Challenges in environmental spatio-temporal modelling

Part of the EPSRC-funded project GEOBEx

16 June 2025 University of Glasgow, UK

School of Mathematics & Statistics

Mathematics & Statistics Building, 32 University Place, Glasgow, G12 8QQ

Table of contents

1 About the workshop

This one-day workshop marks the conclusion of the EPSRC-funded project GEOBEx: Geostatistics Binary Models for Extremes, led by Carolina Euan (Lancaster University) and Daniela Castro-Camilo (University of Glasgow). The project has focused on the development of novel methodologies for modelling spatial binary extremes, with applications to environmental problems. It serves as a compelling illustration of areas in spatio-temporal environmental statistics that require further development, providing valuable insights into current challenges and potential solutions.

The workshop aims to bring together a small group of researchers with shared interests in spatio-temporal environmental statistics and the environmental sciences more broadly. It offers a space for open, collaborative discussion and the exchange of ideas across disciplines and career stages.

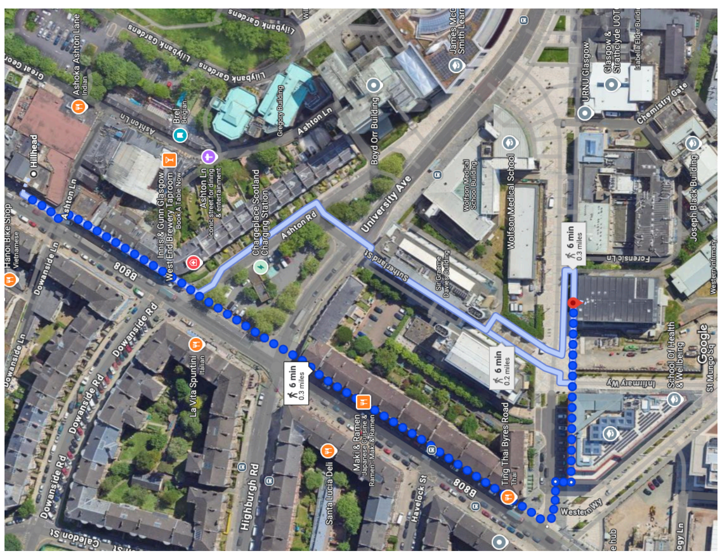
Goals

- Facilitate dialogue on challenges and opportunities in modelling spatial and spatio-temporal environmental extremes.
- Encourage collaboration by bringing together researchers from academia and applied fields.
- Identify future directions for methodological and applied development.
- Foster knowledge exchange, with participants sharing a one-slide summary of their work and selected 15-minute talks on current research related to environmental problems.

We hope that this collective effort will pave the way for stronger interdisciplinary connections and more impactful applications of environmental statistics.

2 Location

We will meet in Seminar Room 311B, located on the third floor of the Mathematics and Statistics Building (Google map). The room is accessible by both stairs and lift. The building is just a 6-minute walk from Hillhead subway station.



3 Programme overview

10:00-10:20 — Opening and summary of the GEOBEx project

10:20-10:50 — 3-minute introductions from each participant

10:50-11:20 — Coffee break

11:20-12:20 — Morning Talks

- **11:20-11:35** — Marian Scott (University of Glasgow): *Environmental digital twins — challenges and opportunities*
- **11:35-11:50** — Claire Miller (University of Glasgow): *Challenges in spatiotemporal modelling of national scale river water catchments*
- **11:50-12:05** — Claire Risley (EA): *Current Challenges for the Environment Agency in Spatial Statistics*
- **12:05-12:20** — Ben Marchant (BGS): *Space-time challenges at the British Geological Survey*

12:20-13:30 — Lunch

13:30-14:30 — Afternoon Talks

- **13:30-13:45** — Israel Martínez-Hernández (Lancaster University): *Modelling large spatio-temporal data: A short, medium, and long term wind forecast*
- **13:45-14:00** — Dave Miller (BioSS): *Adding structure to regression-like ecological models*
- **14:00-14:15** — Michael Tso (UKCEH): *Adaptive sampling for high-frequency nutrient sensors*
- **14:15-14:30** — Craig Wilkie (University of Glasgow): *Data fusion approaches for environmental data*

14:30-15:00 — Coffee break

15:00-16:30 — Group discussion

17:30-19:45 — Dinner at The Bothy, 11 Ruthven Ln, Glasgow, G12 9BG (click here for directions from The Maths & Stats Building)

4 Abstracts

The grouping and sequence of talks aim to gradually build context—from general frameworks and national challenges to specific methods and applications—while maintaining variety and thematic coherence.

