Full list of Abstracts

1. Keynote Talks

Keynote 1

Statistical Methods for Assessing Water Health in England

Dr Eleanor D'Arcy, Environment Agency

At the Environment Agency, we lead the delivery of the Natural Capital and Ecosystem Assessment (NCEA) programme for the water environment. This talk will demonstrate how statistical methods underpin decision making across the programme, from designing new monitoring networks to large-scale data analysis.

Our data team develops reproducible analytical pipelines to generate consistent insights across all water domains, including rivers, lakes, groundwater bodies, estuaries and coasts. Our statistically designed monitoring networks are fundamental for reliable and meaningful analysis. More recently, we've expanded this work by integrating newly collected NCEA data with historic Environment Agency data, enabling retrospective predictions and extending the utility of the NCEA programme.

Keynote 2

People and birds: Analytical methods for citizen science biodiversity data

Dr Alison Johnston, director of CREEM (University of St Andrews)

Citizen science data have become a key source for understanding ecological systems and informing conservation, but working with these data brings many statistical challenges. Key challenges to address analytically include spatial and temporal biases, observer preferences, and observer experience. I'll outline how we've approached these challenges to analyse eBird data, to learn more about avian ecology. eBird is the largest biodiversity citizen science project in the world and contains over 1 billion bird observations contributed by over a million participants. We can use these data to estimate bird species distributions, migratory movements, demographics, and population trends. Each of these targets of ecological inference requires different consideration of the datasets and different analytical methods. I'll outline how we use a spatio-temporal ensemble of machine learning models to estimate bird distributions and migratory movements, and use double machine learning methods to account for confounding factors in estimating bird population trends. I'll also outline how we can for the first time use large-scale observational data to estimate demographic parameters. All of these methods allow us to create new ecological knowledge from unstructured citizen science data.

Keynote 3

Flexible Spatial Modelling with inlabru: From Data to Ecological Inference

Dr Jafet Belmont Osuna, University of Glasgow

Recent technological advances and data-sharing platforms have radically transformed the availability of data for modelling species distributions, presenting both opportunities and challenges for monitoring the spatial distributions of wildlife populations. One major challenge in spatial modelling is translating mathematical models into code and using computational package interfaces effectively.

In recent years, the R-INLA package has become a powerful tool for fitting complex spatial models in ecology. However, its implementation for spatial latent models can be cumbersome, particularly for complex observation processes.

The inlabru package addresses these limitations by automating data preprocessing, simplifying model specification, and extending support for non-linear predictors, all while preserving INLA's computational efficiency. Its seamless integration with widely used spatial data formats makes advanced analyses more accessible to ecological researchers and opens new avenues for addressing complex ecological questions.

In this talk, I showcase inlabru's versatility through two applications. First, I illustrate how the iterate fixed-point INLA scheme implemented in inlabru can be used to determine how landscape features spatially structure species distributions when the spatial scale effects of the landscape variables are unknown. Then, I introduce a new approach to integrating opportunistic citizen science (CS) observations with data from planned surveys while accounting for observational errors inherent in both sources. These examples highlight inlabru's capacity to address pressing challenges in spatial ecological research.

Keynote 4

End-to-end data-driven environmental prediction: *from observations to decision-information*

Scott Hosking, Alan Turing Institute and British Antartic Survey

In a changing climate, precise environmental prediction is essential for protecting people and economies. Recognising this, the Alan Turing Institute has formed a team dedicated to creating an advanced multi-modal AI prediction system that is flexible enough to integrate various data sources (satellites, weather stations, radiosondes). We are leading the development of the "Aardvark Weather" model, the first *end-to-end* AI weather model which streamlines the entire forecasting process - *from data intake to decision-ready information* - into a single AI model that's efficient enough to operate on a desktop computer. This holistic approach offers significant advantages, especially for developing countries where access to supercomputers and digital infrastructure is limited. Our aim is to develop end-to-end forecasting systems that can generalise for weather, sea ice movement and ocean currents.

2. Invited Sessions

IS11 - Challenges in modelling environmental data

Organiser: Emiko Dupont

Demystifying spatial confounding

Emiko Dupont, University of Bath

Spatial regression models are commonly used to model data that have been collected across a geographical region. In such models, spatial random effects are included to approximate unmeasured spatial variation in the response variable. However, as spatial random effects are typically not independent of the covariates in the model, this can lead to significant bias in covariate effect estimates of interest. This fundamental problem, which makes covariate effect estimation in any spatial regression potentially unreliable, is referred to as spatial confounding. In recent years, there has been much interest in spatial confounding, not least because the most established methods for dealing with the problem were proven to be ineffective. However, research into the topic has sometimes led to puzzling and seemingly contradictory results. Here, we develop a broad theoretical framework that brings mathematical clarity to the mechanisms of spatial confounding, providing explicit and interpretable analytical expressions for the resulting bias. From these, we see that it is a problem directly linked to spatial smoothing, and we can identify exactly how the features of the model and the data generation process affect the size and occurrence of bias. Our framework can be used to understand and generalise existing results on spatial confounding, including dependency on spatial frequency information and suggested methods for bias adjustment.

Challenges of modelling presence-only data: developing methods to predict the distribution of European mammals, and their application in Wildlife Management

Simon Croft or Daniel Warren, Animal and Plant Health Agency

Presence-only data, often derived from citizen science, opportunistic sightings, and historical records, has become increasingly central to ecological modelling, particularly for species distribution modelling (SDM). However, such data presents unique methodological challenges, including sampling bias, spatial autocorrelation, and the lack of absence information, all of which can lead to misleading inferences if not properly addressed. In this talk, I will explore recent advances in modelling techniques tailored to presence-only data, with a focus on European mammals—a group for which distribution knowledge is often patchy but urgently needed for conservation and management.

I will review key methodological frameworks, comparing their performance and suitability for different data contexts. Special emphasis will be placed on how bias correction techniques, spatial filtering, and the incorporation of environmental covariates can improve model reliability and ecological interpretability. Drawing on case studies involving mammal species across Europe, I will illustrate how robust SDMs can be developed even in data-limited scenarios.

Finally, the talk will highlight how these predictive models are informing wildlife management practices. By bridging methodological innovation with practical application, this work aims to support evidence-based decision-making in biodiversity conservation, particularly in the face of rapid environmental change and increasing human-wildlife interactions.

A Bayesian approach for species distribution modelling of plants with presence only data

Kabiru Abubakari, Queen Mary University of London

Records of species data available on online portals, such as the Global Biodiversity Information Facility, are collected through the citizen science program. This kind of data is often referred to as presence-only data since only presence records are available. Species distribution models (SDMs) require presence-absence data. In previous studies, SDMs were developed using background data. This approach tends to favor nonparametric models intended for prediction. Thus, the effect of the parameters of interest in the SDMs is ignored.

We propose a Bayesian approach to modeling species distribution with R-INLA. We incorporate uncertainty about the absence of species in places without records using the combination of missing data imputation and Bayesian model averaging. We recognize that misclassification can attenuate the estimated parameters. So, an adjusted logit link function is used to correct the effect of measurement error on the estimated parameters. We present results for the parameters of interest using simulated and real data. Our approach performs better than the alternative parametric method and achieves satisfactory predictive accuracy.

IS12 - Design and Assessment Issues in Environmental and Ecological Studies

Organisers: Mark Kaiser, Abdel H. El-Shaarawi

Chair: Sylvia Esterby

Grasshopper Sparrow Breeding Status: Adaptive selection of neighborhoods in a binary Markov random field

Mark S. Kaiser, Department of Statistics, Iowa State University

Co-authors: Eva Biswas, Andee Kaplan and Daniel J. Nordman

Binary Markov random field models have great application in environmental and ecological problems, including representing the breeding status of Grasshopper Sparrows (Ammodramus svannarum) at sampling locations in the Midwestern United States. Spatial dependence in Markov random field models is formulated in terms of neighborhoods – locations that directly affect the conditional probability mass function at a focal location. Random variables are conditionally independent of values at locations that do not belong to their neighborhood. Neighborhoods are traditionally chosen as part of model formulation, and assessing a choice for such selection is a difficult problem. This is compounded by the fact that there is no unique decomposition for representing data structure as part of a large-scale model component (i.e., expected values) and small-scale model structure (i.e., dependence, which is governed by neighborhood specification). Within the context of modeling the breeding status of Grasshopper Sparrows, we illustrate the use of a goodness of fit criterion based on conditional covariances in choosing neighborhood structures when there are and are not additional covariates incorporated into the model.

Poisson Models for Assessing the Impact of Electromagnetic Fields on Earth Worms

Abdel H. El-Shaarawi, National Water Research Institute (retired)

The Before-After-Control-Impact (BACI) design was used in the U.S. Navy Ecological Monitoring Program to evaluate the impact on biota exposed to electromagnetic fields produced by their

communications system. This was a major monitoring program that began in 1982 and ended in 1993 in which biota were exposed to Extremely Low Frequency (ELF) signals. Research teams from several universities measured biological and ecological variables prior to beginning the operation of the transmitting facilities, during intermittent operation, and during full operation. The study involved treatment and control sites. The analysis of correlated temporal count data on earth worms in soil samples will be examined for possible ELF effects using Poisson regression models.

Modelling the Non-Negative Integer-Valued Time Series of Pseudo-nitzschia spp. Counts in Cascais Bay: A Comparison of Zero-Inflated Generalised Poisson and Zero-Inflated Negative Binomial Regression Models

Helena Mouriño, Faculdade de Ciências, Universidade de Lisboa

Phytoplankton are microscopic, primarily unicellular algae that inhabit the ocean and form the foundation of the marine food chain. In upwelling regions such as the Western Iberian Coast, phytoplankton biomass is largely influenced by a range of meteorological and oceanographic factors, including water column stratification, nutrient availability, and the intensity and persistence of upwelling events.

Diatoms of the genus *Pseudo-nitzschia* are known to produce the neurotoxin domoic acid, which is responsible for amnesic shellfish poisoning (ASP) outbreaks worldwide. Nearly half of the identified *Pseudo-nitzschia* species are considered toxic. Consequently, modelling the temporal dynamics of *Pseudo-nitzschia* is essential, as harmful algal blooms pose significant threats to fisheries and related industries.

In this talk, I will evaluate regression models to describe the temporal and seasonal patterns of Pseudo-nitzschia abundance, based on weekly water samples collected in Lisbon Bay, Portugal, from 2001 to 2006. Given the count nature of the data, characterised by overdispersion and a high proportion of zeros, I will compare the performance of Zero-Inflated Generalised Poisson and Zero-Inflated Negative Binomial regression models. To account for the strong autocorrelation in Pseudo-nitzschia abundance between two consecutive weeks, a partial adjustment structure will be incorporated into the mean function of the models. Explanatory variables include sea surface temperature measured at Lisbon Bay and the upwelling index derived from wind data recorded at the Cabo Carvoeiro meteorological station.

