaying this signal (the symbol $\Delta = 1$ denotes a delay of one sample) and adding different noise processes, we obtain the channels X_2 and X_3 and so on. We start with the bivariate analysis:

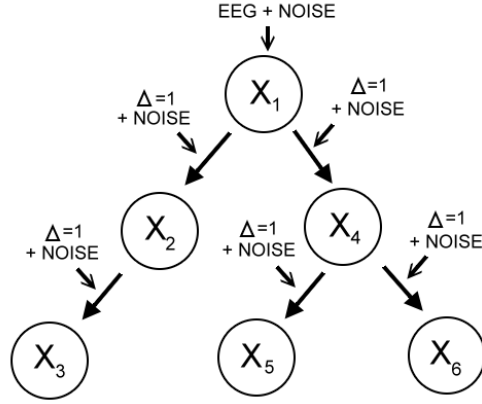


Figure 3.1: **Structure of the simulation:** In each step the signal is delayed by one sample (denoted with $\Delta = 1$) and noise is added.

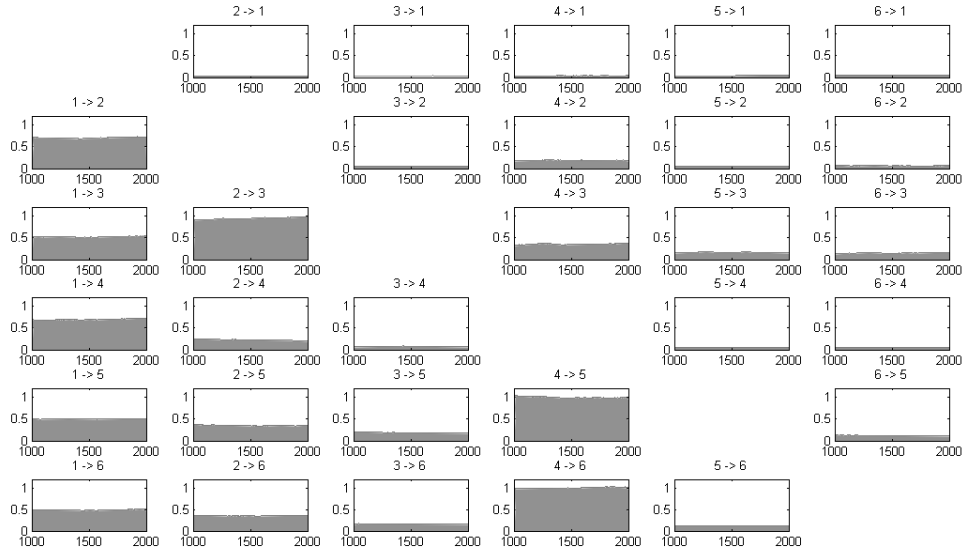


Figure 3.2: **Pair-wise bivariate Granger causality index:** The graph under $a \rightarrow b$ shows the index of a Granger causing b . If and only if the index is near zero, we have a non-causal relationship.

In figure 3.2 the bivariate Granger causality index for each sample is plotted against the time. One channel is said to Granger cause another, if and only if the index is significant higher than zero.

Using this method, too many causal relations are found. It seems impossible to identify the true connections. The channels with a high delayed influence of channel X_1 become terminals of activity. For a better overview the bivariate result is presented in figure 3.3, where an arrow represents a Granger causal relationship.

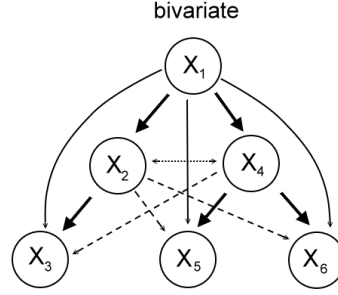


Figure 3.3: **The result of the bivariate analysis:** The bold arrows show the correct connections, the thin ones are resulting from indirect effects and the dashed arrows indicate causality falsely.

Now we are comparing this result with the multivariate analysis. Figure 3.4 presents the conditional Granger causality index plotted against the time. A causal relation is identified if the index is significantly greater than zero.

With this method the causal relations are found correctly. A high index (index at a level around 0.5) is observed for the corresponding plots of $1 \rightarrow 2 \rightarrow 3$, $1 \rightarrow 4$, $4 \rightarrow 5$ and $4 \rightarrow 6$. All other indices stay at a level near zero.

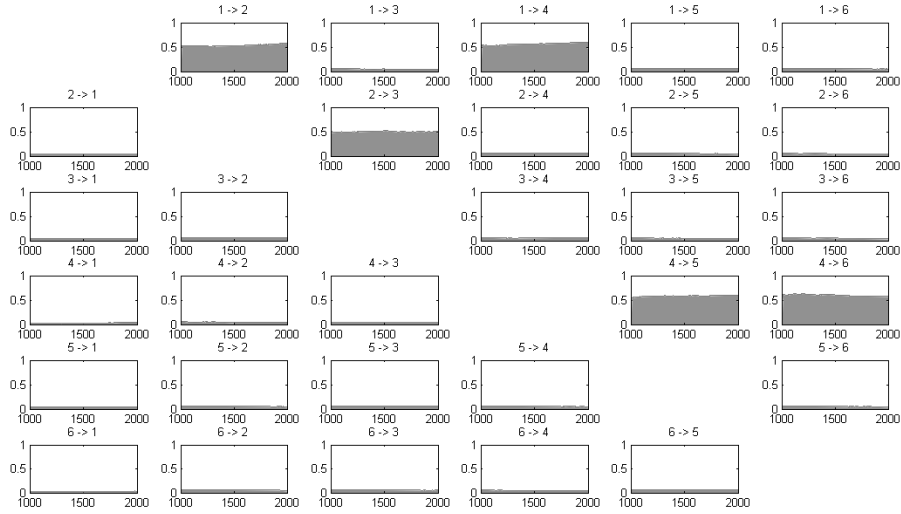


Figure 3.4: **Conditional Granger causality index:** The causal relations are found correctly ($1 \rightarrow 2 \rightarrow 3$, $1 \rightarrow 4$, $4 \rightarrow 5$ and $4 \rightarrow 6$), while the indices not corresponding to them stay near zero.

For every situation we are going to present an extra analysis to be able to concentrate on the core of the problems.

At first we want to know, what happens between channel X_1 and channel X_3 in a bivariate context?

X_1 is Granger causing X_2 and X_2 is Granger causing X_1 . Of course X_1 is causing the variable X_3 , but in an indirect way. The occurring problem is, that in a bivariate context, we cannot distinguish between a direct and an indirect

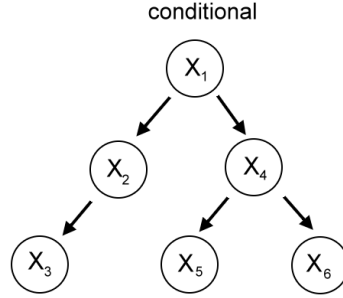


Figure 3.5: *The conditional Granger causality index finds the correct causal relations.*

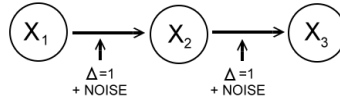


Figure 3.6: *Structure of the partitioned problem: Indirect cause*

connection. In a multivariate analysis the structure is obtained correctly, because X_2 is added in both models (set up for analyzing X_1 causing X_3) and consists of all information of X_1 necessary for a prediction of X_3 .



Figure 3.7: ***Difference of using the bivariate and the conditional Granger index:*** *Contrary to the conditional analysis, the pair-wise case indicates an additional arrow, showing the problem of indirect influence.*

Similar situations appear considering the other variables: $X_1 \rightarrow X_4 \rightarrow X_5$ and $X_1 \rightarrow X_4 \rightarrow X_6$.

A different problem arises between the variables X_2 and X_6 : A causal relation is identified although there exist no direct connection. Figure 3.8 helps to concentrate on the main structure of this situation.

The reason may be that even though X_2 is disturbed by a not negligible noise, there may exist some information, which is also present in X_4 (the delayed variables of X_1). Hence a causal relation is identified.

In the multivariate case this problem is absent. Variable X_4 (or in this shortened example X_1) allows a better explanation of X_6 than X_2 .

An overview of the results is given in figure 3.9.

Before finishing the comparison between the bivariate and the conditional definition, we are presenting a second interesting example, where a strange situation of an inverted relationship exists.

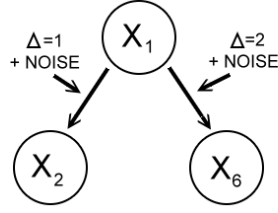


Figure 3.8: *Structure of partitioned problem: Causality without connection*

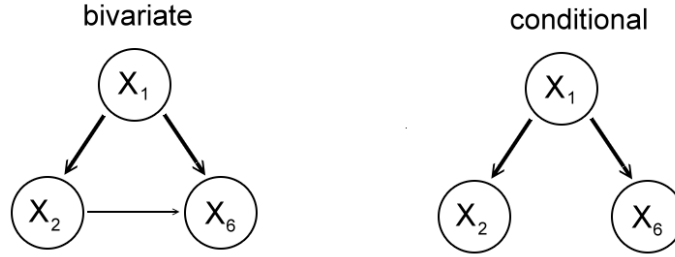


Figure 3.9: **Results of bivariate and conditional Granger causality index:** *In the bivariate case a wrong relation is identified (X_2 Granger causing X_6).*

Example 3.15 (Inversion of causal relationship using bivariate methods). The next simulation is an example of a problem that can easily happen using experimental ECoG data and it shows that using only a bivariate analysis is quite risky.

To introduce this problem we refer to figure 3.10, where the structure of a three dimensional system is presented in the same compact way as used before.

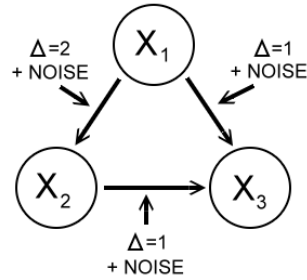


Figure 3.10: **Example: Inversion of causal relationship when using bivariate methods.** *The signals X_2 and X_3 were generated by adding a noise process and delaying the signal X_1 ($\Delta = 1$ denotes a delay by one sample and $\Delta = 2$ denotes a delay by two samples).*

The signal X_2 is obtained by delaying the signal X_1 for two samples and adding a noise process (i.e. a normal distribution with mean zero and variance 400). Delaying this signal for one sample, adding an analogous noise process and further adding X_1 , delayed for one sample signal and disturbed by another

noise process, X_3 is received.

Figure 3.11 represents the results of the bivariate and the conditional analysis. On the left hand side the corresponding Granger causality indices (above the bivariate case and below the conditional one) are plotted. The right hand side presents the structure, which visualizes the results of the plots on the left hand side. (A high level of the index indicates a causal relationship.)

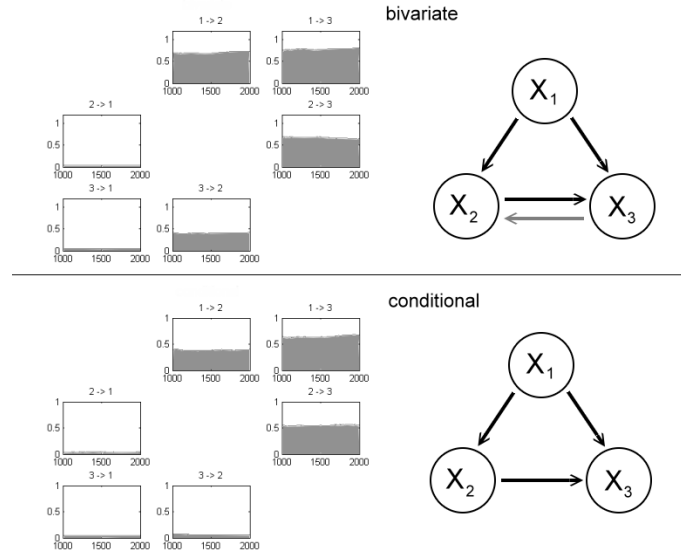


Figure 3.11: **Results of bivariate and conditional analysis:** In the bivariate case an additional causal relation between X_2 and X_3 arises, which has the inverse direction of the real one.

Using the bivariate method, this simulations shows, that a causal relation may appear in the inverse direction. Although only X_2 should be causing X_3 , a feedback occurs.

Because the delay between the signals X_1 and X_2 is larger than between X_1 and X_3 , there is some information in X_3 (X_1 delayed by one sample) which is useful for the prediction of the signal X_2 . Hence X_3 causes X_2 in a bivariate analysis.

3.3 Partial Granger causality

Now we are introducing a second multivariate expansion of the original definition of Granger causality, which is inspired by the idea of partial correlation in statistics. This attempt can be found in Guo et al. [16].

Earlier in this work, when analyzing conditional Granger causality, we defined one variable causing another, if and only if the prediction error of the first is reduced by adding the second variable to the model (with respect to the remaining variables). The critical point is, that all information has to be measured and taken into account for the model to get a correct relationship.

In practice this condition often is violated and latent (unmeasured) variables or exogenous inputs may distort the analysis of Granger causality.

3.3.1 Definition

3.3.1.1 Linear model

We start with a model, which is similar to the one used for the discussion of conditional Granger causality.

Let $\mathbf{x}[n]$ be a $(K = L + M + N)$ -dimensional multivariate system, which consists of three parts: $\mathbf{x}_I[n]$ is a M -dimensional set of variables, $\mathbf{x}_J[n]$ a N -dimensional partition of the system and $\mathbf{x}_R[n]$ the L -dimensional set of the remaining variables.

We are concentrating on the linear case and therefore introducing the vector autoregressive model for the multivariate system

$$\mathbf{A}(z)\mathbf{x}[n] = \mathbf{u}[n] = \epsilon^E[n] + \epsilon[n].$$

$\mathbf{A}(z)$ represents the polynomial coefficient matrix with the lag operator z . To obtain a unique representation, $\mathbf{A}(0)$ is claimed to be the K -dimensional identity matrix ($\mathbf{I}_{K \times K}$). The exogenous not measured input is denoted by $\epsilon^E[n]$, whereas $\epsilon[n]$ are the independent remaining errors. Both errors are combined in the term $\mathbf{u}[n]$. Compared to the work of Guo et al. we are not distinguishing between exogenous and latent variables.

To identify if the time series $\mathbf{x}_J[n]$ Granger causes $\mathbf{x}_I[n]$, we present a two-dimensional system projecting $\mathbf{x}_I[n]$ on the space spanned by the past values of $\mathbf{x}_I[n]$ and $\mathbf{x}_R[n]$.

$$\begin{aligned} \mathbf{x}_I[n] &= \sum_{s=1}^{\infty} \mathbf{A}_{II}[s] \mathbf{x}_I[n-s] + \sum_{s=1}^{\infty} \mathbf{A}_{IR}[s] \mathbf{x}_R[n-s] + \mathbf{u}_{1,I}[n] \\ \mathbf{x}_R[n] &= \sum_{s=1}^{\infty} \mathbf{A}_{RI}[s] \mathbf{x}_I[n-s] + \sum_{s=1}^{\infty} \mathbf{A}_{RR}[s] \mathbf{x}_R[n-s] + \mathbf{u}_{1,R}[n] \end{aligned}$$

The corresponding covariance matrix of the system errors is defined by

$$\mathbf{S}[n] = \begin{pmatrix} \mathbb{V}(\mathbf{u}_{1,I}[n]) & \text{Cov}(\mathbf{u}_{1,I}[n], \mathbf{u}_{1,R}[n]) \\ \text{Cov}(\mathbf{u}_{1,R}[n], \mathbf{u}_{1,I}[n]) & \mathbb{V}(\mathbf{u}_{1,R}[n]) \end{pmatrix}.$$

The partial covariance (see Section 2.20), where the effects of common latent variables (common variables of $\mathbf{x}_I[n]$ and $\mathbf{x}_R[n]$) are eliminated, is given by

$$\mathbf{S}_{x_I|x_R} = \mathbf{S}_{II} - \mathbf{S}_{IR} \mathbf{S}_{RR}^{-1} \mathbf{S}_{RI}.$$

For the analysis of Granger causal relations we need to expand this concept to a three dimensional system containing variables $\mathbf{x}_I[n]$, $\mathbf{x}_J[n]$, and $\mathbf{x}_R[n]$ as

usually:

$$\begin{aligned}
\mathbf{x}_I[n] &= \sum_{s=1}^{\infty} \mathbf{B}_{II}[s] \mathbf{x}_I[n-s] + \sum_{s=1}^{\infty} \mathbf{B}_{IJ}[s] \mathbf{x}_J[n-s] + \sum_{s=1}^{\infty} \mathbf{B}_{IR}[s] \mathbf{x}_R[n-s] + \mathbf{u}_{2,I}[n] \\
\mathbf{x}_J[n] &= \sum_{s=1}^{\infty} \mathbf{B}_{JI}[s] \mathbf{x}_I[n-s] + \sum_{s=1}^{\infty} \mathbf{B}_{JJ}[s] \mathbf{x}_J[n-s] + \sum_{s=1}^{\infty} \mathbf{B}_{JR}[s] \mathbf{x}_R[n-s] + \mathbf{u}_{2,J}[n] \\
\mathbf{x}_R[n] &= \sum_{s=1}^{\infty} \mathbf{B}_{RI}[s] \mathbf{x}_I[n-s] + \sum_{s=1}^{\infty} \mathbf{B}_{RJ}[s] \mathbf{x}_J[n-s] + \sum_{s=1}^{\infty} \mathbf{B}_{RR}[s] \mathbf{x}_R[n-s] + \mathbf{u}_{2,R}[n]
\end{aligned}$$

The corresponding covariance matrix of this system is defined as

$$\mathbf{\Sigma}[n] = \begin{pmatrix} \mathbb{V}(\mathbf{u}_{2,I}[n]) & \text{Cov}(\mathbf{u}_{2,I}[n], \mathbf{u}_{2,J}[n]) & \text{Cov}(\mathbf{u}_{2,I}[n], \mathbf{u}_{2,R}[n]) \\ \text{Cov}(\mathbf{u}_{2,J}[n], \mathbf{u}_{2,I}[n]) & \mathbb{V}(\mathbf{u}_{2,J}[n]) & \text{Cov}(\mathbf{u}_{2,J}[n], \mathbf{u}_{2,R}[n]) \\ \text{Cov}(\mathbf{u}_{2,R}[n], \mathbf{u}_{2,I}[n]) & \text{Cov}(\mathbf{u}_{2,R}[n], \mathbf{u}_{2,J}[n]) & \mathbb{V}(\mathbf{u}_{2,R}[n]) \end{pmatrix}.$$

Eliminating the columns and rows corresponding to $\mathbf{x}_J[n]$ we get the only necessary information of this covariance matrix used to analyze if $\mathbf{x}_J[n]$ Granger causes $\mathbf{x}_I[n]$.

$$\mathbf{\Sigma}_1[n] = \begin{pmatrix} \mathbf{\Sigma}_{II} & \mathbf{\Sigma}_{IR} \\ \mathbf{\Sigma}_{RI} & \mathbf{\Sigma}_{RR} \end{pmatrix} = \begin{pmatrix} \mathbb{V}(\mathbf{u}_{2,I}[n]) & \text{Cov}(\mathbf{u}_{2,I}[n], \mathbf{u}_{2,R}[n]) \\ \text{Cov}(\mathbf{u}_{2,R}[n], \mathbf{u}_{2,I}[n]) & \mathbb{V}(\mathbf{u}_{2,R}[n]) \end{pmatrix}$$

Compared with the covariance matrix of the two dimensional system ($\mathbf{S}[n]$), we now find similar information in $\mathbf{\Sigma}_1[n]$. Again the effects of common latent variables (common variables of $\mathbf{x}_I[n]$ and $\mathbf{x}_R[n]$) are filtered out with the help of the partial covariance matrix

$$\mathbf{\Sigma}_{x_I x_J | x_R} = \mathbf{\Sigma}_{II} - \mathbf{\Sigma}_{IR} \mathbf{\Sigma}_{RR}^{-1} \mathbf{\Sigma}_{RI}.$$

Definition 3.16 (Partial Granger causality). $\mathbf{x}_J[n]$ *partially Granger causes* $\mathbf{x}_I[n]$, if and only if the partial covariance matrix of the model including past values of $\mathbf{x}_J[n]$ ($\mathbf{\Sigma}_{x_I x_J | x_R}$) is smaller than the one if the information of past values from $\mathbf{x}_J[n]$ ($\mathbf{S}_{x_I | x_R}$) is not used.

