Second International Workshop on "Statistical inference for assessing and monitoring natural resources and biodiversity"

Abstracts

Keynote talks

RB O'Hara, Norwegian University of Science and Technology, Trondheim, Norway

Developing the One True Model(s) for the Distributions of Species

A major problem in ecology is the reconstruction of the distribution of species, and prediction of how these will change in the future. This requires data over a large spatial scale, so we cannot normally rely on single surveys. But we are gathering a lot of data in other ways: citizen scientists are recording their observations and uploading their data to portals such as eBird and iNaturalist. This data will massively expand the information we can use, but it also comes with problems: the data are not from randomly selected sites, and there will be biases in which species are observed and recorded. I will outline the approach we are taking to dealing with these problems: integrating different types of data on species' occurrences into single models and dealing with the biases in each data set. From these models of the sampling process, we hope to generate reconstructions of distributions, and use models of the ecology of the species to both inform these reconstructions and predict the future distributions of species and communities of species.

C Rondinini, Sapienza University of Rome, Italy

Past, present, and projected extinction risk of the world's mammals

The conservation status of the > 6000 species of mammals described so far is monitored by the IUCN Red List through the work of the Global Mammal Assessment program at Sapienza University of Rome, one of the Red List Partners. I will present global data on the distribution, extinction risk, and threats faced by mammals, using models of the Area of Habitat of terrestrial mammals, the IUCN Red List Index, and scenarios of global change in the coming decades. Currently, ca. 25% of the mammal species is threatened with extinction, a figure that has increased steadily since the 1970s. Current conservation actions can be effective in reducing the risk or averting extinctions of some species but are insufficient to reverse the trend. Scenarios demonstrate that options for slowing down or halting mammal losses exist but depend on major changes in patterns of production and consumption of food and energy.

A Fassò, University of Bergamo, Italy

Environmental data production and uncertainty: the statistician's role

Open data are becoming more and more important. In fact, on the one side, researchers, decision-makers, and, more generally, people are asking for more information about climate and the environment. On the other side, research project calls increasingly request open data products and, more generally, open science. But data are not perfect. Indeed, self-selection, importance sampling, and other biases are often hidden in modern (big) data. Modern environmental data sets often come

from large-scale monitoring networks and/or remote sensing for which the sampling design is often already defined. Nonetheless, there is a demand for high-level data built on the raw data. The statistician may contribute to uncertainty assessment, data aggregation, missing data imputation, and related uncertainty propagation in this phase. In this talk, I will discuss these topics and review recent examples of the statistician's role in environmental data production.

S Saarela, Norwegian University of Life Sciences, Ås, Norway

Statistical methods in connection with NASA's GEDI mission

NASA's Global Ecosystem Dynamics Investigation (GEDI) mission deployed its full-waveform lidar instrument on the International Space Station in November 2018. Since then, it has been collecting 3D data about vegetation around the globe. From GEDI waveforms, four types of structure information can be extracted: surface topography, canopy height metrics, canopy cover metrics, and relative height (RH) metrics. The latter is used for producing the GEDI L4A product, the footprint level aboveground biomass (AGB), and the GEDI L4B product, the gridded aboveground biomass density (AGBD). In this presentation, I will present and discuss statistical challenges faced by the GEDI mission while producing these two biomass related products. I will specifically address:

i. Effects of multicollinearity in the GEDI data on AGBD prediction uncertainty;
ii. Hybrid inference and its adaptation for the GEDI mission;
iii. Further developments of hybrid inference including incorporation of non-parametric biomass modelling and bootstrap variance estimation;
iv. Hierarchical model-based (HMB) inference developed for the mission.