IS21 - Statistical methods in trait-based ecology: using organism traits to predict ecological outcomes

Organiser: Marta Shocket

Novel methods to extract trait measurements from spider webs and model foraging strategies

Francis Windram, Imperial College London

Spiders are prolific predators, exerting control over prey populations throughout terrestrial ecosystems. All spiders produce silk, but only some weave webs. The morphology of these webs constitutes functional foraging traits of the spider. The study of these traits can provide deep insight into the performance, preference, and internal condition of a given spider. It is, however, not a trivial task to measure or infer these traits. Spider webs can be thought of as spatially explicit networks of

silk on an (approximately) 2-dimensional plane, and the twinned processes of digitisation and measurement of these networks are an interesting exercise in raster-vector inference, heuristic path-following, and Moore neighbourhoods. In this talk I explore the traditional approaches to extracting spider web traits, present a new method of digitisation of physical networks from their raster representations, and demonstrate how to use these methods to extract spider web traits more accurately and in greater detail than previously possible. For one key trait (mesh gap size), we can now generate the full distribution over the entire web, rather than a coarse estimate of the mean. Finally, I demonstrate how the digitised networks could be used to simulate how prey interact with webs.

Image-based trait and demographic analysis for mechanistic modelling of invasive species

Erola Fenollosa, Oxford University

Mechanistic models represent a critical tool for understanding and forecasting species dynamics, as they explicitly incorporate physiological traits, demographic processes, and environmental interactions. These models are particularly well suited to forecasting invasive species impacts under changing conditions. Since invasive species are often in disequilibrium with their environment, correlative approaches to predict potential distributions are likely to fail. In a systematic review, we reveal that, despite their potential, the application of mechanistic models has been limited by the extensive and often prohibitive data requirements necessary for accurate parameterisation across multiple spatial and temporal scales. By harnessing imagery acquired through ground-based sensors, unmanned aerial vehicles (UAVs), and satellite platforms with varying spectral resolutions, we argue that it is possible to extract key functional traits and demographic parameters relevant to species performance and spread. Image-based approaches enable the systematic monitoring of individuals, populations, and landscapes, thereby offering unprecedented opportunities to characterise variation in vital rates and functional attributes across environmental gradients. Here, we illustrate how trait and demographic data derived from high-throughput image analysis can be incorporated into mechanistic niche models, improving their ability to predict invasive species distributions under current and future environmental conditions. By proposing images as a central data source, this work contributes to expanding the applicability and scalability of mechanistic models, supporting more robust ecological forecasting and more informed management strategies for invasive species.

Testing mathematical methods for predicting trait values in fluctuating environments

Marta Shocket, Lancaster University

Trait-based, mechanistic models are often used to predict how variation in temperature affects ecological processes like the transmission of infectious diseases. These models are typically parameterised with trait values measured across a gradient of constant temperatures. However, most organisms live in environments with temperatures that fluctuate at daily time scales. Rate summation––non-linear averaging of trait values measured at constant temperatures—is commonly used to infer performance in fluctuating environments, but its accuracy is rarely validated. Here, we use trait data from a malaria mosquito (Anopheles stephensi) measured across gradients of constant and daily fluctuating temperatures to demonstrate how rate summation can introduce systematic distortions into predicted thermal responses. We found that rate summation partially captured decreases in performance caused by fluctuations near thermal optima, yet incorrectly predicted increases near thermal limits (where performance also decreased). Accordingly, while thermal

suitability for transmission based on constant temperatures did not perfectly capture fluctuating environments, it was better than rate summation for estimating and mapping thermal limits, which are commonly used to predict the impact of climate change on disease at global scales. This work provides insight into methods for predicting ecological processes under climate change and emphasizes the need to improve understanding of organismal performance under fluctuating temperatures.

Harnessing species-level animal behavioural traits - the fuzzy issue of plasticity and intraspecific variation

Sally Keith, Lancaster University

The use of animal behavioural traits in macroecological studies to understand species responses to environmental change is in its infancy. For instance, the way behavioural variation is distributed across space, time and taxa at large scales is broadly unknown. One reason for our lack of knowledge in this area is that behaviours are a combination of genetic predisposition and environmental factors that govern whether they are expressed, making it difficult to quantify. Behaviour is also inherently plastic and variable within species, raising conceptual challenges when distilling it into fixed, species-level trait values. To reveal generalisable principles at global scales, we must therefore figure out how to simplify behavioural traits enough to allow feasible manipulation, without losing ecological meaning. I will highlight some of the key challenges we are currently facing, accompanied by initial thoughts on how new (or existing) statistical methods might help ecologists navigate incorporating behavioural traits into broad-scale ecological analyses without losing sight of their inherent complexity.

IS31 - Real world challenge inspiring the next generation of environmental and ecological statistics researchers

Organiser and Chair: Rachel McCrea

Statistical modelling of sea surface temperatures and marine heatwaves

Kajal Dodhia, Lancaster University

Sea Surface Temperatures (SSTs) around the UK have increased on average 0.3°C per decade over the last 40 years. EDF's priority is safety hence their Nuclear Power Plants (NPPs) should be resilient to natural hazards such as Marine Heatwaves. Sea water is used as a coolant for NPPs. Therefore, a rise in SST compromises safety and can lead to reduced efficiency of NPP cooling systems, equipment failures or temporary shutdowns and impact marine ecosystems.

Some challenges faced with in-situ SST data are the strong seasonal behaviour combined with inter-year variability and the dependence between extreme temperatures (high/low SSTs). To address these issues, we compare several methods which preprocess the data by modelling the seasonality before examining the extremes. We find that a non-parametric approach generally overestimates the magnitudes of the extremes whereas, a parametric approach tends to underestimate them. Regardless of the preprocessing method, the distribution of the extremes is still bimodal with seasonal modes. Therefore, our objectives are to establish a robust method for identifying the extreme values and a definition for Marine Heatwaves by determining whether they should be defined as an absolute extreme or relative extremes based on the time of year. We hope to improve our understanding of changes in high/low SSTs, in coastal regions around the UK.

Statistical modelling of earthquake occurrence and extreme magnitudes for seismic risk assessment

Wanchen Yue, Lancaster University

The rise of human-induced seismicity, driven by industrial activities, such as gas extraction and carbon storage, poses significant risks to public safety and infrastructure. These problems affect communities globally, with severe consequences for urban development and environmental safety. At the Groningen gas field in the Netherlands, earthquakes triggered by gas extraction have damaged infrastructure and raised concerns about seismic events affecting local residents' livelihoods. Accurate statistical modelling is crucial for assessing risks and guiding public policy. Our research leverages statistical methodologies to address key challenges in seismic risk assessment. We develop novel techniques to estimate the upper tail of earthquake magnitude distributions, incorporating expert knowledge of the maximum possible magnitude. This approach enhances the reliability of extreme magnitude predictions, which are vital for effective risk preparedness and mitigation. We also investigate the spatio-temporal occurrence of earthquakes using Epidemic-Type Aftershock Sequence (ETAS) models. Our approach improves upon existing methods by addressing challenges such as incorporating stress variations due to resource extraction, which allow us to better approximate the occurrence of earthquakes with intensity function varying spatially and temporally. Additionally, we enhance computational efficiency, making ETAS inference more robust for practical applications

Challenges of survival estimation for populations impacted by disease under conditions of declining data collection rates

Max Howell, Lancaster University

Integrated population models provide a framework to simultaneously analyse multiple types of data. They are well used in the ecological literature to enable estimation of parameters which may not be estimable from data types analysed in isolation. We will describe why it is important to make use of different types of data when trying to detect declines in survival probability, especially when data may be sparse due to declining detection rates. We will outline the possible types of data that can be collected, and the relative information contained in each. The work conducted is motivated by a case study of Green finches, Chloris chloris, which are experiencing a rapid decline in the UK, primarily driven by the spread of disease.

Multivariate extreme value models for environmental science

Kristina Bratkova, Lancaster University.

Extreme events such as a combination of high values of air pollutants exhibit multivariate dependence but current methods typically assume asymptotic dependence (largest values can occur together) or asymptotic independence (largest values cannot occur together), which significantly affects the estimated probabilities of these extreme events. Conditional extreme value models provide a versatile framework that does not make this assumption about asymptotic (in)dependence, but we face the curse of dimensionality due to the empirical distribution used for modelling the residuals. Therefore, we propose parametric methods for the margins and vine copula for the dependence structure of the residuals. The benefits of the proposed extensions are illustrated on a range of simulation examples and a five variable pollution dataset.

Extreme value theory for long-term risk assessment of offshore structures

Matt Speers, Lancaster University

Any ocean structure will have forces induced on it by the surrounding ocean environment. The modelling of the extremes of this environment is therefore crucial when performing structural risk analysis, as extreme environmental conditions will naturally induce the largest forces. When combined with physical models for the short-term environment and wave-structure interactions, extreme value models for the evolving ocean environment can be used to estimate structure failure probabilities.

Environmental contours are used as an alternative to assessing structure failure probabilities when full simulation from the above 'forward approach' isn't feasible. In one respect, contour methods are advantageous over the forward approach in that they characterise the environment only and so do not require modelling of the structural response. Combined with appropriate assumptions for the ocean-structure interaction, these can be used to assess any structure in that environment.

Unfortunately, the assumptions made by design contours are often unrealistic for certain structure types. Alternative techniques include Gaussian emulation and efficient sampling of draws from physical models. These methods select optimal values of environmental parameters at which to perform full-scale physical simulation, reducing the computational burden of structural analysis whilst preserving physical information.

We describe how to estimate the long-term behaviour of structure response over a period of 1000 years, induced by an environment for which we have less than 100 years of data. These estimates can then be used to evaluate long-term structural failure probabilities and thus assess the lifetime risk to the structure. We discuss differences in estimates obtained using the above approaches, assessing the suitability of each for different structure scenarios.

