Workshop on Recent Advances in Changepoint Analysis

Programme

All activities will take place in Maths and Stats, Zeeman Building, with talks in room MS.04.

Monday 26 March

11:30 Registration opens

12:00 Buffet lunch

13:00 Welcome

13:15 David Stoffer (University of Pittsburgh)
Adaptive Spectral Estimation for Nonstationary Time Series
14:00 Rebecca Killick (University of Lancaster)
Detecting changes in second order structure within oceanographic time series

14:30 Break

15:15 Rainer von Sachs (Université Catholique de Louvain)
Shrinkage Estimation for Multivariate Hidden Markov Mixture Models
16:00 Birte Muhsal (Karlsruhe Institute of Technology)
Change-Point Methods for Multiple Structural Breaks and Regime Switching Models
16:30 Dafyd Jenkins (University of Warwick)
A multiple switch point model for analysing time series gene expression data
17:00 Session Ends

17:15 Drinks Reception

19:00 Conference Dinner

Tuesday 27 March

09:00 Sherzod Tashpulatov (CERGE-EI)
Estimating the Volatilty of Electricity Prices

09:30 Chigozie Utazi (University of Manchester)
Bayesian estimation of changepoints in a partially observed latent process Poisson model

10:00 Melissa Turcotte (Imperial College London)
Real time changepoint detection with applications in dynamic networks

10:30 Break

11:15 The Minh Luong (Université Paris Descartes)
Fast estimation of posterior probabilities in change-point models through a constrained hidden Markov model

11:45 Achim Zeileis (Universität Innsbruck)
strucchange: Model-Based Testing, Monitoring, and Dating of Structural Changes in R

12:30 Christopher Nam & Rebecca Killick (Warwick & Lancaster)
changepoint.info

12:45 Lunch

14:00 Paul Eilers (Erasmus University)
Sharp piece-wise linear smoothing and deconvolution with an L_0 penalty

14:30 Christopher Nam (University of Warwick)
Quantifying the uncertainty of activation periods in fMRI data via changepoint analysis

15:00 Guillem Rigaill (Université d'Evry Val d'Essone / URGV INRA)
Exact posterior distributions and model selection criteria for multiple change-point detection problems

15:30 Break

16:00 Paul Fearnhead (University of Lancaster)
PELT: Optimal detection of changepoints with a linear computational cost

16:45 Alice Cleynen (Inra AgroparisTech)
A Generic Implementation of a Fast Algorithm for Multiple Change-Point Detection and Application to NGS Data

17:15 Chun-Yip Yau (Chinese University of Hong Kong)
LASSO for structural break estimation in time series

Wednesday 28 March

09:30 Claudia Kirch (Karlsruhe Institute of Technology)
Evaluating stationarity via epidemic change-point alternatives for fMRI data

10:15 Alain Celisse (CNRS-Lille)
Kernel change-point detection

10:45 Break

11:30 Dean Bodenham (Imperial College London)
Adaptive Forgetting Factors and Average Run Length

12:00 Richard Davis (Columbia University)
Detection of Structural Breaks and Outliers in Time Series

12:45 Closing Remarks

13:00 Lunch

Abstracts

(in session order)

David Stoffer (University of Pittsburgh)
Adaptive Spectral Estimation for Nonstationary Time Series

We propose a method for analyzing possibly nonstationary time series by adaptively dividing the time series into an unknown but finite number of segments and estimating the corresponding local spectra by smoothing splines. The model is formulated in a Bayesian framework, and the estimation relies on reversible jump Markov chain Monte Carlo (RJMCMC) methods. For a given segmentation of the time series, the likelihood function is approximated via a product of local Whittle likelihoods. Thus, no parametric assumption is made about the process underlying the time series. The number and lengths of the segments are assumed unknown and may change from one MCMC iteration to another.
(joint work with Ori Rosen and Sally Wood)

Rebecca Killick (University of Lancaster)
Detecting changes in second order structure within oceanographic time series