In other words we speak about *partial Granger causality*, if the estimation of $\mathbf{x}_I[n]$ is improved after adding the information obtained by past values of $\mathbf{x}_J[n]$.

Definition 3.17. The *partial Granger causality index* is given by

$$pGCI_{\mathbf{x}_J \rightarrow \mathbf{x}_I} = \ln \left(\frac{\mathbf{S}_{x_I | x_R}}{\mathbf{\Sigma}_{x_I x_J | x_R}} \right) = \ln \left(\frac{\mathbf{S}_{II} - \mathbf{S}_{IR} \mathbf{S}_{RR}^{-1} \mathbf{S}_{RI}}{\mathbf{\Sigma}_{II} - \mathbf{\Sigma}_{IR} \mathbf{\Sigma}_{RR}^{-1} \mathbf{\Sigma}_{RI}} \right) \quad (3.15)$$

Remark. Again if the partial Granger causality index equals zero, there is no improvement in the prediction and no partial Granger causal relation.

3.3.2 Comparison between conditional and partial Granger causality

Before starting with the comparison between conditional and partial Granger causality, we give an example to see what might happen with the conditional index if strong influences of latent variables are present.

Example 3.18 (Conditional Granger causality index under strong influence of latent variables). Let us consider the following situation: Given three variables x_1 , x_2 and x_3 , we want to find out if x_2 Granger causes x_1 .

The variables are fit into an AR-model, disturbed by individual white noise errors ϵ_i , $i = 1, \dots, 3$ and an independent exogenous variable ϵ^E .

$$\begin{aligned} x_1[n] &= \sum_{i=1}^{\infty} b_{11}x_1[n-i] + \sum_{i=1}^{\infty} b_{12}x_2[n-i] + \sum_{i=1}^{\infty} b_{13}x_3[n-i] + \underbrace{a_1\epsilon^E[n] + \epsilon_1[n]}_{u_1[n]} \\ x_2[n] &= \sum_{i=1}^{\infty} b_{21}x_1[n-i] + \sum_{i=1}^{\infty} b_{22}x_2[n-i] + \sum_{i=1}^{\infty} b_{23}x_3[n-i] + a_2\epsilon^E[n] + \epsilon_2[n] \\ x_3[n] &= \sum_{i=1}^{\infty} b_{31}x_1[n-i] + \sum_{i=1}^{\infty} b_{32}x_2[n-i] + \sum_{i=1}^{\infty} b_{33}x_3[n-i] + a_3\epsilon^E[n] + \epsilon_3[n] \end{aligned}$$

For a Granger causality analysis between x_1 and x_2 we have to consider the two dimensional system, where x_1 is projected on the space spanned by past values of x_1 and x_3 .

$$x_1 = \sum_{i=1}^{\infty} c_{11}x_1[n-i] + \sum_{i=1}^{\infty} c_{13}x_3[n-i] + \underbrace{a_1\epsilon^E[n] + \tilde{\epsilon}_1[n]}_{\tilde{u}_1[n]}$$

The term $\tilde{\epsilon}_1[n] = \epsilon_1[n] + \epsilon_{x_2}[n]$, where $\epsilon_{x_2}[n]$ describes the error in the space of past values of x_2 . Hence $\epsilon^E[n]$ and $\tilde{\epsilon}_1[n]$ are independent.

As known, the conditional Granger causality index of x_2 causing x_1 is defined by

$$cGCI_{x_2 \rightarrow x_1} = \ln \left(\frac{\mathbb{V}(\tilde{u}_1[n])}{\mathbb{V}(u_1[n])} \right).$$

Because $\epsilon^E[n]$ is independent of $\epsilon_1[n]$ and of $\tilde{\epsilon}_1[n]$ we can convert the variances.

$$\begin{aligned} \mathbb{V}(u_1[n]) &= \mathbb{V}(a_1\epsilon^E[n] + \epsilon_1[n]) = a_1^2 \mathbb{V}(\epsilon^E[n]) + \mathbb{V}(\epsilon_1[n]) \\ \mathbb{V}(\tilde{u}_1[n]) &= \mathbb{V}(a_1\epsilon^E[n] + \tilde{\epsilon}_1[n]) = a_1^2 \mathbb{V}(\epsilon^E[n]) + \mathbb{V}(\tilde{\epsilon}_1[n]) \end{aligned}$$

The conditional Granger causality index becomes

$$cGCI_{x_2 \rightarrow x_1} = \ln \left(\frac{a_1^2 \mathbb{V}(\epsilon^E[n]) + \mathbb{V}(\tilde{\epsilon}_1[n])}{a_1^2 \mathbb{V}(\epsilon^E[n]) + \mathbb{V}(\epsilon_1[n])} \right).$$

If we speak about a strong influence of the exogenous variable, this corresponds with an increase in a_1 . The fraction converts to one, whereas the index converts to zero.

We have found an example where strong influences of latent variables distort the conditional index and push it to zero.

Example 3.19. To compare partial Granger causality and conditional Granger causality we are going to present an example, found in a modified way in Guo et al.[16]. To simplify the system, it is disturbed by an exogenous, not measured variable $\epsilon^E(t)$ and a white noise process $\epsilon_i(t)$ ($i = 1, \dots, 5$) only. The following equations represent the simultaneously generated system:

$$\begin{aligned} x_1(t) &= 0.95\sqrt{2}x_1(t-1) - 0.9025x_1(t-2) + \epsilon_1(t) + a_1\epsilon^E(t) \\ x_2(t) &= 0.5x_1(t-2) + \epsilon_2(t) + a_2\epsilon^E(t) \\ x_3(t) &= -0.4x_1(t-3) + \epsilon_3(t) + a_3\epsilon^E(t) \\ x_4(t) &= -0.5x_1(t-2) + 0.25\sqrt{2}x_4(t-1) + 0.25\sqrt{2}x_5(t-1) + \epsilon_4(t) + a_4\epsilon^E(t) \\ x_5(t) &= -0.25\sqrt{2}x_4(t-1) + 0.25\sqrt{2}x_5(t-1) + \epsilon_5(t) + a_5\epsilon^E(t) \end{aligned} \quad (3.16)$$

In the simulation of this model an independent standard normal distributed process is used to model the errors $\epsilon_i(t)$, $i = 1, \dots, 5$ and the exogenous variable $\epsilon^E(t)$. Because the starting values are not important for the simulations, they are chosen manually and held constant for all simulations.

Looking at the coefficient matrix of the system a stationary time series can be identified.

Studying the equations above we may see that $x_1(t)$ causes the time series $x_2(t)$, $x_3(t)$ and $x_4(t)$. A feedback situation is modeled between the time series $x_4(t)$ and $x_5(t)$. To keep track of the system, figure 3.12 represents the relations between the variables

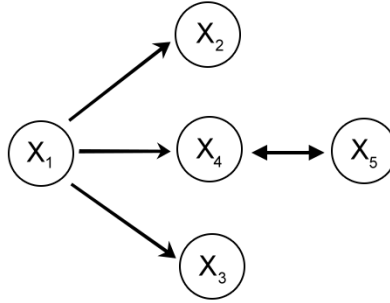


Figure 3.12: **Structure of model 3.16:** One variable causing another is represented by an arrow.

Changing the influence of the exogenous variable (by varying the coefficients a_i , $i = 1, \dots, 5$), the robustness of both multivariate extensions of Granger causality is tested. For each case a time series with 1000 samples is simulated.

First a model without common exogenous variables is considered. All disturbing

processes $\epsilon_i(t)$, $i = 1, \dots, 5$ are independent. The coefficients $a_i(t)$ of model 3.16 are set to zero for $i = 1, \dots, 5$.

The results of the analysis are presented in the figure 3.13. The left hand side shows a diagram, where the Granger causality indices are plotted for each possible causal relation. The upper diagram denotes the conditional Granger causality index, whereas the lower one indicates partial Granger causality.

For example the first bar, called “ $1 \rightarrow 2$ ” (caption on the x -axis) presents the value of the Granger causality index of variable $x_1(t)$ causing $x_2(t)$. In both diagrams the value is clearly higher than zero. Hence a Granger causal relation is identified in the case of conditional Granger causality as well as in the case of the partial extension.

The right hand side of the figure shows the causal relations in the neatly arranged way we have already used before. In both analysis the true structure is identified.

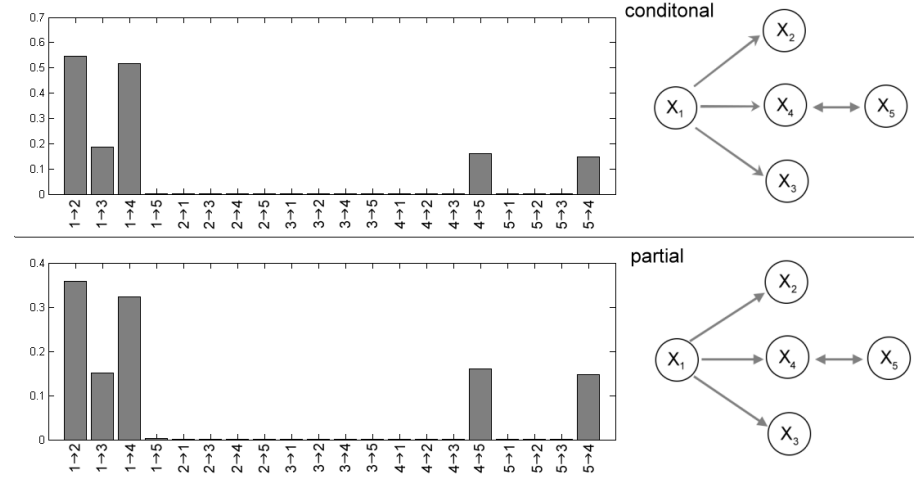


Figure 3.13: **Conditional and partial Granger causality index** of model 3.16 without latent variables. In both analysis the correct system is identified.

Next a system is considered where an exogenous variable $\epsilon^E(t)$ has the same influence on each time series. The corresponding coefficients a_i , $i = 1, \dots, 5$ are equal to 5. The results of this analysis are presented in figure 3.14.

As we may see the conditional Granger causality index fails to find the correct causal relations, whereas the partial Granger causality seems to be the right way for dealing with such data. We notice that the conditional Granger causality index is nearly equal to zero. This phenomenon is found when exogenous variables have a strong influence on the data. For a justification we want to refer to Guo et al. [16] or to example 3.18.

The last case to be analyzed is, where the exogenous variable affects different time series with different strength. Some time series will be even independent of this variable, others will be influenced strongly. Hence we choose the corresponding coefficients (of model 3.16) as $a_1 = 0$, $a_2 = 4$, $a_3 = 6$, $a_4 = 3$ and

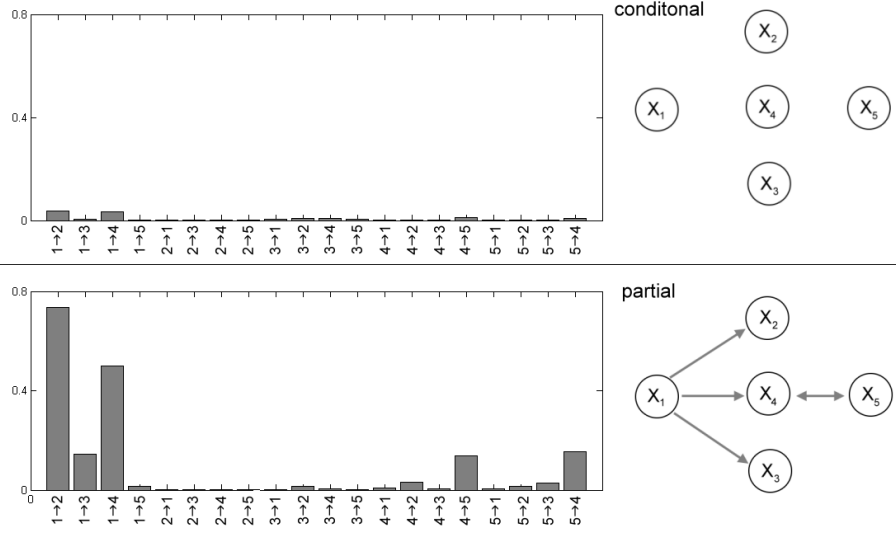


Figure 3.14: **Analysis of system 3.16 with strong influencing latent variables:** Only the partial Granger causality index identifies the correct structure.

$a_5 = 0$. Again the result is summarized in figure 3.15.

We see that, in contrary to the results from conditional Granger causality, the partial Granger causality index delivers correct results. In the case of the conditional Granger causality index the system finds only the relation $x_4(t)$ Granger causing $x_5(t)$ clearly. Also $x_1(t)$ Granger causing $x_2(t)$ and $x_4(t)$ is identified, but both relations are not as strong as in the partial case. (The conditional indices show values around 0.05 and 0.06.)

3.4 Implementation

3.4.1 Recursive Least Square algorithm

In this chapter we introduce the *Recursive Least Square algorithm (RLS)*, which copes quite good with non-stationary signals. Therefore it seems to be a reasonable concept to model the ECoG-data. We follow the derivation of Haykin [18].

3.4.1.1 Preliminary remarks

The basic idea behind this algorithm is an implementation of the ordinary least squares estimation, where the information is updated from sample to sample. This model is not time independent any more: For every point of time the information and the coefficient matrices differ.

Hence this concept is able to cope with non-stationary data. It seems to be useful to work with a recursive model, which adapts the former information. Initial conditions are necessary to start the recursion.

Let $T \in \mathbb{N}$ be the number of observations and let $\mathbf{X} \in \mathbb{R}^{T \times k}$ be the matrix

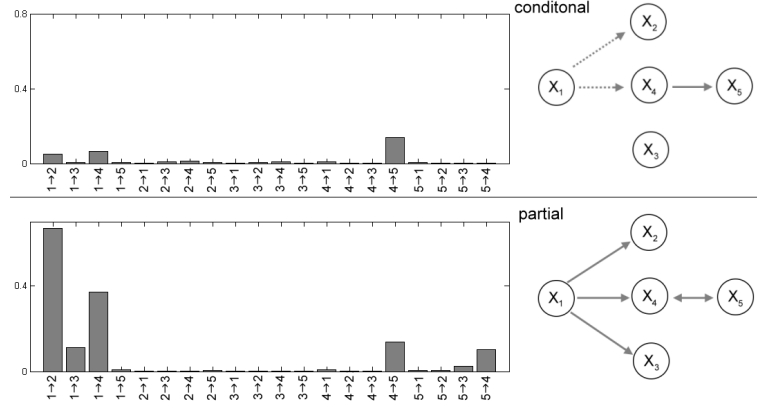


Figure 3.15: **Analysis of system 3.16 with latent variables, having different influences:** The conditional Granger causality index seems to have difficulties to deal with this situation.

of data points, where $\mathbf{x}[n] = (x_1[n], \dots, x_k[n])$ is the k -dimensional signal, observed in time n for $n = 1, \dots, T$.

Because of instationarity we try to fit the ECoG-data in a multivariate and *time-invariant* $AR(p)$ -model

$$\mathbf{x}[n] = \sum_{i=1}^p \mathbf{A}_i[n] \mathbf{x}[n-i] + \epsilon[n],$$

where $\epsilon[n]$ describes a white noise process with covariance matrix Σ_ϵ .

To rewrite this in a more compact way, we assemble the coefficient matrices $\mathbf{A}_i[n]$ (for $i = 1, \dots, p$) and get the matrix

$$\mathbf{A}[n] = (\mathbf{A}_1[n], \dots, \mathbf{A}_p[n]) \in \mathbb{R}^{k \times kp}.$$

Furthermore the p past values of the signal are stacked together in an observation matrix $\bar{\mathbf{X}}[n]$.

$$\bar{\mathbf{X}}[n] = (\mathbf{x}[n-1], \dots, \mathbf{x}[n-p]) \in \mathbb{R}^{1 \times kp}$$

Now the $AR(p)$ -model can be described shortly as

$$\mathbf{x}[n] = \mathbf{A}'[n] \bar{\mathbf{X}}[n] + \epsilon[n].$$

For an ordinary least squares estimation we have to compute the following normal equations⁷

$$\hat{\mathbf{A}}[n] = (\bar{\mathbf{X}}'[n] \bar{\mathbf{X}}[n])^{-1} (\bar{\mathbf{X}}'[n] \mathbf{x}[n]),$$

⁷For more information see Schönfeld [29].

where $(\bar{\mathbf{X}}'[n]\bar{\mathbf{X}}[n])$ is the multivariate covariance matrix $\mathbf{R}_{x,p}$ and $(\bar{\mathbf{X}}'[n]\mathbf{x}[n])$ is the multivariate covariance vector $\mathbf{r}_{x,p}$.

Contrary to the ordinary least squares computation we want to weight the errors $e[i]$ with $\beta(n, i) \in (0, 1]$. With this modification the cost function, which has to be minimized, becomes

$$C[n] = \sum_{i=1}^n \beta(n, i)(e[i])^2,$$

where $e(i)$ denotes the estimation error.

$$e[i] = x[i] - \hat{x}[n] = x[i] - \mathbf{A}'[i] \bar{\mathbf{X}}[i]$$

If the weighting factor $\beta(n, i)$ is smaller than one, the errors of the past become less important - or will become more or less forgotten (after some time). An exponential forgetting factor $\beta(n, i) = \lambda^{n-i}$ is used to weight the errors.

The special case where λ equals one, corresponds to the ordinary least squares estimation. Here we have an infinite memory: each past value is as important as each present value.

There exist two contrary effects to be considered when choosing the forgetting factor: On the one hand a small λ leads to a better adaption, but on the other hand the estimation gets worse. The problem of finding the best factor will be analyzed in detail in Section 3.4.3. For our instationary data a λ smaller, but close to one will be reasonable.

To choose an appropriate λ it is crucial to know how long errors will play a role in the estimation, according to the chosen forgetting factor. Hence we consider values with a weight less than e^{-1} (≈ 0.17) to be “forgotten”.