Contributed talks

Inferential strategies for biodiversity analysis

M Cervellini, University of Camerino, Italy, L Fattorini, S Franceschi, University of Siena, Italy, A Chiarucci, M Di Musciano, P Zannini, University of Bologna, Italy

A sampling strategy for assessing habitat coverage at a broad spatial scale

The quantitative assessment of habitat conservation status is a major task for EU member states in compliance with Habitat Directive (HD). While various approaches have been adopted, there is no consensus on how to achieve common statistically sound estimates of EU terrestrial habitat types coverage, which is one of the four criteria indicated by the HD. We developed a two-phase sampling strategy to estimate the area (approximated at 1 ha cell size) of 9 terrestrial habitats distributed among different EU member states choosing Italy as a case study. Simulation study was performed to check the precision of the coverage estimators accounting for the lack of sampled data (nonresponse treatment), subregions and sustainable sampling effort. Adopting a small sampling fraction ($\leqslant$0.04%) our sampling approach has the potential to increase precision (7%$\leqslant$RSE$\leqslant$15% for common habitats) in estimating the coverage of habitat types with respect to the precision achieved by simple random sampling without replacement.

F Baccini, University of Pisa, Italy

Integration of similarity networks for combining multiple relations between biological species

Multiplex networks are a class of multilayer networks commonly adopted for modelling multi-type relations in a variety of real-world phenomena, including social, economic, and biological systems. A possible approach to analyse the structure of multiplex networks is to aggregate the different layers in a single network which is a good representative of the various layers. In this talk, two alternative methods for aggregating multiplex networks will be presented. The first method is based on the idea that strong and common links must be enhanced in the aggregated network; the second method is built on recent advances in the theory of barycenters of positive definite matrices. To highlight the relevance of the methodology in the ecological field, the two methods are tested on ecological data to analyse functional diversity among different species based on multiple traits.

S Fattorini, University of L'Aquila, Italy

Use of the species-area relationship for biodiversity hotspot identification: statistics and biology

Areas with an exceptionally large number of species are considered as biodiversity hotspots. However, when the compared areas are of different size, their species richness values cannot be compared directly, because larger areas tend to have more species, a pattern known as the species–area relationship (SAR). Since the SAR is not necessarily linear, one cannot divide the total species richness by the area and report the result as species per unit area. To overcome this problem, various approaches have been proposed based on SAR modelling, including the use of the coefficient of the power function (under the assumption that the SAR is best modelled by this function) and the magnitude of residuals (with areas having larger residuals being considered as hotspots – an approach that can be used with any function). Use of residuals may involve model selection to identify the best fit function or model averaging procedures when the data provide support for several models. All these approaches have important, but usually overlooked statistical issues, involving the impact of the error structure of the data, the choice between OLS and Model II regressions, the way in which models are selected as best fit(s), and the paradoxical situation that models that fit the data the poorest are also the best at identifying hotspots. Failure in considering these issues may lead to inconsistent results.

R Cazzolla Gatti, University of Bologna, Italy

The new universal index of absolute and effective biodiversity (AED)

The development of a unique index to represent the "true diversity" of a community is still a challenge. The difficulties to find a single, reliable, diversity metric is mostly due to the fact that the indices allowing an estimation of the number of species actually present in a sample do not provide information on how species abundances are distributed within a community. To compare and interpret the effective number of species with one single metric and analyze the diversity data with techniques that do not depend on a traditional non-parametric index, we developed an absolute measure of diversity based on the effective number of species and derived from the Hill numbers of order 0, 1, and 2. We tested the new index against previously proposed absolute diversity estimators and the first three Hill numbers taken alone. We simulated the new index's behavior with different gradients of richness, abundance and evenness and we, finally, empirically tested it on tree communities of three biomes (boreal, temperate, and tropical forests) in the United States, on a coral reef community of Cuba and during an ecosystem monitoring in Basilicata (Italy). This new index proved to be the first reliable and comparable measure, which combines both the absolute richness and the evenness of a community related to the most used traditional indices in a unique, simple, and comprehensive

numerical value that would represent the absolute effective diversity (AED) of any biological community.

