# 'Neural data analysis: Learning From Other Disciplines' Workshop

# Programme

## 21st-23rd September 2010
## Newcastle University

# Table of Contents

**'Neural data analysis:**
**Learning From Other Disciplines'**
**Workshop**

21st-23rd September 2010

Newcastle University, UK

## Workshop Venue

Daysh Building, Room 1.29

Newcastle University

## Host

Stuart Baker

Professor of Movement Neuroscience and

Wellcome Trust Senior Fellow

## Sponsor

Engineering and Physical Sciences Research Council (EPSRC)

# Invited Speakers

Siddharth Arora – University of Oxford

Jim Austin - The University of York

Rajendra Bhansali - University of Liverpool

Pierfrancesco Cacciola - University of Brighton

Isabella Capellini - Durham University

Pablo Gleiser - Centro Atomico Bariloche

Ioannis Lestas - University of Cambridge

Gorka Zamora López - Humboldt University

John McWhirter - Cardiff University

Antonio Politi - Institute for Complex System

Jian Shi - Newcastle University

Rolf Turner - The University of Auckland

# Invited Facilitators

Sonja Gruen - Brain Science Institute

Leslie Smith - University of Stirling

Daniel Wójcik - Nencki Institute of Experimental Biology

# Workshop Programme

## Tuesday, 21st September

08:30 **Registration** in entrance lobby of the Daysh with Coffee & Tea

09:00 Introduction and objectives of the meeting - ***Stuart Baker***

10:00 In search for correlation in noisy and non-stationary multi-channel spike trains - ***Sonja Gruen***

11:00 A Miscellany of Statistical Modelling Techniques - ***Rolf Turner***

12:00 **Lunch at the Courtyard Restaurant**

13:00 Fundamental limits on the suppression of molecular fluctuations - ***Ioannis Lestas***

14:00 Phylogenetic comparative methods in evolutionary biology - ***Isabella Capellini***

15:00 **Coffee & Tea**

15:30 Use of artificial binary neural networks for pattern matching - ***Jim Austin***

16:30 Ordinary vs. Stable chaos - ***Antonio Politi***

## Wednesday, 22nd September

08:30 **Coffee & Tea**

09:00 Graph Analysis of Complex Networks: Theory and Application - ***Gorka Zamora López***

10:00 Community Structure in Complex Networks: Marvel Superheroes, Sleeping Brains… and all that Jazz - ***Pablo Gleiser***

11:00 Curve learning and clustering using Gaussian process prior - ***Jian Shi***

12:00 **Lunch at the Courtyard Restaurant**

13:00 Convolutive signal separation - the cocktail party problem - ***John McWhirter***

14:00 An Index of Linear Determinism for Spatial data - ***Rajendra Bhansali***

15:00 **Coffee & Tea**

15:30 On the modelling of seismic signals through the theory of stochastic processes - ***Pierfrancesco Cacciola***

16:30 Nonlinear and Nonparametric Modelling Approaches for Time Series Forecasting - ***Siddharth Arora***

18:00 **Assemble at the Northern Stage** and 18:30 walk to the restaurant

19:30 **Six Baltic Restaurant**

## Thursday, 23rd September

08:30 **Coffee & Tea**

09:00 Group Workshop

12:00 **Lunch at the Courtyard Restaurant**

13:00 Group Workshop

15:00 **Coffee & Tea**

15:30 Group Workshop

17:00 **Workshop close**

# Abstracts

## Introduction and objectives of the meeting

**Stuart Baker**, Newcastle University, UK

In this talk, I will provide a brief introduction to the nature of neuroscience data, and enumerate some of the problems which neuroscientists face in data analysis.