3. Contributed Sessions

CS11

Modeling spatial dependence through latent Gaussian models with spatial copulas

Brynjólfur Gauti Guðrúnar Jónsson, University of Iceland

This research explores an approach to modeling spatial dependence in extreme precipitation data through the integration of latent Gaussian models and spatial copulas using the Max-and-Smooth methodology. We implement a Matérn-like copula structure within a stepwise inference framework to address the computational challenges inherent in analyzing large spatial datasets.
Our approach combines Generalized Extreme Value (GEV) marginal distributions with a Gaussian copula transform, employing a precision matrix structure with Kronecker products that facilitates efficient computation when working with data on a regular grid. The methodology follows a carefully designed sequence: site-wise GEV parameter estimation, copula parameter optimization, joint estimation incorporating data-level spatial dependence, and Bayesian spatial smoothing of parameters.
Initial applications to UKCP precipitation projections across the UK, with and without a spatial copula, suggest this approach may offer practical advantages in certain contexts, including coherent spatial patterns in parameter estimates and computational feasibility for high-dimensional datasets. Comparisons of site-wise GEV parameter estimates with and without the spatial copula indicate that the inclusion of the spatial copula is important for proper uncertainty quantification.

A clustering framework for conditional extremes models

Patrick O'Toole, University of Bath

Co-authors: Christian Rohrbeck, Jordan Richards

Conditional extreme (CE) value models have proven useful for analysing the joint tail behaviour of random vectors. One drawback of these methods is that model estimates tend to be highly uncertain due to the inherent scarcity of extreme data. This motivates the development of clustering methods for this class of models; pooling similar within-cluster data drastically reduces parameter estimation uncertainty.
An extensive amount of work to estimate CE models exists in multivariate and spatial applications, but clustering for models of this type has not been explored. To this end, we introduce a clustering framework for conditional extremes models which provides a novel and principled, parametric methodology for summarising multivariate extremal dependence.
In a first step, we define a dissimilarity measure for CE models based on the Jensen-Shannon divergence. One key advantage of our measure is that it can be applied in arbitrary dimension and, as opposed to existing methods for clustering extremal dependence, is not restricted to the bivariate setting. Clustering is then performed by applying the k-medoids algorithm to our novel dissimilarity matrix, which collects the dissimilarity between all pairs of spatial sites.
A detailed simulation study shows our technique to be superior to the leading competitor in the bivariate case across a range of possible dependence structures and uniquely provides a tool for clustering in the multivariate extremal dependence setting. At the end of the talk, we apply our clustering framework to meteorological data from Ireland.

Modeling Nonstationary Extremal Dependence via Deep Spatial Deformations

Xuanjie Shao, KAUST

Co-authors: Jordan Richards, Raphael Huser

Modeling the nonstationarity that often prevails in extremal dependence of spatial data can be challenging, and typically requires bespoke or complex spatial models that are challenging to estimate. Inference for stationary and isotropic models is considerably easier, but the assumptions that underpin these models are rarely met by data observed over large or topographically-complex domains. A possible approach for accommodating nonstationarity in a classical spatial model is to warp the spatial domain to a latent space where stationarity and isotropy can be reasonably assumed. This approach has seen developments in both the classical Gaussian-based geostatistics and spatial extremes contexts. Although this approach is very flexible, estimation of the warping function can be computationally expensive and the transformation is not always guaranteed to be bijective, which can lead to physically unrealistic transformations when the domain folds onto itself. We overcome these challenges by developing deep compositional spatial models to capture nonstationarity in extremal dependence, thereby extending deep compositional Gaussian processes. Specifically, we focus on modeling high threshold exceedances of process functionals by leveraging efficient inference methods for limiting r-Pareto processes. A detailed high-dimensional simulation study demonstrates the superior performance of our model with respect to existing methods, in estimating the warped space, which leads to a more accurate characterization of the highly nonstationary extremal dependence structure. We illustrate our method by applying it to UK precipitation extremes and show that we can efficiently estimate their extremal dependence structure with data observed at thousands of locations. The model is programmed with the R language and tensorflow v2.

Approximate Bayesian inference for analysis of spatiotemporal flood frequency data

Birgir Hrafnkelsson, University of Iceland

Co-authors: Árni V. Jóhannesson, Stefan Siegert, Raphaël Huser, Haakon Bakka

Extreme floods cause casualties and widespread damage to property and vital civil infrastructure. Predictions of extreme floods, within gauged and ungauged catchments, are crucial to mitigate these disasters. A Bayesian framework is proposed for predicting extreme floods, using the generalized extreme-value (GEV) distribution. One of the methodological challenges involves finding a suitable parametrization for the GEV distribution when multiple covariates and/or latent spatial effects are involved and a time trend is present. Another challenge involves balancing model complexity and parsimony, using an appropriate model selection procedure. We propose a latent Gaussian modeling framework with a novel multivariate link function designed to separate the interpretation of the parameters at the latent level and to avoid unreasonable estimates of the shape and time trend parameters. Structured additive regression models, which include catchment descriptors as covariates and spatially correlated model components, are proposed for the four parameters at the latent level. To achieve computational efficiency with large datasets and richly parametrized models, we exploit Max-and-Smooth, an accurate and fast approximate Bayesian inference approach that can also be used to efficiently select models separately for each of the four regression models at the latent level. We applied our proposed methodology to annual peak river flow data from 554 catchments across the United Kingdom. The results show that the time trend is important. Posterior estimates of the

time trend parameters correspond to an average increase of about 1.5% per decade with range 0.1% to 2.8% and reveal a spatial structure across the United Kingdom.

CS12

Invisible Years Before Breeding: How Prey Dynamics and Population Pressure Influence Natal Dispersal in Tawny Owls

Seyngyeon Lee, University of Aberdeen

Coauthors: Xavier Lambin, Deon Roos, Chris Sutherland, Paul Caplat

Understanding how far individuals disperse during their first breeding attempt is critical for many species, particularly due to the high cost of dispersal. Dispersal distance is strongly influenced by food availability, which individuals tend to disperse farther when prey is abundant. Population density also plays a key role in settlement decisions, as high densities can lead to increased competition for mates and nesting vacancies. In long-lived territorial species, natal dispersal is shaped by a complex interplay of these factors. In this study, we utilize 40 years of long-term ringing data on Tawny Owls (Strix aluco) from Kielder Forest, Northumberland, to disentangle the interacting effects of prey abundance and population density on natal dispersal. Tawny Owls in this area primarily prey on field voles, and their breeding decisions are closely tied to the voles' 3–4-year abundance cycles. Although Tawny Owls are physiologically mature to breed at age one, many delay reproduction and remain non-territorial during the early years of life. We hypothesize that individuals who begin breeding at an older age, especially one's experienced periods of high population density, will disperse farther during increasing phases of vole abundance. Using capture–recapture models, we estimate the number of non-breeders (both territorial and non-territorial) and compare these dynamics with natal dispersal distances across different vole phases. This study offers new insights into how environmental cycles and population pressure interact to shape dispersal strategies in territorial raptors, contributing to a broader understanding of population dynamics and spatial structure in long-lived species.

A Bayesian approach to species distribution modeling with Presence-only data

Kabiru Abubakari, Queen Mary University of London

Co-authors: Silvia Liverani, Andrew Leitch and Ilia Leitch

Records of species data available on online portals, such as the Global Biodiversity Information Facility, are collected through the citizen science program. This kind of data is often referred to as presence-only data since only presence records are available. Species distribution models (SDMs) require presence-absence data. In previous studies, SDMs were developed using background data. This approach tends to favor nonparametric models intended for prediction. Thus, the effect of the parameters of interest in the SDMs is ignored.
We propose a Bayesian approach to modeling species distribution with R-INLA. We incorporate uncertainty about the absence of species in places without records using the combination of missing data imputation and Bayesian model averaging. We recognize that misclassification can attenuate the estimated parameters. So an adjusted logit link function is used to correct the effect of measurement error on the estimated parameters. We present results for the parameters of interest using simulated and real data. Our approach performs better than the alternative parametric method and achieves satisfactory predictive accuracy.

Statistical inference for velocity-jump models in movement ecology

Paul Blackwell, University of Sheffield

Animal movement data are often characterised in terms of the step lengths and turning angles between observations. While sometimes a convenient summary, this does not directly correspond to a model of the actual movement process. An alternative is to model a path as a sequence of straight-line segments, with turns between them at times that do not necessarily match the observation times. This kind of continuous-time step-and-turn or velocity-jump model poses challenging problems for statistical inference, especially when the timescale of observations is close to that of the turns. I will describe a statistical approach to reconstructing such trajectories and simultaneously estimating the parameters of the underlying movement process, and illustrate it with some real applications. I will also discuss the connection between such models and so-called bouncy particle particle samplers in continuous-time Markov chain Monte Carlo algorithms. This connection gives one way of developing step-and-turn models that incorporate resource selection in a coherent way.

Comparing ecosystem models using dynamic distances

Matthew Spencer, School of Environmental Sciences, University of Liverpool,

Co-authors: Ned Wontner

Methods for quantitative comparison of ecosystems include those based on controversial theories about energy flow, and those based on more or less theory-free calculation of dissimilarities in observed relative abundances. Here, we develop a mathematical framework for comparing the dynamical properties of ecosystem models, using ecological theory and ideas from functional analysis and topology. We apply the framework to an experimental study of competition between polyps of the moon jellyfish Aurelia aurita and other marine sessile organisms. Four differential equation models, representing different assumptions about the nature of the interaction between jellyfish polyps and other organisms, were fitted using a Bayesian approach to data from communities developing on settlement panels. We show that one of the four models has relatively distinct dynamics from the others. This is consistent with differences in fit to experimental data, but measures different properties of the models. Our approach is applicable even when models have different sets of species, and yields topological results showing that under reasonable axioms, there is only one kind of ecosystem.

CS13

Statistical Analysis of Ocean Turbulence

Max Coten, University of Western Australia

Co-authors: Lachlan Astfalck

Large-scale ocean behaviour is commonly represented with deterministic physics-based numerical models, discretised on a coarse spatiotemporal grid. Whilst this captures certain large-scale oceanographic features, we know that sub-grid processes such as ocean turbulence are also major contributors in the spatiotemporal distribution of important ocean tracers such as temperature, sediment, and nutrients. Sub-grid features are naturally modelled with stochastic boundary conditions, and to best represent the oceanic forcings, we are required to accurately quantify the statistical properties of the sub-grid process.

Ocean turbulence is known to be highly left-skewed such that the majority of turbulent energy is generated by relatively rare events. Furthermore, turbulence is correlated in time and space; this is rarely acknowledged in any statistical modelling in the oceanographic literature. A sensible stochastic representation of ocean turbulence thus requires appropriate representation of the marginal distribution and time-dependency of the turbulent process.
This project builds a stochastic representation of ocean turbulence using data from a recent offshore experiment, conducted by oceanographers at UWA. The dataset comprises of high-frequency measurements of boundary turbulence over a six-week campaign and is singular in its fidelity and length of record. We jointly capture the marginal and temporal structure of the turbulent stochastic processes. This project is conducted alongside the Ocean Dynamics group at UWA and is being used to (1) inform data-collection methodology for an upcoming voyage to the Ningaloo Reef and (2) inform parameterisations of stochastic boundary conditions in numeric ocean models.