Inferential strategies for animal populations

F Ferretti, A Rossi, F Bazzoni, C Riggio, R Oliveira, I Leggiero, University of Siena, Italy, B Esattore, Czech University of Life Sciences, Prague, Czech Republic

Fallow deer and wolves: multiple antipredator responses to a recolonising apex predator

There is a growing interest in assessing behavioural responses of prey to predators, especially considering the recent recovery of large carnivores in Europe and the potential consequences ensuing at the ecosystem level. Through intensive camera trapping, we analysed antipredator responses by fallow deer Dama dama to wolf Canis lupus, in a protected area recently recolonised by this apex predator. We found support for a spatial association between wolf detection rates and those of fallow deer, with no evidence for fallow deer avoiding sites with higher predator detection rates. There was strong support for differences in temporal activity patterns of the two species: the wolf was mainly nocturnal and the fallow deer was largely active during daylight, with an increase in diurnal activity along with wolf stabilisation in the area. Moreover, vigilance rate and duration of female fallow deer increased with the local frequency of wolf activity. We suggest an antipredator response based on temporal avoidance and increased vigilance, but not spatial avoidance.

R Ambrosini, University of Milano, Italy, N Fattorini, University of Siena, Italy

Quantifying migratory connectivity: current methods and future perspectives

Migratory connectivity refers to the spatiotemporal linkage between breeding and nonbreeding ranges where individuals spend different phases of their annual cycle. However, several different definition and methods for quantifying this property have been proposed in the literature, and no general agreement on a measure of migratory connectivity exists. We propose here a theoretical framework that unifies the two most commonly used measures for quantifying connectivity i.e. the Mantel correlation coefficient between the position of the individuals in different phases of their annual life-cycle and the so-called MC coefficient, a weighted average correlation between distances among separate areas among which the population of a specie move. We also propose the solution of some practical problems that arise in particular cases for which the MC coefficient in undefined. We also propose further theoretical development of this concept aiming at linking more tightly the concepts of migratory and ecological connectivity.

N Fattorini, L Lazzeri, F Ferretti, University of Siena, Italy

Camera trap-derived detection rates as relative abundance indices: an empirical test for wild ungulates

Species detection rates achieved through motion-activated cameras have been adopted as relative abundance indices (RAIs) to monitor mammal populations. Yet, group size often is not taken into account by this metric, therefore biases may arise when gregarious species are involved. By considering 3 sympatric ungulate species inhabiting a protected area, between 2019-2021, we compared two RAIs obtained by camera-traps with independent density estimates derived through faecal counts. The RAI incorporating group size (1) showed a stronger linear correlation to density and (2) was the only one correctly ranking interspecific differences in abundance. For the two most

abundant study species, the precision of both RAIs was acceptable, in terms of practical management value (CVs ≤ 25%), albeit it resulted lower for the least abundant one. Generally, our findings support the use of camera trapping-based RAIs as a promising tool to monitor ungulate populations, and that adjusting for group size would improve such metrics.

V La Morgia, B Franzetti, ISPRA, Rome, Italy, F Cagnacci, Fondazione Edmund Mach, S Michele all'Adige, Italy, S Focardi, CNR, Firenze, Italy

Techniques and methods for wildlife monitoring: towards a practical synthesis

Biodiversity can be explored at different levels, from genetic to ecosystem variation, with all levels and their interactions contributing to the overall functional diversity. The study of animal species has always been central in this framework. In particular, to properly manage animal communities and protect biodiversity, much research has been dedicated to the assessment of the abundance and distribution of animal populations, and several attempts have been made to synthesize the available approaches. However, most of the reviews adopts an empirical approach. Focusing on the assessment's methods proposed for ruminants, a well-studied group in this respect, we rather follow a theoretical framework. We review and classify the main methods using the concepts of detection probability and area coverage. This allows to identify similarities among methods, even when they are based on different techniques and devices for data collection. Taking into account these similarities, we group methods, and we introduce the SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis as a method for a transparent choice of the most appropriate methodology to be used by wildlife managers. Our review and synthesis also highlights a number of open questions in wild ruminant assessment, which are probably transferable to other animal groups. In this context, the identification of local population assessment methods that might facilitate data extrapolation to large spatial scale emerges as a key step for the monitoring of biodiversity at regional and national level.