Single neurons sum many inputs, both excitatory and inhibitory, into their **membrane potential**. If this exceeds a certain threshold value, the cell emits an action potential or **spike**; membrane potential is then reset to a resting level. Spikes are transmitted to other cells, sometimes over long distances; only the occurrence time of a spike is important. Sometimes we are fortunate to record intracellular potentials from one cell, occasionally from cell pairs. More commonly, we record extracellular potentials. By filtering extracellular recordings with a high-frequency bandpass (typically 300Hz-10kHz), we can extract spike waveforms emitted by multiple cells in the immediate vicinity of the recording electrode. Using a lower frequency bandpass setting (typically 1-100Hz), we can extract local **field potentials (LFP)**– these represent membrane potential fluctuations summed over all cells near the recording electrode. In a typical experiment, we will record these measures of neural activity at the same time as details of the time and type of stimulation, or measures which quantify behavioural performance such as force, limb position or muscle activity. Direct measurement of spikes, local field potential and intracellular potential is usually only possible in animals. In humans, non-invasive methods include **electroencephalography (EEG)** and **magnetoencephalography (MEG)** (akin to local field potentials, but averaged over a larger spatial area). **Functional magnetic resonance imaging (fMRI)** measures brain blood flow, which increases in regions of high activity.

Over the last 20 years, methods for gathering data have become ever more sophisticated. We are now capable of recording spiking activity from hundreds of electrodes simultaneously; however, methods to interpret these data have lagged behind. We need methods which capture the rich structure of high dimensional data, are open to rigorous statistical testing, and will work with recording lengths which are small compared with the possible parameter space to be explored (sparse data). Our aim is to understand the spatiotemporal evolution of neural spike patterns which encode and process stimuli, leading to a behavioural response. Currently, methods are often dedicated to analysis of only one or two channels. They commonly assume rate stationarity, and that spikes are drawn from a

Poisson process; neither assumption is even close to reasonable for real data. We need wholly novel approaches to this problem, which do not simply build incrementally on what we have.

# In search for correlation in noisy and non-stationary multi-channel spike trains

**Sonja Gruen**, RIKEN Brain Science Institute, Wako-Shi, Japan

It has been proposed that cortical neurons organize dynamically into functional groups ('cell assemblies') by the concerted activity of the member neurons. Based on the efficacy of synchronized presynaptic spikes to reliably generate output spikes, the temporal coordination of spike timing is a commonly accepted signature of assembly activity. Consequently, approaches to detect assembly activity have focused on the detection of correlated spiking. While initially mainly technical problems limited the experimental surge for support of the assembly hypothesis (state-of-the-art electrophysiological setups in the last century allowed to record only few neurons simultaneously), the recent advent of multi-electrode array allowing to record tens to hundreds of neurons simultaneously reveals fundamental shortcomings of available analysis tools.

The reasons are manifold:
1) Data from behaving animals, required to show the relevance of neuronal correlations to behaviour, exhibit highly non-stationary (in time and across trials) and noisy spike sequences.
2) The brain in action is a dynamical system, thus neuronal interactions change as a function of time, requiring time dependent analysis approaches. Solutions to some of these aspects are meanwhile available, however, mostly for small numbers of simultaneously recorded neurons.
3) To detect the concerted activity of large neuronal groups requires the analysis of parallel spike sequences for higher-order correlations, i.e. correlations that cannot be explained/predicted by the firing rate and pair wise correlations alone. Mathematical frameworks for the estimation of such correlations run into combinatorial difficulties, and we require new creative routes for approaching such high-dimensional data.

# A Miscellany of Statistical Modelling Techniques

**Rolf Turner**, The University of Auckland, New Zealand

I shall describe three areas of statistical and probabilistic modelling that I have worked on in recent years. Most of my work has been focused on the area of spatial point processes. I shall give a rough overview of such processes and explain a technique for fitting models to Poisson point processes. This technique involves a numerical trick for approximating the log likelihood of the process by an expression which has the form of weighted log likelihood of a generalized linear model. This expression can then be optimized using standard software. I shall then briefly discuss some of the intricacies in modelling dependence ("interaction") structures for non-Poisson processes.

Another area to which I have contributed is that of hidden Markov models. These constitute a versatile means of analyzing serial dependence in data, and are often applied to time series of discrete data. The general idea is that the observations are influenced by a sequence of unobserved ("hidden") states of a Markov chain. I shall illustrate the ideas by looking at a model for counts of lesions in the brain and spinal chord of multiple sclerosis patients. In this example the hidden states are determined by the patient's being either in a "remission" or "relapse" condition.

