Proceedings of the

DSNE Summer Conference 2019

8th - 9th July 2019

DSNE

Data Science of the Natural Environment

#DSNEConf19

EPSRC Engineering and Physical Sciences Research Council

CEH Centre for Ecology & Hydrology NATURAL ENVIRONMENT RESEARCH COUNCIL

Lancaster University

NERC SCIENCE OF THE ENVIRONMENT

Proceedings of the DSNE Conferene 2019

Lancaster University

Editor:
María Isabel Borrajo García

ISBN: 978-1-86220-361-7

Organising committee

Paul Young (chair)

Maria Isabel Borrajo

Simone Gristwood

David Leslie

Luigi Sedda

Venue information

The Storey
Meeting House Ln
Lancaster
LA1 1TH

Conference program

DAY 1: MONDAY 8TH JULY	
1000 - 1100	**ARRIVAL & REGISTRATION** (with tea, coffee and fruit)
1100 - 1110	**Welcome by DSNE PI** David Leslie, Lancaster University / DSNE, UK
1110 - 1210 **Chair:** D. Leslie	**Keynote talk: High-dimensional Bayesian geostatistics: model-based solutions for the laptop** Sudipto Banerjee, University of California, Los Angeles, USA
1210 - 1250 **Chair:** M. Hollaway	**The collaborative species distribution model: increasing trust in models through co-design** Hamish Holewa, CSIRO, Melbourne, Australia (via Skype)
	Virtual Labs: Realisation concepts and software architecture Maria Salama, Lancaster University / DSNE, UK
1250 - 1400	**LUNCH**
1400 - 1500 **Chair:** D. Clarkson	**Earth system data science activitied at NCAR** Jean-François Lamarque, NCAR, Boulder CO, USA
	Virtual labs: applying data science innovations to environmental science challenges Mike Hollaway, CEH Lancaster / DSNE, UK
	Using extreme value statistics to better understand Greenland ice sheet Maribel Borrajo-García, Lancaster University / DSNE, UK
1500 - 1530	**BREAK** (with tea and coffee)
1530 - 1630 **Chair:** M. Salama	**Keynote talk: New frontiers in cloud and edge/fog computing for big data and internet-of-things applications** Rajkumar Buyya, University of Melbourne, Australia
1630 - 1715 **Chair:** G. Blair	**Panel discussion: data science and the digital environment** Gordon Blair (Lancaster U), Ron Corstanje (Cranfield U), Simon Gardner (NERC), Faiza Samreen (Lancaster U / Ensemble)
1715	**END OF DAY 1**

1900	**DINNER** Royal Kings Arms Hotel, opp. The Storey

DAY 2: TUESDAY 9TH JULY	
0930 - 1030 **Chair:** M.I. Borrajo	**Keynote talk:** Janine Illian, University of St Andrews, UK
1030 - 1100	**BREAK** (with tea, coffee and fruit)
1100 - 1230 **Chair:** O. Oyebamiji	**An integrated approach for co-estimating multiple global geophysical surface processes within a coupled Earth system statistical framework** Jonathan Bamber, University of Bristol, UK
	Using deep neural networks for radar based nowcasting Niall Robinson, Met Office Informatics Lab, UK
	Ecosystem resilience: exploring a complex subject through new data science Jim Latham, Natural Resources Wales, UK
	Why trust matters Bran Knowles, Lancaster University, UK
1230 - 1400	**LUNCH AND NETWORKING**
1400 - 1500 **Chair:** T. Pinder	**Recognizing ice sheet melting using machine learning algorithms** Qingying Shu, Lancaster University / DSNE, UK
	Monitoring plastic on the bottom of European seas Jon Barry, Cefas, UK
	Bayesian hierarchical spatial models for uncertainty quantification in high-dimensional landscape problems Oluwole Oyebamiji, Lancaster University / DSNE, UK
1500 - 1600 **Chair:** Q. Shu	**Keynote talk: Detection of local differences between two spatiotemporal random fields** Bo Li, University of Illinois at Urbanna Champaign, USA
1600	**CLOSE OF MEETING**

Keynote speakers

Dr **Sudipto Banerjee** is a world leader of Bayesian hierarchical modelling and inference for spatial data analysis. His work spans from analysis of geographically referenced datasets to Bayesian statistics (theory and methods), hierarchical modelling, statistical computing and related software development. He is currently Professor and Chair of the Department of Biostatistics in the School of Public Health at the University of California, Los Angeles (UCLA). Dr Banerjee obtained an MS and PhD in Statistics from the University of Connecticut in 2000 after obtaining his B.Sc. (Honours) from Presidency College and his M.Stat from the Indian Statistical Institute in Kolkata, India.