New methodologies for inference on natural and environmental phenomena

C Calculli , University of Bari, Italy S Arima, University of Salento, Italy, A Pollice, University of Bari, Italy

A Poisson model for over dispersed spatial counts with misreporting

We propose a Poisson model for zero-inflated spatial counts contaminated by measurement error. We accommodate the excess of zeroes in the counts, consider the possible under/over reporting of the response and account for the neighbouring structure of spatial areal units. Bayesian inferences are provided by MCMC implementation through the R package NIMBLE. The modelling approach is proposed to investigate the relation between the counts of wildfire occurrences in municipal areas and several potential socio-economic and environmental-driven factors, considering two neighbouring regions in southern Italy (Apulia and Basilicata). Multiple sources of data with different spatial support are used and data were pre-processed in order to re-conduct the analysis to the municipal units. Results suggest the appropriateness of the approach and provide some insights on the features of wildfire occurrences.

L Ippoliti, University of Chieti-Pescara, Italy, N Golini, R Ignaccolo, University of Turin, Italy, N Pronello, University of Chieti-Pescara, Italy

Functional zoning of biodiversity profiles

Spatial mapping of biodiversity is crucial to investigate spatial variations in natural communities. In the literature, several indices have been proposed to represent biodiversity as a single statistic. However, such indices do not provide useful information on the richness and uniformity of the number of species and may lead to misleading results. In this work we focus on biodiversity profiles which provide a more flexible framework to express biodiversity through non-negative and convex curves which can be analysed by means of a functional data analysis (FDA). By treating the whole curves as single entities, we propose a penalized model-based clustering technique to achieve a functional zoning of the region of interest. This provides a spatial mapping of the biodiversity profiles which is useful for policy-makers both for conservation and management of natural resources and to reveal patterns of interest. Our approach is discussed through the analysis of Harvard Forest Data which provide information on the spatial distribution of woody stems within the Harvard Forest.

T Di Battista, SA Gattone, University of Chieti-Pescara, Italy, F Fortuna, University of Roma Tre, Rome, Italy, F Maturo, University of Campania, Caserta, Italy

Functional design-based estimation of diversity profiles

In the ecological literature, there is broad consensus that the diversity profile is a useful tool for diversity evaluation. Since the diversity profile is a positive, decreasing, and convex curve, it is possible to cast the problem of its estimation into a constrained functional context. In this work, a functional design-based estimation of diversity profiles is considered by taking into account the constrained nature of the profile. Indeed, a naive direct application of the functional data analysis methodology can be misleading, both theoretically and practically. To tackle this issue, the constrained estimation problem is redefined into an unconstrained one by defining the diversity profile in terms of a differential equation. An approximation of the bias and the variance of the estimator is derived using the delta method.

RM Di Biase, F Mecatti, University of Milano-Bicocca, Italy

Sequential adaptive strategies for sampling rare, clustered populations

Adaptive sampling strategies are particularly effective to perform population-based surveys targeting a rare trait unevenly spread over an area of interest. In this paper, we propose a new class of sampling strategies that integrates an adaptive component into a sequential selection, which aims both to intensify the detection of positive cases, upon exploiting the spatial clustering, and to provide a flexible framework to manage logistics and budget constraints. In particular, we detail two of these strategies, based on the Poisson sampling. Moreover, we propose a class of estimators, which are proven unbiased for the population mean, consistent and asymptotically Normal distributed. Unbiased variance estimation is also provided. A simulation study is performed on a blue-winged teal population in Florida (USA) to illustrate strengths and weaknesses of the proposed sequential adaptive sampling strategies with respect to traditional methods.

