

53rd Gregynog Statistical Conference 21st–23rd April 2017



Contents

1 Administrative Details	4
1.1 About	4
1.2 Webpages	4
1.3 Key Dates & Times	4
1.4 Internet Access	4
1.5 Gregynog Young Musicians Conference	5
1.6 Powis Castle Excursion	5
2 Getting to Gregynog	6
2.1 Venue Details	6
2.2 Getting there by Minibus	6
2.3 Getting there by Car	6
2.4 Getting there by Train	7
2.5 Local Taxi Companies	7
3 About Gregynog	8
3.1 History	8
3.2 Walks	8
3.3 Library	9
4 Timetable	12
4.1 Friday 21st April	12
4.2 Saturday 22nd April	13
4.3 Sunday 23rd April	13
5 Abstracts	14
5.1 Short Course	14
5.2 Talk	15
5.3 Poster	17
6 Participant List	22



TREGYNON

B4389

B4389

B4389

B4389

BETTWS
CEDEWAIN

B4389



1 Administrative Details

1.1 About

- The “53rd Gregynog Statistical Conference” will take place from Friday, 21st – Sunday, 23rd April 2017 at Gregynog Hall, in the fantastic Welsh countryside. The conference will comprise short courses, a number of research talks and a poster session aimed at a general mathematical audience. There will be lots of opportunity (particularly for research students) to get to know one another, explore the numerous walks (or hikes!) around the hall, take an excursion to Powis Castle and Garden and even attend the final of the Gregynog Young Musicians Competition.
- Organising Committee: Jane Hutton (Warwick) and Murray Pollock (Warwick).

1.2 Webpages

- Conference: www.warwick.ac.uk/gregynog
- Past Conferences / History: www.warwick.ac.uk/gregynog/past
- Gregynog Young Musicians Conference: https://en.wikipedia.org/wiki/Gregynog_Young_Musicians_Competition

1.3 Key Dates & Times

- **Arrival & Check-In:** From 2pm, Friday (Gregynog Hall Shop).
- **Welcome:** 2.55pm, Friday (Seminar Room, 2nd Floor).
- **First Talk:** 3pm, Friday (Seminar Room, 2nd Floor).
- **Poster Session:** 4.30pm, Friday (Joicey Room).
- **Breakfast:** 8am (Saturday / Sunday – Dining Room).
- **Coffee Breaks / Afternoon Tea:** 11am / 4pm (Blayney Room).
- **Lunch:** 12.30pm (Saturday / Sunday – Dining Room).
- **Dinner:** 7pm on Friday; 6pm on Saturday (Dining Room).
- **Bar:** From 9pm (Friday / Saturday – Basement).
- **Departure:** 1.30pm, Sunday.

1.4 Internet Access

- Wireless access is available in the lecture rooms and public areas. *Note that from experience the wireless can be extremely poor.*

1.5 Gregynog Young Musicians Conference

- Gregynog Young Musicians Competition began in 2005, originally as part of the Gregynog Festival, and more recently as a free-standing event. It is held at Gregynog Hall and is open to instrumentalists under the age of 18.
- The competition will take place on Saturday evening, and those interested in attending (time has been made in the schedule) ought to refer to the following website to purchase a ticket:
- Webpage: <http://www.gregynogymc.co.uk>