We propose a test for changes in general autocovariance structure motivated by a problem arising from Ocean Engineering. The approach is founded on the locally stationary wavelet (LSW) process model for time series which has previously been used for classification and segmentation of time series. The proposed test can be performed when there is little information of the form of the autocovariance structure. The method will be illustrated with an application to oceanographic data from the North Sea where we seek to identify seasons when storms are prevalent.
(joint work with Idris Eckley and Philip Jonathan)

Rainer von Sachs (Université Catholique de Louvain)
Shrinkage Estimation for Multivariate Hidden Markov Mixture Models

Motivated from a changing market environment over time, we consider high-dimensional data such as financial returns, generated by a hidden Markov model which allows for switching between different regimes or states. To get more stable estimates of the covariance matrices of the different states, potentially driven by a number of observations which is small compared to the dimension, we apply shrinkage and combine it with an EM-type algorithm. The final algorithm turns out to reproduce better estimates also for the transition matrix. It results

into a more stable and reliable filter which allows for reconstructing the values of the hidden Markov chain. In this talk, in addition to some numerical examples, we will also present some theoretical results behind the use of shrinkage in this particular context.
(joint work with J. Franke, J. Tadjuidje and M. Fiecas).

Birte Muhsal (Karlsruhe Institute of Technology)
Change-Point Methods for Multiple Structural Breaks and Regime Switching Models

We introduce a method based on moving sums, which tests for at least one structural break at significance level α and simultaneously estimates the number and locations of change points. We investigate asymptotic properties, such as consistency of the change-point estimators, in case of changes in the mean in an otherwise independently and identically distributed sequence of random variables. Moreover we consider the generalisation of the results to corresponding Regime Switching Models, which in contrast to the classical change-point situation, allow for random change points and an unbounded number of structural breaks.

Dafyd Jenkins (University of Warwick)
A multiple switch point model for analysing time series gene expression data

Many biological systems implicitly contain states and discrete changes between them. Gene expression is one such system; the transcription activity of a gene is typically either binary, “on” or “off”, or a finite number of discrete states, with the current state changing through time. Gene expression can also be quantified using now standard high-throughput approaches, such as microarrays. Therefore, the time-series data derived from these quantitative approaches can be analysed in terms of a change-point model. However, the number and time of these change-points are not likely to be known, so these components of a model need to be approximated from the data. To do this, we have developed an algorithm utilising Markov chain Monte Carlo (MCMC) to sample model parameters for a piece-wise ODE describing the birth-death process of mRNA expression dynamics. However, as the number of change-points, or “transcriptional switches”, within the time-series is variable, we use an implementation of a reversible jump (RJ-MCMC) sampler, originally introduced by Green in 1995, to allow moves between models of different dimensionality, whilst maintaining balance. Output from the RJ-MCMC includes estimates of mRNA degradation rate, noise in the data, and time, variance and strength of transcriptional switches. The estimated time of transcriptional switches can then be used in a number of different ways, including gene-regulatory interaction identification, and clustering of multiple expression time-series.
(joint work with Bärbel Finkenstädt and David Rand)

Sherzod Tashpulatov (CERGE-EI)
Estimating the Volatilty of Electricity Prices

Price fluctuations that partially comove with demand are a specific feature inherent to liberalized electricity markets. The regulatory authority in Great Britain, however, believed that sometimes electricity prices were significantly higher than what was expected and, therefore, introduced price-cap regulation and divestment series. In this study, I analyze how the introduced institutional changes and regulatory reforms affected the dynamics of daily electricity prices in the England and Wales wholesale electricity market during 1990-2001. This research finds that the introduction of price-cap regulation did achieve the goal of lowering the price level at the cost of higher price volatility. Later, the first series of divestments is found to be successful at lowering price volatility, which however happens at the cost of a higher price level. Finally, this study also documents that the second series of divestments was more successful at lowering both the price level and volatility.