Leveraging massive opportunistically collected datasets to study species communities in space and time

Maxime Fajgenblat, KU Leuven / UHasselt

Co-authors: Robby Wijns, Geert De Knijf, Robby Stoks, Pieter Lemmens, Marc Herremans, Pieter Vanormelingen, Luc De Meester & Thomas Neyens

Online portals have facilitated collecting extensive biodiversity data by naturalists, offering unprecedented coverage and resolution in space and time. Despite being the most widely available class of biodiversity data, opportunistically collected records have remained largely inaccessible to community ecologists since the imperfect and highly heterogeneous detection process can severely bias inference. We present a novel statistical approach that leverages these datasets by embedding a spatiotemporal joint species distribution model within a flexible site-occupancy framework. Our model addresses variable detection probabilities across visits and species by modelling phenological patterns and by extending the use of latent variables to characterise observer-specific detection and reporting behaviour. We apply our model to an opportunistically collected dataset on lentic odonates, encompassing over 100,000 waterbody visits in Flanders (N-Belgium), to show that the model provides insights into biological communities at high resolution, including phenology, interannual trends, environmental associations and spatiotemporal co-distributional patterns in community composition.

Incorporating Crowdsourced Data to Improve Wind Speed Modelling

Eamonn Organ, University of Limerick

Co-authors: James Sweeney

Access to accurate high resolution spatial and temporal wind speed data is critical for estimating wind energy resource potential. For real-time prediction, statistical or hybrid models that combine numerical weather prediction (NWP) outputs with statistical methods depend on high-quality real-time or near-real-time data to improve forecasting accuracy.
Personal weather stations (PWS), which allows for real-time data with broader spatial coverage, offers a potential additional data source. However, unlike official meteorological stations, PWS data are not subject to rigorous quality control and may exhibit bias or increased measurement errors. This talk addresses the challenges of incorporating PWS data into prospective models in two ways.

First, a bias correction is performed on PWS wind speed data by leveraging reanalysis data to correct errors. Second, we apply a flexible Bayesian hierarchical spatio-temporal statistical model that allows for varying levels of measurement error at personal weather stations to mitigate the influence of noisy data. The proposed model is applied to an Irish wind dataset, combining crowdsourced wind data with official meteorological station data to enhance prediction at unobserved locations. Results highlight that the inclusion of the bias corrected personal weather stations leads to a significant improvement in accuracy compared to using official weather stations only, with an 11% reduction in prediction error across all sites. The results are comparable with popular reanalysis products, but unlike these numerical models the proposed modelling framework is available in real-time and has improved uncertainty quantification characteristics., CS13

Bayesian analysis of weighted regression models for spatiotemporal data and its application to PM2.5 mapping using satellite-derived data

Yovna Junglee, Department of Statistical Sciences, University of Toronto

Co-authors: Meredith Franklin, Vianey Leos Barajas

Spatiotemporal statistical methods are commonly used to analyze data that vary across both space and time, typically assuming that the response variable depends on predictors observed at the same location and time. However, traditional approaches face challenges such as spatiotemporal misalignment and missing data, issues that are particularly pronounced in environmental science. Moreover, these methods generally overlook the potential influence of predictors from both nearby locations and preceding time points, which may carry important information about spatiotemporal dependencies.

Approaches that have been developed to address these limitations include weighted regression models that assign weights to covariates based on their spatial and temporal distances. These models provide a flexible framework to handle spatiotemporal misalignment, but their robustness to missing data and sensitivity to parameter choices remains insufficiently understood. We thus propose a Bayesian framework to evaluate how spatiotemporal predictors influence the response process, enabling formal uncertainty quantification of parameters that govern these effects. Through extensive simulation studies, we assess the method's effectiveness to recover the true data-generating process under different missing data mechanisms and its sensitivity to parameter choices.

This study is motivated by the challenges of mapping ground-level PM2.5 concentrations using satellite-derived aerosol optical depth (AOD). Frequent cloud coverage leads to missing satellite data near ground-level monitoring stations, making it difficult to model relationships between AOD and PM2.5. Our approach explores whether incorporating information in a structured manner rooted in spatiotemporal statistics from locations over a broader spatial extent and previous time points, enhances spatial mapping.

CS21

A novel statistical approach for designing representative NCEA river flow and temperature monitoring networks

Cameron Bullen, APEM Ltd.

Co-authors: Alex Lumsdon, Mike Summers, Eleanor D'Arcy, Brittany Heap, Alberto Scotti, Andrew Davey

As part of the Defra funded Natural Capital and Ecosystem Assessment (NCEA) programme, the Environment Agency is developing dedicated monitoring networks for both river flow and temperature. The ambition for these networks is to yield unbiased and precise estimates of long-term trends at a national scale, using only a subset of the Environment Agency's existing hydrometric stations. However, the existing hydrometric stations are non-randomly distributed and are not representative of the target population of English rivers, making it challenging to design a monitoring network capable of estimating accurate national trends.

In order to overcome this challenge, APEM and the Environment Agency developed a statistical network design approach using logistic regression models to calculate inclusion probabilities, and stepwise selection of stations to optimise the monitoring networks. The use of Gini coefficients to quantitatively assess representation enabled comparison between alternative designs. Despite the constraints on available stations, this novel design process allowed the NCEA flow and temperature monitoring networks to achieve maximum representation, and combined with post-stratification enables improved trend estimates for the target population of English rivers. Combining this clear statistical framework for station selection with expert review of the candidate stations also made it possible to balance the goal of representation with other desirable characteristics such as length of record, co-location with other monitoring programs, and data quality. This network design process can be extended to optimise the process of establishing new or replacement monitoring stations, designing 'bolt-on' networks to characterise trends in priority river types, and could be applied to a range of other monitoring programs.

Marine Environmental Monitoring

Jon Barry, Cefas

I briefly describe some current marine environmental projects that I and the Centre for Fisheries and Aquatic Science (CEFAS) are involved in. I then give more details on two specific projects.

The first is monitoring and mapping of carbon stocks in the seabed and our attempts to generate a carbon map for the UK as part of the Natural Capital and Ecosystem Assessment programme (described by one of the keynote speakers at this conference).

The second project is monitoring of seafloor litter. I outline previous attempts at seafloor litter assessments and some of the current statistical decisions and dilemmas being faced to both monitor seafloor litter and to set compliance thresholds.

Understanding diffuse environmental effects on hydrology

Thomas Cornulier, BioSS - Biomathematics & Statistics Scotland

Co-authors: Dave L. Miller, Zhou Fang

Predicting environmental and ecological processes and patterns in rivers is complex, due to the interplay of processes operating at multiple spatial and temporal scales. Indeed, drivers originating in the river catchment or beyond, may strongly interact with hydrological or ecological

processes operating during transport towards and along the river network, combining both short and long timescales.

However, we still lack statistical models, that can capture such rich interactions at multiple spatial and temporal scales, while considering the specific constraints imposed by network topology and flow.

Signal (vector on scalar) regression provides a helpful framework for identifying the spatial or temporal extent of the ""zone of influence"" of predictors on environmental and ecological observations, by estimating the joint contribution of predictors measured at different spatial and temporal lags. We show how this general framework can help address questions about complex spatio-temporal lagged processes, including linear and non-linear or lagged interactions, both in terrestrial applications and, with the addition of suitable constraints, in a hydrological context.

CS31

Producing robust abundance trends using sparse long-term citizen-science data

James Clarke, British Trust for Ornithology

Co-authors: Philipp Boersch-Supan, Nina O'Hanlon, Niall Burton, Sarah Harris, Dawn Balmer

Long-term ecological monitoring is important for determining, for example, temporal trends in abundance for species to allow for well-informed policy decisions and conservation actions. One such group subject to long-term monitoring in the UK are seabirds. This group is facing multiple challenges such as the effects of climate change and avian flu, making the production of accurate abundance trends incredibly important. The Seabird Monitoring Programme (SMP), established in 1986, aims at obtaining abundance trends for seabirds by collecting counts at their colonies throughout the UK.

However, there are numerous challenges to analysing SMP data, with a particular aspect of concern being the sparsity of the data. Many seabird colonies are only intermittently and irregularly visited making the determination of colony level trends over time particularly difficult. To obtain an estimate of the total UK population within each year the missing data are currently imputed based on sparse available count data. This is problematic because high levels of imputed counts can create bias and overconfidence in subsequently calculated trends. This talk will discuss the model-based solutions reached to produce more robust long-term abundance trends for UK seabird species, namely using Hierarchical Generalized Additive Models and state-space models, which use information from well-monitored colonies to influence trends at colonies with poor coverage. Additional challenges encountered when analysing such sparse count data will also be discussed, such as with the production of confidence intervals.

Estimating camera malfunction times from detection histories.

Milly Jones, University of Kent

Co-authors: Eleni Matechou, Diana Cole, Nicolas Deere

Camera trap surveys are a popular and efficient method of surveying an environment. Multiple cameras are set up in an array in the study area, and take an image when an individual passes by. The camera detection histories are informative about which species were present at a given time-point and location. However, cameras within the array can malfunction in several ways. The cameras may break or run out of battery, and some cameras may malfunction and continuously take images despite no individuals being present. For long term camera deployments, records about when cameras malfunction and are repaired are not often kept. As such it can be difficult to analyse detection histories when it is not known when cameras are operating normally. The detection histories contain time-stamps for each image taken by a camera, and images can be 'marked' if they indicate the presence of an individual. We model these data using a Markov Modulated Mark Poisson Process (MMMPP) in a Bayesian framework. The MMMPP models the latent camera state as a continuous time Markov Process, and then detection rates and mark distributions conditional on the underlying state of the camera. Given the posterior distributions for model parameters, we then estimate for each camera the probability that it was operational, non-operational, or mis-firing at each time-point throughout the survey period. This provides a flexible framework for camera trap practitioners to estimate the malfunction times of cameras during the survey.

The statistical challenges in tackling persistent climate model uncertainty through model-observation comparisons.

Jill Johnson, School of Mathematical and Physical Sciences, University of Sheffield

Co-authors: Jonathan Owen, Iain Webb, Jeremy Oakley (MPS, University of Sheffield); Leighton Regayre, Kunal Ghosh, Lea Prevost, Ken Carslaw (ICAS, University of Leeds)

The effects of aerosols on the Earth's energy balance since pre-industrial times (aerosol radiative forcing) has significantly and repeatedly dominated the uncertainty in reported estimates of global temperature change from IPCC. The magnitude of aerosol radiative forcing of climate over the industrial period is estimated to lie between about -2 and -0.4 W m-2, compared to a much better understood forcing of 1.6 to 2.0 W m-2 due to CO2.