During his career, Dr Banerjee has received many honours, including the Abdel El-Shaarawi Awarmm from The International Environmetric Society (TIES), the Mortimer Spiegelman Award from the American Public Health Association, elected membership of the International Statistical Institute, elected fellowships in the Institute of Mathematical Statistics (IMS), the American Statistical Association (ASA), and the International Society for Bayesian Analysis, a Distinguished Achievement Medal from the ASA's Section on Statistics and the Environment, the ASA's Outstanding Statistical Application Award, and the 2019 George W. Snedecor Award from the Committee of Presidents of Statistical Societies (COPSS). Dr Banerjee is author of seminal literature including two textbooks "Hierarchical Modeling and Analysis for Spatial Data" (Banerjee, Carlin, Gelfand, 2014) and "Linear Algebra and Matrix Analysis for Statistics" (Banerjee, and Roy, 2014), an edited volume "Handbook of Spatial Epidemiology" (Lawson, Banerjee, Haining, Ugarte, 2016), and over 130 peer-reviewed journal articles, including several with over 100 citations.

Dr **Rajkumar Buyya** is a Redmond Barry Distinguished Professor and Director of the Cloud Computing and Distributed Systems (CLOUDS) Laboratory at the University of Melbourne, Australia. He is also serving as the founding CEO of Manjrasoft, a spin-off company of the University, commercializing its innovations in Cloud Computing. He served as a Future Fellow of the Australian Research Council during 2012-2016. He has authored over 625 publications and seven text books including "Mastering Cloud Computing" published by McGraw Hill, China Machine Press, and Morgan Kaufmann for Indian, Chinese and international markets respectively. He is one of the highly cited authors in computer science and software engineering worldwide (h-index=126, g-index=271, 80,600+ citations). "A Scientometric Analysis of Cloud Computing Literature" by German sci-

entists ranked Dr Buyya as the World's Top-Cited (#1) Author and the World's Most-Productive (#1) Author in Cloud Computing. Dr Buyya is recognized as a "Web of Science Highly Cited Researcher" for three consecutive years since 2016, a Fellow of IEEE, and Scopus Researcher of the Year 2017 with Excellence in Innovative Research Award by Elsevier for his outstanding contributions to Cloud computing. He served as the founding Editor-in-Chief of the IEEE Transactions on Cloud Computing. He is currently serving as Co-Editor-in-Chief of Journal of Software: Practice and Experience, which was established about 50 years ago.

Dr **Janine B Illian** is currently Senior Lecturer in Statistics and Head of Statistics in the School of Mathematics and Statistics at the University of St Andrews, Scotland, and will take up a position as Chair in Statistical Sciences at the University of Glasgow in July 2019. Dr Illian's work focuses on spatial point process methodology with the specific aim to transform the field of spatial point process modelling from a theoretical to a practically relevant area of statistics. She has recently worked on spatial modelling in the context of environmental sciences, cancer research, health sciences, terrorism studies and earthquake research.

Dr **Bo Li** received her PhD in Statistics from Texas A&M University in 2006, and then became a Post-Doc at National Center for Atmospheric Research before joining the Purdue faculty 2008. In 2013 she moved to University of Illinois at Urbana-Champaign and is now a Professor with Data Science Founder Scholar in the Department of Statistics. Dr Li's research mainly focuses on spatial and spatio-temporal statistics and environmental statistics concerning problems in climatology, atmospheric sciences, public health, forestry and agriculture. Dr Li has served on the editorial boardsof several journals including the Journal of the American Statistical Association, Journal of Agricultural, Biological and Environmental Statistics and Environmetrics, and was guest editor for a special issue in Statistica Sinica and a special issue in Journal of Agricultural, Biological and Environmental Statistics.