Chigozie Utazi (University of Manchester)
Bayesian estimation of changepoints in a partially observed latent process Poisson model

This work proposes a parameter-driven model for the estimation of changepoints in the analysis of time series data involving a discrete-valued manifest variable (or observed counts) and a real-valued latent variable. We focus attention on the case where the observed counts follow the Poisson distribution and the latent variable is only partially observed. Our interest lies majorly in making valid inferences about the partially observed latent process by garnering useful information from the observed counts and partially in estimating the association between the two variables. The methodology developed relies on the assumption that the observed counts are dependent upon the underlying latent process and therefore informative about it. The model, which is a variant of the parameter-driven model used in Zeger (1988), Chen and Ledolter (1995) and Davis et al (2000), specifies a dynamic model for the continuous latent variable, with the dynamics of the latent variable transferred to the observed counts through its conditional mean function. In contrast to the traditional use of latent variables in parameter-driven models to account for serial dependence in the observed counts, our work focuses on the reverse problem of using the observed counts to facilitate the estimation of the parameters of the partially observed latent variable. Several scenarios where a changepoint could occur in all the model parameters, the latent variable or some exogenous covariates are considered. We adopt a fully Bayesian approach for parameter estimation. Simulation studies illustrating the applicability of the model are also presented.

Melissa Turcotte (Imperial College London)
Real time changepoint detection with applications in dynamic networks

The talk discusses a Bayesian continuous time, multiple changepoint model used to detect anomalous behaviour in communications between actors in a dynamic network. Assuming the communications from an actor follow an imhomogeneous Poisson process, an actor is considered anomalous when there has been a recent changepoint, which represents a change in their communication intensity. We propose a novel Sequential Monte Carlo (SMC) algorithm to sample from the posterior distribution on the number and position of the changepoints in a sequential setting, where the communications in the network are being observed in real time. The performance of the algorithm is demonstrated on a synthetic network data set, but can be applied to a wider class of changepoint problems.

The Minh Luong (Université Paris Descartes)
Fast estimation of posterior probabilities in change-point models through a constrained hidden Markov model

The detection of change-points in heterogeneous sequences is a statistical challenge with applications across a wide variety of fields. In bioinformatics, a vast amount of methodology has been developed to identify an ideal set of change-points for detecting Copy Number Variation (CNV). While numerous efficient algorithms are currently available for finding the best segmentation of the data in CNV, few papers consider the important problem of assessing the uncertainty of the change-point location. Most approaches have quadratic complexity which are intractable for large datasets of tens of thousands points or more. We describe a constrained hidden Markov model for segmentation that corresponds to the classical segmentation model of heterogeneous segments, each consisting of contiguous observations with the same distribution. Forward-backward algorithms from this model estimate quantities of interest with linear complexity. The methods are implemented in the R package postCP, which uses the results of a given change-point detection algorithm to estimate the probability that each observation is a change-point. We present the results of the package on a sequence of Poisson generated data, and on a publicly available smaller data set (n=120) for CNV. Due to its frequentist framework, postCP obtains less conservative confidence intervals than those from previously published Bayesian methods. On another data set of high-resolution data (n=14,241), the implementation processed high-resolution data in less than one second on a mid-range laptop computer.

Achim Zeileis (Universität Innsbruck)

strucchange: Model-Based Testing, Monitoring, and Dating of Structural Changes in R

An overview of the capabilities of the R package "strucchange" for assessing parameter stability in statistical models is given, highlighting past developments, the present status, and some thoughts on future challenges.

Initially, the package focused on assessment of the coefficients of linear regression models and collected functions for (1) *testing* for structural change in a given sample, (2) sequentially *monitoring* structural changes in new incoming data, and (3) *dating* structural changes, i.e., estimating breakpoints in a given sample. Subsequently, a general object-oriented framework for *testing* for structural changes in a general class of parametric models was added, containing many of the linear regression model tests as special cases. Some building blocks for object-oriented monitoring and dating are also available in "strucchange" (and its sister package "fxregime"), but more general tools are still needed.