Here we will discuss past efforts to quantify the range of possible aerosol forcings predicted from an aerosol-climate model that are caused by parametric uncertainty, and to constrain that forcing uncertainty through model-observation comparison using extensive aerosol and cloud-based measurements from ships, flight campaigns, satellites and ground stations. We find that despite a very large reduction in plausible parameter space and reasonable constraint on observable properties, the observational constraint based on this comprehensive set of measurements only partially reduces the range of aerosol radiative forcings from our model.

In a new project called 'Towards Maximum Feasible Reduction in Aerosol Forcing Uncertainty' (Aerosol-MFR), we are addressing several key statistical challenges highlighted from this work in order to improve our model-observation comparison process for uncertainty constraint. This includes optimising the way observational constraints are applied, designing new approaches for reducing error compensation effects and using the PPE to identify and characterise model structural errors. Here we will explore preliminary results from the project so far and further plans to tackle this important problem.

Bayesian Emulation and History Matching to Constrain Aerosol Radiative Forcing Uncertainty

Jonathan Owen, School of Mathematical and Physical Sciences, University of Sheffield

Co-authors: Jill Johnson, Iain Webb, Jeremy Oakley, Leighton Regayre, Kunal Ghosh, Lea Prevost, Ken Carslaw

Computer models are used extensively in the study of environmental and ecological systems to improve knowledge of the underlying processes and predict their future evolution. An essential application in climate modelling is to study drivers of climate change; quantify uncertainties in future climate projections; and guide policy decisions. However, their direct use is inhibited by: their complex structure; high-dimensional input and output spaces, including spatial-temporal fields; and their long evaluation times, often taking weeks to months on high-performance computers.

Earth System Models (ESMs) are computer model representations integrating atmosphere, ocean, land, ice, and biosphere. They are used to investigate interactions of both natural and anthropogenic aerosol emissions with clouds which yield large Aerosol Radiative Forcing (ARF; temporal change in Earth's energy balance due to aerosols) induced uncertainty in historical climate change, hence are important for predicting future climate. ARF is unobservable with previous research yielding little reduction in its uncertainty within 30-years of IPCC reports.

We employ history matching to robustly constrain ARF for the Met Office's UKESM1 model using satellite and field campaign measurements. Bayesian emulators are employed as fast statistical approximations for a carefully selected set of observable model outputs enabling an efficient global parameter search. These are embedded within an uncertainty quantification framework incorporating structural model discrepancy linking ESM output and the real-world, as well as representation and observation errors, to meaningfully compare observations with the model. History matching results are used to resolve parametric uncertainty and identify sources of structural deficiencies to aid further ESM development.

CS32

Causal discovery in ecology

Rebecca Supple, Centre for Research into Ecological & Environmental Modelling, University of St Andrews

Co-authors: Hannah Worthington, Ben Swallow

High dimensionality largely prevents the use of experiments to understand the underlying causes and mechanisms of biodiversity change. To identify key points of intervention in preventing biodiversity loss, we must find innovative ways to derive causal information from observations. Causal discovery algorithms are designed to uncover causal relationships from observational data, whilst simultaneously accounting for confounding variables that may mask or distort causal effects. If adapted to the spatiotemporal structure of much ecological data, causal discovery algorithms could prove powerful tools for ecological model development and forecasting.

In this talk, we explain the theoretical underpinnings of causal discovery algorithms as they are used in fields such as machine learning and epidemiology. We then present our adapted algorithm for spatiotemporal data and its performance on simulated data. Finally, we discuss the strengths and challenges of an application to systems of landcover, agricultural impact, climate, and bird abundance

data across Europe from 1992-2016. Our work demonstrates potential applications of novel causal statistical methods in ecology.

Multispecies sites occupancy modeling with replicate to analyze eDNA data

Marwah Soliman, NTNU, Trondheim, Norway

Co-authors: Robert Brian O'Hara, Bert van der Veen, Glenn John Dunshea

Environmental DNA (eDNA) has emerged as a powerful tool for biodiversity monitoring. The composition of the DNA in environmental mediums varies due organism presence/absence and organism- and environment-specific rates of DNA shedding, transport and degradation, as well as laboratory processes. Repeated sampling at locations, and repeated technical replicates within samples are thus required to account for biological- and laboratory-derived variation. Thus, analytical approaches for eDNA data should accommodate hierarchical repeated sampling designs that account for biological and technical replication, but few such approaches exist.

We extended a hierarchical modelling framework to determine multi-species occurrence at the site-level, from repeated technical replicates. The framework leverages developments in zero-inflated modelling to account for imperfect detection of DNA and repeated observations of taxonomic units at sites refine estimates of their true occurrence and its uncertainty. When all replicates of a taxonomic unit at a site are non-zero, it is naturally predicted to occur at a site. However, when all replicate counts are zero, taxonomic units can still be predicted to occur. As such, zeros in observations can be generated either by the sampling process due to imperfect detection, or by the ecological process of species absence.

Real-world marine eDNA datasets were used to demonstrate that these hierarchical models can greatly improve inference from eDNA metabarcoding data. This method also provides an analytical approach for incorporating eDNA-based detections into species' distribution modelling frameworks, contributing to the advancement of robust eDNA-based ecological monitoring approaches useful for biodiversity conservation and management.

A parametric model to explain isotopic richness

Alcaraz Angelo, LMBA, Université Bretagne Sub (UBS) , Vannes.

Co-authors: Durrieu Gilles, Poterie Audrey

Computation of the diet diversity of certain animal species can be based on the determination and study of carbon and nitrogen isotopic signatures. This estimation of trophic diversity has its methodological origins in quantifying functional diversity, which has given rise to key concepts such as trait probability densities. Using this formalism, we propose a parametric model based on quantile regression to explain the variability of isotopic signatures. We choose to lose precision in estimating the trait probability density compared with the non-parametric approaches conventionally used to explicitly link our variables of interest to environmental or biotopic covariates. A robust measure of trophic diversity can also be proposed based on our model. Trophic diversity, like behavioral diversity in general, is linked to adaptation to environmental change, and identifying generalist (or specialist) species is a conservation issue. Quantifying the effect of environmental or biotopic parameters on

functional diversity could help identify the most crucial variables to take into account in conservation strategies.

Distributed lag non-linear models with Laplacian-P-splines for analysis of spatially structured time series

Sara Rutten, Hasselt University

Co-authors: Bryan Sumalinab, Oswaldo Gressani, Thomas Neyens, Elisa Duarte, Niel Hens, Christel Faes

Modelling nonlinear lagged relationships is important in environmental and ecological sciences, as variables often have delayed and complex effects over time. Distributed lag non-linear models (DLNMs) are widely used to capture these dynamics, from climate impacts on health to ecosystem resilience. However, applying DLNMs to spatially referenced data might introduce spatial dependence, increasing model complexity and computational demands, especially for large spatio-temporal datasets.

We propose a Bayesian DLNM-Laplacian P-splines (DLNM-LPS) approach that integrates spatial dependence using conditional autoregressive (CAR) priors within a penalized DLNM framework. This combination enhances the accuracy and robustness of modelling nonlinear and lagged effects through P-splines, offering a more effective representation compared to traditional B-splines or natural splines, while also addressing spatial variability. Laplace approximation is used to approximate the conditional posterior distribution of the regression parameters.

Simulations demonstrate that our approach achieves good coverage and low RMSE values overall. Moreover, the methodology is applied to analyse the temperature-mortality relationship in London, showcasing its practical application.

CS33

Modelling dynamics in soil moisture time series with an extended Bayesian changepoint detection algorithm

Mengyi Gong, Lancaster University

Co-authors: Prof Rebecca Killick, Prof Christopher Nemeth

Continuous monitoring of soil moisture content via modern sensors has opened up a unique opportunity to investigate the dynamics of soil health. The abundance of data and the complexity in the time series data requires the development of statistical methodologies to analyse the data appropriately. For example, the soil moisture time series can display an exponential decay pattern when the soil is relatively dry and a fluctuating pattern when the soil is saturated.

In this study, we proposed an extension to the Bayesian online changepoint detection (BOCPD) method to model soil moisture time series with complex patterns. In particular, the proposed method segments the time series data and assigns each segments with potentially different types models in order to better describe the patterns in data. It combines sequential Monte Carlo or online gradient ascent with the classic BOCPD to estimate the key unknown parameters in the models. Simulation study was carried out to compare the performance of the proposed method to the classic BOCPD

method. The possibilities of reducing computational cost has been investigated. The method was applied to the soil moisture time series from the US National Ecological Observatory Network and the field study conducted in the Yorkshire Dale.

A Generalized Cross Entropy Model for soil biological quality

Anna Simonetto, Università degli Studi di Brescia

Co-authors: Gianni Gilioli, Isabella Ghiglieno, Marco Sandri, Maurizio Carpita, Enrico Ciavolino, Mario Angelelli

Soil is a living and complex ecosystem inhabited by a diversity of organisms playing crucial roles in key ecosystem services, e.g., soil fertility, carbon cycle, water quality. Arthropods are a fundamental component of the dense edaphic trophic network, contributing significantly to organic matter decomposition. The Soil Biological Quality–Arthropods (QBS-ar) index is widely used to assess edaphic biodiversity by quantifying specific arthropod taxa and their biological forms. The QBS-ar index can be used to identify the main drivers of edaphic biodiversity, which is essential for developing management strategies to preserve and enhance soil biological quality.

This study investigates edaphic arthropod communities in several Italian vineyards. We apply a regression model to evaluate the response of QBS-ar to different soil types, weather conditions, and agronomic management practices. Since the initial dataset presents some critical issues, including multicollinearity among potential regressors, it is necessary to adopt statistical tools capable of addressing these issues. Entropy-based regression methods provide a robust alternative to classical techniques, particularly in ill-posed problems, multicollinearity, or limited data contexts. These methods do not rely on distributional assumptions for errors and offer built-in regularization, leading to stable estimates. Generalized Cross Entropy (GCE) allows for incorporating prior information, making it useful in Bayesian-like frameworks. The analysis is conducted using GCEmodels, an R package under development that implements maximum entropy-based regression approaches. GCEmodels is open-source and available on GitHub at https://github.com/GCEmodels/GCEmodels.