To express when a sample becomes unimportant, we would like to compute a relationship between λ and the number M of non-negligible values.

$$\lambda^M < e^{-1} \Leftrightarrow M > \frac{-1}{\ln \lambda}$$

For a given frequency f_s we get the memory time constant:

$$\tau_\lambda = \frac{M}{f_s}.$$

The choice of an appropriate value for λ will be based on the analysis of the one-step prediction error. (For this discussion see Section 3.4.3.)

3.4.1.2 Derivation of the RLS

After getting some idea about this algorithm we want to go into more detail.

Before we are able to concentrate on the derivation, we give a technical lemma.

Lemma 3.20 (Matrix Inversion Lemma). *Let A and B be two positive-definite $p \times p$ matrices, D is a positive-definite $T \times p$ matrix and let C be an $p \times T$ matrix. The matrix A should satisfy the relation*

$$A = B^{-1} + CD^{-1}C'.$$

The inverse of the matrix A is defined by

$$A^{-1} = B - BC(D + C'BC)^{-1}C'B.$$

Back to our problem finding a recursive algorithm, we have to minimize the cost function with the weighted estimation errors $\mathbf{e}[i] = \mathbf{x}[i] - \mathbf{A}'[i]\bar{\mathbf{X}}[i]$:

$$\begin{aligned} C[n] &= \sum_{i=1}^n \lambda^{n-1} (\mathbf{e}[i]) (\mathbf{e}[i])' \\ &= \sum_{i=1}^n (\lambda^{\frac{n-1}{2}} \mathbf{x}[i] - \mathbf{A}'[i] \lambda^{\frac{n-1}{2}} \bar{\mathbf{X}}[i]) (\lambda^{\frac{n-1}{2}} \mathbf{x}[i] - \mathbf{A}'[i] \lambda^{\frac{n-1}{2}} \bar{\mathbf{X}}[i])' \end{aligned}$$

As for a ordinary least squares estimation with the descriptive variable $\lambda^{\frac{n-1}{2}} \bar{\mathbf{X}}[n]$ and the dependent variable $\lambda^{\frac{n-1}{2}} \mathbf{x}[n]$, we obtain the following normal equation:

$$\sum_{i=1}^n \lambda^{n-i} \mathbf{x}[i] \mathbf{x}'[i] \hat{\mathbf{A}}[n] = \sum_{i=1}^n \lambda^{n-i} \mathbf{x}[i] \mathbf{y}'[i].$$

As we want to write this equation in a compact form, we define

$$\begin{aligned} \Phi[n] &= \sum_{i=1}^n \lambda^{n-i} \mathbf{x}[i] \mathbf{x}'[i] \\ \mathbf{z}[n] &= \sum_{i=1}^n \lambda^{n-i} \mathbf{x}[i] \mathbf{y}'[i]. \end{aligned} \tag{3.17}$$

We receive

$$\Phi[n] \hat{\mathbf{A}}[n] = \mathbf{z}[n]. \tag{3.18}$$

To get the estimation for the coefficients $\mathbf{A}[n]$ we have to invert the positive definite $p \times p$ covariance matrix $\Phi[n]$.

Because we like to receive a recursive algorithm, the definition of the matrix $\Phi[n]$ (equation 3.17) is rewritten as:

$$\Phi[n] = \lambda \Phi[n-1] + \mathbf{x}[n] \mathbf{x}'[n].$$

To invert $\Phi[n]$ for every $n = 1, \dots, T$ the Matrix Inversion Lemma is used (to reduce the calculation effort). $\Phi[n]$ is positive definite and non singular. Hence the inversion exists. For a better application of this lemma we define $A = \Phi[n]$, $B^{-1} = \lambda \Phi[n-1]$, $C = \bar{\mathbf{X}}[n]$ and $D = 1$ and get

$$\Phi^{-1}[n] = \lambda^{-1} \Phi^{-1}[n-1] - \frac{\lambda^{-2} \Phi^{-1}[n-1] - \bar{\mathbf{X}}[n] \bar{\mathbf{X}}'[n] \Phi^{-1}[n-1]}{1 + \lambda^{-1} \bar{\mathbf{X}}'[n] \Phi^{-1}[n-1] \bar{\mathbf{X}}[n]}. \tag{3.19}$$

For simplification reasons we rename

$$\mathbf{P}[n] = \Phi^{-1}[n] \quad (3.20)$$

$$\mathbf{k}[n] = \frac{\lambda^{-1}\mathbf{P}[n-1]\bar{\mathbf{X}}[n]}{1 + \lambda^{-1}\bar{\mathbf{X}}^T[n]\mathbf{P}[n-1]\bar{\mathbf{X}}[n]}. \quad (3.21)$$

Reminding equation 3.19 and using definitions 3.20 and 3.21, an updating relationship for the inverse correlation matrix $\Phi^{-1}(n)$ (respectively $\mathbf{P}[n]$) is found:

$$\mathbf{P}[n] = \lambda^{-1}\mathbf{P}[n-1] - \lambda^{-1}\mathbf{k}[n]\bar{\mathbf{X}}'[n]\mathbf{P}[n-1]. \quad (3.22)$$

With this equation we are able to interpret the vector $\mathbf{k}[n]$ as a gain vector:

$$\begin{aligned} \mathbf{k}[n] &\stackrel{(3.21)}{=} \lambda^{-1}\mathbf{P}[n-1]\bar{\mathbf{X}}[n] - \lambda^{-1}\mathbf{k}[n]\bar{\mathbf{X}}'[n]\mathbf{P}[n-1]\bar{\mathbf{X}}[n] \\ &= (\lambda^{-1}\mathbf{P}[n-1] - \lambda^{-1}\mathbf{k}[n]\bar{\mathbf{X}}'[n]\mathbf{P}[n-1])\bar{\mathbf{X}}[n] \\ &\stackrel{(3.22)}{=} \mathbf{P}[n]\bar{\mathbf{X}}[n] \end{aligned} \quad (3.23)$$

To establish a recursion for the estimated coefficients, equation 3.18 is used. The right hand side can be rewritten as

$$\mathbf{z}[n] = \sum_{i=1}^n \lambda^{n-i}\bar{\mathbf{X}}[i]\mathbf{x}'[i] = \lambda\mathbf{z}[n-1] + \bar{\mathbf{X}}[n]\mathbf{x}'[n].$$

The *a-priori estimation error*⁸ is defined as

$$\hat{\xi}[n] = \mathbf{x}[n] - \hat{\mathbf{A}}'[n-1]\bar{\mathbf{X}}[n]. \quad (3.24)$$

Contrary the *a-posteriori error*

$$\mathbf{e}[n] = \mathbf{x}[n] - \hat{\mathbf{A}}'[n]\bar{\mathbf{X}}[n]$$

shows the error made by fitting the signal \mathbf{x} in an AR-model using all information up to time n . It represents the estimated white noise in n .

For the estimation of the coefficients $\hat{\mathbf{A}}[n]$ we take in mind that we have defined $\mathbf{P}[n]$ as $\Phi^{-1}[n]$. To come from the second row to the third row of the calculation we substitute the recursion of $\mathbf{P}[n]$. We obtain the following recursion:

$$\begin{aligned} \hat{\mathbf{A}}[n] &= \Phi^{-1}[n]\mathbf{z}[n] \\ &= \mathbf{P}[n]\mathbf{z}[n] \\ &= \lambda\mathbf{P}[n]\mathbf{z}[n-1] + \mathbf{P}[n]\bar{\mathbf{X}}[n]\mathbf{x}'[n]. \end{aligned}$$

Using the recursive representation of $\mathbf{P}[n]$ (equation 3.22) we get

⁸The a-priori estimation error conforms with the one step prediction error. The real value $\mathbf{x}[n]$ is estimated using the information obtained up to time $n-1$: $\hat{\mathbf{x}}[n] = \hat{\mathbf{A}}'[n-1]\bar{\mathbf{X}}[n]$.

$$\begin{aligned}
\hat{\mathbf{A}}[n] &= \underbrace{\mathbf{P}[n-1]\mathbf{z}[n-1]}_{\hat{\mathbf{A}}[n-1]} + \mathbf{k}[n]\bar{\mathbf{x}}^T[n] \underbrace{\mathbf{P}[n-1]\mathbf{z}[n-1]}_{\hat{\mathbf{A}}[n-1]} + \underbrace{\mathbf{P}[n]\bar{\mathbf{x}}[n]\mathbf{x}'[n]}_{\mathbf{k}[n]} \\
&= \hat{\mathbf{A}}[n-1] + \mathbf{k}[n] \underbrace{(\mathbf{x}^T[n] - \bar{\mathbf{x}}'[n]\hat{\mathbf{A}}[n-1])}_{\hat{\xi}[n]} \\
&= \hat{\mathbf{A}}[n-1] + \mathbf{k}[n]\hat{\xi}[n].
\end{aligned} \tag{3.25}$$

Thus we have found a recursion for the estimated coefficient matrix $\hat{\mathbf{A}}[n]$.

To be able to start with this recursive adaption we have to find a useful initialization. A first inverted covariance matrix $\mathbf{P}[0]$ has to be defined.

One way to compute such a first inverse covariance matrix is to use the first $(-n_0, \dots, 0)$ values:

$$\mathbf{P}[0] = \left(\sum_{i=-n_0}^0 \lambda^{-i} \bar{\mathbf{x}}[i] \bar{\mathbf{x}}'[i] \right)^{-1}.$$

Another way of initialization is to define $\mathbf{P}[0]$ with the $p \times p$ identity matrix ($\mathbf{I}_{p \times p}$) and a positive small constant δ

$$\mathbf{P}[0] = \frac{1}{\delta} \mathbf{I}_{p \times p}.$$

We are using the second way setting $\delta = 1$ and getting the identity matrix for $\mathbf{P}[0]$.

At the beginning of the algorithm, where no information is available, it seems to be reasonable to set the starting coefficient matrix $\mathbf{A}[0]$ to the zero matrix.

3.4.1.3 Summary of the RLS algorithm

To get an overview which steps have to be done for each sample we want to summarize the RLS-algorithm. The four equations (3.21), (3.22), (3.24) and (3.25) are enough to find the coefficient matrices $\hat{\mathbf{A}}[n]$ and further the estimated AR-model (algorithm 3.1).

The one step prediction error (a-priori error) is used to find the coefficient matrices. Later this error will be used to determine the correct forgetting factor λ (Section 3.20) and to calculate the Granger causality indices (in Section 3.5).

3.4.2 Channel selection

The ECoG data we are dealing with, contains between 28 and 32 different variables (channels). Using all of them for the calculations would mean on the one hand a big calculation effort and on the other hand we would face serious problems with the stability of the algorithm.

Therefore we have to choose a set of channels to work with. This can be done in two ways:

- (i) Finding an algorithm which automatically chooses a set of variables to work with:

A detailed description might found in [12, 17]: The set of channels is chosen

Algorithm 3.1 Recursive Least Squares (RLS) algorithm

Initialization:

$$\begin{aligned}\hat{\mathbf{A}}[0] &= 0 \\ \mathbf{P}[0] &= \frac{1}{\delta} \mathbf{I}_{p \times p}.\end{aligned}$$

Compute for $n = p, \dots, N$:

$$\begin{aligned}\mathbf{k}[n] &= \frac{\lambda^{-1} \mathbf{P}[n-1] \bar{x}[n]}{1 + \lambda^{-1} \bar{x}'[n] \mathbf{P}[n-1] \bar{x}[n]} \\ \mathbf{P}[n] &= \lambda^{-1} \mathbf{P}[n-1] - \lambda^{-1} \mathbf{k}[n] \bar{x}'[n] \mathbf{P}[n-1] \\ \hat{\xi}[n] &= \mathbf{x}[n] - \hat{\mathbf{A}}'[n-1] \bar{\mathbf{x}}[n] \\ \hat{\mathbf{A}}[n] &= \hat{\mathbf{A}}[n-1] + \mathbf{k}[n] \hat{\xi}[n]\end{aligned}$$

by an algorithm, which decides for every channel whether an additional channel yields enough gain to add it or not.

(ii) Choosing the set of channels manually.

We have decided to choose the set of channels manually. On one side it seems to be reasonable, because the channel selection algorithm takes much calculation effort. On the other side we keep the set constant during the whole seizure, contrary to the channel selection algorithm. Furthermore we keep it constant for every seizure of a patient.

For our selection we tried to consider all “interesting” channels (these which indicate a lot of epileptic activity and these located near to them). The problem of finding these channels will be discussed in detail in Section 3.5.

Example 3.21 (Reduction of the information). To get an idea how important the channel selection might be, we are presenting an example, where an incomplete multivariate model may lead to wrong Granger causal relations. The corresponding model can be found in Eichler [8].

$$\begin{aligned}x_1(t) &= \alpha x_4(t-2) + \epsilon_1(t) \\ x_2(t) &= \beta x_3(t-1) + \gamma x_4(t-1) + \epsilon_2(t) \\ x_3(t) &= \epsilon_3(t) \\ x_4(t) &= \epsilon_4(t)\end{aligned}$$

We claim the coefficients α , β and γ to differ from zero and claim the errors $\epsilon_i(t)$, $i = 1, \dots, 4$ to be independent and identically normally distributed with mean zero and variance σ^2 . (For our numerical example we will set $\alpha = \beta = \gamma = 0.9$ and use a standard normal distribution to calculate the white noise $\epsilon_i(t)$, $i = 1, \dots, 4$.)

Reminding the equivalence (found in lemma 3.13) and rewriting the system

$\mathbf{x}(t) = \mathbf{A}(z)\mathbf{x}(t-1) + \epsilon(t)$: The variable $x_i(t)$ Granger causes $x_j(t)$, if and only if the coefficient $\mathbf{A}_{ij}(z)$ differs from zero.

Using this information we are analyzing the system above: Considering the coefficients, we see that variable $x_4(t)$ is Granger causing $x_1(t)$ and $x_2(t)$. Additionally $\mathbf{A}_{23}(1) = \beta$ indicates a causal flow from $x_3(t)$ to $x_2(t)$.

For a better overview a graphical representation is presented in figure 3.16. As in earlier sections we denote the variables $x_i(t)$ with a node and a causal relationship between two components using arrows.

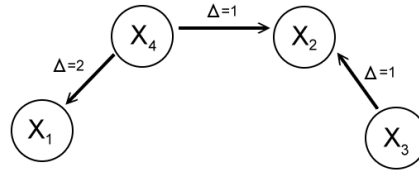


Figure 3.16: **Structure:** The arrows denote the influence of a signal to another. ($\Delta = i$, $i = 1, 2$ denotes a delay by one or by two samples.)

The numerical result is given in figure 3.17, where the bars on the left hand side represent the corresponding conditional Granger causality indices and on the right hand side the system is presented in the usual way. The system is correctly identified.

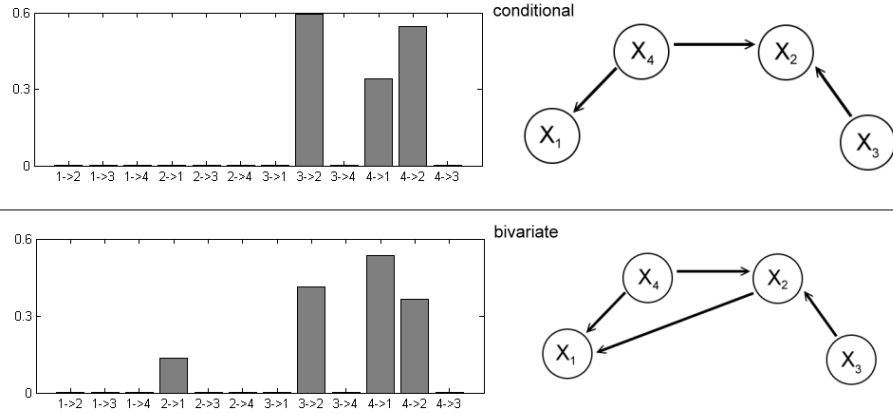


Figure 3.17: **Conditional and bivariate Granger causality index:** While using the multivariate method, the whole system is correctly identified, in the bivariate case an additional relations (X_2 causing X_1) is received.

In the next step we show what will happen, if not all variables are measured or selected. So for example what does happen if we only select the channels 1 to 3? It can be shown (see Eichler [8]) that therefore the following reduced model is obtained:

$$\begin{aligned}
x_1(t) &= \frac{\alpha\beta}{1+\beta^2} x_2(t-1) - \frac{\alpha\beta\gamma}{1+\beta^2} x_3(t-2) + \tilde{\epsilon}_1(t) \\
x_2(t) &= \gamma x_3(t-1) + \tilde{\epsilon}_2(t) \\
x_3(t) &= \tilde{\epsilon}_3(t)
\end{aligned}$$

It can be shown easily that $\tilde{\epsilon}_1(t) = \epsilon_1(t) - \frac{\alpha\beta}{1+\beta^2} \epsilon_2(t-1) + \frac{\alpha}{1+\beta^2} x_4(t-2)$, $\tilde{\epsilon}_2(t) = \beta x_4(t-1) + \epsilon_2(t)$ and $\tilde{\epsilon}_3(t) = \epsilon_3(t)$ are independent and identically normally distributed with mean zero.

($\text{Cov}(\tilde{\epsilon}_1(t), \tilde{\epsilon}_2(t-1)) = -\frac{\alpha\beta}{1+\beta^2} \mathbb{V}(\epsilon_2(t-1)) + \frac{\alpha}{1+\beta^2} \beta \mathbb{V}(x_4(t-2)) = 0$, all other covariances are zero, because $\epsilon_i(t)$, $i = 1, \dots, 3$ are independent normally distributed.)

Again we want to analyze the multivariate system and search for conditional Granger causality. Because the coefficients $a_{12}(1)$, $a_{13}(2)$ and $a_{23}(1)$ differ from zero, we have $x_3(t)$ Granger causing $x_1(t)$ and $x_2(t)$, and further $x_2(t)$ Granger causing $x_1(t)$. Compared to the true relationship (we had in the four dimensional case), the multivariate systems tells us that $x_3(t)$ causes $x_1(t)$, although in the original system there was no connection between the two.

The numerical result (presented in figure 3.18) leads to the same result as in the theoretical consideration.

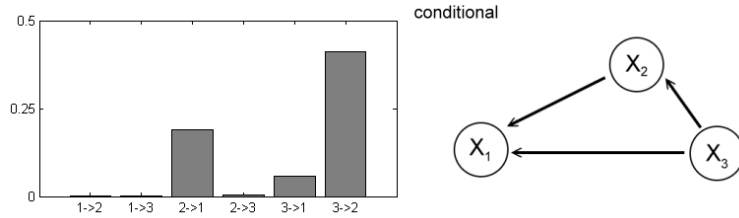


Figure 3.18: **Conditional Granger causality index for the partitioned system:** Contrary to the original system the signal X_3 is Granger causing X_1 .

Something quite interesting in this example is that the bivariate analysis for the 3-dimensional system leads to better results than that using conditional Granger causality.

In figure 3.17 the result of the analysis from the whole system is presented.

We see that it correctly marks a causality between $x_3(t)$ and $x_2(t)$, between $x_4(t)$ and $x_1(t)$ respectively $x_2(t)$. The relationship between $x_1(t)$ and $x_2(t)$ arises, because both variables are influenced by the third variable $x_4(t)$.