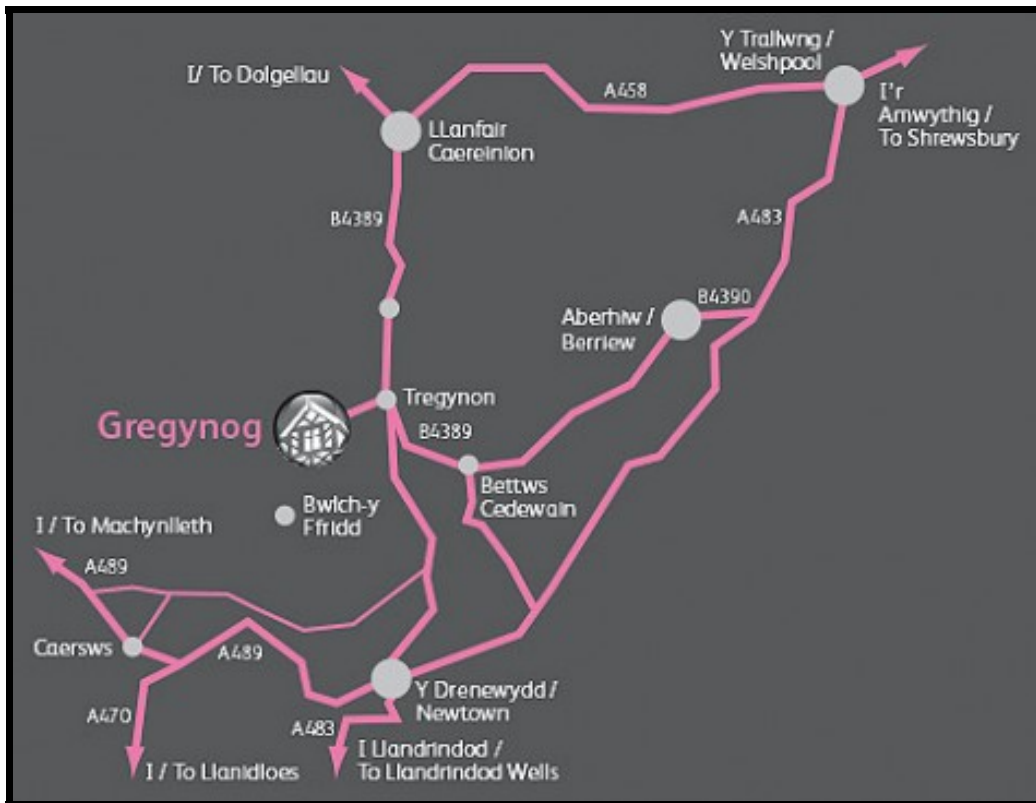
1.6 Powis Castle Excursion

- An optional excursion to Powis Castle will take place on the Saturday after lunch. Please speak to Jane Hutton for further details and to register your interest.
- Powis Castle (Welsh: Castell Powis) is a medieval castle, fortress and grand country mansion located near the town of Welshpool, in Powys, Mid Wales. The residence of the Earl of Powis, the castle is known for its extensive, attractive formal gardens, terraces, parkland, deerpark and landscaped estate. The property is under the care of the National Trust, who operate it under the name "Powis Castle and Garden". Princess Victoria (later Queen Victoria) visited the castle as a child when her mother took her to tour England and Wales in 1832.
- Webpage: www.nationaltrust.org.uk/powis-castle

2 Getting to Gregynog

2.1 Venue Details

- **Address:** Gregynog Hall, Tregynon, Nr. Newtown, Powys, SY16 3PW
- **Telephone:** 01686 650224
- **Webpage:** www.gregynog.org
- **Contact / Travel Information:** www.gregynog.org/contact/



2.2 Getting there by Minibus

- A minibus will leave from Warwick Statistics common room at 11.45am *sharp* on the Friday going to Gregynog Hall, and leave from the Gregynog Hall dining room at 1.30pm *sharp* on the Sunday returning to Warwick. Individuals must request a place on the minibus. There will be limited space on the minibus, so please do not overpack.

2.3 Getting there by Car

- Gregynogs location near the quiet village of Tregynon, 6 miles north of Newtown in Powys, makes it reachable within 3 hours from all parts of Wales, within 2 hours from Birmingham, Manchester, Chester and Liverpool and just 50 minutes from Shrewsbury.
- **From Newtown**

- Entering Newtown from the South, keep on the A489 until you reach the traffic lights at McDonalds. Turn left at the traffic lights (keeping McDonalds on your left).

Go over the river bridge following signs for the hospital. Take the fifth turning on the right (opposite the Bell Hotel). Carry on up the hill out of Newtown for approx. 6 miles.

The entrance to Gregynog is sign-posted on the left just before the village of Tregynon.

- **From Welshpool**

- Head towards Newtown on the A483 for approx. 4 miles. Turn right towards Berriew (B4390).

In Berriew village take the second turning on the left, sign posted Bettws Cedewain 5 miles.

In Bettws follow the road round to the right (keeping the New Inn pub on your right) sign-posted Tregynon 2.5 miles.

At the next T junction the entrance to Gregynog is sign posted straight opposite.

- **For satellite navigation**

- Use the postcode SY16 3PL, which will bring you into the Hall grounds via the main Estate entrance. From the Berriew direction, it may also direct you to turn right towards Brooks, which is a steep single track road. Please ignore this and continue onto Bettws Cedewain.

2.4 Getting there by Train

- Rail links are via the Birmingham Aberystwyth line. The local train station is Newtown (Powys), approximately a 12 taxi journey from Gregynog Hall. There are direct trains to Newtown (Powys) from Birmingham Int'l and Birmingham New Street.

