

1st Workshop on Ecological Data Analysis

Department of Mathematics and Statistics and Hamilton Institute
Maynooth University, Ireland

May 24, 2019

Welcome

Dear participant,

It is with great pleasure that we welcome you to the 1st Workshop on Ecological Data Analysis. We hope this will be an interesting and productive day, and we are looking forward to the talks and discussions on different types of work on the interface between Ecology and Statistics. We conceptualised this Workshop so that it works as a platform for establishing new interdisciplinary connections (both within and outside of Maynooth University) for future collaborations among researchers in our departments / institutions.

This inaugural edition of the Workshop gathers participants from several institutions in Ireland and Brazil, namely, Maynooth University, NUI Galway, University College Dublin, University of Limerick, Teagasc, and the University of São Paulo. Researchers with a vast array of fields of expertise are attending, such as Statistics and Data Science, Applied Mathematics and Probability, Biology, Geography, Computer Science, Agriculture, Ecology and Paleoecology, and Biogeography.

The morning session will consist mainly of short talks on different areas of ecology and methods for data analysis (animal abundance, biomass data, biodiversity, methods for count data, and sea level modelling, as well as lightning talks by PhD students on their recent developments). The afternoon session will consist of three short talks and a brainstorming session, where the participants will be split into groups to discuss possible collaboration opportunities at the interface of Ecology and Statistics. Finally, each group will briefly present the main opportunities identified and next steps for developing collaboration will be discussed.

We wish that this Workshop will serve as a springboard for developing future collaborations among this network of participants and thank you for your contributions.

With best wishes,

Rafael Moral, Caroline Brophy, and Estevão Prado

The Organising Committee of the 1st Workshop on Ecological Data Analysis

Acknowledgement

We would like to thank the Department of Mathematics and Statistics, Hamilton Institute, and especially Gráinne O'Rourke, Janice Love, Tony Waldron, Rosemary Hunt, and Kate Moriarty for all the support.

Contents

Welcome	ii
Acknowledgement	iii
Abstracts	1
The use of vestiges on population density estimate of terrestrial mammals in human-altered environments (<i>Luciano M. Verdade and Rafael A. Moral</i>)	1
Modelling the impact of biodiversity on ecosystems in ecology and agronomy (<i>Caroline Brophy</i>)	2
Bayesian implementation of N-mixture models (<i>Estevão B. Prado, Rafael A. Moral and Andrew C. Parnell</i>)	2
Assessing the effects of plant diversity in intensive grasslands (<i>Guylain Grange</i>)	3
Count data: over/under-dispersion, zero-inflation and some extended models (<i>John Hinde</i>)	3
A Bayesian statistical model for reconstructing and analysing former sea levels (<i>Niamh Cahill, Andrew Kemp, Benjamin Horton and Andrew Parnell</i>)	4
Analysing ecological and evolutionary patterns in bird songs (<i>Rafael A. Moral</i>)	5
The Application of Random Forests on Ecology Data (<i>Alan N. Inglis</i>)	5
Effects of potassium silicate application on growth of pickling cucumber plants in competition with weeds (<i>Maira Fatoretto</i>)	6
Diversity-Interaction models applied to a grassland biodiversity experiment (<i>Ahmed Ali</i>)	6
Does grassland diversity reduce N ₂ O emissions? (<i>S. Cummins</i>)	7
Proceeding with caution: exploiting multiple statistical approaches in biogeographical analysis (<i>Conor Meade</i>)	7
Developing climate information for use in decision making (<i>Rowan Fealy</i>)	8
Spectral and signal analysis methods applicable to Ecological analysis (<i>Joe Timoney</i>)	8
Author Index	9

Abstracts

The use of vestiges on population density estimate of terrestrial mammals in human-altered environments