Abstracts

Monday 8th July, 11.10 - 12.10

Title: *High-dimensional Bayesian geostatistics: model-based solutions for the laptop*
Presenter: Sudipto Banerjee, UCLA, USA
Abstract: With the growing capabilities of Geographic Information Systems (GIS) and user-friendly software, statisticians today routinely encounter geographically referenced data containing observations from a large number of spatial locations and time points. Such data arise in diverse disciplines within the environmental and physical sciences. Over the last decade, hierarchical spatiotemporal process models have become widely deployed statistical tools for researchers to better understand the complex nature of spatial and temporal variability in the environmental and physical sciences. However, fitting hierarchical spatiotemporal models often involves expensive matrix computations with complexity increasing in cubic order for the number of spatial locations and temporal points. The computational bottleneck renders such models unfeasible for large data sets. I will present a focused review of some approaches for constructing well-defined highly scaleable spatiotemporal stochastic processes. Such processes can be used as "priors" for spatiotemporal random fields within a hierarchical latent process setting and deliver full Bayesian inference. These approaches can be described as model-based solutions for spatiotemporal BIG DATA. The models ensure that the algorithmic complexity has n floating point operations (flops), where n is the number of spatial locations (per iteration). We compare these methods and provide some insight into their methodological underpinnings through a variety of applications in the environmental sciences.

Monday 8th July, 12.10 - 12.30

Title: The Collaborative Species Distribution Model: Increasing trust in models through co-design
Presenter: Hamish Holewa, CSIRO, Australia
Abstract: Effective management of our natural world under current and future conditions requires efficient, collaborative and complementary planning and decision-making processes with clear lines of accountability. While there has been significant progress in establishing national databases for the management of species observation data, these only represent samples of a species total distribution. The need and challenge therefore is to model these point-based observation data to obtain estimates or projections of the total range and distribution of the species. Such Species Distribution Models (SDMs), also known as Environmental Niche Models (ENMs), and the geographic data (or ?maps?) they generate provide vital information needed by governments at all levels to meet various policy and statutory responsibilities and obligations. SDMs quantify the response of species occurrence to environmental conditions described by variables such as climate, substrate, productivity and vegetation. The outcomes of an SDM can be used to identify locations and regions with potentially suitable environmental conditions for a species as well as assess how species may respond to projected future climate changes or habitat loss. While SDMs are widely used in many decision and policy making programs, investment in species distribution information has been fragmented and limited.

In Australia, three different government departments have collaborated with national computational research programs, the Atlas of Living Australia and the Biodiversity and Climate Change Virtual Laboratory to co-design and develop a standard framework for modelling threatened species distributions for use in policy and environmental decision making. The program includes three complementary pillars:

- An expert panel with both researchers and government practitioners who will review current SDM practices used in government and develop a set of best-practice methods.
- A technology program that includes the development of a new modelling platform that implements the best-practice methods for transparent and reproducible SDMs for decision making as established by the expert panel. Additionally, there will be an online portal for publishing ecological model outputs in a searchable catalogue to enhance cross-jurisdiction collaborations.
- Establishment of a training and skill development program to upskill decision makers using the new tools and methodology in practice.+

This presentation will showcase the outcomes of this program and highlight how digital infrastructure can enhance decision making. It will also showcase how taking a codesign approach to developing digital models and workflow tools with decision makers increases trust and adoption. In this case specifically, the collaboration across government departments ensures a) a consistent approach across jurisdictions, b) an increase in model quality, thereby leading to decrease in unnecessary survey or consultation efforts, c) an increase in suitability, robustness and reproducibility of SDMs, and d) increased advocacy and coordination in national programs and resources.

Monday 8th July, 12.30 - 12.50
Title: Virtual Labs: Realisation concepts and software architecture
Presenter: Maria Salama, Lancaster University, UK
Abstract: Virtual Labs are emerging as a key component of Data Science for the Natural Environment. Virtual Labs are online environments that support collaborative and integrative science across institutions and disciplines. They offer access to a range of data, analytical, process and data models, as well as assessment tools. Such environments offer secure collaborative space around data and data science techniques that supports Open Science. Virtual Labs are deployed in the cloud built on recent advances in cloud computing. Given the computational complexities of data and environmental models, appropriate software architectures and system principles are crucial for building virtual labs for data and environmental science. Self-adaptive and self-aware architectures are prominent for benefiting from the elasticity of the cloud environments and allowing green computing. The recently emerged microservices are also promising as a more fine-grained and scalable architecture. This talk will present an overview of Virtual Labs in Data Science for Natural Environment. The focus will be on the realisation of the concept and underlying software architecture.

Monday 8th July, 14.00 - 14.20
Title: Earth system data science activitied at NCAR
Presenter: Jean François Lamarque, NCAR, USA
Abstract: This presentation will cover a range of topics on data sciences that are currently under investigation at the National Center for Atmospheric Research (NCAR), an NSF-funded climate and environmental science research centre headquartered in Boulder, Colorado. In particular, it will focus on the analysis of very large outputs from global climate models (the same as archived for the internationally-coordinated climate experiment CMIP6) in order to quantify and understand global change, as well as the use of machine learning for physics parameterizations within these models. Finally, the presentation will include personal views on future directions for research and

applications, informed from my experiences as heading the development of NCAR?s global climate model and (currently) leading the Climate and Global Dynamics Laboratory at NCAR.