Modelling topsoil organic carbon declines under ash dieback using a Bayesian beta mixture modelling approach

Fiona Seaton, UK Centre for Ecology & Hydrology

Co-authors: David Robinson, Claire Wood, Clare Benskin, Rebecca Rowe, Karen Hornigold, Keith Kirby, Chris Nichols, Simon Smart

Tree diseases are increasingly affecting woodland ecosystems across the world. However, the impact of these diseases upon the soil, and in particular soil carbon, are still poorly understood. Here we present the results of a field survey of ~100 woodlands across Great Britain measured in 1971, 2001, and 2022 and evaluate the fifty year trend in topsoil (0-15cm) carbon based upon measurements of soil organic matter (SOM) and the impact of Hymenoscyphus fraxineus (ash dieback). To better represent the full SOM distribution, including the extremely high SOM measurements, we adopt a Beta mixture modelling approach within a Bayesian framework in Stan. Across all woodlands, comprising ~1500 plots per survey, average SOM remained constant across the fifty years time series. However, the 311 plots with ash dieback experienced a decline in SOM in the most recent survey while it remained constant under the 328 plots with ash trees present but no dieback recorded. This resulted

in plots with ash dieback having a modelled mean SOM of 12.2% compared to 13.4% in plots without ash dieback. Ash dieback was more likely to be recorded in plots that had higher soil pH pre-ash dieback invasion, but the decline in SOM under ash dieback was not explained by changes in soil pH, or changes in the ground flora composition. Our results show the importance of understanding the impacts of tree disease when considering current and future woodland carbon dynamics.

Understanding tree water usage and stress via sap flux density and dendrometer time series

Rebecca Killick, Lancaster University

Co-authors: Mengyi Gong, Andrew Hirons

Understanding the water usage patterns of trees and their responses to drought or flood are crucial to the irrigation management of tree nurseries, in order to produce high-quality trees and provide resilience to an increasingly warmer and drier climate. Modern sensors provide an opportunity to monitor the behaviour of trees at a higher frequency in continuous time. In Hillier nursery in the south of England, sap flux density time series and dendrometer time series were collected using sensors installed on different species of tree over the last three growing seasons.

Using these time series data and related weather data, such as vapour pressure deficit, solar radiation and air temperature, we are able explore the relationships between sap flux density, dendrometer and weather patterns in greater detail. In particular, we conducted a changepoint analysis to investigate the stress responses of trees under extreme weather conditions, and we developed an ensemble prediction approach based on additive models to forecast the sap flux density time series and hence estimate the water usage of trees. Various aspects that can affect the sap flux density, such as the impact of water-logging and heatwaves, were discussed. Some of the analyses have been integrated into an online monitoring dashboard, with the aim of increasing water-use-efficiency and informing strategies to manage periods of water deficit.

CS41

More data lead to worse predictions of seabird population change

Jazz Rhoades, British Trust for Ornithology (BTO)

Co-authors: Sophie Bennett (joint first author), Jacob Davies, Elizabeth M. Humphreys, Andrew Upton, Philipp Boersch-Supan

Quantitative ecology is a data-hungry discipline, and we usually expect more data to result in better model performance. In 2021, Davies et al. utilised data from two seabird censuses, the Seabird Colony Register (1985-1988) and Seabird 2000 (1998-2002), to predict changes in seabird presence and abundance in Great Britain and Ireland under future climate change. An INLA framework was used alongside spatio-temporal climatic and spatial oceanographic variables. Although the models were generally poor at predicting the magnitude of observed changes in abundance between the censuses, the direction of change was correctly predicted for the majority of species. We updated the modelling conducted by Davies et al. (2021) to

incorporate the most recent census (Seabirds Count 2015-2021), alongside new oceanographic data. The updated model benefited from one additional timestep of census data and inclusion of spatio-temporal (as opposed to only spatial) oceanographic variation. However, the updated model had poorer performance at predicting change in abundance over time, and the fitted covariate relationships had greater uncertainty. This was contrary to our original expectation that more data would result in improved model performance. In the work we present, we explore several possible reasons for this poorer model performance: 1) the fitted relationships in the original study may have been spurious; 2) spatial and temporal variation in oceanography may have different effects on seabirds; and 3) the relationship between seabird presence/abundance and oceanography may not be constant through time. Our findings highlight that caution is required when using space-for-time substitution for predicting seabird presence/abundance.

Modelling spatiotemporal trends in Belgian spider communities by temporally disaggregating pitfall trap data through interval-integrated basis functions

Thomas Neyens, Hasselt University & KU Leuven

Co-authors: Maxime Fajgenblat

Data collection practices often yield records with temporal resolutions that are insufficient to effectively model mechanisms of interest. For example, pitfall traps constitute the most frequently used technique to study and monitor invertebrate populations across the globe. They typically yield temporally aggregated count data of multiple species. Analysing such data can be challenging as the number of trapped individuals is only available at the resolution of the sampling interval (often multiple days or weeks), preventing straightforward inference at a different resolution and impeding the combined analysis of data collected under temporally misaligned sampling intervals. Importantly, seasonal patterns in species' activities (i.e. the phenology of species) further complicate the analysis of such data as they interact non-additively with the sampling interval duration. We introduce a temporal disaggregation approach using B-splines to spatiotemporally analyse data collected over misaligned sampling intervals while accounting for phenological influences. We present a computationally efficient approach to approximate the integral under the latent phenological curve throughout the sampling interval. We demonstrate that our method outperforms naive approaches and provides improved inference on phenology and other parameters of interest. We present a spatiotemporal joint species distribution model to model invertebrate communities through time and space using temporally aggregated data. As a case study, we apply this model to a large pitfall trapping dataset, spanning almost 50 years and over 10,000 trapping events in the Belgian province of Limburg, to infer phenological, spatiotemporal and co-distributional patterns for over 300 spider species.

Spatial-Temporal models for investigating solar potential in Ireland using a range of data sources

Maeve Upton, University of Limerick

Co-authors: Eamonn Organ and James Sweeney

In today's changing climate, one of the primary challenges facing Ireland is achieving the goal of decarbonising the energy grid by 2050. Increasing the capacity for renewable energy sources,

such as harnessing solar power, is playing a crucial role in this transition. As of December 2024, nearly 100,000 Irish homes had installed solar panels. However, the power generation capacity of renewables is highly variable. To maximise the integration of additional renewables into the grid, accurate real-time power generation forecasts are essential.

As a result, there is a need for high-resolution quantification of the current solar potential in Ireland. Met Éireann data offers valuable insights into varying solar irradiance across multiple Irish weather stations. However, the spatial distribution of these stations is limited. To address this, we have collaborated with industry partners and homeowners to source additional data, using solar panels as proxy meteorological stations. By employing a Bayesian framework, we can account for the additional uncertainty associated with this solar panel data.

By incorporating these data sources, our objective is to develop a range of data-driven Bayesian spatio-temporal statistical models for solar resources at 15-minute level, achieving refined spatial resolution nationwide. We present a range of models including a Gaussian Process, an autoregressive based approach and a Generalised Additive Model. We carry out a rigorous model comparison using techniques such as cross-validation. Through our interactive R Shiny app, we aim to forecast potential solar production, offering insights not only for solar panels at new locations but also for predicting solar output hours or days ahead.

CS51

Discovering the hidden sources of particulate matter

Israel Martinez Hernandez, Lancaster University

Particulate matter (PM) is well known to be detrimental to health, and it is crucial to apportion PM into the underlying sources to target policies. Particle number size distribution (PNSD) is the most accessible data to identify these sources, providing information on the PM sizes. However, due to high dimensionality and complex correlation structure, traditional methods usually require aggregating data (aggregation over time and over the different sizes of particles). The latter leads to a loss of information and increases the uncertainty of the different source estimates. In this talk, I will present a new method with a functional data approach, namely, a dynamic functional factor model. This new model allows us to disentangle the PM into sources and contributions while considering the complex dependencies of the data across different sizes and periods, and it is designed to handle large datasets. The outputs of this functional model are (1) the source profile present in the PM data and the time series contribution of each source to the total PM, which help to understand the dynamic of each source. The data analysed is hourly PNSD measurements collected at an urban background monitoring station located in a residential area of London. The data represents seven years and contains different particle sizes, ranging from 16.55 to 604.4 nm (fine and ultrafine particles). Six well-defined sources were found.

A step into the unknown - improving the characterisation of long-term water quality trends in England using modern statistical tools.

Michael Dunbar, Environment Agency (England)

Co-authors: Tullus Bergmann, Rebecca Gray, Tania Iglesias, James Delaney, Hugh Potter, Lindsey Sturdy

Governments typically have an interest in broad-scale environmental monitoring programmes to understand exposure, gauge the success of policy and to warn of emerging risks. The Environment Agency of England monitors concentrations of dissolved nutrients, metals and other chemicals in rivers and estuaries for many purposes including national reporting. In the past, our reporting of long-term trends has been hampered by changes to our monitoring programmes and the recording of some observations as being below a limit of detection (LoD).

These issues are relatively straightforward to resolve using modern statistical tools implemented in R scripts, however these tools do not have a history of application in regulatory bodies. Without such tools, reporting is limited to site-by-site assessments without a consideration of uncertainty and which are hampered by substitution of <LoD values.

As a case study in the application of modern tools we will outline the reporting undertaken for the H4 and B3 indicators for the Department of Environment Food and Rural Affairs (Defra) 25 year plan.

The approach can be summarised as follows:

1. Use R scripts to directly extract and process data; avoid creating ephemeral shapshots as part of the extraction process;

2. Individual monitoring sites are identified in the model structure as random intercepts;

3. (generalised) additive models (R-mgcv package) are used to describe potential non-linearities in an efficient manner;

4. Concentrations of most determinands typically exhibit strong seasonality which we decompose into trend and seasonal components;

5. The model can be used to produce trends for regions or individual sites. Optimal site-level predictions can be derived, even for sites with minimal data;

Chemical concentrations in water typically follow a lognormal distribution, in the absence of a censored lognormal distribution family we log-transform our raw data and treat as interval censored normal.

Analyses for both projects have indicated strong long-term downward trends in nutrients and metals and a more mixed pictured for organic chemicals which have only been monitored in recent years.

An additional benefit of undertaking data extraction and analysis in R scripts is the creation of a reproducible analytical pipeline which can be repurposed for other tasks. We'll conclude with some observations on the introduction of such a 'statistical' approach into a regulatory body.

Navigating data scarcity and surplus: An analytics framework for national water quality assessment

Ionut Paun, University of Glasgow

Co-authors: Marian Scott, Claire Miller, Ruth O'Donnell, Stephen Thackeray, Don Monteith, Eleanor Mackay, Philip Taylor, Erica Zaya

Recently, our rivers have been impacted by complex mixtures of contaminants including pharmaceuticals, pesticides, and microplastics, leading to declining water quality levels. The

MOT4Rivers project (UKRI-NERC, NE/X01620X/1), aims to explore how prolonged exposure to mixtures of pollutants are impacting freshwater ecosystems.