2.5 Local Taxi Companies

- **Station Taxis:** 01686 621818
- **Pauls Taxis:** 01686 624314
- **Ross Taxis:** 01686 627600

3 About Gregynog

3.1 History

Gregynog has existed for 800 years. By the 16th century it was the home of the Blayney family, local gentry who claimed descent from the early Welsh princes and whose courage and benevolence were praised by the court poets. Their coat of arms is the centrepiece of the fine oak carvings in what we now call the Blayney Room.

For hundreds of years Gregynog was one of Montgomeryshires leading landed estates, at the heart of the community and the local economy. The Blayney squires gave way to the Lords Sudeley, then Lord Joicey.

After several hundred years of private ownership, in 1913 a huge estate sale saw Gregynogs farms, cottages and woodlands sold off, many to their tenants. Gregynog Hall might have been demolished had not the wealthy Davies sisters acquired it in 1920 to become the headquarters of their enterprise to bring art, music and creative skills to the people of Wales in the aftermath of the First World War.

For twenty years the house was full of music, fine furniture and ceramics, hand-printed books from the Gregynog Press and, most extraordinary of all, the sisters collection of paintings by artists such as Monet, Cezanne and Van Gogh. Leading lights, such as George Bernard Shaw and Gustav Holst visited during these years for musical concerts or simply to enjoy the beautiful gardens and woodland walks.

At the end of the 1950s, after wartime use as a Red Cross convalescent home, Gregynog was bequeathed to the University of Wales as a conference centre. It welcomed its first students in 1963 and theyve been coming ever since! But the old Gregynog lives on the music, the art, the printing press and the gardens. It is still a magical, timeless place where you can walk in the grounds on a quiet evening and listen to the birdsong just as the Davies sisters did many decades ago.

3.2 Walks

The gardens at Gregynog are unrivaled, offering a mixture of formal and woodland walks.

To assist our visitors in fully appreciating the beauty and diversity of the estate, we have created a variety of colour-coded woodland walks. The walks are of varying length and difficulty, weaving their way through the estate to offer tantalising views of both the Hall and the stunning Montgomeryshire countryside.

The new Lily Lake Walk, Warren Walk, Great Wood Walk and Valley Walk have been created to offer something of interest to everyone.

Attractions on the walks include the secluded Mellors cottage, the Davies sisters painting shed and Quackers Hall, perched in the middle of the lily lake, and a birdwatching hide located deep in the Garden House Wood. Simultaneously striking and amusing is the stone statue of a giant hand protruding from the earth, a particular favourite of passers-by taking a woodland stroll. Against this backdrop, the meandering Bechan Brook flows through the estate attracting birds, including kingfishers.



The Bee Apiary, acknowledged to be the prettiest in Wales, is located in the Dell. Visitors can see the bees flying from their hives and coming back again after collecting pollen from the gardens. The attractive viewing shelter has been designed to allow close but safe access to the bees: there are over one million of them, and contains interpretation boards describing the importance of bees, their life cycle and the various types of hives within the apiary. Find out about when the beekeepers will be in the apiary, as they will bring frames of bees close enough for you to see and smell, by visiting the Monty Bees website.

With support from Natural Resources Wales, a number of wildlife interpretation boards are installed throughout the estate, enabling visitors to understand the importance of the natural environment within Gregynog, recently designated a National Nature Reserve.

Our walks are naturally maintained, mainly by people's feet and dogs' paws with minimal interference in this unspoilt environment. You may find yourself bashing through bracken and wading through muddy patches at times .. just a perfect escape in the wilds of Wales, but bring your boots!

3.3 Library

A unique collection of books

The fine arts, Gregynog Press books, Welsh history, literature, culture and language

The books on open access in the west corridor and in the Thomas Jones Library and Dora Herbert Jones Library are the most visible part of a substantial collection of books and



archive material held at Gregynog. Many of the books once belonged to the Davies sisters, although most of the general non-fiction collection has been acquired since the 1960s when the University of Wales took over the hall. The policy behind the development of the library over the years has been firstly, to complement the activities which take place at Gregynog; secondly, to offer insights into its history and its special significance to music, art and fine printing; and thirdly, to reflect its nature as an institution at the heart of Welsh cultural life.

In addition, a considerable number of documents and other items relating to the history of the house have been collected over the years and these are now listed and stored securely. This includes a full set of Gregynog Press and Gwasg Gregynog publications which can be consulted on application.