In the bivariate case the three dimensional reduction leads to figure 3.19.

Comparing this result with that of Section 3.2.2, it leads to a contrary result. In this example the bivariate case copes better with a reduction of the dimension. What we like to show with this example is, that we have to be very careful when selecting the channels for multivariate calculations. Maybe the bivariate analysis should be used to reduce the set of channels.

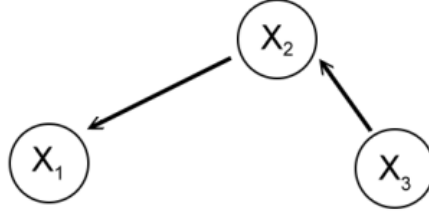


Figure 3.19: *In this case the bivariate Granger causality indicates the correct relations.*

3.4.3 Selection of the forgetting factor λ

Next an optimal forgetting factor λ is needed. Before starting with the calculations we give a short review:

To get a better adaptation of ECoG data to an AR model, less weight is put on residual errors of the past, than on that nearer to the present. The cost function, which has to be minimized, is defined as

$$C[n] = \sum_{i=1}^n \lambda^{n-i} e^2[i],$$

with $\lambda < 1$ (to let errors in the past get less important).

To find the best value of the forgetting factor, we have to consider a trade off between two opposite tendencies:

- (i) Small values of λ lead to a “short memory” (residual errors of samples far in the past become quickly unimportant for the estimation). This results in a good adaptation.
- (ii) A “short memory” means a few samples for the estimation and that the prediction becomes worse.

One advantage of the concept (using a forgetting factor) is to be able to deal with a sudden change in the system. For example if one channel indicates an epileptic activity we want to be able to adapt the AR-system as quickly as possible. Only then we can identify the epileptic activity quickly by using the concept of Granger causality.

Hence the ECoG signal of a patient during an epileptic seizure is analyzed and the best adaption (the corresponding λ) of the point of time, where the seizure is starting to effect this signal is searched.

Figure 3.20 presents the situation, which has to be modeled using the recursive least squares algorithm with a forgetting factor λ . The epileptic activity starts at second 11, where we seem to have a change in the system: The amplitude of the signal is getting larger.

To determine the forgetting factor, the one step prediction error seems to be a useful instrument to solve this problem. On the one hand a quite good

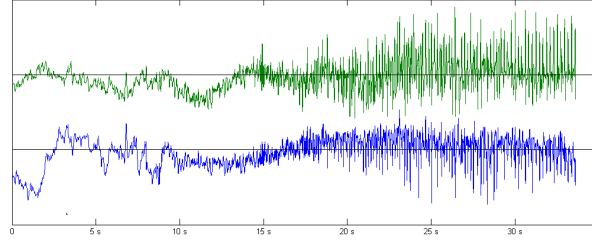


Figure 3.20: **Channels used to find an optimal lambda:** The epileptic activity starts at second 11. We see that after this time the systems changes and starts to have a higher amplitude.

adaption is necessary to be able to predict the next value and on the other hand a longer memory is needed to be able to predict the system. We plot the variances of the one step prediction error against the corresponding λ (figure 3.21). The forgetting factor with the smallest variance ($\lambda = 0.995$) is taken for our later analysis of the data in Section 3.5.

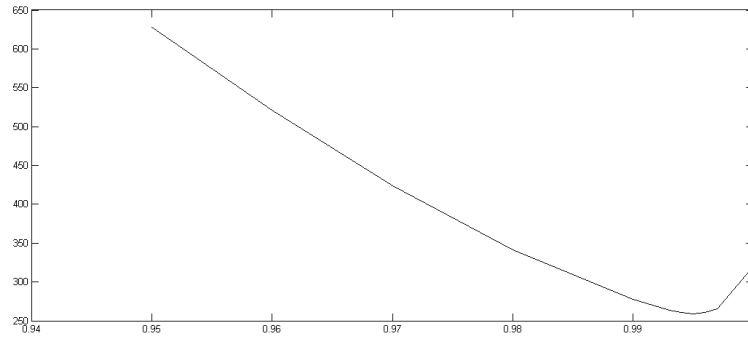


Figure 3.21: **Forgetting factor lambda vs. variance of the one-step prediction error:** We can see that the minimum is around $\lambda = 0.995$.

To get an idea about the effect of choosing this special forgetting factor, we remind the connection between λ and the number M of non-negligible values in Section 3.4.1.

$$\lambda^M < e^{-1} \Leftrightarrow M > \frac{-1}{\ln \lambda}$$

For a given frequency f_s the memory time constant is

$$\tau_\lambda = \frac{M}{f_s}.$$

Using frequency $f_s = 128\text{HZ}$, we obtain the following result for some interesting values of the forgetting factor λ .

Using $\lambda = 0.995$ for the analysis in the later work means that we have to consider around 200 samples for the estimation. This corresponds to a time window of 1.6 seconds.

forgetting factor (λ)	M non-negligible samples	memory time constant (τ_λ)
0.950	20	0.15 s
0.970	33	0.26 s
0.990	100	0.78 s
0.995	200	1.56 s
0.999	1000	7.81 s

Table 3.2: *Connection: forgetting factor λ , the number M of non-negligible samples and the memory time constant (τ_λ)*

3.4.4 Order selection

Next a reasonable lag order for the RLS-algorithm is needed. To be able to decide how large this lag should be, we consider the value of the Akaike's Information Criterion (AIC).⁹

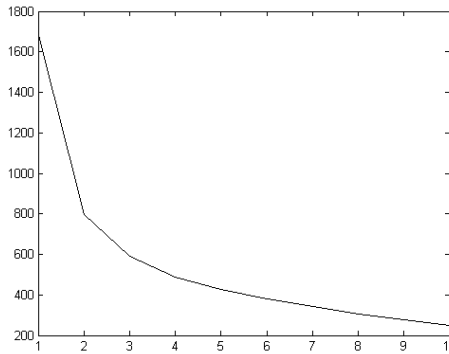


Figure 3.22: **AIC criterion for the ECoG data:** After order $p = 6$ the gain of an higher order sinks rapidly and the question arises if the higher effort of an higher order is necessary.

Hence the ECoG data of a patient during an epileptic seizure is analyzed. As in the section before we are going to use the data of patient 2, seizure 3 to find the correct order. Considering figure 3.22, we have a low loss of the AIC value after order $p = 6$. Therefore it has to be decided if it is reasonable to use a much higher order, because the calculation effort rises too much. (Contrary to a usually OLS-estimation, the RLS-algorithm is more complex). Hence we use the order $p = 6$ in the following analysis.

⁹This information criterion is a measure of the quality of a fit of an estimated statistical model, which reacts infictive to the number of parameters.

$$AIC = \ln(\hat{\sigma}^2) + 2 \frac{M}{T}$$

$\hat{\sigma}^2$... variance of the residual errors
 M ... number of estimated parameters
 T ... number of samples

3.5 Results

To conclude this chapter, the results of the ECoG data analysis of three patients, suffering from temporal lobe epilepsy, are presented. All three Granger causality definitions are used and the pros and cons will be shown.

As mentioned in the introduction, one significance of channels affected with epileptic activity is the simultaneous oscillation. As this similar rhythm is caused directly by the seizure's initial focus, it should help us to track the propagation of the seizure and to find its focus. Granger causality measures seem to be a promising tool.

3.5.1 Bivariate Granger causality

Before analyzing the ECoG data, we want to remind the definition of bivariate Granger causality (for more details see Section 3.1).

Analyzing if x_j Granger causes x_i , we try to fit these channels into two AR-models (one with x_j and another without x_j):

$$(i) \quad x_i(n) = \sum_{s=1}^6 a_s x_i(n-s) + \epsilon_1(n)$$

$$(ii) \quad x_i(n) = \sum_{s=1}^6 b_s x_i(n-s) + \sum_{s=1}^6 c_s x_j(n-s) + \epsilon_2(n).$$

Channel x_j (bivariate) Granger causes x_i , if and only if adding the information obtained from past values of x_j improves significantly the prediction of x_i . The measure, which should help in this decision is defined by

$$bGCI_{x_j \rightarrow x_i} = \ln \left(\frac{\mathbb{V}(\epsilon_1(n))}{\mathbb{V}(\epsilon_2(n))} \right).$$

Approaching the problem of finding Granger causal relations for the given instationary data means to fit the channels in an AR-system with the help of the RLS-Algorithm (Section 3.4.1).

Because instationary data may change the causal relations, we calculate the bivariate Granger causality index for each sample and try to find a threshold for each patient, indicating whether the Granger causality index is significantly different from zero or not. To avoid the indication of "spikes", significant interaction is given, if the index exceeds this threshold for 32 times (corresponding to a quarter of a second at the given sampling frequency of 128Hz) within one time window.

3.5.1.1 Why Common Average data are necessary

To explain why using transformed data for the analysis is reasonable (i.e. data, which is redefined with reference to the common average), we want to present the bivariate result of seizure 3 of patient 2 with the original ECoG data - measured with reference to channel 6.

The threshold for the original data is chosen to be 0.25 (which is the same as used for the common average data).

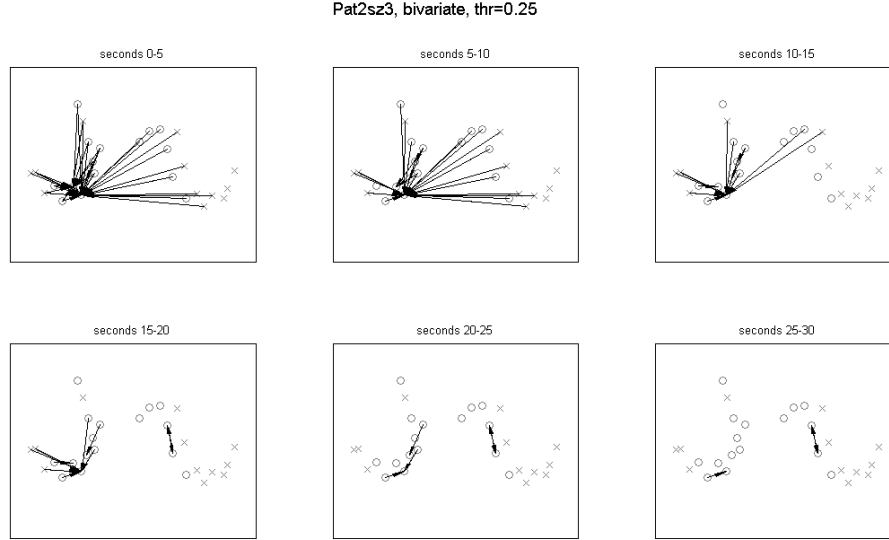


Figure 3.23: *Bivariate analysis of seizure 3, patient 2 - without common average: Considering the result, many channels seem to cause channels 26 and 17.*

We are going to compare the result, obtained from the data measured with reference to channel 6 (figure 3.23) with that measured with reference to the common average of all channels (figure 3.27):

Many channels are causing channel 26 and channel 17. Based on the medical description, channel 26 is declared to be a focus channel. It seems that the arrows (presenting the causal relation) are pointing at the focus (or near to it). Furthermore many arrow starting points seem to be channels, which do not show any epileptic activity.

If we consider the result of the data measured with reference to the common average, this wrong relations vanish.

In figure 3.24 we compare graphs of both data (starting five seconds before the onset of the seizure and ending five seconds after). We consider one channel of the focus (channel 26) and one showing no epileptic activity (channel 1).

While the red line represents the graph of the data, measured with reference to the common average, the blue line shows the data, measured in reference to channel 6. In the first five seconds (before the onset of the seizure), we seem to have a similar situation in both channels. Using common average, many peaks seem to be reduced.

Contrary this effect only is remaining for the data showing no epileptic activity in the first five seconds of the seizure (second 5 to second 10). Channel 26 seems to have more peaks in both ways of presenting the data. Hence we assume that the different presentation might have an effect on the variances of the data.

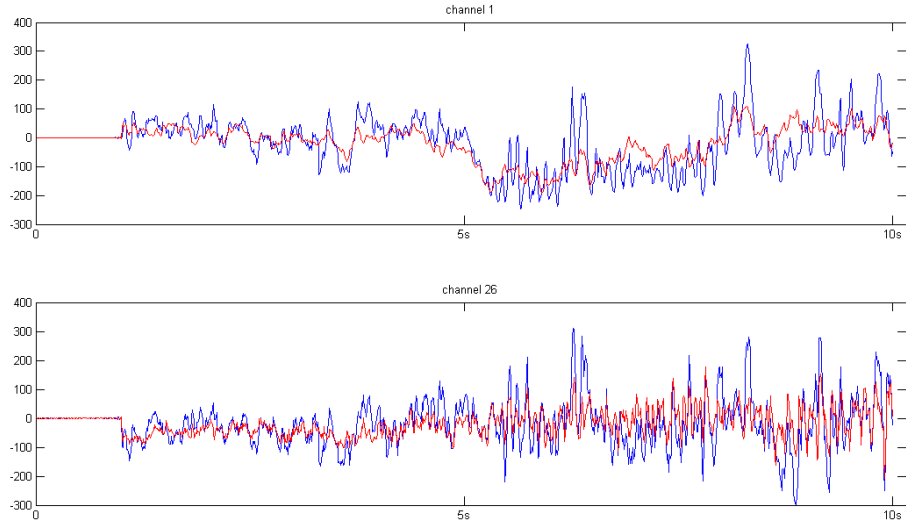


Figure 3.24: *Graph of data measured with reference to a channel (blue line) respectively to the common average (red line).*

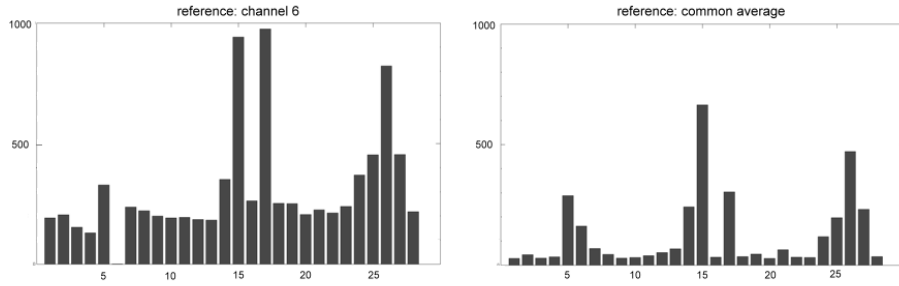


Figure 3.25: **Variances of data with reference to a channel and to the common average:** *Common average reduces the variances.*

Comparing the variances of the data measured with reference to channel 6 and of the data with the common average as reference in figure 3.25, a reduction of the variance is identified in the case of common average. Furthermore we see that especially channels 15, 17 and 26 still have a very high variance.

For a better comparison we want to look at the percentage of the reduction of the variances in figure 3.26.

Channels 15, 17 and 26 (the channels which had a huge variance in the case of channel 6 as reference electrode) are the ones with a relatively low reduction.

We want to give an assumption, why this situation may arise: There could exist some “basic signal”, which is part and parcel of all channels. This basic signal could be a behavior of the brain in situations of resting, but it could also result from external influences. Thus the variance of non-focus channels decreases if common average is used, because that filters the basic signal.

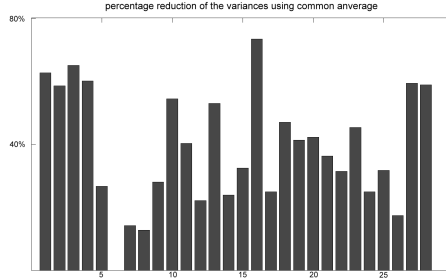


Figure 3.26: **Percentage of variance reduction** between channel 6 as reference and common average as reference.

The channels delivering information in a seizure, are these showing the epileptic activity. Hence these channels cannot be explained with the basic signal - variances stay high, even if the common average is used.

So what could be the reason for the indication of wrong causal influences? Considering the variances in figure 3.25, we notice extremely high variances for channels 15, 17 and 26. High variances correspond to bad estimations (projecting each channel on itself). Hence we assume that the information of any channel is able to improve the estimation significantly, because the basic signal is strongly present in this channel.

For example the variance of channel 17 is reduced quite dramatically if common average is used. Consequently most arrows pointing towards channel 17 disappear.

3.5.1.2 Patient 2, seizure 3

The first data to analyze is seizure 3 of patient 2. According to the clinical description the epileptic activity appears at channels 26 and 27, two neighboring electrodes on the right side of the brain, at 12:31:41. After about eleven seconds the activity propagates to the other hemisphere and affects channels 10, 11 and 12, where the seizure ends after about one minute.

To be able to find a threshold for the bivariate Granger causality index of patient 2, we first plot this index against the time. In figure 3.27 we present some selected graphs of the index, which show a typical behavior. The red line presents the bivariate causality index of channel 10 causing channel $i = 2, 11, 21$, whereas the blue one shows the inverse causal relation.

We may discover that between an affected electrode and any non-affected the index stays smaller than 0.25 (see for example the graph of the causality index between channels 10 and 21 in figure 3.27). Sometimes we could see a short peak reaching this value, but normally this is significantly lasting less than a quarter of a second. An analogous situation may be identified between two not affected channels.

The index between two neighboring channels might be a bit higher than usually (see for example the causality indices between channels 10 and 2 - for the

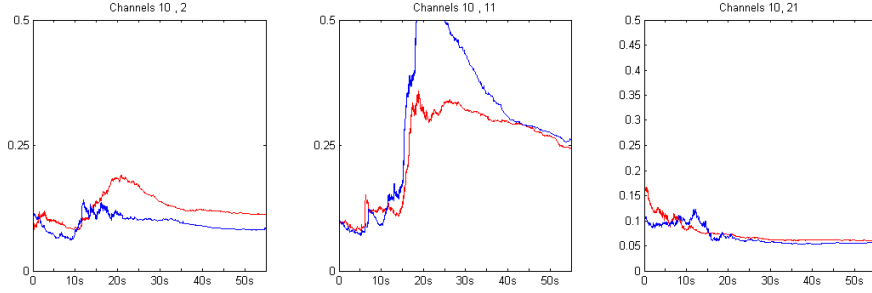


Figure 3.27: **Bivariate Granger causality index:** The red line shows the index of channel 10 causing channel $i = 2, 11, 21$ and the blue line shows channel $i = 2, 11, 21$ Granger causing 10.

localization of the corresponding electrodes see A.2). This might be based on the fact, that neighboring brain regions act similarly - whether there is epileptic activity or not.

Furthermore we recognize that between two affected, neighboring electrodes (channel 10 and 11) the index is high (significantly over 0.25). The indices between these channels increase significantly after about 15 seconds. Comparing this with the clinical description (a propagation to channels 10 and 11 after 11 seconds) the start of the epileptic activity can be indicated pretty well using the bivariate method (only some seconds later as we would expect it).

From these considerations a threshold of value 0.25 seems to be reasonable. Using this for calculation of the first 30 seconds, we receive the 5 second windows of figure 3.28.

Before we are going to analyze the result we want to present the clinical description: The onset of this seizure is at channels 26 and 27 at 12:31:41. After 11 seconds the epileptic activity propagates to channels 10, 11 and 12. At 12:32:39 the seizure ends.

The bivariate analysis does not seem to fit perfectly with the clinical description. Further channels causing another are found.