¹Luciano M. Verdade and ²Rafael A. Moral

¹University of São Paulo, Brazil ; ²Maynooth University, Ireland

27th May
9:40 - 10:05am
Room: 317

Distance-based methods tend to underestimate the abundance of terrestrial mammals in special game species because their assumption of detecting all individuals on the trail is usually broken by the avoidance behaviour of those species in relation to humans. In addition, studies including the space use by wildlife (e.g., habitat use, detectability, and dispersal) in agricultural landscapes tend to be biased as their matrix (i.e., their main plantation) is usually erroneously considered as non-habitat. We propose the spatialization of vestiges (e.g., tracks, scats and marks in trees) as well as sights of living animals around which we plot the minimum and maximum adequate home ranges found in literature for each local species. As adequate we mean those home ranges determined in environments as comparable as possible to the study area in question. The basic assumption for this approach is that if there is a vestige like a track, there was at least one individual in the same spot in a near past. Although we assume round home ranges are only a rough approximation of actual ones, their overlap might be used as a criterion for the estimate of detected individuals or groups in such a way that partial overlaps can generate an interval of the estimated number of individuals detected (e.g., for five isolated home ranges and three overlapped ones the estimated number of individuals would be from six to eight). The population size of each species can be estimated by an asymptotic model adapted from species incidence curve to the individual level. For social species camera-traps can be used to determine group size, assuming the maximum group detected in the same photography as the minimum actual group size. The present approach is compatible with occupancy and detectability methodology. Future studies should prioritize the determination of home range and habitat use of local species. In addition, modeling might be useful to improve accuracy and precision estimates of the present method.

Modelling the impact of biodiversity on ecosystems in ecology and agronomy

27th May
10:05 - 10:30am
Room: 317

Caroline Brophy

Maynooth University, Ireland

Agri-ecosystems are under serious pressure to increase production to feed rising world populations, yet are simultaneously faced with challenges because of more frequent extreme climatic events. Biodiversity and ecosystem function (BEF) research seeks to explain the relationship between the diversity of the species in an ecosystem and the outputs that the ecosystem will produce. Decades of BEF research has shown that increasing species diversity often promotes ecosystem outputs, such as biomass yield in grasslands; however, many statistical challenges arise when modelling the data from biodiversity experiments. In recent years, the family of Diversity-Interactions models were developed to improve understanding of the BEF relationship by assessing how changing the proportions of species in the ecosystem will affect the outputs produced. In this talk, these recent statistical developments in BEF research are introduced and illustrated. The statistical challenges include complexities related to the handling of the simplex space of predictors (the species proportions), large-scale geographical and plot-scale localised spatial patterns, repeated and multivariate responses, and the modelling of ecosystems with large numbers of species.

Bayesian implementation of N-mixture models

27th May
10:30 - 10:40am
Room: 317

Estevão B. Prado, Rafael A. Moral and Andrew C. Parnell

Maynooth University, Ireland

In this work, we provide a Bayesian implementation of the N-mixture model as well as the comparison of the classical and Bayesian estimates. The N-mixture model (Royle, 2004) was proposed in the context of ecology to estimate the size of animal populations in situations of imperfect detection. Due to the imperfect detection, traditional techniques, such as Generalised Linear Models, cannot be applied.

Assessing the effects of plant diversity in intensive grasslands

Guylain Grange

Maynooth University, Ireland

27th May
10:40 - 10:50am
Room: 317

Most of Irish grasslands are based on ryegrass (*Lolium perenne*) monocultures, requiring an important fertilisation. However, with the triple challenge of increasing production, mitigating environmental issues and facing climate changes, sustainable intensification seems the only orientation to be taken. A field experiment were sown to investigate how plant diversity could help increasing yield, reducing greenhouse gases emissions and coping extreme climate events. To do so, a simplex design were used to establish monocultures and combinations of 6 grassland species, over two environments: an undisturbed control and an 8-weeks summer drought. A diversity interaction model, with repeated measures was used to analyse data for each indicator.