Monday 8th July, 14.20 - 14.40
Title: Virtual labs: applying data science innovations to environmental science challenges
Presenter: Mike Hollaway, CEH Lancaster, UK
Abstract: The natural environment is a complex system with many feedbacks and interactions. In a constantly changing world key challenges are faced to gain an understanding as to how the environment will respond to changing drivers. The fields of data science and computing offer a wealth of expertise and methods to help address these challenges and therefore provide a strong opportunity for an inter-disciplinary approach to the problem. Virtual Labs offer a prime collaboration space where data scientists and environmental scientists can come together and start to develop techniques which can be used by a wide range of stakeholders. The collaborative nature of Virtual Labs allows the access to and development of common resources (E.g. Environmental Datasets) and their integration with various analytical methods to tackle a range of environmental science challenges. E.g. Uncertainty propagation through model chains or change-point detection in climate data. This talk will focus on some of the data science and methodological challenges faced in the Data Science of the Natural Environment (DSNE) project along with an example application of a virtual lab to help address these.

Monday 8th July, 14.40 - 15.00
Title: Using extreme value statistics to better understand Greenland ice sheet
Presenter: Maribel Borrajo-García, Lancaster University, UK
Abstract: Different climate models have been developed to forecast the future behaviour of climate in polar regions. While these models perform well at reproducing past climates in a broad sense, they have troubles with identifying extreme events. This talk focuses on extreme high temperature events on the Greenland ice sheet. An understanding of the location, frequency, duration and magnitude of these events is therefore necessary to understand the ice sheet?s response to climate change, interpret contemporary measurements of ice sheet volume change and constrain predictions of future ice sheet state. In this talk, we use advanced statistical techniques for extreme event identification to compile a statistical climatology of extreme temperature events on Greenland based on different regional and global climate models.

Monday 8th July, 15.30 - 16.30
Title: New frontiers in cloud and edge/fog Computing for big data and internet-of-things applications
Presenter: Rajkumar Buyyais, University of Melbourne, Australia
Abstract: Computing is being transformed to a model consisting of services that are commoditised and delivered in a manner similar to utilities such as water, electricity, gas, and telephony. Several computing paradigms have promised to deliver this utility computing vision. Cloud computing paradigm has turned this vision of ”computing utilities” into a reality. It offers infrastructure, platform, and software (application) as services, which are made available as subscription-based services in a pay-as-you-go model to consumers. Cloud application platforms need to offer (1) APIs and tools for rapid creation of elastic applications and (2) a runtime system for deployment of applications on geographically distributed computing infrastructure in a seamless manner.

The Internet of Things (IoT) paradigm enables seamless integration of cyber-and-physical worlds and opening up opportunities for creating new class of applications for domains such as smart cities. The emerging Fog computing paradigm is extends Cloud computing model to edge

resources for latency sensitive IoT applications.

This keynote presentation will cover (a) 21st century vision of computing and identifies various IT paradigms promising to deliver the vision of computing utilities; (b) innovative architecture for creating elastic Clouds integrating edge resources and managed Clouds, (c) Aneka, a Cloud Application Platform, for rapid development of Cloud/Big Data applications and their deployment on private/public Clouds with resource provisioning driven by SLAs, (d) a novel FogBus software framework with Blockchain-based data-integrity management for facilitating end-to-end IoT-Fog(Edge)-Cloud integration for execution of sensitive IoT applications, (e) experimental results on deploying Cloud and Big Data/Internet-of-Things (IoT) applications in engineering, and health care, satellite image processing, and smart cities on elastic Clouds; and (f) directions for delivering our 21st century vision along with pathways for future research in Cloud and Fog computing.