However it should be noted that most surviving archive material relating to Gregynog, including items such as the Visitors Book kept here in the 1920s and 1930s, is now in the National Library of Wales in Aberystwyth.

Gregynog Library books are not available for external loan, but we welcome Visiting Readers. By becoming a Gregynog Member you can apply for a Visiting Readers ticket which will entitle you to visit Gregynog on most occasions when the house is open, to research, study or just browse in the library.

Books on open access for browsing and private reading are arranged as follows:

1 The Library Corridor

The books on open access in the library corridor are general non-fiction books and literature. Subjects include philosophy, religion, history and literature. There is a large collection of books on the fine arts, including Impressionist and Post-Impressionist painters, also printing, binding and the book arts. Journals include *The Studio* magazine dating back to the early 20th century, also *The Burlington Magazine* and other art related journals.

Gregynogs collection of material relating to Irish language and literature is shelved here, and at the far end of the corridor is a separate collection of material relating to Arthurian myth and legend as it spread from its Celtic roots to German, France and beyond.

2 The Thomas Jones Library

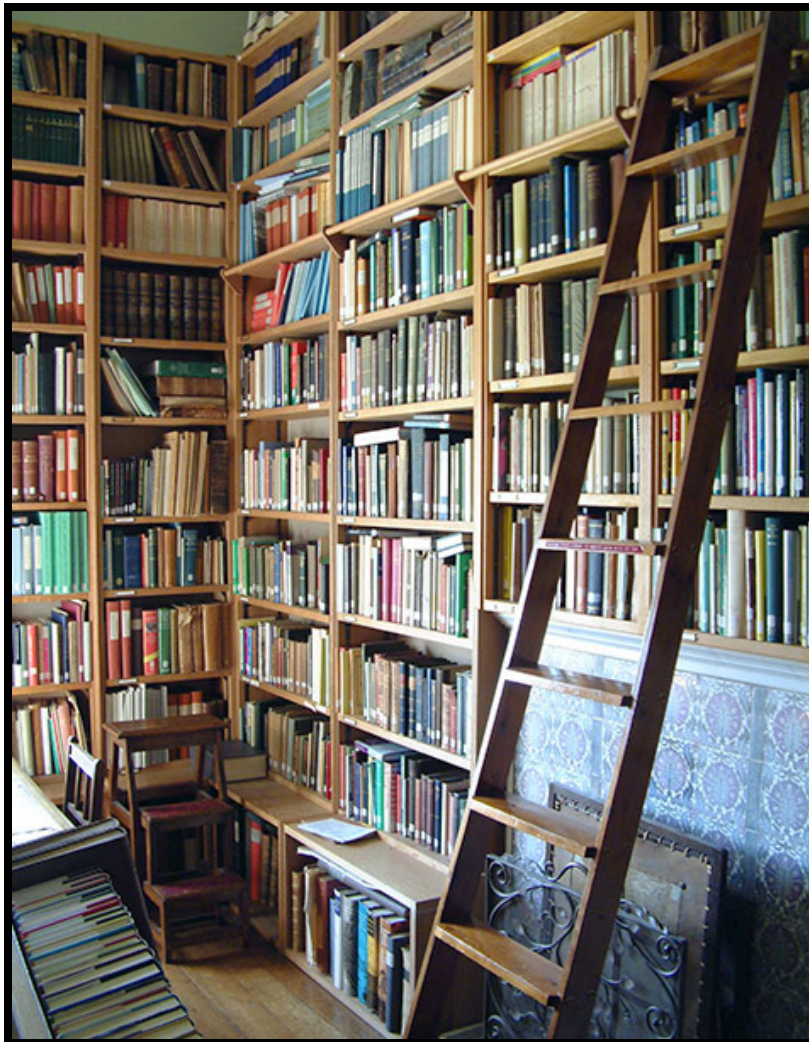
The Thomas Jones library houses a collection of reference books, encyclopaedias, dictionaries, atlases etc., including some useful horticultural reference books. A section of the Fine Art Collection is also housed in this room, which is in regular use for meetings and seminars.

3 **The Music Library**

This is a collection of books shelved in the corridor next to the Music Room. It includes an early edition of Groves Dictionary of Music, and a large collection of biographies of musicians and composers.

4 **The Dora Herbert-Jones Library**

This is what is known as the small library at the far end of the library corridor, where the Librarians desk and computer are also located. All the books and journals in this library relate to Wales and the Celtic countries, either in Welsh or other Celtic languages, or about Wales and the Celtic countries, their history, literature and culture.