Count data: over/under-dispersion, zero-inflation and some extended models

John Hinde

National University of Ireland, Galway, Ireland

27th May
11:10 - 11:35am
Room: 317

The standard distribution for the analysis of count data is the Poisson. Frequently, in practice, it is too restrictive in that the variability in the data is either significantly greater (overdispersed) or less (underdispersed) than that implied by the models variance function. For the analysis of count data, Nelder and McCullagh (1989) says that overdispersion is the norm and not the exception and this has been well studied, see Hinde and Demétrio (1998) and many subsequent articles presenting a wide range of distributions. An associated phenomena is zero-inflation where the data exhibit more zeros than expected under the Poisson model. Models allowing for zero-inflation include the zero-inflated mixture model and the two-stage hurdle model. In this talk I will discuss these basic ideas and consider some generalisations of the Poisson distribution that provide greater flexibility, including COM-Poisson, discrete Weibull, and Poisson-Tweedie models. These will be illustrated with various datasets and I will attempt to make some general points about modelling and computation.

A Bayesian statistical model for reconstructing and analysing former sea levels

27th May
11:35 - 12:00pm
Room: 317

¹Niamh Cahill, ²Andrew Kemp, ³Benjamin Horton and ¹Andrew Parnell
¹Maynooth University, Ireland; ²Tufts University, USA; ³Nanyang Technological University, Singapore

In order to understand the present we must first gain insight into the past. Therefore, to understand and have historical context for current rates of sea-level rise we need to be informed about past changes that have occurred. Sea-level reconstructions can provide this information by giving us insight into the magnitude and rates of past sea levels.

We have produced sea-level reconstructions along the U.S Atlantic East coast using biological and geochemical sea-level indicators preserved in dated cores of salt-marsh sediment. I have developed statistical models that can help us to bridge the gap between the information held in these raw proxy data and a high-resolution sea-level reconstruction. Using a Bayesian framework for these models aids in the understanding and quantification of the uncertainty that is inherent in these data and the resulting records of former sea levels.

I present a Bayesian transfer function modeling approach that is used to produce reconstructions of past sea level through the calibration of a biological proxy (e.g., foraminifera) into tidal elevation. The first step in the transfer function approach is building a model that captures the relationship between a biological proxy and tidal elevation in a modern environment. The second step uses this relationship to produce estimates of paleo-tidal elevation with uncertainty for each layer in a sediment core. Additional proxies (e.g., $\delta_{13}\text{C}$) can be used to further constrain these estimates and potentially reduce uncertainty.

Combining output from the Bayesian transfer function with a core chronology provides us with a reconstruction of relative sea level through time. With the aim of estimating rates of sea-level change, reconstructions are analyzed using an errors-in-variables integrated Gaussian process model. Ultimately, through the combination of these statistical models we can capture the continuous and dynamic evolution of rates of RSL change with a full consideration and propagation of available uncertainties. Results show that 20th century sea-level rise along the U.S. Atlantic coast is the highest its been in at least the last 15 centuries.

Analysing ecological and evolutionary patterns in bird songs

27th May
12:00 - 12:25pm
Room: 317

Rafael A. Moral

Maynooth University, Ireland

Biomusicology is the study of animal singing in biological populations. It is an increasingly growing interdisciplinary area of science, especially as a new branch of ecological studies. Sound traits such as frequency, amplitude, period (among many others) can be used to study evolutionary and ecological processes related to the emission and reception of acoustic signals. In this work, we model the mean and dispersion of the fundamental frequency of perching bird songs to study how their phylogeny and functional ecology influence their singing. We then propose a joint model for the duration, minimum and maximum frequencies of the bird songs, based on the multivariate covariance generalized linear modelling framework. Our results suggest that modelling the mean alone would not reveal the contribution of musical pitch variability to microevolutionary and ecological processes of Neotropical perching birds. We discuss model implementation problems and present ideas for further studies. This is work in collaboration with Wagner Bonat (Federal University of Paraná, Brazil), Joe Timoney (Maynooth University), Mateus Mendes (University of Campinas, Brazil), and Luciano Verdade (University of São Paulo, Brazil).