This result would lead to a focus near electrode 15. Especially in the first part of the seizure it seems that channel 15 might be the cause of the epileptic activity and the activity propagates to channels 26 and 27.

On the other hand we like to remember that in Section 3.2.2 we have found several examples, showing that only relying on the bivariate method can be very risky. One example showed that even an inversion of the causal relation is possible.

Further a focus at channel 26 and 27 cannot be excluded: Causalities between them are indicated from the very beginning.

Looking at seizure 4 of this patient (see Section A.2), the start is found at channel 15. Hence a chance to have the real focus situated near this electrode exists. As we have mentioned before one problem of this data is, that even the

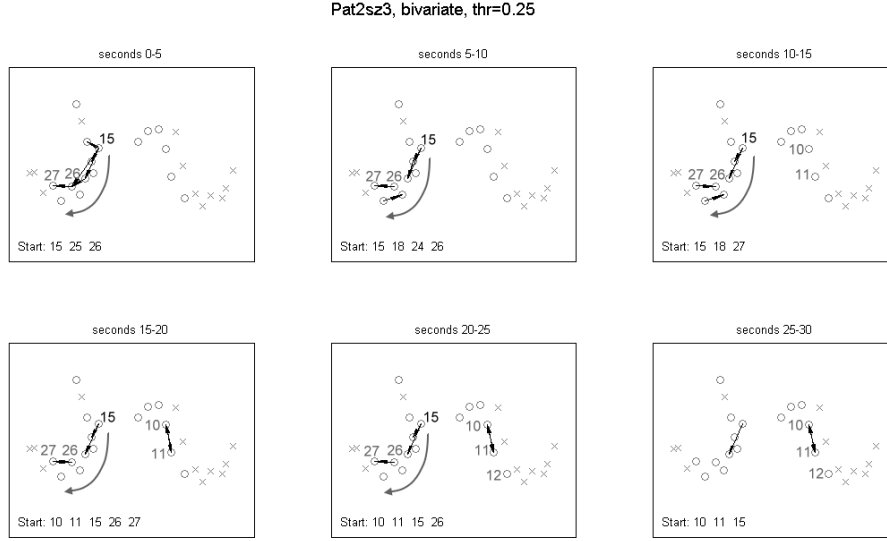


Figure 3.28: *Bivariate Granger causality for seizure 3 of patient 2: We denote the channels to select in the multivariate case with a circle (symbol \circ) whereas the neglected ones are denoted with a cross (\times). In the bivariate case all channels are considered.*

medics were not sure about the real focus.

3.5.1.3 Patient 1, seizure 1

To justify the use of the bivariate method we present a second analysis. Therefore we take another patient, where we found out that a threshold near 0.3 seems to be reasonable. This threshold yields to similar results for each seizure of patient 1.

As before we present the result of the bivariate Granger causality index in several windows, representing 5 seconds. We draw an arrow, if the threshold (0.3) is exceeded for more than 0.25 seconds. The result is presented in figure 3.29.

In the first seconds there is hardly any arrow - except for the two wrong ones at the beginning. The problem could be that no other electrode is affected, except for number 25 on the right hemisphere. Hence no information could flow from one channel to another.

From second 10 to 15 the bivariate method indicates a first causality. This one (from channel 17 to channel 18) is not found in the clinical description at this moment. But it is quite near the starting point of epileptic activity, and otherwise channel 18 is one of the starting points for the other seizures of this patient (Section A.1).

The last graphs show the causal relations we have expected. Bivariate Granger causality marks channels 17, 18, 25 and 26. No wrong relations are found.

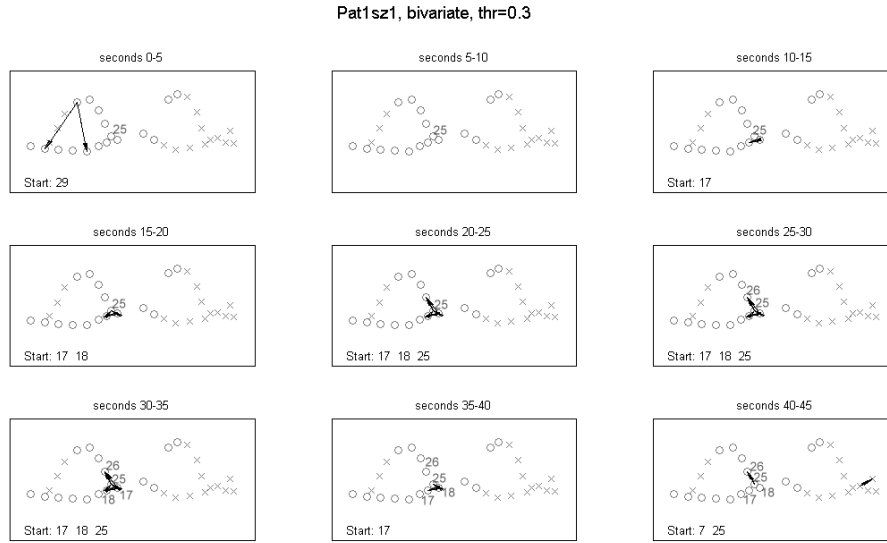


Figure 3.29: **Bivariate analysis of seizure 1, patient 1:** As in the figure above, the electrodes, selected in the multivariate case, are denoted with a circle, the rest with a cross. The channels, which should be affected, are marked.

3.5.1.4 Patient 1, seizure 3

The next example shows that the bivariate analysis could lead to some problems. As said before the threshold is chosen to be 0.3 again.

Using bivariate Granger causality we receive the following result (presented in figure 3.30):

The first plots seem to fit quite good. Comparing them with the clinical description we should identify an epileptic activity at channels 18 and 25: In fact they are in causal relations. Furthermore it seems that they already are influencing channels in the neighborhood.

After the first five seconds, the causal relations seem to be not significant any more. Meanwhile (in second 11) the seizure should propagate to channel 26. This effect can be observed some seconds later.

The last two graphs expose some problems: At the first sight it seems that many arrows point at the focus. But if we examine it in more detail we recognize that none of them points at channels 18 or 25 (i.e. at the focus channels for all seizures).

Hence we suppose that there is either the situation where an arrow changes the direction in the bivariate case (for more details see example 3.15) or that there is the situation where two not connected channels cause a bivariate causal relation (see example 3.14).

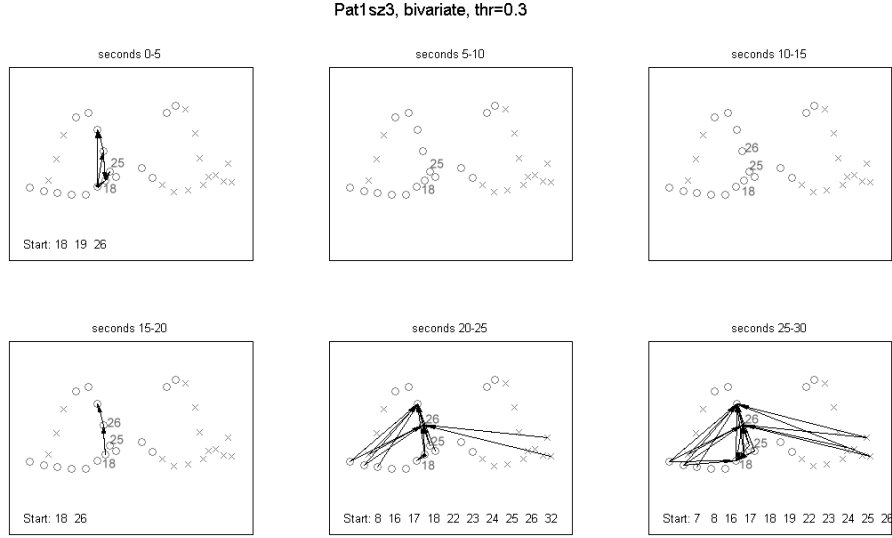


Figure 3.30: **Bivariate Analysis of seizure 3, Patient 1:** The channels, which should be affected according to the medical description are marked. We recognize that after second 20 many arrows point at the neighborhood of the focus.

3.5.2 Conditional Granger causality

First we would like to give a short summary of the definition of Conditional Granger causality.

x_j Granger causes x_i in the conditional definition, if and only if the estimation of x_i is significantly improved when adding the information coming from x_j (conditioned to a set of channels (\mathbf{x}_R)). Again a linear model is used:

$$(i) \quad x_i[n] = \sum_{s=1}^p a_s x_i[n-s] + \sum_{s=1}^p \mathbf{b}_s \mathbf{x}_R[n-s] + \epsilon_1[n]$$

$$(ii) \quad x_i[n] = \sum_{s=1}^p c_s x_i[n-s] + \sum_{s=1}^p d_s x_j[n-s] + \sum_{s=1}^p \mathbf{f}_s \mathbf{x}_R[n-s] + \epsilon_2[n].$$

The quality of the prediction is measured using the variances of $\epsilon_1[n]$ and of $\epsilon_2[n]$: The conditional Granger causality index (cGCI) is:

$$cGCI_{x_j \rightarrow x_i} = \ln \left(\frac{\mathbb{V}(\epsilon_1[n])}{\mathbb{V}(\epsilon_2[n])} \right).$$

Because of instationarity of the ECoG-data, we model the AR-systems with the help of the RLS-algorithm (Section 3.4.1).

3.5.2.1 Channel selection

At first a reasonable set of electrodes has to be found. Therefore we tried to consider as many of the affected channels as possible. To avoid instability of

the algorithm, we have to restrict the number of selected variables to 16.

As mentioned in Section 3.4.2, we are manually selecting the set and keep it constant for each patient. On the one hand we consider the clinical description and try to select the channels, where the doctors found epileptic activity. On the other hand we take a look at the bivariate analysis and try to select many channels found in bivariate causal relations. (This might be one possible step for an automatic algorithm.)

3.5.2.2 Patient 2, seizure 3

From the bivariate results of Patient 2 (results of all four seizures) it seems to be reasonable to select channels 7 to 12, 15 to 18, channel 21 and channels 23 to 27.

After finding a useful set of channels we want to start analyzing the conditional Granger causality index.

Because we are dealing with non-stationary data, we use the same method to find causal relations as in Section 3.5.1. We define one channel causing another if the causality index stays over a threshold for more than a quarter of a second.

Again the problem arises finding such a threshold. For this we considered the conditional Granger causality indices (similar to the bivariate case). We want to present some typical cases in figure 3.31.

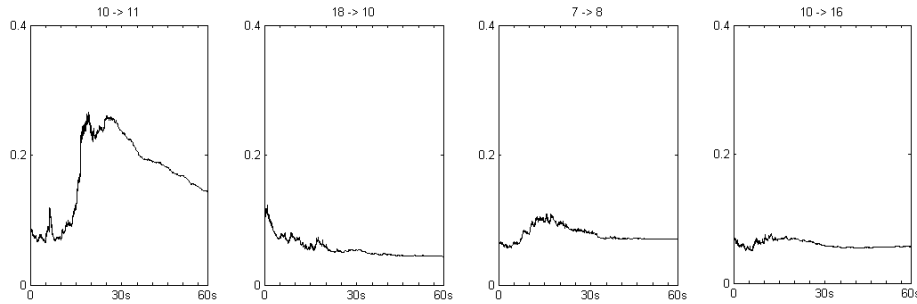


Figure 3.31: **Conditional Granger causality index:** The index increases significantly, if two affected channels (10 and 11) are analyzed. Therefore we are going to choose a threshold near 0.2.

Reminding the clinical description, we identify a significantly higher causality index between two affected and neighbored electrodes (channel 10 and 11) than between two not affected ones (channel 7 and 8) or between an affected and a non-affected one (channel 18 and 10, or 10 and 16).

After considering all seizures of this patient it seems to be reasonable to choose the threshold to be 0.2.

Now we try to concentrate on the first 30 seconds of this seizure in figure 3.32 and compare the result with the bivariate case.

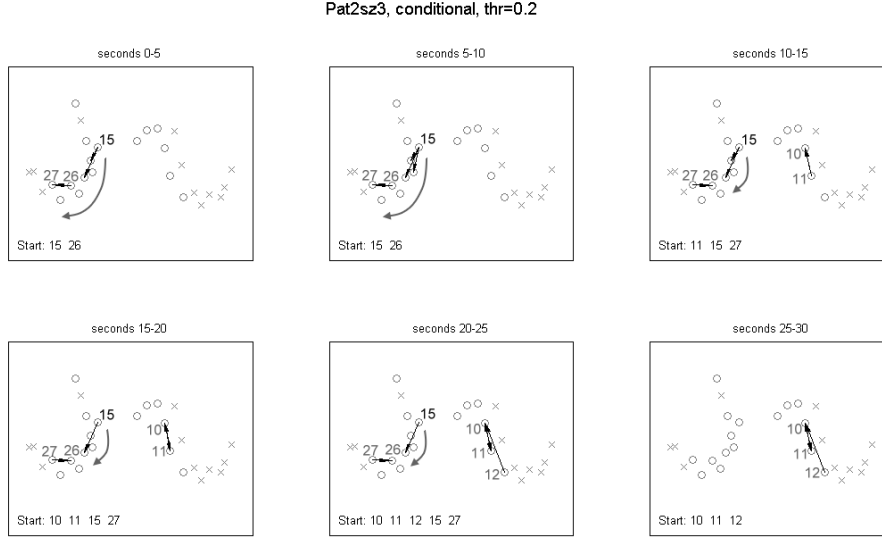


Figure 3.32: *Conditional analysis of seizure 3, patient 2: Although the clinical description (the corresponding channels are marked) would not agree with this, the other seizures of this patient affirm this assumption.*

Following the clinical description of the first 10 seconds, we should find epileptic activity at channels 26 and 27 only. For both channels causal relations are indicated. But as in the bivariate case this method indicates a flow starting at channel 15.

Although the medics could not find epileptic activity at this electrode, we cannot exclude that this method is right. Looking at the other seizures of this patient we find out that in some of them channel 15 is affected from the onset on.

Comparing the rest of the seizure with the medical description the conditional method is able to identify the propagation to channel 10 and 11 correctly. Further even the affection of channel 12 is found.

Contrary to the bivariate result (Section 3.5.1.2), we obtain less causal relations. The indirect causality between the channels (for example the arrow from 15 to 26 in the first five seconds-window) have vanished.

3.5.2.3 Patient 1, seizure 1

The next seizure to analyze and to compare with the bivariate result is seizure 1 of patient 1. Using the same methods as in the case before we obtain a threshold around 0.15 which seems to fit quite good for all seizures.

Next we have to decide, which set of channels to select. Therefore we are considering the bivariate results and are trying to use many channels having a bivariate causal relation for a multivariate analysis. We decided to work with the set $\{1, 2, 9, 10, 11, 17 - 28\}$.

Using this for the analysis of conditional Granger causality, we receive the result represented in figure 3.33.

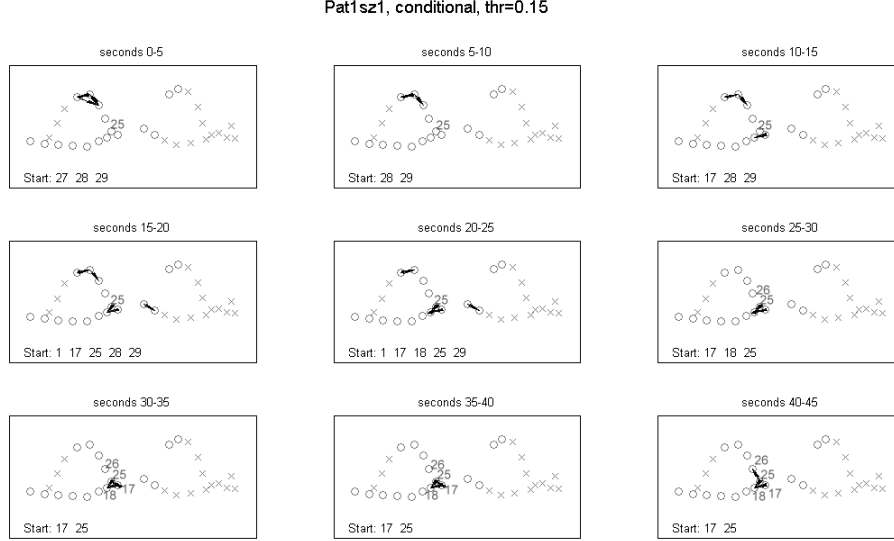


Figure 3.33: **Conditional analysis:** This method leads to some confusing results. A focus around channel 27, 28 or 29 is indicated. The correct focus at channels 25, 26, 17 and 18 is found about ten seconds after seizure onset.

The medics assume the seizure's initial focus to be located at channel 25. Unfortunately a causal relationship based on conditional Granger causality near this focus is found primary more than 10 seconds later.

One quite misleading effect appears in the first ten seconds: A strong causal relationship between channels 27, 28 and 29. If we would neglect the results of the other seizures and ignore the medical description, we would come to the decision that the focus has to be located there.

The reason why the real focus is identified so late, could be the fact that the medics assume only one channel to be affected in the first 25 seconds (until 03:01:14). No causal flow has to happen earlier. Now there arises the problem, why we are able to indicate a flow earlier than expected (after about 15 seconds)? It could be that the influence is not as obvious. Similar to patient 2 we are able to identify more causal relations compared to the opinion of the medics.

Comparing both methods, the bivariate and the conditional Granger causality, we recognize the same misleading effect of identifying a focus near channel 29. This wrong impression is stronger in the multivariate case. The identification after ten seconds of the focus can be observed in both cases. The only problem with the multivariate case is that the relations between channels 27, 28 and 29 are remaining and even if we neglect the first ten seconds at our analysis, we would get no unique interpretation of an initial focus.

3.5.2.4 Patient 1, seizure 3

As already mentioned in the introduction of this section, we like to keep both constant, the threshold and the set of channels for each patient.

Therefore the multivariate analysis for patient 1 uses a threshold with value 0.15 and the same set of channels as in seizure 1 ($\{1, 2, 9, 10, 11, 17 - 28\}$). The result of the conditional Granger causality is presented in figure 3.34.

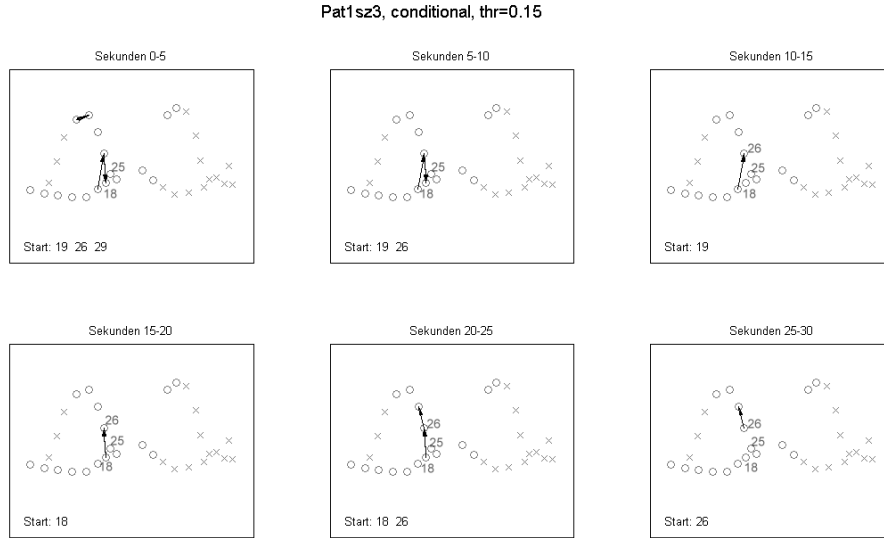


Figure 3.34: **Conditional analysis of seizure 3, patient 1:** Comparing the result with the clinical description (the corresponding channels are marked), we receive a pretty good match.