Finally I shall discuss a probabilistic model that can be used to determine optimal prices for "perishable" assets (such as seats on a particular airline flight). The model assumes that groups of customers arrive according to a Poisson process, with a known probability distribution for the group size. It also assumes a known form of a time-varying elasticity of demand function. These assumptions permit the setting up of a coupled system of differential equations which can be solved numerically to determine the optimum price. In a simple example, the increase in expected revenue from the optimum pricing scheme, as compared with a plausible but suboptimal scheme, was over 40%.

# Fundamental limits on the suppression of molecular fluctuations

**Ioannis Lestas**, University of Cambridge, UK

The derivation of hard performance bounds in the presence of noisy feedback has been extensively addressed by control theorists over the last century, with many results going back to the early work of Wiener, Bode and Kalman. In a biochemical reaction network, however, where the noise is associated with the spontaneous births and deaths of individual molecules, many of the more conventional methodologies come to a halt, and need to be appropriately refined and extended. The approach in this talk is to address such complications by quantifying the way these spontaneous fluctuations restrict the ability to transmit information reliably, i.e. they lead to communication channels with finite capacity in the sense of Shannon, which can be explicitly calculated. We show that

such restrictions in information transmission fundamentally limit the ability to suppress fluctuations in molecular concentrations, thus leading to hard bounds for noise suppression that hold for arbitrary feedback policies. The limits derived illustrate the difficulty in maintaining biochemical accuracy, and show how an interface between control and information theory can be used to rigorously analyse poorly characterized biological systems.
Parts of the results that will be presented can be found in the article:
Nature 467(7312), 174-178, Sep 2010.

# Phylogenetic comparative methods in evolutionary biology

**Isabella Capellini**, Durham University, UK

We often wish to understand how and why species differ in morphology, physiology and behaviour. When examining macroevolutionary patterns of trait variation across species, however, evolutionary history has to be taken into account because closely related species tend to be more similar to one another than expected by chance; this leads to 'phylogenetic pseudoreplication' and violation of independence in statistical tests. Therefore evolutionary biologists have developed comparative methods that explicitly account for phylogeny in statistical tests. I will present the rationale of the most recent and powerful method, Phylogenetic Generalised Least Squares Models (PGLS). PGLS can be used to investigate the correlated evolution between continuous traits or between continuous and discrete traits, study the tempo and mode of evolution of continuous traits, and estimate the strength of their 'phylogenetic signal'.

# Use of artificial binary neural networks for pattern matching

**Jim Austin**, The University of York, UK

Binary neural networks are a class of artificial neural networks for pattern matching on large and complex data sets. We have developed methods to find patterns in signal, graph and text data. These have been applied to various problems including images of trademarks, engine data for Rolls-Royce, train data and power units etc. There use in the analysis of neural data has not been fully explored. The talk will overview the methods and possible applications.

# Ordinary vs. Stable chaos

**Antonio Politi**, CNR – Institute for Complex System (Istituto dei Sistemi Complessi), Italy

"Stable chaos" is a still little know phenomenon which means "irregular but linearly stable dynamics". This statement seems to be an oxymoron since ordinary chaos is associated with positive Lyapunov exponents (i.e., linear INstability). It is not an absurdity for a series of reasons that I can indeed explain.
Quite recently this kind of behaviour has been found also in some models of neural networks and I suspect it can play an important role in understanding how a brain works. Talking on this subject is also a good excuse to revisit some tools used to characterize chaotic systems.