The Application of Random Forests on Ecology Data

27th May
12:25 - 12:35pm
Room: 317

Alan N. Inglis

Maynooth University, Ireland

Random Forests models are an ensemble machine learning method that are comprised of many independent decision trees. Random Forests are used for classification and regression problems and are generally considered to be superior to parametric models due to their predictive accuracy on a broad range of data structures. The predictions from the ensemble of trees are averaged (for regression) or employ a majority voting system (for classification). Random forests have multiple applications, including classification and regression analysis, survival analysis, and variable importance, all of which can be applied to ecological data. Other aspects of random forests that can be particularly useful in the field of ecology (which often has large complex data, missing values or outliers) is that random forests have the ability to impute missing values, whilst maintaining high accuracy and can handle multiple types of response variable (e.g., categorical, numerical, survival etc.). In this work, we present the basic principles behind random forests and show how random forests can be applied to complex ecological data sets.

Effects of potassium silicate application on growth of pickling cucumber plants in competition with weeds

27th May
12:35 - 12:45pm
Room: 317

Maira Fatoretto

University of São Paulo, Brazil

The cucumber *Cucumis sativus* is one of the most important raw materials for the manufacture of pickles. However, it is influenced by *cyperus rotundus* L. (*Cyperaceae*), a type of grass that is the main weed in the region, which can drastically reduce yield and interfere with harvest. Potassium silicate (K_2SiO_3) has been reported as an element capable of stimulating the induction of resistance in plants, however, the effect of silicate in cucumber plants is still not known. To study how potassium silicate influences the development of cucumber plants under weed competition, an experiment was carried out in the greenhouses of the Instituto Federal Goiano in Gias State. This dataset present a tricky structure. The first difficulty in its modeling is due to the fact that be a longitudinal data set, hence hierarchical. Besides, the two outcomes present different behavior over time. To analyze jointly number of leaves and height we may consider a bivariate model. However, as outcome we have continuous and count data which becomes more complicated than a bivariate normal structure. The main in this work is to develop suitable models considering the biological process that may be help to researchers to explain the impact of the Potassium silicate (K_2SiO_3) on the growth of the cucumber plant under biotic stresses.

Diversity-Interaction models applied to a grassland biodiversity experiment

27th May
12:45 - 12:55pm
Room: 317

Ahmed Ali

Maynooth University, Ireland

Diversity-Interaction models are a family of statistical models used to learn about the relationship between biodiversity and the outputs of an ecosystem. It allows statisticians working in ecology to understand the contributions that the interactions between specific species provide to the overall diversity. Diversity-Interaction models tend to have the form:

$$y = \sum_i^s \beta_i P_i + \alpha A + \sum_{i < j}^s \delta_{i,j} P_i P_j + \epsilon,$$

where y is the functional response in a community, P_i is the proportion of the i -th species in the community and A includes all density and blocking/treatment effects. β_i are the expected performances of the species in monoculture $i = 1, \dots, s$, when $P_i = 1$, at mean levels of terms in A . $\delta_{i,j}$ is the coefficient of the interaction between the i -th and j -th species which measures the strength of the interaction, the sign allows us to see whether the interaction between the two species contributes to the overall diversity or otherwise. The residual terms ϵ_i are distributed normally with a constant variation of σ^2 .

These models are applied to a dataset from a grassland biodiversity experiment where fertiliser is added to different plots with different combinations of proportions of species. We wish to understand how does manipulating fertiliser and species diversity (monocultures up to seven-species mixtures) affect the yield. We will also touch on some further work including a simulation study on estimating a parameter θ for a generalised version of the models.

Does grassland diversity reduce N₂O emissions?