Monday 8th July, 16.30 - 17.15
Title: Data science and the digital environment
Presenters: Gordon Blair (Lancaster University, chair), Ron Corstanje (Cranifield University), Simon Gardner (NERC), Faiza Samreen (Lancaster University / Ensemble)
Abstract: Last year, the UK?s research and innovation funding agency (UKRI) announced the "Constructing a Digital Environment Strategic Priorities Fund programme" to develop a digitally enabled environment through creating an "integrated network of sensors (in situ and remote sensing based), and associated methodologies and tools for assessing, analysing, monitoring and forecasting the state of the natural environment". This is an exciting and potentially groundbreaking programme aiming at enhancing decision-making for a wide range of stakeholders. This also mirrors concerns and research priorities that are being realised the world over. This panel will examine the potential role of data science in supporting a Digital Environment, raising and discussing questions such as:

- How important is data science to the success of the programme?
- What are the particular opportunities and challenges for data science in this area?
- What barriers do we need to overcome to realise this vision?

Tuesday 9th July, 9.30 - 10.30
Title:
Presenter: Janine Illian, St Andrews University, UK
Abstract: asked

Tuesday 9th July, 11.00 - 11.20
Title: An integrated approach for co-estimating multiple global geophysical surface processes within a coupled Earth system statistical framework
Presenter: Jonathan Bamber, University of Bristol, UK
Abstract: Correctly identifying and partitioning the sources of sea level change is crucial for improving future sea level rise (SLR) projections. Traditionally, changes in each component of the integrated signal have been tackled separately, which has often lead to inconsistencies between the sum of these components and the integral as measured by satellite altimetry. This is termed closing the Sea Level Budget.

To address these issues, the European Research Council has funded a five year project aimed at producing the first physically-based and data-driven solution for the complete coupled land-ocean-solid Earth system that is consistent with the full suite of observations, prior knowledge and fundamental geophysical constraints. This project is called "GlobalMass" based at the Bristol Glaciology Centre and Dept of Maths, University of Bristol.

Observed mass movement from the Gravity Recovery And Climate Experiment (GRACE) satellites, vertical land motion from a global network of permanent GPS stations are combined

with other satellite and in-situ observations such as the global network of Argo buoys, satellite radar and laser altimetry and tide gauge data. A Bayesian Hierarchical Model (BHM) has been developed to combine the observations alongside prior information that incorporates the physics of the coupled system such as conservation of mass and salt and characteristic length scales of different processes in both space and time, derived from models and/or observations. We use a GMRF approach for dimensional reduction of the observations so that a simultaneous solution can be obtained at a global scale. The scale of the problem negates the use of MCMC methods for the inference and we have adopted the Inverse Nested Laplace Approximation (INLA) approach for computational efficiency. Despite this, the computational challenges in solving multiple latent processes (five) globally, at adequate spatial and temporal resolution, are significant.

The BHM was developed and tested on Antarctica, where it has been used to separate surface, ice dynamic and vertical land motion signals simultaneously. We illustrate the approach and concepts with examples from this test case and present some early results using the BHM globally for a time-invariant test of closing the sea level budget.

Tuesday 9th July, 11.20 - 11.40
Title: Using deep neural networks for radar based nowcasting
Presenter: Niall Robinson, Met Office Informatics Lab, UK
Abstract: Conventional nowcasting is able to make use of the most up-to-date observations by neglecting to resolve the physics of the atmospheric system. However, machine learning may provide the opportunity to combine the latest observations with prior knowledge of atmospheric physics. Here we describe initial results from an ongoing project to use deep neural networks to perform radar based rain nowcasting over the UK.

Tuesday 9th July, 11.40 - 12.00
Title: Ecosystem resilience: exploring a complex subject through new data science
Presenter: Jim Latham, Natural Resources Wales, UK
Abstract: Ecosystem Resilience (ER) is a concept that has become politically topical and embedded in environmental legislation in Wales; Natural Resources Wales (NRW) and other public bodies have a duty to “seek to maintain and enhance biodiversity and in so doing promote the resilience of ecosystems”. We have defined ER as the ability of ecosystems to resist, recover from, or adapt to pressures and disturbances such as habitat loss, climate change or pollution, and so maintain their function and delivery of ecosystem services. However, ER is an immensely complex topic and remains largely theoretical; its underlying mechanisms are not well understood and it has rarely been demonstrated in practice. We therefore fall back on pragmatic interventions that can be ‘reasonably assumed’ to be beneficial for ER, which are structured around a suite of attributes derived from published literature: diversity, extent, condition and connectivity. We can map elements of these to understand broad spatial patterns, but more elaborate analyses are required into both the relationships between the attributes themselves, and between the attributes and long-term changes in land-use and biodiversity that might actually demonstrate ER. Consequently, we are interested in the potential of DSNE to apply cutting-edge methodologies to our extensive spatial and temporal datasets to better understand ER. We will outline some avenues for collaborative research that we hope will be both academically appealing and have significant applied impact.