The functionality will be presented along with code and empirical examples.

Paul Eilers (Erasmus University)

Sharp piece-wise linear smoothing and deconvolution with an L_0 penalty

In many applications the goal of smoothing is to estimate piece-wise constant or piece-wise linear segments and their breakpoints. An interesting and useful solution is to modify the Whittaker smoother (also known as the Hodrick-Prescott filter). The objective function consists of two parts: 1) the sum of squares of differences between the data and a smooth series, and 2) a penalty on the roughness of the latter. Normally the differences are of order two and the penalty is expressed as the sum of their squares. No segmentation will occur this way.

However, the picture changes when we use sums of absolute values of differences. With first-order differences the segments will be piece-wise constant and with second-order differences piece-wise linear. The penalty can be implemented easily using iteratively re-weighted squares. Even better results are obtained when the iterative weighting scheme is modified to approximate an L_0 penalty, the number of non zero components. This algorithm is fast and works remarkably well, even though the objective function is non-convex. It also works well in a deconvolution setting, when we observe a filtered version of a segmented signal.

I will explain the statistical model and the estimation procedure and illustrate it with real-life applications, like copy-number estimation in genomics.

Christopher Nam (University of Warwick)
Quantifying the uncertainty of activation periods in fMRI data via changepoint analysis

Changepoint analysis has recently been used in fMRI data in determining activation periods of brain regions when the exact experimental design is unknown. This is particularly common in psychological experiments where different subjects react in different ways and different times to an equivalent stimulus. This talk will discuss a general changepoint method which can consider multiple changepoints (activation periods), and quantify the uncertainty regarding these periods. The methodology combines recent work on evaluation of exact changepoint distributions conditional on model parameters via finite Markov chain imbedding in a hidden Markov model setting, and accounting for parameter uncertainty and estimation via Bayesian modelling and sequential Monte Carlo. The methodology will be applied to fMRI data from a psychological experiment, where the methodology also allows for the inclusion of an autoregressive error process assumption and detrending within a unified analysis. In particular, we demonstrate how detrending and error process assumptions can affect changepoint conclusions.
(joint work with John Aston and Adam Johansen)

Guillem Rigaill (Université d'Evry Val d'Essone / URGV INRA)
Exact posterior distributions and model selection criteria for multiple change-point detection problems

In segmentation problems, inference on change-point position and model selection are two difficult issues due to the discrete nature of change-points. In a Bayesian context, we derive exact, explicit and tractable formulae for the posterior distribution of variables such as the number of change-points or their positions. We also demonstrate that several classical Bayesian model selection criteria can be computed exactly. All these results are based on an efficient strategy to explore the whole segmentation space, which is very large. We illustrate our methodology on both simulated data and a comparative genomic hybridization profile.
(joint work with E. Lebarbier and S. Robin)

Paul Fearnhead (University of Lancaster)
PELT: Optimal detection of changepoints with a linear computational cost

We consider the problem of detecting multiple changepoints in large data sets. Our focus is on applications where the number of changepoints will in- crease as we collect more data: for example in genetics as we analyse larger regions of the genome, or in finance as we observe time-series over longer pe- riods. We consider the common approach of detecting changepoints through minimising a cost function over possible numbers and locations of changepoints. This

includes most common procedures for detecting changing points, such as penalised likelihood and minimum description length. We introduce a new method for finding the minimum of such cost functions and hence the optimal number and location of changepoints, that has a computational cost which, under mild conditions, is linear in the number of observations. This compares favourably with existing methods for the same problem whose computational cost can be quadratic, or even cubic. The method, called PELT, is evaluated on real and simulated data. Code implementing PELT is available within the R package changepoint.
(joint work with Rebecca Killick and Idris Eckley)