Considering these results, we identify causal relations near the seizure's initial focus (channels 18 and 25). Unfortunately, based only on the multivariate result of this seizure, we might also identify channel 19 to be the focus, because it is one of the arrows' starting points and no arrow points at it.

Examining the graph and comparing it with the mentioned clinical description we will find out, that more channels than in the description are affected. But although we obtain no perfect fit, an approximate location of the seizure's focus is localized correctly. Because the focus may lay between some electrodes, the graphs don't have to be completely wrong.

Considering the bivariate analysis (figure 3.30) we assess that the relations pointing near the focus (for example the one from channel 18 to channel 26) disappear.

Comparing this seizure with the analysis of seizure 1 of this patient, we would have an affirmation that the focus is located near channels 18 and 25 and not around channels 27 to 29 as we may have assumed after the analysis of seizure 1.

3.5.3 Partial Granger causality

In this section, we like to present all seizures of all patients to demonstrate, that this method works quite good and in fact is able to localize the initial focus. Before we start with the results using Partial Granger causality, we give a short summary.

Similar to the conditional Granger causality, x_j is partially Granger causing x_i , if and only if the estimation of x_i is significantly improved when adding the information coming from x_j (conditioned to a set of channels (\mathbf{x}_R) and common latent errors). Again we only use a linear concept to model the variables:

$$\begin{aligned}
 (i) \quad x_i[n] &= \sum_{s=1}^{\infty} \mathbf{A}_{ii}[s] x_i[n-s] + \sum_{s=1}^{\infty} \mathbf{A}_{iR}[s] \mathbf{x}_R[n-s] + \epsilon_{1,i}[n] \\
 \mathbf{x}_R[n] &= \sum_{s=1}^{\infty} \mathbf{A}_{Ri}[s] x_i[n-s] + \sum_{s=1}^{\infty} \mathbf{A}_{RR}[s] \mathbf{x}_R[n-s] + \epsilon_{1,R}[n] \\
 (ii) \quad x_i[n] &= \sum_{s=1}^{\infty} \mathbf{B}_{ii}[s] x_i[n-s] + \sum_{s=1}^{\infty} \mathbf{B}_{ij}[s] x_j[n-s] + \sum_{s=1}^{\infty} \mathbf{B}_{iR}[s] \mathbf{x}_R[n-s] + \epsilon_{2,i}[n] \\
 x_j[n] &= \sum_{s=1}^{\infty} \mathbf{B}_{ji}[s] x_i[n-s] + \sum_{s=1}^{\infty} \mathbf{B}_{jj}[s] x_j[n-s] + \sum_{s=1}^{\infty} \mathbf{B}_{jR}[s] \mathbf{x}_R[n-s] + \epsilon_{2,j}[n] \\
 \mathbf{x}_R[n] &= \sum_{s=1}^{\infty} \mathbf{B}_{Ri}[s] x_i[n-s] + \sum_{s=1}^{\infty} \mathbf{B}_{Rj}[s] x_j[n-s] + \sum_{s=1}^{\infty} \mathbf{B}_{RR}[s] \mathbf{x}_R[n-s] + \epsilon_{2,R}[n]
 \end{aligned}$$

Contrary to the first multivariate extension of the bivariate definition we want to filter out the effects coming from external influences (for example channels, which are not selected). For this we use the covariance matrices

$$\mathbf{S}[n] = \begin{pmatrix} \mathbb{V}(\epsilon_{1,i}[n]) & \text{Cov}(\epsilon_{1,i}[n], \epsilon_{1,R}[n]) \\ \text{Cov}(\epsilon_{1,R}[n], \epsilon_{1,i}[n]) & \mathbb{V}(\epsilon_{1,R}[n]) \end{pmatrix}.$$

The partial covariance, where the effects of common latent variables (common between $x_i[n]$ and $\mathbf{x}_R[n]$) are eliminated is given by

$$\mathbf{S}_{x_i|x_R} = \mathbf{S}_{ii} - \mathbf{S}_{iR} \mathbf{S}_{RR}^{-1} \mathbf{S}_{Ri}.$$

The corresponding covariance matrix of the system with channel x_j is defined as

$$\mathbf{\Sigma}[n] = \begin{pmatrix} \mathbb{V}(\epsilon_{2,i}[n]) & \text{Cov}(\epsilon_{2,i}[n], \epsilon_{2,j}[n]) & \text{Cov}(\epsilon_{2,i}[n], \epsilon_{2,R}[n]) \\ \text{Cov}(\epsilon_{2,j}[n], \epsilon_{2,i}[n]) & \mathbb{V}(\epsilon_{2,j}[n]) & \text{Cov}(\epsilon_{2,j}[n], \epsilon_{2,R}[n]) \\ \text{Cov}(\epsilon_{2,R}[n], \epsilon_{2,i}[n]) & \text{Cov}(\epsilon_{2,R}[n], \epsilon_{2,j}[n]) & \mathbb{V}(\epsilon_{2,R}[n]) \end{pmatrix}.$$

Eliminating the columns and rows, corresponding to $\mathbf{x}_J[n]$, we obtain the only information needed to analyze if $\mathbf{x}_J[n]$ Granger causes $\mathbf{x}_I[n]$.

$$\mathbf{\Sigma}_1[n] = \begin{pmatrix} \mathbf{\Sigma}_{ii} & \mathbf{\Sigma}_{iR} \\ \mathbf{\Sigma}_{Ri} & \mathbf{\Sigma}_{RR} \end{pmatrix} = \begin{pmatrix} \mathbb{V}(\epsilon_{2,i}[n]) & \text{Cov}(\epsilon_{2,i}[n], \epsilon_{2,R}[n]) \\ \text{Cov}(\epsilon_{2,R}[n], \epsilon_{2,i}[n]) & \mathbb{V}(\epsilon_{2,R}[n]) \end{pmatrix}$$

To eliminate all external and common effects, we use the partial covariance matrix

$$\Sigma_{x_i x_j | x_R} = \Sigma_{ii} - \Sigma_{iR} \Sigma_{RR}^{-1} \Sigma_{Ri}.$$

$x_j[n]$ partially Granger causes $x_i[n]$, if and only if the partial covariance matrix of the model including past values of $x_j[n]$ ($\Sigma_{x_i x_j | x_R}$) is smaller than that without using the information of past values from $x_j[n]$ ($\Sigma_{x_i | x_R}$).

The partial Granger causality index is given by

$$pGCI_{x_j \rightarrow x_i} = \ln \left(\frac{\Sigma_{x_i | x_R}}{\Sigma_{x_i x_j | x_R}} \right) = \ln \left(\frac{\Sigma_{ii} - \Sigma_{iR} \Sigma_{RR}^{-1} \Sigma_{Ri}}{\Sigma_{ii} - \Sigma_{iR} \Sigma_{RR}^{-1} \Sigma_{Ri}} \right).$$

To get a stable algorithm we have to reduce the set of channels. Therefore we choose the same set as used for the the conditional Granger causality analysis. This set is chosen manually by considering the bivariate results and the clinical description. In the figures presenting the results the selected channels are marked with a circle, whereas the others are marked with a cross.

3.5.3.1 Patient 2, seizure 1

As mentioned above, we are using the same set of channels as in Section 3.5.2 (which were channels 7 to 12, 15 to 18, channel 21 and channels 23 to 27).

After considering all seizures with different thresholds, a threshold near the value 0.15 seems to lead to results, which agree pretty good with the clinical description.

We now want to interpret the result presented in figure 3.35.

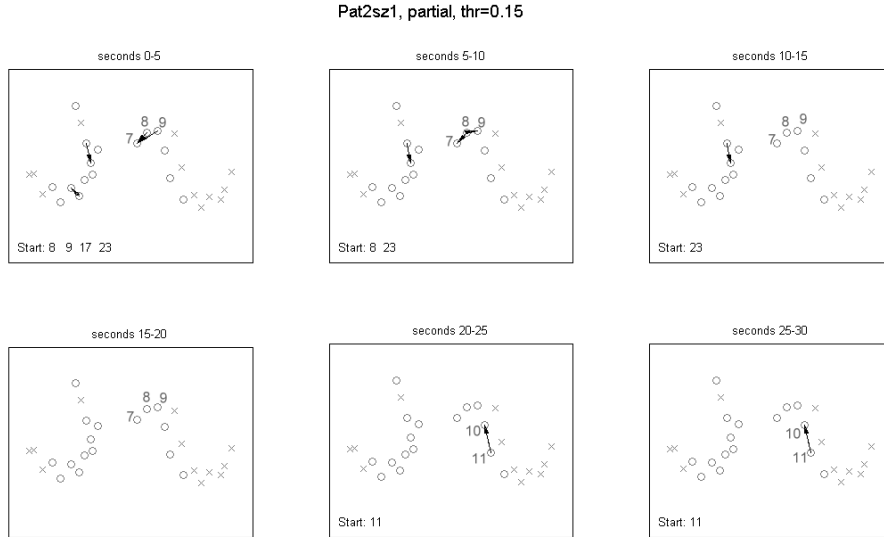


Figure 3.35: *Partial analysis of seizure 1, patient 2: Comparing the seizure with the clinical description (start at channels 7, 8, 9 and propagates to channels 10 and 11 after 21 seconds), this method is able to find all activities.*

Reminding the clinical description: The medics locate seizure's onset at channels 7, 8 and 9 at 03:04:36 (second 0 of the graphical representation).

Considering the result obtained by partial Granger causality analysis, we are finding causal relations between all three channels. Two arrows are pointing at channel 7, one from channel 8 and one from channel 9. We discover that this causal relation seems to become weaker after ten seconds and the arrows vanish.

Following the clinical description there is a propagation of the epileptic activity to channels 10 and 11 after 21 seconds.

Comparing this with the obtained result we recognize that channel 11 seems to cause channel 10 after the time window between second 20 and second 25. The propagation is found correctly.

The figure and the description of the medics seem to fit. An initial focus near channels 7, 8 and 9 may be assumed.

3.5.3.2 Patient 2, seizure 2

We continue using the set of channels $\{7 - 12, 15 - 18, 21, 23 - 27\}$ with threshold 0.15.

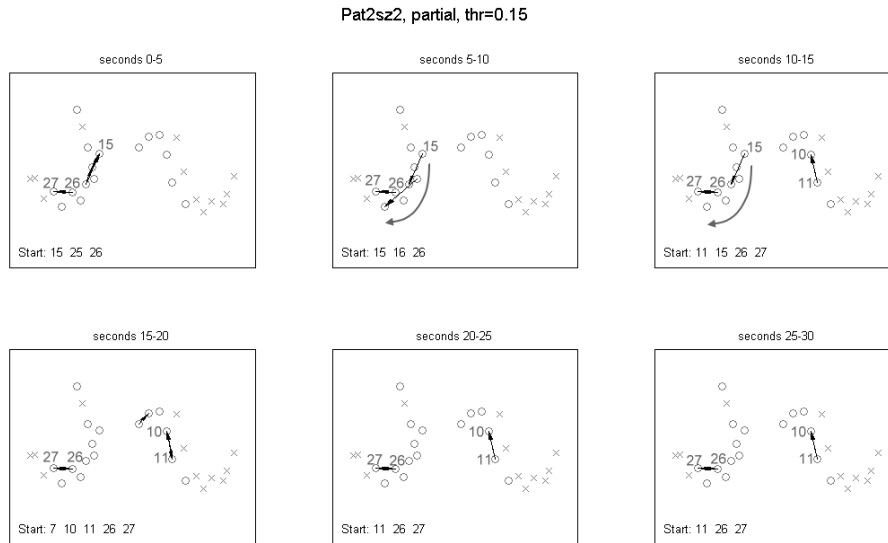


Figure 3.36: **Partial analysis of seizure 2, patient 2:** The result fits pretty well to the clinical description. The onset should be at channels 26 and 27, and the epileptic activity should simultaneously propagate to channels 15 to 25. After 13 seconds the activity proceeds to the left hand side at channels 10 and 11.

In the first five seconds strong causal relations between channels 15, 24 and 25 are identified, which are later reduced to 15 causing 25 (see figure 3.36).

Further a feedback situation between channels 26 and 27 is observed. These two channels seem to be the ones, which are stronger affected than all the others. Both of them are declined to be the focus.

Following the clinical description, after 13 seconds the seizure should proceed to the left side to channels 10 and 11. Comparing this with our result a causal relationship between channels 10 and 11 is found after the time window from second 10 to 15.

Further we find causal relations with channel 12 involved. If we consider seizure 3 of this patient, the clinical description tells us that a propagation to channel 12 in fact is possible.

If we search for a focus we would suppose to find it on the right hand side of the brain (which corresponds to left half of the plots). Because a flow from channel 15 to channels 26 and 27 seems to be identified, the focus might be assumed at channel 15.

Obviously channels 26 and 27 might also be focus channels as interactions between them are indicated in every time window.

3.5.3.3 Patient 2, seizure 3

Now we want to analyze the third seizure of patient 2. This one was already analyzed using bivariate and conditional Granger causality methods. Hence we are going to compare all three methods to find advantages and disadvantages of the methods.

Again we continue using the set of channels $\{7 - 12, 15 - 18, 21, 23 - 27\}$ and the threshold 0.15 for our analysis.

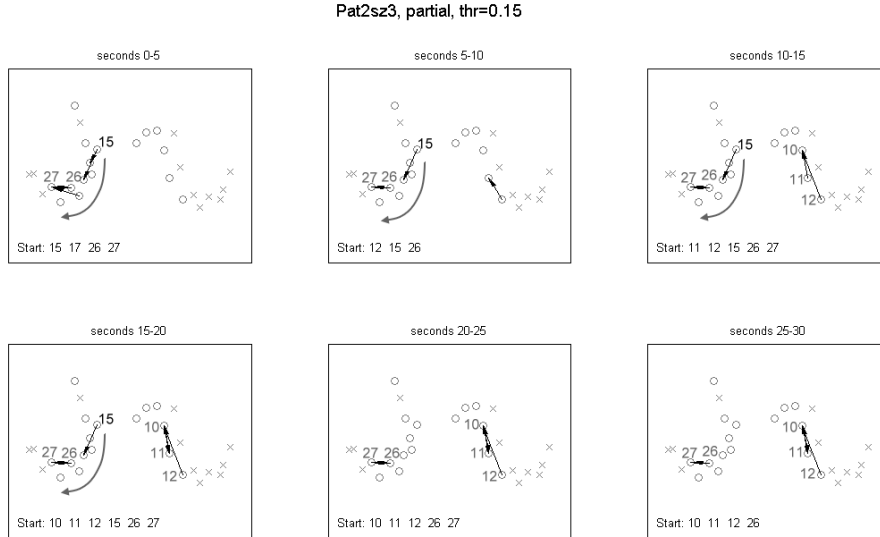


Figure 3.37: **Partial analysis of seizure 3, patient 2:** Considering this result we would assume the focus at channel 15, which is a discrepancy to the clinical description (start at channels 26, 27, propagation to channels 10 to 12 after eleven seconds).

Comparing the result (presented in figure 3.37) with the clinical description we discover that channels 26 and 27 are marked from the very beginning. Ac-

cording to the medics, both channels should in fact be focus channels.

Otherwise channel 15 seems to be the initial focus. Like for the other patients a flow of information might start there and proceed to channels 26 and 27. But because the medics do not find any epileptic activity there we have to be careful. Maybe the causality is not caused by the seizure.

Following the further clinical description, channels 10 to 12 should be affected after second 11. Again the corresponding arrows are found in the plot after the first seconds.

The task of tracking a focus channel cannot be solved easily. On the one hand channels 26 and 27 have very strong causal relations (even a feedback situation is obtained). Otherwise the flow starting at channel 15 seems to be an indicator of a focus.

Comparing this result with that from the bivariate (Section 3.5.1.2) and the conditional approach (Section 3.5.2.2) we notice only small differences. In all three results we would prefer channel 15 to be the focus and find the propagation to the other channels.

Reminding the GPDC result in Section 2.5.3 it seems to be nearly identical. Similar to the Granger causality result an activity is identified at channel 15. Both methods identify the causal relations at channels 26, 27, 10, 11 and 12 correctly.

3.5.3.4 Patient 2, seizure 4

Now we want to analyze the last seizure of patient 2. As usually we use the set of channels $\{7 - 12, 15 - 18, 21, 23 - 27\}$ and the threshold 0.15. The result is presented in figure 3.38.

Following the clinical description of seizure 4, we expect to see epileptic activity starting at channel 15. Comparing this seizure with the other ones of this patient, there is no such clear flow starting at channel 15 in the first seconds. But although the result is not as clear as before, we see that channel 15 is always the cause and never the effect. Considering the plots obtained between second 15 and 25, a flow is marked.

Similar to the seizures 2 and 3 a strong causal relation between channels 26 and 27 is obtained. Again it seems to be possible that focus is situated there.

The propagation to channels 24 up to channel 27 after about five seconds is identified clearly. Further even the epileptic activity at channels 10 and 11 appears correctly after 20 seconds.

Conclusion of Patient 2 Comparing the results with the clinical description we seem to have a satisfying fit. On the other hand an exact location of the seizure's initial focus could not be located. Either channel 15 or channel 26 could be the focus. Because at channel 15 the propagation to the other channels

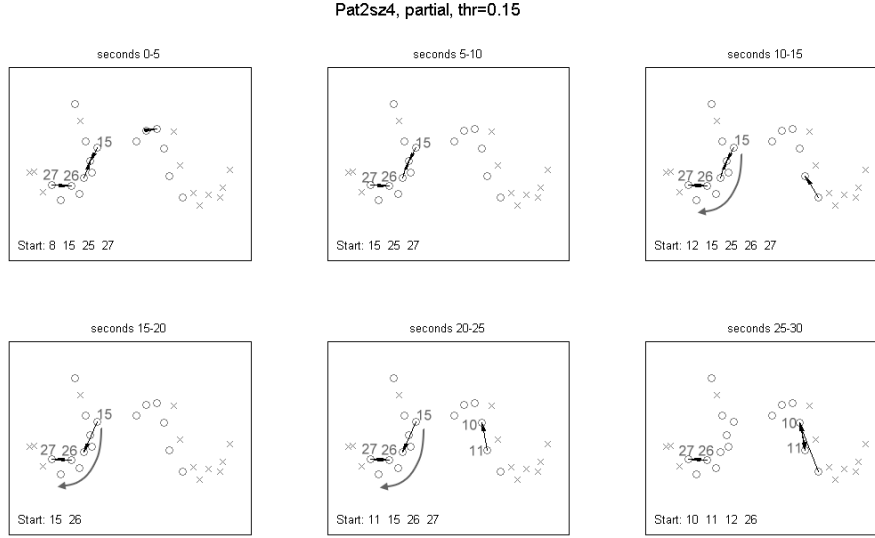


Figure 3.38: **Partial analysis of seizure 4, patient 2:** Comparing the result with the clinical description (start at channel 15, propagation to channels 26 and 27 after about five seconds, further affecting channels 9, 10 and 11 after the first 20 seconds) we obtain a pretty good fit.

is identified, it seems to be more likely to be the focus. Contrary the medics have the preference that the epileptic center is near channels 26 and 27.