# Graph Analysis of Complex Networks: Theory and Application

**Gorka Zamora López**, Humboldt University, Germany

Complexity emerges in real systems due to a variety of factors, but in general, it is tightly related to the manner in which the components of a system interact with each other. Very often it happens that the topology of this interactions are complex, meaning that the patterns of interactions between the components are neither regular, nor random. Graph analysis provides the tools to characterise and interpret the properties of these complex networks, and is flexible enough to be applied to a wide range of scientific disciplines. I will summarise the graph analysis methods that permit to extract useful information out of networked systems as-well-as the statistical issues that arise such that the measured properties can be classified as relevant or trivial. The overview will include several examples that show how real systems of very different nature may indeed share common topological properties, suggesting common rules of organization.

# Community Structure in Complex Networks: Marvel Superheroes, Sleeping Brains...and all that Jazz

**Pablo Gleiser**, Centro Atomico Bariloche, Argentina

A vast diversity of systems can be described and modelled as complex networks, formed by a set of nodes and their corresponding links. Some of these networks present groups of nodes that are more connected between themselves than with the rest of the network. In this talk I will give an introduction to the problem of finding and characterizing these groups of nodes, which are usually referred to as communities. In order to present some techniques and their application I will focus on three examples. First I will consider a simple example of a collaboration network, where the nodes of the network are jazz musicians, and they are connected if they have recorded together. Then, as an example of a weighted network, I will analyze an artificial social network, known as the Marvel Universe, which is formed by characters that appear in comic magazines published by Marvel Comics. In this case the weighted links between the characters are set according to the number of times that they appear together. Finally I will show an example of overlapping community structure in brain networks built using functional magnetic resonance imaging. I will show how these networks can be built and also discuss mechanism that allow for the emergence of networks with overlapping communities.

# Curve learning and clustering using Gaussian process prior

**Jian Shi**, Newcastle University, UK

In this talk, I will first introduce a Gaussian process functional regression (GPFR) model on modelling functional response curves with a set of functional covariates. The two main features of the model are: modelling nonlinear and nonparametric regression relationship and modelling covariance structure and mean structure simultaneously. The method gives very good results on curve fitting for functional data or spatial data. I will then discuss how to extend it to a mixture model to address the problem of heterogeneity. A new method will be presented for modelling functional data with 'spatially' indexed data, for example, the heterogeneity is dependent on factors such as region and individual patient's information. For data collected from different sources, we assume that the data corresponding to each curve (or batch) follows a Gaussian process functional regression model as a lower-level model, and introduce an allocation model for the latent indicator variables as a higher-level model. This higher-level model is dependent on the information related to each batch. This method takes advantage of both GPFR and mixture models and therefore improves the accuracy of predictions. The mixture model has also been used for curve clustering, but focusing on the problem of clustering functional relationships between response curve and covariates. I will also present some applications on simulated data and real data (but unfortunately we have not applied it to any neural data yet).

# Convolutive Signal Separation - The Cocktail Party Problem

**John McWhirter**, Cardiff University, UK

The term "cocktail party problem" is used to describe the difficult task of listening to an individual speaker in a noisy reverberant room in the presence of other conversations which constitute sources of interference. The human brain and auditory system perform this task remarkably well using a combination of signal processing and cognitive techniques which are not well understood and prove very difficult to replicate. One important feature is the binaural auditory system which helps to provide a useful degree of directional response. There is much active research in the context, for example, of video conferencing into the use of multiple microphone arrays with adaptive signal processing to emulate this type of directional response.

The problem of adaptively separating signals with different spatial location is particularly difficult in a highly reverberant environment. In such circumstances, the impulse response from an individual source to an individual sensor can involve significant broadening and multiple delays. In other words, a single spike emitted from the source would appear at the sensor as one or more broadened spikes arriving over a significant period of time. A more general signal such as a musical note or passage of speech would arrive at the sensor convolved with the corresponding impulse response which has a distorting effect and compromises intelligibility. In the case of multiple sources and sensors in a reverberant environment, the signal processing problem involves adaptively separating multiple signals which have not only been temporally distorted but also spatially mixed. This entire mixing process may be described mathematically in terms of a matrix of filters where each filter generates the corresponding point-to-point impulse response and takes the form of a polynomial in the delay operator z. The term polynomial matrix is used to describe such a structure.