4 Timetable

4.1 Friday 21st April

All talks will take place in the Seminar Room, 2nd Floor.

Time	Presenter	Title	Pg
11:45	Warwick Minibus	Departure from Statistics Common Room 11:45 sharp	-
14:00	Arrival & Check-In	Gregynog Hall Shop	-
14:55	Murray Pollock	Welcome and Information	-
15:00	Jonathan Tawn	Extreme Value Methods and their Applications (Talk 1 of 3)	14
16:00	Afternoon Tea	Blayney Room	-
16:30	Poster Session	Joicey Room	-
	Nanda Aryal	Fitting the Bartlett-Lewis rainfall model using Approximate Bayesian Computation	17
	Alejandra Avalos	Batch effect adjustment using Bayesian factor analysis	17
	Cyril Chimisov	Adaptive MCMC	18
	Mathias Cronjager	An almost infinite sites model	18
	Jon Cockayne	Bayesian Probabilistic Numerical Methods	19
	Nathan Cunningham	Particle Monte Carlo methods to cluster multiple datasets with applications to genomics data	19
	Jack Jewson	Robust Bayesian Decisions in the M-open world	20
	Matt Moores	Bayesian modelling and computation for Raman spectroscopy	20
	David Selby	PageRank and the Bradley-Terry model	21
	Rachel Wilkerson	Exploring bespoke causal graphs	21
	Daniel Wilson-Nunn	A Rough Path Signature Approach to Arabic Online Handwritten Character Recognition	21
18:00	Paul Embrechts	Topics in Quantitative Risk Management (Talk 1 of 3)	14
19:00	Dinner	Dining Room	-
20:15	Alexander McNeil	More Powerful Backtests for Market Risk Models Using Realized p-Values	15
21:15	Bar	Basement	-

4.2 Saturday 22nd April

All talks will take place in the Seminar Room, 2nd Floor.

Time	Speaker	Title	Pg
08:00	Breakfast	Dining Room	-
09:00	Jonathan Tawn	Extreme Value Methods and their Applications (Talk 2 of 3)	14
10:00	Paul Embrechts	Topics in Quantitative Risk Management (Talk 2 of 3)	14
11:00	Coffee Break	Dining Room	-
11:30	Jenny Wadsworth	Non-limiting spatial extremes	16
12:30	Lunch	Dining Room	-
13:30	<i>Free Afternoon</i>	<i>Optional Powis Castle trip 13:00-16:45</i>	-
16:00	Afternoon Tea	Dining Room	-
17:00	Jonathan Tawn	Extreme Value Methods and their Applications (Talk 3 of 3)	14
18:00	Dinner	Dining Room	-
19:00	<i>Free Time</i>	<i>Optional Gregynog Young Musicians Conference</i>	-
21:00	Bar	Basement	-

4.3 Sunday 23rd April

All talks will take place in the Seminar Room, 2nd Floor.

Time	Speaker	Title	Pg
08:00	Breakfast	Dining Room	-
09:00	Paul Embrechts	Topics in Quantitative Risk Management (Talk 3 of 3)	14
10:00	Ioannis Papastathopoulos	Modelling Clusters of Extreme Values	16
11:00	Coffee Break	Blayney Room	-
11:20	Group Photo	Location weather dependent	-
11:30	Vicky Henderson	Prospect Theory in a Dynamic Context	15
12:30	Lunch	Dining Room	-
13:30	Departure	-	-
13:30	<i>Warwick Minibus</i>	<i>Departure from dining room immediately after lunch</i>	-

5 Abstracts

5.1 Short Course Abstracts

Topics in Quantitative Risk Management

Paul Embrechts

ETH Zurich

In this short course I will discuss three problems from the realm of Quantitative Risk Management. These topics are based on the following recent papers:

- Topic 1: Embrechts, P., Wang, B., Wang, R. (2015): Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics* 19(4), 763-790.
- Topic 2: Chavez-Demoulin, V., Embrechts, P., Hofert, M. (2016): An extreme value approach for modeling Operational Risk losses depending on covariates. *Journal of Risk and Insurance* 83(3), 735-776.
- Topic 3: Embrechts, P., Hofert, M., Wang, R. (2016): Bernoulli and tail-dependence compatibility. *Annals of Applied Probability* 26(3), 1636-1658.