Alice Cleynen (Inra AgroparisTech)
A Generic Implementation of a Fast Algorithm for Multiple Change-Point Detection and Application to NGS Data

Change-point detection problems arise in genome analysis for the discovery of CNVs or new transcripts. In both contexts, we believe that the signal is affected by (K - 1) abrupt changes caused by biological factors such as amplification of a region of the genome in tumoral cells, or expression of particular genes in certain growing conditions. The length of those signals vary from 10^6 to 10^9 so the usual methods such as the Dynamic Programming Algorithm (DPA, Bellman 1961), which recovers the optimal segmentations in 1 to K_{max} segments w.r.t. the quadratic loss with a complexity of $\Theta(K_{max}n^2)$ are not computationally efficient. Alternatives exist but have drawbacks: the CART algorithm (1984, Breiman, 2008, Lebarbier), is a heuristic procedure that approaches the best segmentation with a complexity bouded by O(n log n) $\leq$ C(n) $\leq$ O(n^2) but that does not recover the optimal solution. The PELT algorithm (Pruned Exact Linear Time, 2011,) was proposed for applications where the number of changepoints increases linearly with the number of datapoints n. In this context, it recovers the best segmentation with respect to criteria such as the standard AIC or BIC in a linear time. For CNV analysis the number of changepoints is usually considered fixed and in this context the standard AIC or BIC are not theoretically justified and they overestimate the number of breakpoints. (see e.g. Picard 2005 and Zhang and Siegmund 2007). We implemented generically in C++ the exact algorithm introduced by Rigaill in 2010 (PDPA). This algorithm is at worst in $\Theta(K_{max}n^2)$ but empirically faster $\Theta(K_{max}n \log n)$. We implemented the Quadratic, Poisson and Negative Binomial losses. The algorithm recovers the optimal segmentations in 1 to K_{max} segments and can incorporate penalty terms that are theoretically justified and that do not in practice overestimate the number of breakpoints (e.g. Zhang and Siegmund, Lavielle or Lebarbier). The C++ code is available as well as an R package. Our algorithm can be applied to a wide range of domains. The simulation section of the talk will show that our algorithm recovers the same break-points than the DPA but in a much faster time, and compare the performances of the PDPA and CART.
(joint work with Koskas, M., and Rigaill, G.)

Chun-Yip Yau (Chinese University of Hong Kong)
LASSO for structural break estimation in time series

We consider the structural break autoregressive process where a time series has an unknown number of break-points, and the time series follows a stationary AR model in between any two break-points. It is well-known that the estimation of the locations of the break-points involves huge computational challenges. By reformulating the problem in a regression variable selection context, we propose in this paper a group least absolute shrinkage and selection operatior (LASSO) procedure to estimate the number and the locations of the break-points, where the computation can be efficiently performed. We show that the number and the locations of the break-points can be consistently estimated from the data. Furthermore, the convergence rate of the breaks is shown to be nearly optimal. Simululation studies are conducted to assess the finite sample performance.

Claudia Kirch (Karlsruhe Institute of Technology)
Evaluating stationarity via epidemic change-point alternatives for fMRI data

Functional magnetic resonance imaging (fMRI) is now a well established technique for studying the brain. However, in many situations, such as when data are acquired in a resting state, it is difficult to know whether the data are truly stationary or if level shifts have occurred. To this end, change-point detection in sequences of functional data is examined where the functional observations are dependent and where the distributions of change-points from multiple subjects are required. Of particular interest is the case where the change-point is an epidemic change – a change occurs and then the observations return to baseline at a later time. The case where the covariance can be decomposed as a tensor product is considered with particular attention to the power analysis for detection. This is of interest in the application to fMRI, where the estimation of a full covariance structure for the three-dimensional image is not computationally feasible. Using the developed methods, a large study of resting state fMRI data is conducted to determine whether the subjects undertaking the resting scan have non-stationarities present in their time courses. It is found that a sizeable proportion of the subjects studied are not stationary. The change-point distribution for those subjects is empirically determined, as well as its theoretical properties examined.
(joint work with John Aston)

Alain Celisse (CNRS-Lille)
Kernel change-point detection

Nowadays statisticians deal with very different types of data from high dimensional settings (speech recognition, video segmentation,…) to non vectorial objects such as DNA sequences or protein structures for instance. The purpose of the present work is to provide a unified strategy allowing to handle such different objects to perform off-line change-point detection in their distribution.