The convolutive signal separation problem outlined above is the subject of much current research and may be approached in several ways. In this presentation I will briefly outline my own recent research which aims to tackle the problem by developing entirely new mathematical algorithms for extending conventional matrix factorisation techniques such as QR decomposition and eigenvalue decomposition to polynomial matrices. Results obtained by applying these techniques in the context of underwater acoustic arrays and MIMO (multiple input multiple output) wireless communications will be briefly presented.

# An Index of Linear Determinism for Spatial data

**Rajendra Bhansali**, University of Liverpool, UK

An Index of Linear Determinism for lattice data is introduced and its application for judging the adequacy of a parametric model fitted to spatial data is discussed. The work is joint with L. Ippoliti.

# On the modelling of seismic signals through the theory of stochastic processes

**Pierfrancesco Cacciola**, University of Brighton, UK

Ground motion arising from seismic waves is a phenomenon affected by several factors, i.e. source patterns, path, site effects, etc., that generally cannot be described in a deterministic fashion. Consequently, only a probabilistic approach can provide a rigorous representation of the phenomenon. Earthquake ground motion by its nature varies with time and in space, as a consequence the theory of stochastic processes is the natural vehicle for representation the seismic signals. Furthermore, since the ground motion on the free field can be seen as the result of the superposition of a number of symmetric waves propagating in the soil underneath it is commonly assumed (invoking the Limit Central Theorem) that zero-mean Gaussian processes reliably model the uncertainty in earthquake ground motion. Therefore, considering the ground motion acceleration in a given point on the earth surface as a realization of a zero-mean Gaussian random process this can be fully defined from a probabilistic point of view by the knowledge of the Power Spectral Density function (PSDf) or alternatively by the Correlation function (Cf). Power Spectral Density function and Correlation function are strictly related by the Wiener–Kintchine relationships. At this juncture it is worth mentioning that in literature exist a number of seismic models belonging to the theory of Gaussian random process. The wider classification is based on the stationary behaviour of the process. A stationary process possesses all the statistics that do not vary with respect to time. In this case the PSDf is function only of the frequency. On the other hand a zero-mean non-stationary Gaussian process is fully defined by the Evolutionary Power Spectral Density that in general is a non separable function of both frequency and time (in the case in which time and frequency functions are separable the process is called “quasi-stationary” or “uniformly modulated”). Commonly, in earthquake engineering practice, attention is focused on the time variability component and its impact on the structural response. On the other hand, it is well known that the spatial variability of earthquake ground motion can influence significantly the response of critical civil engineering structures, especially if they are long and/or rigid.

In this context, ground motion spatial variability is usually modelled via the (frequency domain) coherence function, which is the normalized cross-power spectral density of the motion at two stations.