S. Cummins
Teagasc

27th May
12:55 - 13:05pm
Room: 317

Nitrous oxide (N₂O), a potent greenhouse gas, is the greatest contributor towards the depletion of the ozone layer and has a global warming potential 298 that of CO₂. The use of more diverse grasslands could be a mitigation option for reducing N₂O emissions from soils due to their increased nitrogen (N) use efficiency and to their potential to impact both the Carbon and N cycles in comparison to grass monocultures. Increased Nitrogen use efficiency will result in less mineral N available in the soil for loss to the environment through nitrate leaching and denitrification. The objective of this experiment was to investigate the effect of diverse forages on N₂O emissions. The experiment varies levels of grassland diversity with combinations of 1-6 species comprised of 3 functional groups of grass, legumes and herbs. The experiment used a simplex experimental design, allowing the interactions between species which effect N₂O emissions to be established. The plots received 150 kg N ha⁻¹ in 5 splits over the growing season. N₂O fluxes were analysed using the static chamber method combined with gas chromatography. The effect of species mixtures on cumulative N₂O emissions will be evaluated over the experimental period. The over-arching hypothesis is that more diverse combinations of grassland species and functional groups will utilise N more efficiently than monocultures, resulting in reduced N₂O emissions.

Proceeding with caution: exploiting multiple statistical approaches in biogeographical analysis

Conor Meade
Maynooth University, Ireland

27th May
02:00 - 02:25pm
Room: 317

Biogeography is concerned with framing a holistic view of how and why organisms came to their current situation, and for this reason problem solving proceeds through multiple lines of evidence and data. Because of the diversity of data sources, moving between different methods requires knowledge of validity, prerequisites, and the limits of confidence and inference. The challenge is to marry knowledge about historical processes, the current biology and ecology or target organisms, and molecular variance data at local and phylogenetic scales. In this short presentation I will discuss the kinds of methods we have been using in the lab, with examples.

27th May
02:25 - 02:50pm
Room: 317

Developing climate information for use in decision making

Rowan Fealy
Maynooth University, Ireland

Based on global climate model simulations, the location, extent and timing of moisture limited regimes are likely to alter as a consequence of climate change (Seneviratne et al., 2010). These findings are consistent with regional climate model simulations for Ireland by Fealy et al. (2018) who found that temperatures are likely increase in all seasons, over the period 2021-2050; while precipitation simulations are less certain, the range from multiple models suggest decreases are likely during the summer months for the same period (e.g. RCP2.6 -13.6 - +4.8%; RCP8.5 -19.0 - 0%). While Ireland does not experience the extremes in temperature associated with continental Europe, it is subject to the occurrence of rainfall extremes, particularly droughts (Noone et al, 2017); the frequency and magnitude of which are also projected to increase globally, by the end of the century (Seneviratne et al., 2012). These changes, if realised will have consequences for ecosystems here, and in particular grasslands, as the role of drought stress, rather than heat stress, has been highlighted as the primary factor in limiting ecosystem productivity (De Boeck et al., 2011). This talk will explore some of the issues, look at on-going research in the area and propose a potential way to develop climate services and information for use in decision making.

27th May
02:50 - 03:15pm
Room: 317

Spectral and signal analysis methods applicable to Ecological analysis

Joe Timoney
Maynooth University, Ireland

Commercial tools such as Avisoft are well- known in the field of sound analysis of ecological data. They use the spectrogram to transform the captured signal waveforms into a time-frequency representation. This reveals the spectral patterns of the song which can provide interesting features that can be extracted for further analyses. However, this tool has limitations which affect its resolution in time and frequency. Enhancements that have been introduced to the spectrogram will be described along with alternative time-frequency representations with different properties. Following this, some signal analysis techniques will be introduced which treat the signal as a combination of Amplitude and Frequency modulated components. The frequency modulation analysis should, in theory, accurately capture the frequency excursions of the signal over time. The potential benefits of their application to this domain will be discussed.

Author Index

Ahmed Ali, 6

Alan N. Inglis, 5

Andrew Kemp, 4

Andrew Parnell, 4

Benjamin Horton, 4

Caroline Brophy, 2

Conor Meade, 7

Estevão B. Prado, 2

Joe Timoney, 3, 8

John Hinde, 3

Luciano M. Verdade, 1

Maira Fatoretto, 6

Niamh Cahill, 4

Rafael A. Moral, 1, 5

Rafael A. Moral and Andrew C.
Parnell, 2

Rowan Fealy, 8

S. Cummins, 7