Kernels are a convenient way to deal with high dimensional and especially non vectorial objects. Observations are mapped from the initial space (possibly non vectorial) into a Reproducing Kernel Hilbert Space (RKHS) in which statistical inference can be performed. We propose to detect abrupt changes along the time in the *mean element* of the distribution of these "kernelized observations", which result from changes in the distribution of the initial observations. In the homoscedastic and weakly heteroscedastic settings, we derive new oracle inequalities with optimal constants, which assess the performance of the final segmentation. These theoretical results rely on new concentration inequalities specific to the Hilbert space framework. We also present an efficient algorithm based on dynamic programming allowing to deal with large data sets. Finally, the performance of our approach is empirically assessed on synthetic and real data such as speech recognition, functional data,…

Dean Bodenham (Imperial College London)
Adaptive Forgetting Factors and Average Run Length

Detecting change in streaming data provides two significant challenges Ű first, the large amount of data requires our detection algorithms to be online, and second, the initial and final underlying distributions of the data are often unknown, as is the point where they change from one to the other. Standard techniques customarily involve certain tunable parameters ('control' parameters) that control the importance of false alarms over delayed detections, or entirely missed changes.

In this talk we introduce a framework utilizing exponential forgetting factors in an attempt to reduce the dependence on control parameters. Exponential forgetting factors have been used in the changepoint literature (EWMA) as a means of downweighting old observations and thereby placing more weight on recent observations, resulting in detectors capable of maintaining their sensitivity to small values.

Two useful performance measure are ARL0 (the average delay in detecting a false negative) and ARL1 (the average delay in detecting a true positive) defined by Page. While it is useful to have a high ARL0 value and a low ARL1 value, interestingly, changing the value of the fixed forgetting factor generally either raises both or lowers both. Ideally, though, one would want high sensitivity before, and low sensitivity after, a change. This is particularly true in the case of restarting CUSUMs.

This raises the question of considering adaptive forgetting factors (AFF), a technique which has proved useful in other fields, including streaming classification and adaptive filtering.

In this work, we present a novel method for updating the AFF, tailored to the problem of detecting a change in a mean, with a view towards investigating the extent to which AFFs can simultaneously improve both ARL0 and ARL1. We present the method in detail, as well as early results from simulation studies.

Richard Davis (Columbia University)
Detection of Structural Breaks and Outliers in Time Series

Often, time series data exhibit nonstationarity in which segments look stationary, but the whole ensemble is nonstationary. In this lecture, we consider the problem of modeling a class of non-stationary time series with outliers using piecewise autoregressive (AR) processes. The number and locations of the piecewise autoregressive segments, as well as the orders of the respective AR processes are assumed to be unknown and each piece may be contaminated with an unknown number of innovational and/or additive outliers. The minimum description length principle is applied to compare various segmented AR fits to the data. The goal is to find the "best" combination of the number of segments, the lengths of the segments, the orders of the piecewise AR processes, and the number and type of outliers. Such a "best" combination is implicitly defined as the optimizer of a MDL criterion. Since the optimization is carried over a large number of configurations of segments and positions of outliers, a genetic algorithm is used to find optimal or near optimal solutions. Strategies for accelerating the procedure will also be described. Numerical results from simulation experiments and real data analyses show that the procedure enjoys excellent empirical properties. The theory behind this procedure will also be discussed.
(joint work with Thomas Lee and Gabriel Rodriguez-Yam.)