As part of this project, we examine national chemical datasets across the UK with the aim of assessing freshwater quality at the national scale. The scale and complexity of the data present several challenges for data analysis, as the datasets comprise 1.6 million observations, covering 94 priority contaminants of varying types, monitored at numerous sites with differing sampling frequencies. Collected over the past two decades, these datasets are often sparse and exhibit spatial and temporal misalignment.

An analytics pipeline, addressing these challenges for such national scale modelling is proposed. First, we standardise the temporal and spatial resolution of chemical and biological data into uniform reporting units across the UK using temporal aggregation and spatial matching to river reaches. Second, we implement automatic quality assurance, including handling anomalies, censored observations, and filling gaps using multivariate chained imputation. Finally, we apply dimensionality reduction to contaminants using a generalised Principal Component Analysis (PCA) approach. This presentation will showcase the results from each of these steps, along with initial findings from the modelling.

CS52

Bayesian time-varying autoregressive models for aphid abundance data

John Addy, Biomathematics Statistics Scotland

Co-authors: Dave Miller and Katharine Preedy

Time-Varying Autoregressive (TVAR) processes are powerful functions which assess how the autoregression of a time-series changes over time. Usually fitted as a non-parametric smoothed function of time, TVAR processes provide insight into the stability of an autoregressive process and assess if there has been a change in serial correlation of the time-series. For example, periods of increased or decreased autocorrelation usually reflect periods of growth or decline in the time-series. Along with TVAR processes, Generalised Additive Model for Location, Shape, and Scale (GAMLSS) provide a flexible modelling procedure which relaxes model assumptions and allows parameters to be modelled directly as smoothed non-parametric functions. Incorporating TVAR processes with the GAMLSS framework provides a complementary model which models the TVAR process of discrete data while also modelling the overdispersion parameter of a Negative Binomial distribution as a function of time.

To better understand changes in aphid abundance across Scotland over time (1967-2023), we model aphid counts from several Scottish suction trap sites (Ayr, Dundee, Edinburgh) using a TVAR(1) process. We incorporate a TVAR(1) process into a GAMLSS framework and use a Bayesian approach to estimate the overdispersion parameter and the TVAR(1) processes as a function of time. Where smoothed functions are penalised using localised shrinkage via Horseshoe priors. Our results show that increases in aphid abundance over time results in increases in autocorrelation. This suggests as aphid abundance increases, aphids are more likely to survive overwintering, highlighting the appropriate application of TVAR(1) processes to model aphid abundance.

On the identifiability of the stochastic state-space models

Julien Gibaud, Institute of Mathematics of Bordeaux, France

Co-authors: Frédéric Barraquand

State-Space Models (SSMs) are deterministic or stochastic dynamical systems defined by two processes. The state process, which is not observed directly, models the transformation of the states over time. On another hand, the observation process produces the observables on which model fitting and prediction are based. Ecology frequently uses stochastic SSMs to represent the imperfectly observed dynamics of population sizes or animal movement. However, several simulation-based evaluations of model performance suggest broad identifiability issues in ecological SSMs. Formal SSM identifiability is typically investigated using exhaustive summaries, which are simplified representations of the model. The theory on exhaustive summaries is largely based on continuous-time deterministic modelling and those for ecological SSMs have developed by analogy. While the discreteness of time does not constitute a challenge, finding a good exhaustive summary for a stochastic SSM is more difficult. The strategy adopted so far has been to create exhaustive summaries based on a transfer function of the expectations of the stochastic process. However, this evaluation of identifiability does not allow to take into account the possible dependency between the variance parameters and the process parameters. We show that the output spectral density plays a key role in stochastic SSM identifiability assessment. This allows us to define a new suitable exhaustive summary. Using several ecological examples, we show that usual ecological models are often theoretically identifiable, suggesting that most SSM estimation problems are due to practical rather than theoretical identifiability issues.

Efficient dynamic occupancy models for large citizen science datasets

Hannah Worthington, University of St Andrews

Co-authors: Emily Dennis, Butterfly Conservation, Byron Morgan, University of Kent, Takis Besbeas, University of Kent

Three-quarters of the butterfly species that can be found in the UK are showing persistentdeclines in distribution and/or population. Thanks to long-term monitoring schemes and the increasing availability of large-scale citizen science data, we are able to quantify changes in species' occupancy, and butterflies in particular are recognised as valuable indicators of biodiversity. Simple dynamic occupancy models, which account for imperfect detection, are a popular analysis option but often run into difficulties when applied to large datasets such as those collected at Butterfly Conservation. These opportunistic datasets are of the magnitude of several hundred thousand sites, with records spanning several decades. We explore the potential for more efficient model fitting to presence-absence data when standard packagesand functions appear to encounter computational limitations. A hidden Markov model (HMM) formulation offers the desired computational efficiency, with the additional bonus ofthe simplicity in how the large amounts of missing data are handled. We apply the HMM approach to several UK butterfly species covering differing rarities and some that are showing changes in spatial distribution. We provide estimates of species persistence andmovement, visualise the

range contraction or expansion of species over time, as well as potentially revealing some effects on species distributions linked to climate change.

4. Posters

CP01. Statistical Approaches for Modelling Multiple Heavy Metal Pollutants in River Water-Quality Data

Iain Bell, University of Glasgow.

Monitoring heavy metal concentrations in river water plays an important role in tracking and predicting trends that may impact on biota and ecosystem health. Missing observations, and interdependencies between metals, introduce challenges in the statistical approaches to modelling these concentrations. This poster presents a case study, using data sourced from the Environment Agency, on methods used to jointly model temporal trends of multiple heavy metal pollutants, initially at individual spatial locations in the south-west of England. Generalised Additive Mixed Models (GAMMs) are considered, where the concentrations of multiple pollutants are flexibly modelled using random effects. These models allow for flexible temporal trends to be fitted and intra-metal variability to be accounted for. The correlation structure between metals is investigated with the aim of enabling predictions for metals when the concentration is unobserved to be informed by its previous observed concentrations, and those of other metals observed at the same time. Other statistical approaches to achieve this same goal are also considered. There is scope to extend the study to model the concentrations of multiple metals within a river network, allowing spatial correlation between locations to be included. By developing methods that account for both temporal and spatial correlation, as well as intra-metal dependencies, a promising approach to the effective and realistic analysis of multivariate environmental data, such as water-quality data, is proposed.

CP02. Introducing the UK Network for Agricultural, Biological and Environmental Statistics

Eleanor D'Arcy, Environment Agency.

Co-authors: Hayley Carr, Harriet Low, Fiona Underwood. All authors contributed to this poster on behalf of NABES of NABES.

The Network of Agricultural, Biological and Environmental Statistics (NABES) was set-up to support a UK-wide network of quantitative experts, (including statisticians and data scientists) working in applied agricultural, biological and environmental sciences. Our network includes quantitative experts from government, or government funded bodies, research organisations, charities and the private sector. We aim to develop a community of practice for professionals in all stages of their career with a special focus on supporting early career and those who may be quite isolated as the only quantitative expert supporting large numbers of scientists and decision makers within their organisation. Here we describe NABES aims, the challenges many of our network members face, current and planned activities including monthly workshops, benefits and how to get involved. Come and find us – we are a friendly bunch and would be happy to chat with you about what we are doing.

CP03. Predicting the likelihood of land around ditches being drained: an Icelandic case study.

Gylfi Snær Sigurðsson, University of Iceland,

Co-authors: Birgir Hrafnkelsson, Jón Guðmundsson

Wetlands are globally significant ecosystems, providing critical ecological functions such as biodiversity support, water regulation, and climate mitigation. They serve as carbon sinks, storing a substantial portion of terrestrial soil organic carbon (SOC). A key factor in maintaining SOC levels is the anaerobic soil conditions sustained by high water tables. However, when wetlands are drained, these conditions are lost, leading to the gradual release of SOC into the atmosphere as carbon emissions. In Iceland, emissions from drained wetlands account for over one-third of the country's annual carbon output. Despite this, a comprehensive map distinguishing drained from undrained regions is currently unavailable.

To address this gap, we developed a logistic regression model to predict the probability of land drainage using predominantly satellite-derived GIS data. The model is inferred through INLA. Initially, a spatial model was considered, but due to the country's size, the high variability between observational sites, and their limited number, the spatial component was not included. Some of the observational sites have multiple observations, each with different distances with respect to a draining ditch of interest. A mixed-effects INLA model, which takes into account the neighbor structure of the observational sites through a random walk component, was implemented. The final model adequately predicts the likelihood of land being drained, providing a valuable tool for environmental management and carbon accounting. However, further refinement and validation are needed to enhance its accuracy and applicability.

CP04. Comparison of methods to calculate the LCL of the mean and 95th percentile for water quality parameters.

Jillian Delaney, Environmental Protection Agency, Ireland,

Ireland's national Water Quality Monitoring Programme measures the ecological and chemical status of surface waters (rivers, lakes, transitional and coastal waters) and the quantitative and chemical status of groundwaters. Physico-chemical elements such as nutrients, dissolved oxygen, salinity, pH, etc., are measured to provide an indication of how these elements support the achievement of good or better ecological status. For certain physico-chemical elements, regulatory standards for the mean and 95th percentile have been identified that are associated with good ecological status. These standards are set out in Irish legislation.

Lower Confidence Limits (LCLs) of the population parameters are used to measure compliance with the relevant regulatory standard. The physico-chemical data are right skewed with small sample sizes, so that the Central Limit Theorem cannot be assumed to apply. The sample sizes are also too small for bootstrapping methods to be used. Therefore, parametric methods using log transformations are used. Various methods exist to estimate the LCL of the mean, including Land's method, the modified Cox method, the method of Zou et al. and Parkin et al. A simulation study compares these methods for three physico-chemical elements and identifies the best method to use for the Irish data. Two methods to calculate the LCL of the 95th percentile are also compared: the normal approximation and the lower β-content tolerance interval. Again, the better method for Irish data is identified.

CP05. Mathematical Modeling of Covid-19 Dynamics in a West African Context.