4.1 Opening / Framing the Landscape

4.1.1 Marian Scott (University of Glasgow)

Title: Environmental digital twins — challenges and opportunities

Summary. Increasingly more and more data are being generated on environmental systems, from new sensors and also earth observation missions. However the governance of such data sources is fragmented, and the first challenge we have is in discovering and linking these disparate data sources. At the same time, there are a number of integrative models, offering holistic views of the "system". It is in this context that digital twins are being developed. They offer a dynamic system view, informed/learning from data, they are a digital replica of the environmental system and there is a bi-directional flow of intelligence from real-virtual and virtual- real. Their main power lies in their predictive capacity, and the testing of different future scenarios. In this short presentation, I will focus on water and present a framework for catchment (river basin) modelling of water quality and quantity.

4.2 National-Scale Environmental Monitoring Challenges

4.2.1 Claire Miller (University of Glasgow)

Title: Challenges in spatiotemporal modelling of national scale river water catchments

Summary. In recent years, with an increasing number of emerging potential contaminants, the potential combined effect of multiple chemicals on river water quality has become of greater interest. Additionally, there are knowledge gaps around the effects of hydrological, climate and landscape changes, in combination with different mixtures of chemicals, on freshwater species in river water catchments. This presentation will introduce the related statistical and data analytics challenges and developments being explored through projects such as NERC MOT4Rivers.

4.2.2 Claire Risley (Environmental Agency)

Title: Current Challenges for the Environment Agency in Spatial Statistics

Summary. Claire will showcase spatial statistics projects she has contributed to during her time at the EA.

4.3 Institutional Perspectives on Spatio-Temporal Modelling

4.3.1 Ben Marchant (BGS)

Title: Space-time challenges at the British Geological Survey

Summary. Scientists at the British Geological Survey study many earth science systems that vary in space and time. These include the variation of groundwater levels and quality, the occurrence of landslides, vertical ground motion, the temperature of subsurface water that can provide geothermal heating and space-weather which can impact communication systems. There is often a need to make spatial and temporal predictions of key properties of these systems and to test hypotheses regarding the main drivers of variation. However, such analyses can be hampered by the complexity and heterogeneity of the systems, the number of measurements to be analysed and the use of legacy data where sampling locations may have been selected purposively. I will illustrate these points with reference to BGS work on the variation of groundwater levels and more generally discuss our remaining space-time challenges.

4.4 Statistical Methods for Complex Spatio-Temporal Data

4.4.1 Israel Martínez-Hernández (University of Lancaster)

Title: Modelling large spatio-temporal data: A short, medium, and long term wind forecast

Summary. Due to the rapid development of complex, performant technologies, spatio-temporal data can now be collected on a large scale. However, the statistical modelling of large sets of spatio-temporal data involves several challenging problems, particularly regarding computational demands. I will present a new methodology to model complex and large spatio-temporal datasets. This approach invloves estimating a continuous surface at each time point, effectively capturing spatial dependence, possibly nonstationary. As a result, the spatio-temporal data can be seen as a sequence of surfaces. Then, we model this sequence of surfaces using functional time series techniques. The functional time series approach allows us to obtain a computationally feasible methodology and also provides extensive flexibility in terms of time forecasting. We illustrate these advantages using a high-resolution wind speed simulated dataset of over 4 million values.

4.4.2 Dave Miller (BioSS)

Title: Adding structure to regression-like ecological models

Summary. Covariates that occur in ecological and environmental science are often inherently non-scalar, sometimes having a spatial or temporal structure of their own. While aggregating or selectively summarizing such covariates to yield a scalar covariate allows use of standard regression models, exactly how to do so can be problematic, e.g., using a mean or median of some subsequence of a time series. Here I'll talk about three useful extensions that fit in the GAM framework (varying-coefficient, scalar-on-function and distributed lag models) that can be used to add more structure to our models and the connections between these approaches. Although these models are a useful basic set of tools, I'll also discuss further advances that would enhance our modelling of highly structured data further.