Tuesday 9th July, 12.00 - 12.30
Title: Why trust matters

Presenter: Bran Knowles, Lancaster University, UK
Abstract: Trust matters. We assume this to be true, but what is trust? It is a term used in daily life to characterise relationships one has with people, organisations, institutions, systems, services, devices, etc. We know from experience what it feels like when it is present and when it is absent. But trust is deceptively complex. Scholars have been trying in vain for some time to develop a universal theory of trust, but rather than asking "What is trust?" it appears there may be greater traction in asking "Which trust and when?" In this interactive talk, I will explore some of the diversity of perspectives on trust, with a particular emphasis on perspectives that appear most relevant to the problem of uncertainty in data science for the natural environment. I will offer my view on how we might foster trust in this context and why doing so matters.

Tuesday 9th July, 14.00 - 14.20
Title: Recognizing ice sheet melting using machine learning algorithms
Presenter: Qingying Shu, Lancaster University, UK
Abstract: Ice sheet data are available in the form of brightness temperature figure at daily time scale. These ice sheet figures are analysed for recognizing melting. In the exploratory analysis, we started with differentiating melting and non-melting conditions using a constant threshold, and then applied edge detection methods for delineating the shape of the ice sheet. We aim to develop Machine Learning algorithms for better characterization of ice sheet melting. The ultimate goal is to predict ice sheet melting in Antarctic and Greenland.

Tuesday 9th July, 14.20 - 14.40
Title: Monitoring Plastic on the Bottom of European Seas
Presenter: Jon Barry, Cefas, UK
Abstract: I will describe the European seafloor litter assessment. This is one of several tools to see if policy measures (e.g. taxes on plastic bags) are having an impact on the marine environment. This assessment is based on combining litter counts from 12 bottom trawl surveys from seas around Europe. There are several practical and statistical issues to overcome when carrying out this assessment. These include:

- Ensuring consistency in data collection and recording
- Eliminating spatial bias due to non-equal sampling effort
- Standardising for the different trawl equipment and trawl durations.

I use various methods including Generalised Additive Models (frequentist and Bayesian versions) and Gaussian Process smoothing to produce 2-dimensional smoothed maps of the probabilities that trawls contain plastic litter.

Tuesday 9th July, 14.40 - 15.00
Title: Bayesian hierarchical spatial models for uncertainty quantification in high-dimensional landscape problems
Presenter: Oluwole Oyebamiji, Lancaster University, UK
Abstract: We tested a Bayesian hierarchical model for achieving dimension reduction in modelling of large dimensional spatial data. This being a step towards emulating a complex integrated model of land use change. The method uses a combination of Bayesian principal component and Gaussian process based on nearest neighbour approximation. The approach is to first retrieve the low-dimensional underlying patterns from high-dimensional outputs using a Bayesian principal component analysis where the effective dimensionality of the latent space is determined automatically as part of the Bayesian inference procedure. This is followed by the emulation of the resulting low-dimensional data using a composite nearest-neighbour GP based on an assumption

of conditional independence. This reduces model complexity and captures different aspects of the socio-economic scenarios. The approach is computationally efficient and improves the accuracy of estimating the parameters as well as incorporating various sources of uncertainty. The method is being applied to a dataset from the IMPRESSIONS Integrated Assessment Platform (IAP2) model, an extension of the CLIMSAVE IAP, which has been widely applied in climate change impact, adaptation and vulnerability assessment for robust policy analysis.

Tuesday 9th July, 15.00 - 16.00

Title: Detection of local differences between two spatiotemporal random fields

Presenter: Bo Li, University of Illinois, USA

Abstract: Comparing the characteristics of spatiotemporal random fields is often at demand. However, the comparison can be challenging due to the high-dimensional feature and dependency in the data. We develop a new multiple testing approach to detect the local difference in the characteristics of two spatiotemporal random fields by taking the spatial information into account. Our method adopts a two-component mixture model for location wise p-values and then derives a new false discovery rate (FDR) control, called mirror procedure, to determine the optimal rejection region. This procedure is robust to model misspecification and allows for weak dependency among hypotheses. To integrate the spatial heterogeneity, we model the mixture probability as well as allow the alternative distribution to be spatially varying. An EM-algorithm is developed to estimate the mixture model and implement the FDR procedure. We study the FDR control and the power of our new approach both theoretically and numerically, and finally apply the approach to compare the mean and teleconnection pattern between two synthetic climate fields.