Obviously results from partial Granger causality are nearly identical to those from GPDC in Section 2.5.3. Only small unimportant differences are found. Both methods lead to either a focus around channels 26 and 27 or at channel 15.

3.5.3.5 Patient 1, Seizure 1

The next three seizures we are going to analyze belong to patient 1, where a 32-dimensional ECoG-data is given. For stability reasons the data is reduced to a smaller set of channels (channels $\{1, 2, 9 - 11, 17 - 28\}$). This set is identical to that used for the conditional Granger causal analysis in Section 3.5.2. It is chosen manually with the help of the bivariate analysis and with the help of the clinical description.

After considering the results of all seizures with different thresholds, the threshold is chosen to be 0.14.

Analyzing the first seizure we obtain the result presented in figure 3.39.

Comparing the result with the clinical description small differences appear. Whereas in the doctors report only channel 25 is affected in the first 25 seconds, this method even identifies causal relations to channels 17 and 18. These two channels are directly next to the effected one and according to the medics they show epileptic activities some seconds later.

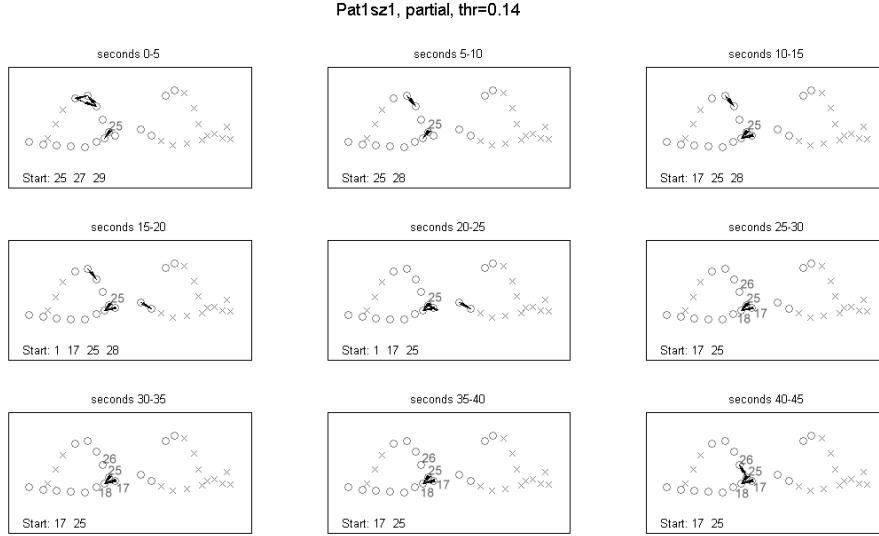


Figure 3.39: *Partial analysis of seizure 1, patient 1: Analyzing the result, obtained by partial Granger causality, an approximate location for the seizure's focus seems to be identified (start at channel 25, propagation to channels 17, 18 and 26 after 25 seconds).*

One problem of a causal analysis in the first 25 seconds is that only one channel is affected according to the medical description. Hence a causal analysis seems to be not reasonable until more channels are involved. There the result coincides with the clinical description. Both describe epileptic activity between channels 17, 18 and 25.

Considering the first plots, an additional causal relation between channels 27, 28 and 29 is found, although the doctors have not found any epileptic activity there. This probably could be explained by an action the patient made during that time (for example rolling the eyes). Reminding the conditional and the bivariate Granger causality analysis this effect was stronger there.

The graphs of these three channels (in figure 3.40) might explain this effect. Channels 27, 28 and 29 seem to have the same rhythm. No typical epileptic symptoms are found (for example no periodic oscillations), which confirms the assumption of a normal behavior.

Reminding the analysis of the bivariate (Section 3.5.1.3) and the conditional Granger causality (Section 3.5.2.3), we want to compare all three results.

Contrary to the bivariate result, in both multivariate methods causal relations are found between channels 27, 28 and 29. But this wrong identification could be explained with the help of figure 3.40.

The advantage of the partial Granger causality is, that it is the only method which is able to identify the focus channels (channels 17, 18 and 25) right after the onset.

Hence either the partial consideration or the bivariate definition lead to the

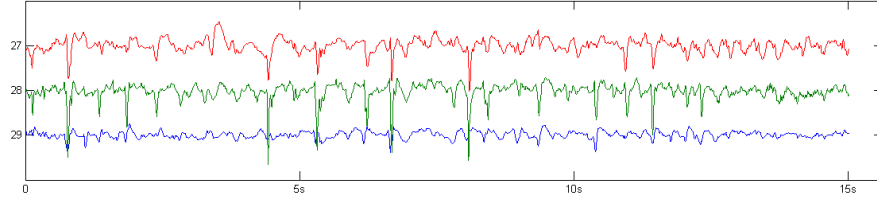


Figure 3.40: **Channels 27 to 29 of seizure1, Patient 1:** Because all three graphs seem to have a similar rhythm in the first seconds, a causal relationship between them is plausible.

best results in this seizure.

3.5.3.6 Patient 1, seizure2

Next we are going to analyze the second seizure of patient 1. Again the threshold is chosen to be 0.14 and the reduced set of channels is chosen to be $\{1, 2, 9 - 11, 17 - 28\}$.

Using this adjustment the result presented in figure 3.41 is obtained.

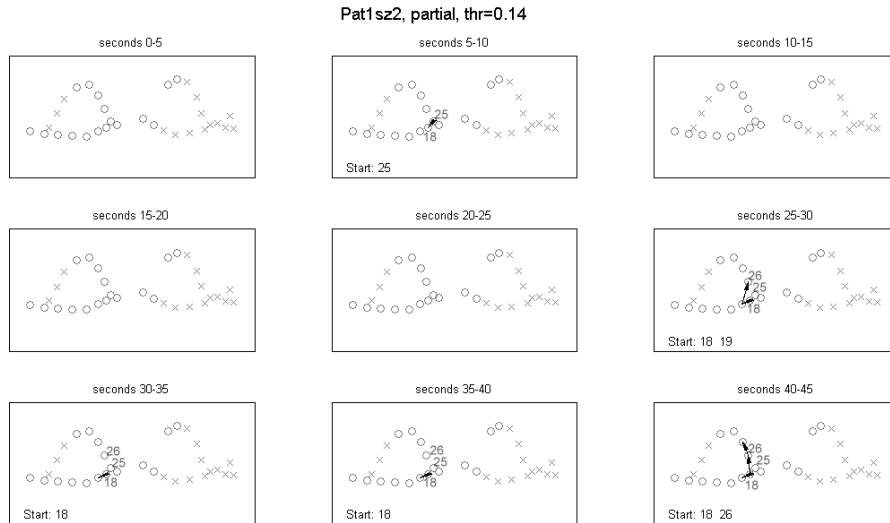


Figure 3.41: **Partial analysis of seizure 3, patient 1:** Comparing this result with the clinical description (start at channel 18 and 25, after a short pause of the seizure channels 18, 25 and 26 are affected) we find the correct channels causing each other.

According to the medics the focus should be at channels 25 and 18. After about seven seconds the epileptic activity propagates to channel 26.

Comparing this description with the result, a causal relation between 18 and 25 is found in the time window between second five and ten, but later this relationship vanishes, because the causality index is too weak to be noticed.

After the first 25 seconds the causal relations return and identify an epileptic activity at channels 18, 19 and 25. Now the causality index seems to stay significantly over the threshold.

Although the result of the partial Granger causality index seems to be delayed in time, we are able to identify an approximate location of the seizure's initial focus..

3.5.3.7 Patient 1, seizure 3

To complete the analysis of patient 1, we present the third seizure. As in both seizures before the set of channels is reduced to $\{1, 2, 9 - 11, 17 - 28\}$ and the threshold is equal to 0.14.

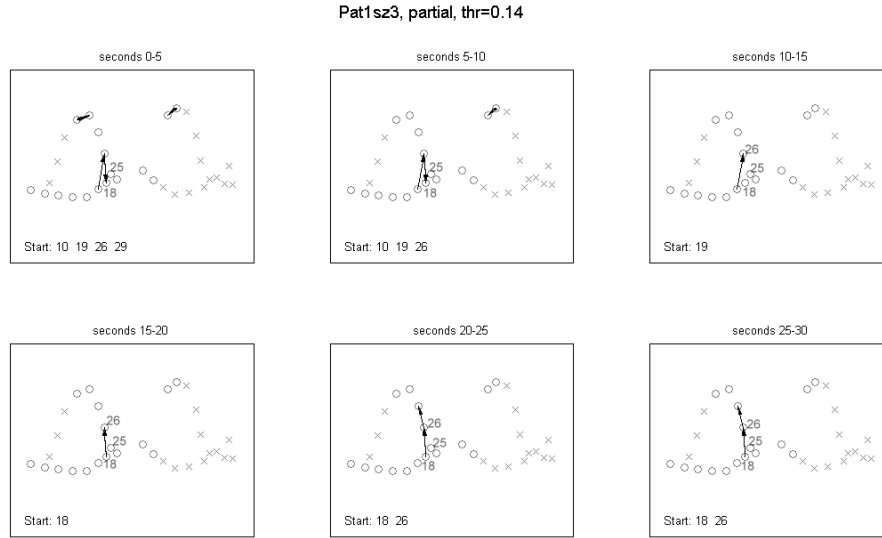


Figure 3.42: **Partial analysis of seizure 3, patient 1**: We seem to get an outcome, similar to the clinical description (start at channel 18 and 25, propagation to channel 26 after 11 seconds).

Following the clinical description, only channels 18 and 25 should be affected in the first eleven seconds. After these eleven seconds channel 26 is showing epileptic symptoms in the ECoG-data.

Comparing this statement with the result presented in figure 3.42 further channels (in the neighborhoods of the starting points) are in causal relations. Some of them become epileptic some seconds later, the others disappear quickly.

Some of the non-affected channels in a causal relation are the same, which were identified for the first seizure (channels 28 and 29). We determined that they seem to have the same rhythm, but no typical epileptic signals are identified. They probably result from a motion of the patient before the seizure started.

After the first ten seconds these causal relations vanish. The only remaining ones are those around the focus.

Reminding the results of the bivariate and conditional Granger causality index (Section 3.5.1.4 and 3.5.2.4) both multivariate methods lead to a nearly identical result. But compared with the bivariate result many of the disturbing causal relations (the ones pointing near of the focus) vanish.

Hence both multivariate methods seem to be superior to the bivariate consideration.

Comparing this result with the one of GPDC (see Section 2.5.3), we are able to identify the approximate location for the seizure's focus earlier. Here we have channels next to the focus involved right after seizure onset. But one similar situation arises: In both methods we were not able to identify the correct focus (channels 18 and 25) at the beginning.

Conclusion of patient 1 Considering all results of patient 1, we discover the seizure's start in the neighborhood of channels 18 and 25, which is in great accordance to the medical meaning.

Comparing the different methods, partial Granger causality seems to be the best for identifying epileptic activity.

3.5.3.8 Patient 4, seizure 1

To conclude the practical part of finding the focus of an epileptic seizure using partial Granger causality, we want to present the results of patient 4, which seems to expose some problems. After some seconds nearly all channels are affected in all three seizures. Thus the choice of the selected channels may be critical here.

In analogy to the other patients the set of channels used for the multivariate methods is chosen manually: Channels next to the starting points (according to the meaning of the medics) are selected as well as channels, which seemed to be quite often in a bivariate causal relation.

We decided to restrict ourselves to channels 3 and 4, channels 7 to 10, channel 12, channel 18 on the left side of the brain (in the right half of the plots), channels 21 to 24 and channels 27 to 30.

A threshold near 0.14 seems to represent the medical description quite good. The result presented in figure 3.43 is obtained for an analysis of the first seizure using these restrictions.

According to the doctors the start of this seizure is at channels 27 and 28. After one second the epileptic activity propagates to channels 21 and 22.

Contrary, Granger causality is not that clear. At the beginning causal relations starting at channels 23 and 29 are found. Although they are near the focus, they could be wrong (because of a discrepancy to the clinical description - these channels should never show epileptic activity).

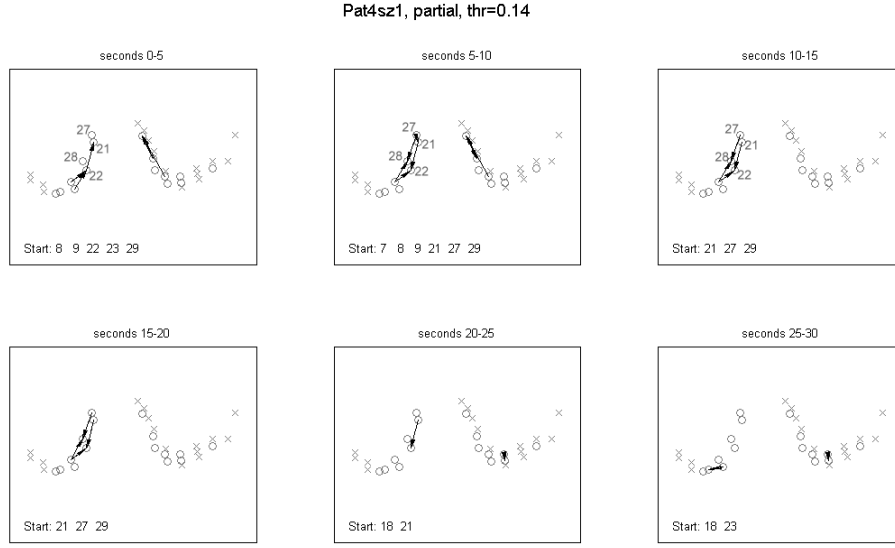


Figure 3.43: *Partial analysis of seizure 1, patient 1: Comparing the result, received by partial Granger causality with the clinical description (start at channels 27, 28, propagation to channels 21 and 22 after one second and after 14 seconds the left hemisphere shows epileptic activity (channels 1 – 20)), there is no perfect fit.*

After five seconds channel 21 causes its neighbor channel 22, and channel 27 causes channel 28. From that moment it is not clear if the focus is located near near channels 21 and 27 or between channels 23 and 29. According to the medical description a focus between channel 21 and 27 would be the correct one.

Furthermore the propagation to the other side of the hemisphere after 14 seconds is neglected. Hardly any causal relation is found. This is probably because not all channels showing epileptic activity are considered any more.

Comparing this result with the one obtained by GPDC in Section 2.5.3 (where we had a delay of the seizure's onset), we are now able to identify the location of epileptic activity from the very beginning.

3.5.3.9 Patient 4, seizure 2

As in both seizures before the set of channels is reduced to $\{3, 4, 7 - 10, 12, 18, 21 - 24, 27 - 30\}$ and the threshold is equal to 0.14.

Analyzing seizure 2 under these assumptions we obtain the result presented in figure 3.44.

Similar to seizure 1 of this patient, there are some discrepancies between the medical description and the obtained result. The doctors affirm the start of this seizure at channels 27 and 28. One second later channels 21 and 22 should show epileptic activity.

Using the assumptions above, causal relations between the truly affected chan-

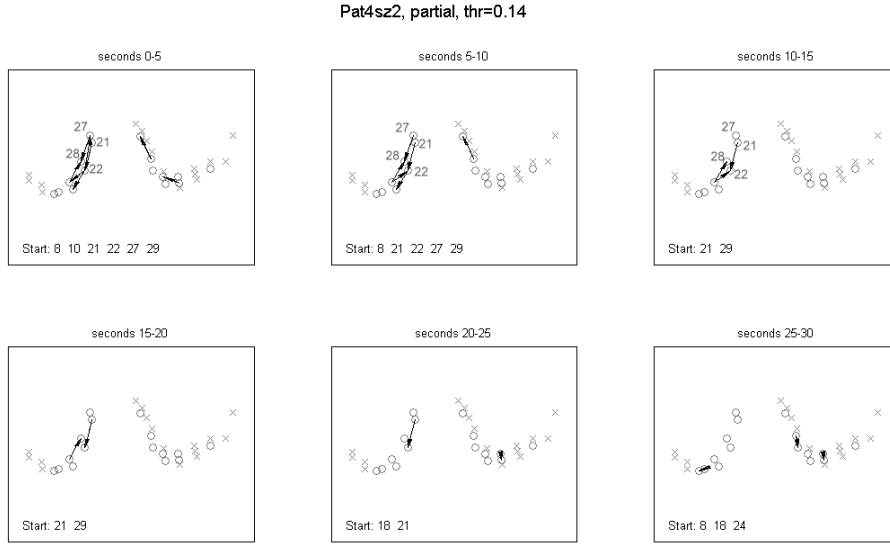


Figure 3.44: **Partial analysis of seizure 2, patient 4:** *Contrary to the meaning of the medics (start at channels 27, 28 and one second later at 21 and 22) it is not that clear where the seizure's focus is located. Channel 27 or channel 29 seem to be the favorites.*

nels are identified on the right hemisphere (starting at 21 and 27 and ending at 22 and 28). In analogy to seizure 1, we see causal relations with channels 23 and 29 involved. Considering the first seconds we could assume the start of this seizure at channel 27, which later infects channels 21, 22 directly and channels 22, 23 and 29 indirectly.

Again the results on the other side of the hemisphere do not fit perfectly with the clinical description. Misleading causal relations appear in the first seconds. While the medics assume the onset after 12 seconds on this hemisphere, we are not able to identify a mathematical affirmation. Some seconds later causal relations return on this side of the brain.

Based on this result the focus might be located in the area between channels 21,22,23 and 27,28,29. Unfortunately a more precise localization cannot be given.

3.5.3.10 Patient 4, seizure 3

The last seizure to analyze is seizure 3 of patient 4. This one seems to be more difficult than all other seizures before. According to the doctors all channels are affected from the very beginning.

Using the channel set $\{3, 4, 7 - 10, 12, 18, 21 - 24, 27 - 30\}$ and the threshold 0.14, we obtain the result presented in figure 3.45.

Contrary to the medical description the onset seems to be identified. In the first seconds the channels in causal relations are the same the medics found to

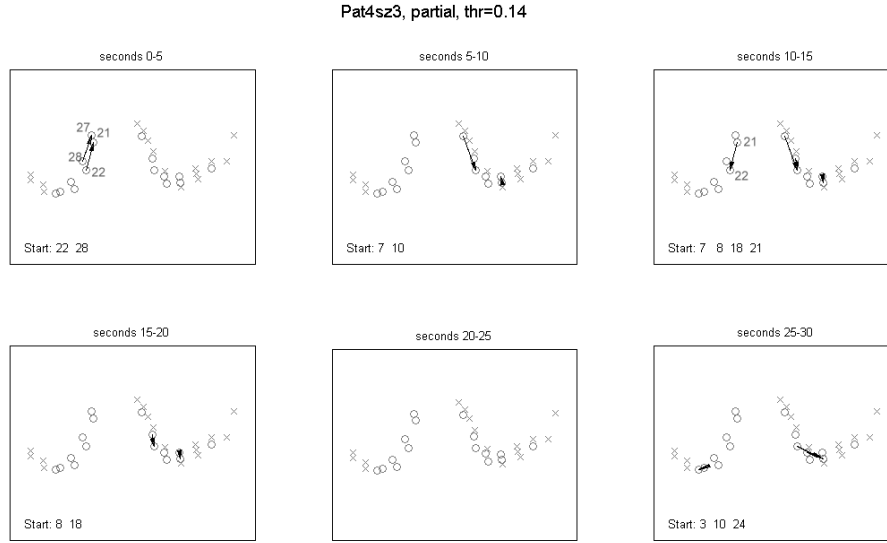


Figure 3.45: **Partial analysis of seizure 3, patient 4:** The medical description for this seizure is, that all channels are affected right after the onset. Comparing this with the result of Granger causality we identify that some of them have a stronger causal relation than others.

be the starting points in the first two seizures.