Extreme Value Methods and their Applications

Jonathan Tawn

Lancaster

This series of talks will introduce the strategy and associated methods of using asymptotically motivated extreme values models for univariate and multivariate extremes value problems and will provide illustrations of how these are used to address substantive problems in environmental risk assessment. In the univariate case I will cover maxima and threshold methods and discuss how to account for non-identically and non-independent variables. In the multivariate case the definition of an extreme is not unique. I will give an overview of the different types of asymptotic approach in terms of the direction of the extrapolation and show different ways of measuring extremal dependence. I will illustrate these methods in their use for the prevention of coastal flooding, for identifying the cause of the sinking of the MV Derbyshire, for setting new global shipping safety standards, for assessing the risk of heatwaves, for finding out how often the UK gets a 1 in 100 years flood event, and for deriving the estimated distribution of flood insurance claims.

5.2 Talk Abstracts

Prospect Theory in a Dynamic Context

Vicky Henderson

Warwick

Prospect theory is one of the most popular non-expected utility models designed to capture biases in our behaviour well documented in experimental studies. In fact, Kahneman and Tversky's (1979, 1992) papers on prospect theory are among the most highly cited papers in economics. One important component of prospect theory is the overweighting of unlikely events, captured by probability weighting. For example, this has been used to explain why people both buy insurance and lottery tickets. In this talk I will present some recent progress on dynamic prospect theory models with probability weighting. We will find people can change their minds, in some settings people might "gamble until the bitter end", and find randomization attractive, but fortunately, in other settings, we can better capture features of investor behaviour including the desire for stop-loss strategies, right-skewed payoffs, and the disposition effect.

More Powerful Backtests for Market Risk Models Using Realized p-Values

Alexander McNeil

York

We present an overarching approach to the statistical backtesting of market risk models over daily or other short time horizons using weighted realized p-values. This approach subsumes VaR exceptions tests at one or more levels as well as other tests that have been proposed in the recent academic literature. It allows new, more powerful tests to be devised that are effective at exposing deficiencies of the predictive models that are used to compute measures of tail risk for capital adequacy purposes. Under the Fundamental Review of the Trading Book (FRTB) banks are required to implement backtesting programmes that go beyond the basic VaR exception tests that are necessary for internal model approval. Our approach sheds light on the kind of tests that banks and regulators should consider to satisfy this requirement. We also show how the methodology can be generalized to construct simultaneous backtests of multiple desks.

Modelling Clusters of Extreme Values

Ioannis Papastathopoulos

Edinburgh

It is well known that short range dependence of random processes leads to clusters of extreme values. In particular, the generating mechanism of extreme events in time or space, has an inhomogeneous Poisson process character that describes the peaks of the events together with a cluster distribution that can vary spatio-temporally. This representation provides an insightful modelling basis according to which asymptotically motivated statistical models can be constructed. Such models are typically fitted to observed data in order to infer summaries of the cluster distribution such as the extent and severity of the event. In this talk, I will present a broad framework for modelling clusters of extreme events based on limiting forms of stationary Markov processes with higher order memory and with special emphasis placed on asymptotically independent stochastic processes. I will present strong solutions of limiting stochastic difference equations that give rise to a new class of tail chains that encompasses a very rich structure that is indispensable for statistical inference. Last, I will discuss extensions to non-stationary processes and statistical estimation together with an application to extremes of North Sea ocean waves.

Non-limiting spatial extremes

Jenny Wadsworth

Lancaster

Many questions concerning environmental risk can be phrased as spatial extreme value problems. Classical extreme value theory provides limiting models for maxima or threshold exceedances of a wide class of underlying spatial processes. These models can then be fitted to suitably defined extremes of spatial datasets and used, for example, to estimate the probability of events more extreme than we have observed to date. However, a major practical problem is that frequently the data do not appear to follow these limiting models at observable levels, and assuming otherwise leads to bias in estimation of rare event probabilities. To deal with this we require models that allow flexibility in both what the limit should be, and in the mode of convergence towards it. I will present a construction for such a model and discuss its application to some wave height data from the North Sea.

5.3 Poster Abstracts

Fitting the Bartlett-Lewis rainfall model using Approximate Bayesian Computation

Nanda Aryal

Melbourne

Presently cluster rainfall models are usually fitted using Generalized Method of Moments (GMM). GMM compares empirical and theoretical moments using weighted least squared criteria. Complex stochastic models may not have theoretical moments. Approximate Bayesian Computation (ABC) fills this gap by simulation techniques. It is a new approach for Poisson cluster rainfall modeling regime. ABC compares the observed data with the simulated data through summary statistics. The method replaces the theoretical moments by simulated data moments. This simulation study shows that ABC is a better option for parameter estimation of the Bartlett-Lewis rainfall model.