Christabel Emaeyak James, University of Glasgow,

Co-authors: Jason Matthiopoulos & Daniel Haydon

The novel Coronavirus Disease 2019 (covid-19), which emerged in Wuhan, China is a highly infectious disease caused by (SARS-CoV-2) and has significantly affected public health and socio-economic well-being worldwide. Its transmission highlights the potentially important role of transmission heterogeneities, requiring better modeling approaches to determine their role in dynamics and control. This study aims to develop a method for detecting heterogeneities in susceptibility or connectivity in covid-19 transmission by fitting a modified Susceptible, Infected Recovered model to incidence data. The parameters of the model are estimated using Markov Chain Monte Carlo techniques. The proposed method is tested on simulated data to ascertain its effectiveness before applying it to real-world incidence time-series from different Nigerian States supplied by their Centre for Disease Control. The best performing models including different sources of heterogeneity is determined using the Deviance Information Criterion and used to make recommendations on possible interventions. By identifying and detecting heterogeneities that act to lower the Herd Immunity Threshold and the effective reproduction number, findings from this study will be useful in providing an improved understanding of disease spread, reducing the epidemic size and the burden of disease to make better-informed decisions for managing emerging and re-emerging infectious diseases like covid-19.

CP06. Ecosystem processes driving deer space-use and Lyme disease hazard in urban greenspaces.
Sara Gandy, University of Glasgow,

Co-authors: Grace Plahe, Jessica Hall, Dave Johnson, Richard Birtles, Lucy Gilbert

Over the past few decades, cities have shifted towards implementing nature-based solutions to increase climate change mitigation and enhance biodiversity. However, these rewilding schemes, which include tree planting, could increase tick-borne disease hazard by causing shifts in tick and pathogen hosts space use. This study aimed to determine ecological processes driving deer space use, tick density and Lyme disease hazard in urban greenspaces using structural equation modelling (SEM). Urban greenspaces were surveyed across 12 towns in the United Kingdom and deer space use, tick density and Lyme disease hazard were estimated at each site. Site-level metrics hypothesised to influence disease hazard, such as urbanisation history, woodland cover and connectivity, were incorporation into the SEM framework to understand direct and indirect relationships. Overall, ticks were detected in 73% of urban greenspaces. The SEM approach revealed that urbanisation around urban greenspaces in the past 30 years, connectivity to a large woodland and woodland cover within a park positively directly and indirectly affected deer space use, tick density and Lyme disease hazard. Built-up cover surrounding parks and age of urban greenspaces, on the other hand, negatively affected Lyme disease hazard. By using SEM to understand causal pathways, this study highlights how rewilding schemes in cities may unintentionally increase tick-borne disease hazard in urban parks. The findings underscore the need for integrated urban planning strategies.

CP07. Sub-daily rainfall simulation using multifractal canonical disaggregation: a parsimonious calibration strategy based on intensity-duration-frequency curves

Francesco Cappelli, Università degli Studi della Tuscia.

Co-authors: Elena Volpi, Andreas Langousis, Roberto Deidda, Anastasios Perdios, Pierluigi Furcolo, Salvatore Grimaldi

Synthetic rainfall scenarios at high temporal resolutions are pivotal in numerous environmental applications. Despite the abundance of available simulation methods, their practical utilization among practitioners remains limited, often due to challenges in model calibration stemming from sample size constraints. We introduce a novel parsimonious approach for estimating parameters of multifractal disaggregation models, based solely on available Intensity-Duration-Frequency curves, which are widely and readily accessible within the practitioner community. The performance of the proposed approach is assessed using three case studies, wherein detailed statistical properties of the simulated time series are compared against observed benchmarks. Our results indicate the potential of our approach to facilitate the straightforward application of complex models.

CP08. Statistical estimation of probable maximum precipitation (PMP)

Anne Martin, Polytechnique Montréal,

Co-authors: Jonathan Jalbert, Ã‰lyse Fournier

In the province of Québec, engineers design infrastructure exposed to hydrometeorological hazards, such as hydroelectric dams, using the estimation of probable maximum precipitation (PMP). The World Meteorological Organization (WMO) defines PMP as 'the greatest depth of precipitation for a given duration meteorologically possible for a design watershed or a given storm area at a particular location and time of year, with no allowance made for long-term climatic trends.' Thus, overestimating PMP can lead to unnecessary costs when developing infrastructure, while underestimating it can have dramatic consequences for neighboring populations and local energy production. Current methods of estimating PMP have several flaws: some variables used are not directly observable and require a series of approximations to estimate; uncertainty is not always accounted for and can sometimes be difficult to determine; and climate change, which exacerbates extreme precipitation events, is challenging to incorporate into the calculations. The goal of this work is to propose a statistical approach for estimating PMP that adheres to the WMO definition and allows for uncertainty estimation and the incorporation of climate change.

CP09. Is an image still worth 1000 words in the era of AI? A comparison of Human and AI responses in urban environmental image analyses

Luigi Cao Pinna, University of Glasgow,

Co-authors: Claire Miller, Marian Scott, Cris Hasan

The £10.2M NERC-funded GALLANT (Glasgow as a Living Lab Accelerating Novel Transformation) project aims to support a transition to a socially and environmentally thriving city. This will require understanding citizens' perceptions of Glasgow's ‘living space’. Such citizen science data are various and complex, potentially including images and text from social media on environmental themes such as flooding and biodiversity. These underutilised data types are nowadays accessible through Artificial Intelligence (AI) embedded pipelines whose results are increasingly used to inform decision-making. However, despite the often reported high AI performance, their black-box nature can make the results difficult to interpret, evaluate and trust.
In this study, we developed a new pipeline to perform image-to-sentiment analysis, which also allow

a transparent and effective comparison of AI and human responses. Human participants responded to a survey by rating and captioning a set of urban landscape images. For the same set of images, three vision-language (AI) models generated captions. Image descriptions were transformed into a numeric sentiment value by the AI models, allowing for a comparison of human sentiment and ratings with AI responses through data science and statistical approaches.
When comparing the extracted sentiment, some AI algorithms performed more similarly to humans than others without a clear or simple relation with AI algorithmic complexity. While AI and human responses largely overlap, mostly because of the similarities in the descriptions, a co-pilot approach is still required to inform effective city-wide environmental planning by combining human insights with AI intelligence from the data.

CP10. A Data Science Toolbox for ecological and environmental statistics

Susan Jarvis, UK Centre for Ecology & Hydrology,

Co-authors: Michael Hollaway, Jeremy Carter

Ecologists and environmental scientists often ask complex questions of messy data, frequently encountering problems related to missing data, biases and complicated experimental setups. Whilst statistical approaches to many of these problems exist, uptake of methodology by applied scientists can be limited. This may be because methods are difficult to discover (e.g. published only in the statistical literature), difficult to re-use (e.g. code is not available) or difficult to understand (e.g. documentation is not written for a wider audience). To address these issues we have started to develop a "Data Science Toolbox" which presents statistical and data science methods to an applied audience. The toolbox currently takes the form of a Jupyter book, with enhanced discoverability through the use of AI. We will present the initial prototype toolbox, explaining how our approach addresses discoverability, re-use and comprehension of methods. We will also discuss the co-design approach around the toolbox, and attendees will have the ability to feed back on our prototype design.

CP11. State of the art in data fusion approaches for environmental data

Craig Wilkie, University of Glasgow

Co-authors: Claire Miller, Marian Scott, Surajit Ray, Daniela Castro-Camilo, Daniela Cuba, Stephen Jun Villejo, Pietro Colombo

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 poster will present a review of state-of-the-art methods for data fusion, with a particular focus on work being carried out by early career researchers at the University of Glasgow.

CP12. Spatio-temporal modelling of bumblebee populations with INLA and inlabru

Cam Milliken, University of Kent,

Co-authors: Eleni Matechou, Christos Efstratiou, Richard Comont

Bumblebees are critically important pollinators, but many species are in decline. Standardised, long-term monitoring schemes are required to assess the extent of this decline and inform conservation efforts, but monitoring bumblebees at the desired spatiotemporal scale is challenging. The Pollinator Monitoring Scheme (PoMS) has been set up for this purpose. Participants in the Flower-Insect Timed Count (FIT Counts) survey record the number of bumblebees they observe landing on flowers in ten minutes within a 50cm x 50cm quadrat, generating count data of the number of bumblebees observed at a particular space and time. In this talk, I will discuss modelling this data using the Integrated Laplace Approximation (INLA) and inlabru to construct a spatiotemporal model that takes into account site and weather covariates. I will then present related findings.

CP13. Performance of Open Data versus Classical Meta-Analysis - a simulation study

Danijela Žanko, Institut Ruđer Bišković

Co-authors: Antica Čulina, Azra Tafro, Sunčana Geček

Meta-analysis is widely used to synthesize research findings from different studies. However, meta-analysis that uses the results of primary studies suffers from issues related to the methodological and reporting quality of the underlying studies, and publication bias. One solution is a meta-analysis of raw data. The growing open science movement has significantly increased the availability of raw research data. The performance of meta-analysis of open data versus results published in studies remains poorly understood.

This study systematically compares open data meta-analysis (ODMA) and classical meta-analysis (CMA) approaches on three key dimensions: accuracy of the effect size estimation, precision of confidence intervals, and assessment of heterogeneity across different effect size conditions. This is done under plausible scenarios of publication bias, p-hacking, study sample sizes, and magnitude and distribution of the values of true effects.

Our simulation framework incorporates parameters based on ecological literature, including variable sample sizes, publication bias rates from environment/ecology studies, and p-hacking prevalence from researcher surveys. We expect that ODMA may provide more accurate estimates than CMA when research deviates from ideal processes, particularly under conditions of publication bias and p-hacking.

This study aims to provide ecological and environmental researchers with empirical guidance on meta-analytic approaches in the context of open science practices.

CP14. Rural education at a threat: assessing the causes for public rural schools closures in the Brazilian Amazon

Hilder André Bezerra Farias, Universidade Federal do Pará

Co-authors: Luke Parry, Leonardo de Sousa Miranda

This research investigates the closure and survival of publicly administered and partnered rural schools in the Brazilian Amazon. These schools, products of historical struggles by rural populations for formal education, including both children and adults, and the "Educação do Campo" social movement, are crucial for literacy and access to scientific knowledge. School closures thus pose a significant threat to the resilience, livelihoods, and territorial rights of rural households. Preliminary findings indicate a widespread deterioration of school quality, particularly in facilities and amenities (e.g., access to drinking water and school food), affecting both closed and open schools. However, schools in specific territories—settlements, indigenous lands, and quilombola areas—demonstrate greater resilience compared to "regular" rural schools. We utilize publicly available data to test hypotheses from the literature regarding the role of specific pressures in school closures, including proximity to the agribusiness frontier, impacts of river floods on school operations, and vulnerability to political cycles.

Room Details

1. Faraday Complex - Main conference Building
2. Library
3. County South LT - Conference Dinner