Joint work with Ken Newman and Thomas Cornulier (BioSS)

4.5 Adaptive & Intelligent Monitoring

4.5.1 Michael Tso (UKCEH)

Title: Adaptive sampling for high-frequency nutrient sensors

Summary. Traditional environmental monitoring employs fixed designs, which do not vary over the survey duration. Adaptive sampling uses previously collected data to inform and vary a sample design over the course of a monitoring period to offer a more optimal and flexible data collection. Existing work on adaptive sampling has mostly been focused on spatial survey designs. Here we highlight the potential of adopting adaptive sampling principles by varying frequency of environmental sensors, which we are motivated by a new generation of high-frequency 'lab-on-a-chip' sensors. Even though these sensors feature low reagent consumption, measuring at very high frequency continuously comes with high energy, computational, data storage, and analysis costs and challenges. We envisage sensors to be configured to maintain a basic level of monitoring and record data continuously, while only being triggered to monitor at a very high frequency when certain data-driven conditions are met (and then return to baseline monitoring). The datasets and criteria to inform the triggering can be highly adaptable. For datasets, with sophistication increases, it can include data on the sensor itself, other collocated sensors, other sensors in the region, remote sensing data, stormwater overflow, or weather forecasts. For criteria, it can be based on user-defined thresholds, domain expert knowledge, or statistical methods such as changepoints, clustering, or Bayesian adaptive sampling. We illustrate our approach with examples and a R package. We welcome contributions to this ongoing work, particularly in addressing stakeholder concerns, such as robustness in statistical analysis of multi-frequency data and the detection of extreme events.

Joint work with: Peter A Henrys, Eleanor Mackay, Xize Niu (UK Centre for Ecology & Hydrology and University of Southampton)

4.6 Data Fusion and Integration

4.6.1 Craig Wilkie (University of Glasgow)

Title: Adaptive sampling for high-frequency nutrient sensors

Summary. Increasing availability of environmental data from multiple sources such as satellites and low-cost sensors provides us with improved understanding of our changing environment. However, data from these disparate sources can be of varying quality, and often on different spatial and temporal scales. Data fusion approaches are designed to combine information from multiple sources to provide an enhanced understanding of environmental variables, with associated uncertainty measures that account for differences in the quality of the information provided by each source. This talk will present statistical data fusion approaches, with possible extensions to modelling extreme values.

5 Proposed topics for discussion

Feel free to use the topics below as a guide—they're entirely optional, so you're welcome to include others or leave them out.

5.1 Theme 1: Data Complexity and Integration

Relevant talks: Marian Scott, Claire Miller, Ben Marchant, Craig Wilkie

Prompt questions:

- What are the most pressing barriers to integrating heterogeneous environmental data sources?
- How can we balance model complexity with interpretability when fusing diverse datasets?
- Are there good examples of success (or failure) in multi-source integration we can learn from?

5.2 Theme 2: Modelling Choices and Scalability

Relevant talks: Dave Miller, Israel Martínez-Hernández, Michael Tso

Prompt questions:

- What modelling approaches scale best to large or high-frequency datasets?
- How do we decide between simpler models (e.g., regression/GAMs) and complex hierarchical or functional models?
- What compromises have you made between model flexibility and computational feasibility?

5.3 Theme 3: Decision-Driven Modelling

Relevant talks: Claire Miller, Claire Risley, Michael Tso

Prompt questions:

- How do we make our models useful for decision-makers and practitioners?
- What statistical uncertainties matter most in practice, and how should we communicate them?
- How do we build adaptive systems that react to real-time information?

5.4 Theme 4: Collaboration and Co-Production

All talks are relevant here

Prompt questions:

- What has helped or hindered successful collaboration across academia, agencies, and industry?
- What would an ideal collaborative workflow look like for tackling a spatio-temporal environmental problem?
- Are there shared needs (e.g., open datasets, tools, training) that we could address together?