Batch effect adjustment using Bayesian factor analysis

Alejandra Avalos Pacheco

Warwick

With the rapidly increasing volume of heterogeneous biological data, available due to high-throughput technologies, it is becoming more and more difficult to integrate and summarise these data for exploratory analyses. Experimental variation, such as batch effects, is present in most large datasets and adds difficulty to the aforementioned task. In order to keep up with this large influx of biological data, new integrative methods are needed, based on dimensionality reduction techniques for effective understanding of these vast amounts of information. There is a need for new dimensionality reduction methods that can integrate heterogeneous data while preventing these technical biases from dominating the results. We provide a model based on factor analysis and latent factor regression, which incorporates a novel adjustment for variance batch effects often observed in bioinformatics data. Our model is directly applicable as data integration technique via a concatenation of heterogeneous datasets, resulting into common latent factors across the datasets.

Adaptive MCMC

Cyril Chimisov

Warwick

The popularity of Adaptive MCMC has been fueled on the one hand by its success in applications, and on the other hand, by mathematically appealing and computationally straightforward optimisation criteria for the Metropolis algorithm acceptance rate (and, equivalently, proposal scale). Similarly principled and operational criteria for optimising the selection probabilities of the Random Scan Gibbs Sampler have not been devised to date. In the present work we close this gap and develop a general purpose Adaptive Random Scan Gibbs Sampler that adapts the selection probabilities. The adaptation is guided by optimising the L2spectral gap for the targets Gaussian analogue, gradually, as targets global covariance is learned by the sampler. The additional computational cost of the adaptation represents a small fraction of the total simulation effort. We present a number of moderately- and high-dimensional examples, including Truncated Normals, Bayesian Hierarchical Models and Hidden Markov Models, where significant computational gains are empirically observed for both, Adaptive Gibbs, and Adaptive Metropolis within Adaptive Gibbs version of the algorithm. We argue that Adaptive Random Scan Gibbs Samplers can be routinely implemented and substantial computational gains will be observed across many typical Gibbs sampling problems.

An almost infinite sites model

Mathias Christensen Cronjager

Oxford

When modelling aligned nucleotide sequences using a coalescent subject to mutation, it is common to presume that every mutation will affect a previously unaffected position on the genome; this is known as the infinite sites assumption. In a number of cases (such as when working with vira with short genomes and high mutation-rates) this assumption is violated which in turn may lead to model misspecification. Our work investigates an alternative to the infinite sites model which admits recurrent mutation, but which still allows for analysing data under an assumption of parsimony (i.e. assuming that the number of mutation-events is minimal or near-minimal). We present a recursive characterisation of the likelihood similar to known formulae for other coalescent-based models along with a dynamical-programming algorithm for computing likelihoods. The results presented are the result of joint work with Alejandra Avalos Pacheco (Warwick), Paul Jenkins (Warwick), and Jotun Hein (Oxford)

Bayesian Probabilistic Numerical Methods

Jon Cockayne

Warwick

The emergent field of probabilistic numerics has thus far lacked rigorous statistical foundations. We establish that a class of Bayesian probabilistic numerical methods can be cast as the solution to certain non-standard Bayesian inverse problems. This allows us to establish general conditions under which Bayesian probabilistic numerical methods are well-defined, encompassing both non-linear models and non-Gaussian prior distributions. For general computation, a numerical approximation scheme is developed and its asymptotic convergence is established. The theoretical development is then extended to pipelines of numerical computation, wherein several probabilistic numerical methods are composed to perform more challenging numerical tasks. The contribution highlights an important research frontier at the interface of numerical analysis and uncertainty quantification, with some illustrative applications presented.

Particle Monte Carlo methods to cluster multiple datasets with applications to genomics data

Nathan Cunningham

Warwick

Cluster analysis is a popular technique for discovering group structure in a dataset. While it has been applied in a number of fields, it is often applied in clustering genomics data where identifying groups of genes can help uncover gene function and discover regulatory networks. Modern high-throughput technologies produce a vast array of data from disparate sources, however classic clustering algorithms, such as k-means and hierarchical cluster analysis, are typically geared towards continuous data and are not suitable for mixed data types. Previous work on integrating these data employed a Bayesian non-parametric model, where observations were assigned to clusters using a one-at-a-time approach. This was found to be able to uncover information which would be difficult or impossible to uncover from approaches considering a single dataset, however it was found to have slow mixing properties. We present an extension to the original MDI algorithm where cluster allocations are determined using a particle Gibbs approach, which has been proposed as a means of escaping this problem.