Finding the exact location of the focus seems difficult. The causality between channel 21 and 22 changes its direction. A start at channel 22 would be a contradiction to the other seizures. But we can assume a focus in the area between channels 21, 22, 27 and 28.

Conclusion of patient 4: Considering the results we do not find an exact location of the focus. Similar to the medics we are only able to identify the correct part of the brain. It can be assumed somewhere between channels 21, 22, 27 and 28.

Comparing the results with the GPDC results (in Section 2.5.3) we obtain very similar results. In both methods the location of the epileptic activity can be identified. Whereas using Granger causality methods no channel is referred being the focus, GPDC prefers channel 28.

3.5.3.11 Conclusion

All three methods lead to satisfying results. As expected, partial Granger causality was the most appropriate approach to find an approximate location of the seizure's focus.

Contrary to the bivariate case less disturbing relations (especially see seizure 3 of patient 1) are identified. On the one hand the multivariate extension reflects the medical description better, but on the other hand we have to restrict our-

selves to a smaller set of channels, contrary to the bivariate Granger causality. The choice of the channels might be critical, especially if the seizure generalizes quickly (like for Patient 4).

The results of the conditional and the partial Granger causality are quite similar. Sometimes (for example seizure 1 of patient 1) the partial extension leads to a better fit.

Chapter 4

Conclusion and Outlook

In the last chapter we want to give a short summary of topics discussed in this diploma thesis. We will mention satisfying results as well as unresolved problems and problematic assumptions. Furthermore, a list of potential improvements will be given.

4.1 Conclusion

In this diploma thesis, we compared a couple of mostly modern measures of multivariate time series analysis and their ability to localize the initial focus of epileptic seizures. All data came from patients suffering from temporal lobe epilepsy, a special case of focal epilepsy.

Our experiments were performed with ECoG recordings consisting of either 28 or 32 channels, whereas each channel corresponds to an electrode attached directly onto the brain's surface.

First of all (in Chapter 2), a frequency domain approach was implemented, based on the assumption of stationarity within windows of 3 or 4 seconds length. This assumption is in fact wrong, but has been justified by the results and by numerous publications.

The first measure used to track epileptic activity was the classical coherency measure, which is well known as a frequency domain representation of correlation between channels of multivariate time series. Unfortunately, coherence was not able to localize the initial focus as well as its partial extension was not. Although some of the results presented were not that bad, (partial) coherence indicated interactions between neighboring electrodes on the brain's surface in most cases, no matter whether there was epileptic activity or not.

Further research in frequency domain was based on parametric models, i.e. an AR-model was fitted for each time window. Measures could then be derived from the AR-model's estimated coefficients.

The best known of these measures is partial directed coherence (PDC), which

could be derived directly from the parametric representation of ordinary partial coherence. Unfortunately, PDC works if and only if all residual error variances of the time series's components are equal. We have shown that the focus channels of epileptic seizures tend to have much higher residual error variances. We further proved that, if arrows represent interactions indicated by high PDC values, arrows tend to point towards channels with higher residual error variances, thus, mostly (but not exclusively) towards focus channels.

Baccalá established a corrected version of PDC, which was called generalized PDC (GPDC). Its definition was based not only on the coefficients of the estimated AR-model, but on residual error variances as well. Therefore, as awaited after analysis of PDC results, GPDC was able to track epileptic activity quite good.

The second part of this diploma thesis dealt with a more modern approach: Granger causality (GC). For all Granger approaches a Recursive Least Squares algorithm was used to estimate the AR-model's coefficients (instead of ordinary least squares). Using that adaptive coefficient estimation we were able to deal with instationarities in our biosignals.

At first the so called bivariate Granger causality was introduced, where the calculation of interactions between two channels really had to be done separately for each pair of channels. Results were surprisingly good, although bivariate GC cannot separate between direct and indirect influences.

The next step of the analysis was an obvious multivariate extension of Granger's concept. Adjusted time series were calculated by partialization, i.e. by eliminating the influences from all other channels. Bivariate GC between these adjusted time series was then called conditional GC. Obviously, as more information is considered for each calculated GC, results were far better if conditional GC was used instead of bivariate GC.

Finally, another partialization was applied to withdraw influences from external variables. In practice this step was necessary, because not all available channels could be used for our calculations due to linear dependencies, which cause badly scaled covariance matrices. As each unselected channel is in fact an external variable for our calculations, results are altered by the influence of those channels. Consequently the presented measure, which was called partial GC, gave the best results for our aim of epileptic seizure tracking and focus localization.

Due to double partialization and RLS, partial GC even has given better results than the frequency domain approach. Figures have looked clearer and less disturbing arrows on uninvolved channels or brain sides have been drawn. In most cases the initial channel has been localized more precisely. On the other hand the RLS is extremely computational intensive. Thus the advantage of better results was based on much higher computational effort.

In general the GPDC approach as well as the partial GC have delivered results which were in great accordance with the findings of medical experts.

In some cases results from the mathematical analysis were even better than from medical analysis, i.e. we could find the correct initial focus for even more seizures than the doctors.

4.2 Outlook

As already discussed several simplifying assumptions have been made in this diploma thesis.

First, we always have used AR-models (i.e. stationary models) to explain our data, although we had to deal with instationary signals. RLS estimation is one simple concept to cope with those instationarities, but definitely not the best, as it is based to some extent on the assumption of stationarity again.

Due to numerical stability reasons we had to chose a set of selected channels for all our analysis (except for the bivariate approach). In this work this has been done manually, either due to information we have got from a quick look onto the graph of the time series, or by fine-tuning due to the results themselves. The chosen set of channels has been held constant over time. Obviously, a better approach would be an automatic channel selection algorithm, whereas the selected channel set might change as the seizure proceeds.

Further research might even deal with dependencies between chosen model orders, forgetting factors and channels sets. In fact, all of them have been held constant in our work for simplicity reasons, i.e. we consequently over- or underestimated the order of our ECoG signal.

To visualize results from any of our approaches, we had to find an appropriate threshold, whereas higher values of the calculated measure indicate interactions and lower values do not. Again, this threshold has been chosen manually and held constant over time. Further research should think about decisions based on automatically calculated confidence intervals and statistical tests.

Finally, we only dealt with data recorded during epileptic seizures. In practice, it will be necessary to find a mechanism to detect the seizure onset automatically. But of course, if the onset could be found by any other tool, our work will help to find the seizure's focus.

Besides all assumptions made, for the exact physiological signification of measures like partial GC and GPDC further research would be necessary. As they are directed measures, we displayed indicated interactions by flashes. Obviously, the direction of these flashes should be linked to physiological processes.

Appendix A

Clinical description

A.1 Patient 1

The electrodes of patient 1 are arranged and numbered as in figure A.1.

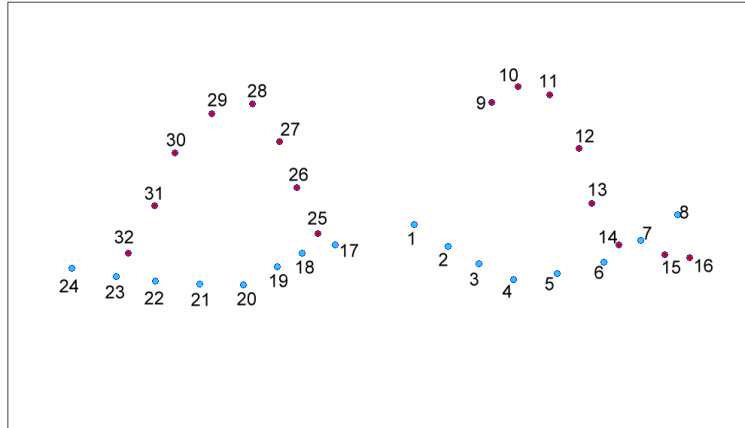


Figure A.1: *Numeration of the electrodes of the ECoG from patient 1: The corresponding channel numbers we use in the analysis are situated next to the electrode nodes.*

Clinical description of seizure 1, patient 1

Time	Activity	Electrodes
03:00:49	start	channel 25
03:01:14 (after 25 sec.)	susp	channel 25,26
03:01:19 (after 30 sec.)	prop	channel 17,18
03:01:46	end	

Table A.1: *Clinical description of seizure 1 from patient 1: By the abbreviation “susp” we mean a suspension of the seizure and “prop” stands for propagation.*

Clinical description of seizure 2, patient 1

Time	Activity	Electrodes
15:50:42	start	channels 18, 25
15:50:49 (after 7 sec.)	susp	
15:50:49 (after 30 sec.)	prop	channels 18, 25, 26

Table A.2: *Clinical description of seizure 2 from patient 1: The abbreviation “susp” stands for suspension of the seizure and “prop” stands for propagation.*

Clinical description of seizure 3, patient 1

Time	Activity	Electrodes
22:30:03	start	channels 18, 25
22:30:14		channels 18, 25, 26

Table A.3: *Clinical description of seizure 3 from patient 1.*

A.2 Patient 2

The arrangement of the electrodes and their numeration is presented in figure A.2.

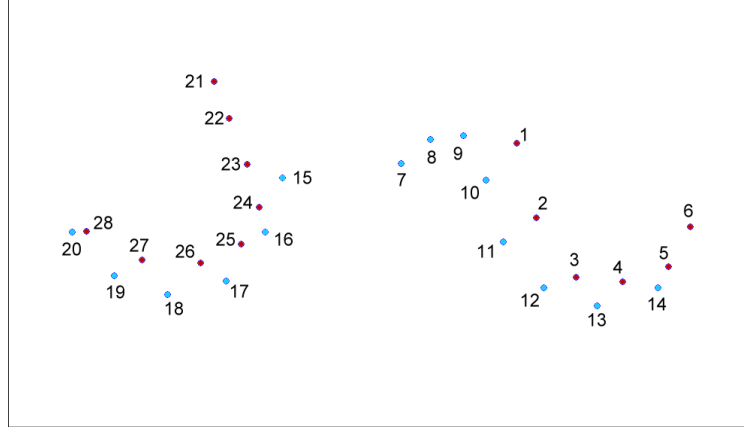


Figure A.2: *Numeration of the electrodes of the ECoG from patient 4: The corresponding channel numbers we use in the analysis are situated next to the electrode nodes.*

Clinical description of seizure 1, patient 2

Time	Activity	Electrodes
03:04:36	start	channels 7, 8, 9
03:04:57	prop	channels 10, 11
03:05:27	end	

Table A.4: *Clinical description of seizure 1 from patient 2 : The abbreviation “susp” stands for suspension and “prop” stands for propagation.*

Clinical description of seizure 2, patient 2

Time	Activity	Electrodes
12:45:51	start	channels 26, 27
	prop	channels 15-21, 24, 25
12:46:04	prop	channels 10, 11
12:47:00	end	channels 10, 11

Table A.5: *Clinical description of seizure 2 from patient 2 : The abbreviation “prop” means propagation.*

Clinical description of seizure 3, patient 2

Time	Activity	Electrodes
12:31:41	start	channel 26, 27
12:31:52 (after 11 sec.)	prop	channels 10, 11
	prop	channels 12
12:32:39	end	channel 10, 11, 12

Table A.6: *Clinical description of seizure 3, patient 2. The abbreviation “prop” stands for propagation.*

Clinical description of seizure 4, patient 2

Time	Activity	Electrodes
15:21:42	start	channel 15
15:21:47		channels 24-27
15:22:02	prop	channel 9, 10, 11
15:22:54	end	channels 10, 11

Table A.7: *Clinical description of seizure 4, patient 2. The abbreviation “prop” stands for propagation.*

A.3 Patient 4

The arrangement of the electrodes and their numeration is presented in figure A.3.

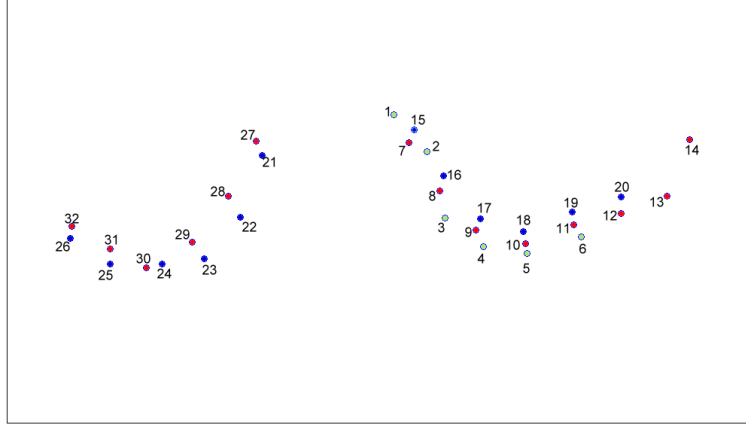


Figure A.3: *Numeration of the electrodes of the ECoG from patient 2: The corresponding channel numbers we use in the analysis are situated next to the electrode nodes.*

Clinical description of seizure 1, patient 4

Time	Activity	Electrodes
07:02:02	start	channels 27, 28
07:02:03	prop	channels 21, 22
07:02:16	prop	channels 1-20
07:03:00	end	

Table A.8: *Clinical description of seizure 1, patient 4. The abbreviation “prop” stands for propagation.*

Clinical description of seizure 2, patient 4

Time	Activity	Electrodes
08:06:34	start	channels 27, 28
08:06:35	prop	channels 21, 22
08:06:46	prop	channels 1-20
08:07:55	end	channels 1-20
08:08:30	end	

Table A.9: *Clinical description of seizure 2, patient 4. The abbreviation “prop” stands for propagation.*

Clinical description of seizure 3, patient 4

Time	Activity	Electrodes
08:55:41	start	
08:55:41	prop	all channels
08:56:58	end	channels 1-20
08:57:12	end	channels 21-32

Table A.10: *Clinical description of seizure 3, patient 4. The abbreviation “prop” stands for propagation.*

Bibliography

- [1] *Stedman's Medical Dictionary for the Health Professions & Nursing*,. Wolters Kluwer, 28 edition, 2005.
- [2] L. Baccalá. Generalized partial directed coherence. 2007.
- [3] L. Baccalá and K. Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics*, 84:463–474, 2001.
- [4] C. Baumgartner. *Handbuch der Epilepsien: Klinik, Diagnostik und psychosoziale Aspekte*. Springer, 2001.
- [5] P. Brockwell and R. Davis. *Time Series: Theory and Methods*. Springer-Verlag, 1987.
- [6] R. Dahlhaus. Graphical interaction models for multivariate time series. *Metrika*, 51:157–172, 2000.
- [7] M. Deistler and W. Scherrer. The prague lectures econometrics ii. CERGE Prague 1992, April 1994.
- [8] Eichler. Graphical modelling of dynamic relationships in multivariate time series. *M. Winterhalder, B. Schelter, J. Timmer (eds), Handbook of Time Series Analysis, Wiley-VCH, Berlin,, pages 335–372., 2006*.
- [9] M. Eichler. Granger-causality graphs for multivariate time series. *Beiträge zur Statistik*, 64, 2001.
- [10] P. Franaszczuk, K. Blinowska, and K. Kowalczyk. The application of parametric multichannel spectral estimates in the study of electrical brain activity. *Biological Cybernetics*, 51:239–247, 1985.
- [11] J. Geweke. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, 77:304–313, 1982.
- [12] A. Graef. Nonstationary autoregressive modeling for epileptic seizure propagation analysis. Master's thesis, Vienna University of Technology, 2008.
- [13] A. Graef, M. Hartmann, and M. Deistler. Dependency measures in multivariate neural signals. AIT Report, 2009.
- [14] A. Graef, M. Hartmann, M. Deistler, and T. Kluge. Regression-based analysis of synchronization in multichannel eeg in epilepsy. *IEEE EMBC*, 2009.

- [15] C.W.J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.
- [16] S. Guo, A. Seth, K. Kendrick, C. Zhou, and J. Feng. Partial granger causality - eliminationg exogenous inputs and latent variables. *Journal of Neuroscience Methods*, 172:79–93, 2008.
- [17] M. Hartmann, A. Graef, H. Perko, C. Baumgartner, and T. Kluge. Method for the characterization of synchronization and coupling in multichannel eeg and ecog. *WASET Proceedings Venice 2008*, 34:6–11, 2008.
- [18] S. Haykin. *Adapive filter theory*. Prentice Hall, 4 th edition, 2002.
- [19] S. Hermann. *Epilepsien: Diagnose und Behandlung*. Thieme, 1999.
- [20] M. Kaminski. Determination of transmission patterns in multichannel data. *Philosophical Transactions of the Royal Society*, 360:947–952, 2005.
- [21] R. Kus, M. Kaminsky, and K. Blinowska. Determination of eeg activity propagation: Pair-wise versus multichannel estimate. *IEEE Transactions on Biomedical Engineering*, 51:1501–1510, 2004.
- [22] H. Lütkepohl. Comparison of criteria for estimating the order of a vector autoregressive process. *Journal of Time Series Analysis*, 8:35–52, 1985.
- [23] H. Lütkepohl. *Introduction to Multiple Time Series Analysis, 2nd edition*. Springer, 1993.
- [24] E. Möller, B. Schack, M. Arnold, and H. Witte. Instantaneous multivariate eeg coherence analysis by means of adaptive high-dimensional autoregressive models. *Journal of Neuroscience Methods*, 105:143–158, 2001.
- [25] B. Schelter, J. Timmer, and M. Eichler. Assessing the strength of directed influences among neural signals using renormalized partial directed coherence. *Journal of Neuroscience Methods*, 179:121–130, 2009.
- [26] B. Schelter, M. Winterhalder, M. Eichler, M. Pfeifer, B. Hellwig, B. Guschlbauer, C. Lücking, R. Dahlhaus, and J. Timmer. Testing for directed influences among neural signals using partial directed coherence. *Journal of Neuroscience Methods*, 152:210–219, 2005.
- [27] T. Schneider and A. Neumaier. Arfit - a matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software*, 27:58–65, 2001.
- [28] T. Schneider and A. Neumaier. Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software*, 27:27–37, 2001.
- [29] P. Schönfeld. *Methoden der Ökonometrie*. Franz Vahlen, 1969.
- [30] A. Schuster. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestrial Magnetism and Atmospheric Electricity*, 3:13–41, 1898.

- [31] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.
- [32] M. Small, K. Judd, and A. Mees. Testing time series for nonlinearity. *Statistics and Computing*, 11:257–268, 2001.
- [33] P. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, AU-15:70–73, 1967.