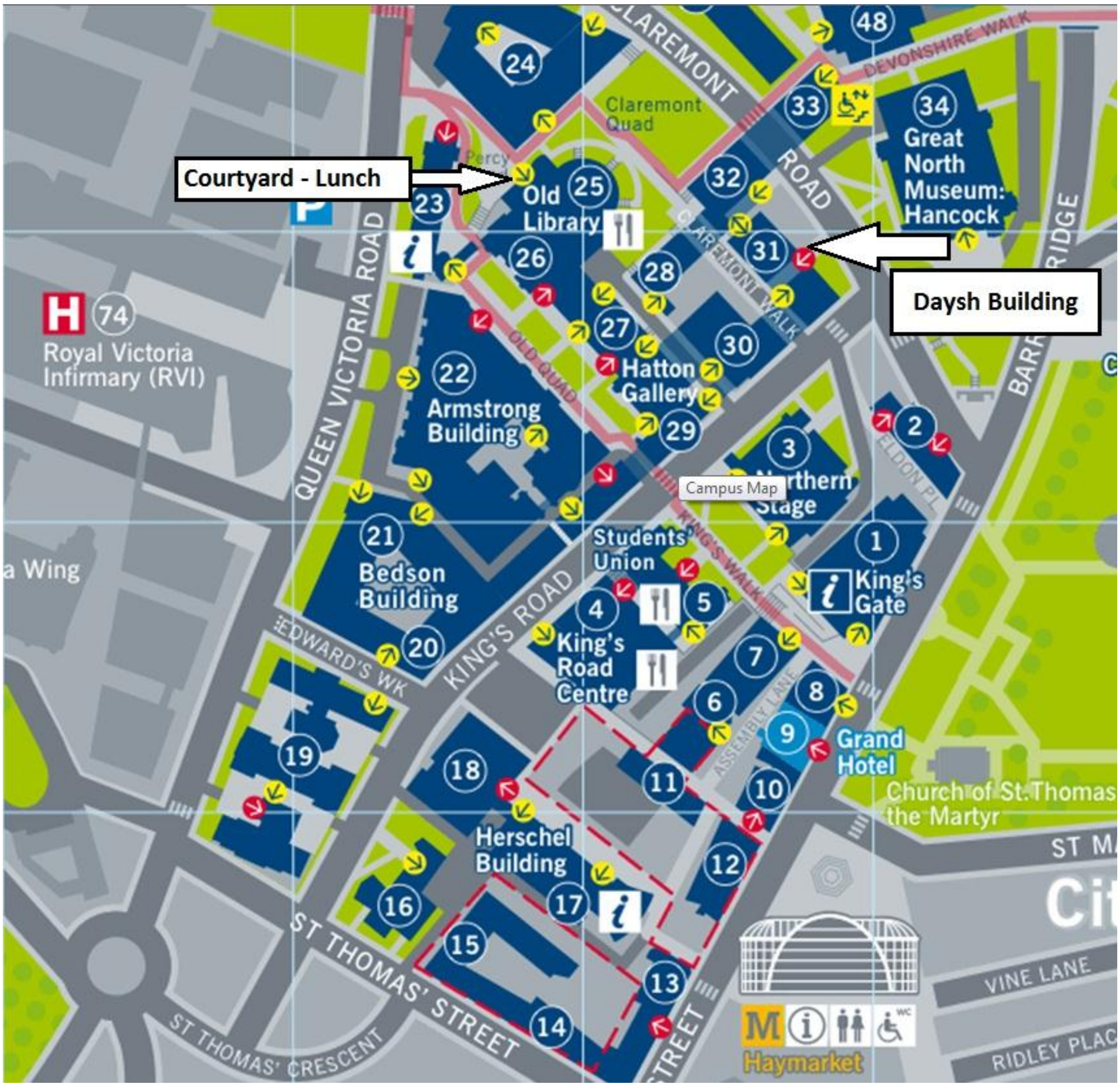
Aim of this talk is to present the main strategies based on the theory of stochastic processes used in earthquake engineering for modelling seismic signals. Preliminary concepts of stationary and non-stationary random processes will be initially addressed. Furthermore, m-variate non-stationary ground motion processes will be also discussed for defining the statistical correlation of ground motion signals recorded in various locations.

# Nonlinear and Nonparametric Modelling Approaches for Time Series Forecasting

**Siddharth Arora**, University of Oxford, UK

Accurate and timely forecasts are vital for efficient policy and decision-making, be it in the field of economics, biomedicine or climatology. Forecasts provide us the opportunity to mitigate potential risk associated with the system. Given that most time series arising from the real-world are known to exhibit nonlinear behaviour, we investigate nonlinear and nonparametric modelling approaches for forecasting. We employ nonparametric models because in principle, they have the flexibility to be able to estimate the true form of the function from the data itself, without needing to make potentially incorrect assumptions about the functional form in advance of seeing the data. The aim of our work is to generate accurate forecasts for time series exhibiting structural shifts. As it is important to quantify the uncertainty in forecasts, leading to well informed policy-making, we generate both point and density forecasts. We evaluate point forecasts using the root mean square error (RMSE) and mean absolute error (MAE), while density forecasts is evaluated using the continuous ranked probability score (CRPS).

# Detailed map of the campus



# Detailed map of the city

# SPIKE TRAIN

data analysis network