Robust Bayesian Decisions in the M-open world

Jack Jewson

Warwick

Given a prior and a model, Bayesian statistics provides the tools to update prior beliefs given some data, and identifies the predictive distribution closest to the true data generating process in terms of Kullback-Leibler divergence. When this model is a correct representation of the true data generating process, Bayesian statistics learns in the optimal way (Zellner, 1988), 'correctly' capturing the information available in the data and finding the true data generating process as n grows. However, in the M-open world, it is not believed that any model correctly represents the real world process, but is instead chosen as a convenient approximation. We show that in this scenario, blindly targeting the predictive distribution minimising the Kullback-Leibler divergence to the truth is not robust to misspecifications in the model. Instead we consider minimising different divergence measures and appeal to general Bayesian updating (Bissiri, Holmes and Walker (2016)), a method to produce a coherent Bayesian update in the absence of a data generating model, in order to produce more robust Bayesian decisions.

Bayesian modelling and computation for Raman spectroscopy

Matt Moores

Warwick

Raman spectroscopy is a technique for detecting and identifying molecules such as DNA. It is sensitive at very low concentrations and can accurately quantify the amount of a given molecule in a sample. Raman scattering produces a complex pattern of peaks, which correspond to the vibrational modes of the molecule. Each Raman-active molecule has a unique spectral signature, comprised by the locations and amplitudes of the peaks. The shift in frequency of the photons is proportional to the change in energy state, which is reflected in the locations of the peaks. The amplitudes of the peaks increase linearly with the concentration of the molecule. However, the presence of a large, nonuniform background presents a major challenge to analysis of these spectra. The estimated amplitudes are completely dependent on the position of the baseline and vice-versa. We introduce a sequential Monte Carlo (SMC) algorithm to separate the observed spectrum into a series of peaks plus a smoothly-varying baseline, corrupted by additive white noise. The peaks are modelled as Lorentzian, Gaussian or Voigt functions, while the baseline is estimated using a penalised cubic spline. This latent continuous representation accounts for differences in resolution between measurements. Our hierarchical model allows for batch effects between technical replicates. We incorporate prior information to improve identifiability and regularise the solution. The posterior distribution can be incrementally updated as more data becomes available, resulting in a scalable algorithm that is robust to local maxima. These methods have been implemented as an R package, using RcppEigen and OpenMP.

PageRank and the Bradley-Terry model

David Selby

Warwick

Network centrality measures are used to compare nodes according to their importance. Applications include ranking sports teams, ordering web pages in search results and estimating the influence of academic journals. Eigenvector-based metrics such as PageRank derive these measures from the stationary distribution of an ergodic Markov chain, whereas techniques such as the Bradley-Terry model treat ranking as a statistical estimation problem. By using a quasi-symmetry representation, we show that the PageRank vector, suitably scaled, is a consistent estimator for the Bradley-Terry model. Scaled PageRanks can therefore be used, for example, to initialise iterative algorithms for Bradley-Terry maximum likelihood estimation, improving their performance on large datasets. We study the variance of scaled PageRank as an estimator, and find full asymptotic efficiency in some balanced situations of practical importance.

Exploring bespoke causal graphs

Rachel Wilkerson

Warwick

The semantics for causal relationships expressed in the study of Bayesian networks are powerful and widely used, but may be ill suited to the dynamics of a problem in a given domain. Extending established causal algebras to new classes of graphical models results in nuanced notions of causation.

A Rough Path Signature Approach to Arabic Online Handwritten Character Recognition

Daniel Wilson-Nunn

Warwick

Arabic handwriting poses a new set of challenges previously unseen in Latin and Chinese handwriting. These challenges include the strictly cursive nature of the script, the changing in shape of each character owing to its position within a word and numerous different ligatures that occur in handwritten Arabic. Building on recent breakthroughs in Chinese handwriting recognition using tools from Rough Path Theory, alongside state of the art Deep Learning techniques, this work introduces the foundations for a Rough Path Signature approach to Online Arabic character recognition. In the preliminary stages of this work, results are exceedingly promising, with a recognition rate approximately 5% higher than the current state of the art tools on the same data set. Keywords: Online handwriting recognition, Arabic handwriting recognition, rough path theory, rough path signature, random forests, decision trees, neural networks, LSTMs.