J Rodeschini, University of Bergamo, Italy, P Maranzano, University of Milano-Bicocca, Italy, P Otto, Leibnitz University, Hannover, Germany

Comparison of variable selection method in space-time setting : an application to the analysis on PM2.5 concentration in Lombardy (Northern Italy)

This article compares two variable selection techniques suitable for geostatistical models in presence of both spatial and temporal correlation. The preliminary analysis of the features in regression models is a crucial aspect of studying complex phenomena in high-dimensional settings, such as the study of air quality indices (AQ). Specifically, we apply the following variable selection techniques: (i) Hybrid (or two-stage) LASSO and (ii) PMLE in a space-time. Both techniques are used in a univariate geostatistical model (HDGM) in which the dependent variable is the daily concentration of PM2.5 in the Lombardy region (Northern Italy). The Hybrid LASSO method is based on a heuristic approach, while the PMLE uses an Adaptive LASSO penalty function. Both approaches are validated with cross-validation techniques: (i) random k-fold location out (K = 5) and (ii) random leave one location out. To capture the physical complexity of the phenomenon, the design matrix includes data provided by Copernicus, European Environmental Agency (EEA), ARPA Lombardy and other agencies and includes a degree 2 polynomial model with all the interaction terms, and subsequently reduced through the feature selection methods.

A Evangelista, A Sarra, T Di Battista, University of Chieti-Pescara, Italy, CJ Acal, AM Aguilera, University of Granada, Spain, S Palermi, ARTA, Pescara, Italy

The Group Lasso selection method in FDA setting: an optimal detection of PM 10 predictors

This study presents the assessment of PM 10 concentration in Abruzzo region in relation to traffic and meteorological variables. We propose a multiple function-on- function linear regression (MFFLR) model, which integrates a group-LASSO method. The main idea is to reduce the MFFLR model to a classical linear regression model for each principal component of the functional response in terms of all principal components of the functional predictors, by assuming the Karhunen-Loève expansion. Next, a group-LASSO penalty is applied to the parameters derived by basis expansion to automatically filter out irrelevant principal components for predictors. The application of the proposed approach to the environmental problem at hand reveals that most of PM 10 variability is explained by the first two PCs. Through the group-LASSO criterion, we selected the leading PCs of functional predictors. We have found that an important role in defining the PM 10 concentrations is played by pressure, rain, temperature, solar radiation, relative humidity, and nitrogen dioxide.

Inferential strategies in forest studies

A Marcelli, University of Tuscia, Viterbo, Italy, RM Di Biase, University of Milano Bicocca, Italy, P Corona, CREA, Research Centre for Forestry and Wood, Arezzo, Italy, SV Stehman, SUNY College of Environmental Science and Forestry, Syracuse, NY, USA, L Fattorini, University of Siena, Italy

Design-based mapping of land use/land cover classes with bootstrap estimation of precision by nearest-neighbour interpolation

Land use/land cover mapping is usually performed by classifying wall-to-wall satellite imagery e.g. Landsat) using classification algorithms implemented with training data but without quantifying the spatially explicit precision of the map. Subsequently, probabilistic samples are usually implemented

to estimate the area of the land use/land cover classes and to assess the accuracy of these maps. For the first time, these probabilistic samples are exploited to estimate the land use/land cover class at any location of the study region in a design-based framework by means of the well-known nearest-neighbour interpolator. The design-based consistency of these resulting maps is theoretically proven and a pseudo-population bootstrap estimator of their precision is proposed. A simulation study is performed on an estimated land use map in Southern Tuscany - taken as the true map - to check the finite-sample performance of the proposal as well as the matching of the coverage estimates arising from the map with those achieved by traditional estimators. The Italian land use map arising from the IUTI surveys and the US land cover map arising from the LCMAP program are considered as case studies.

A Tomao, D Giuliarelli, E Masini, L Portoghesi, M Agrimi, University of Tuscia, Viterbo, Italy, P Corona, CREA, Research Centre for Forestry and Wood, Arezzo, Italy, L Fattorini, University of Siena, Italy

The ecosystem disservices of trees on sidewalks. A study based on the urban forest inventory of the Municipality of Viterbo, Italy

Urban green infrastructures, including street trees, play a key role in providing environmental, economic, and social benefits to urban residents. However, to fully understand the role of trees in urban context, it is necessary to also evaluate the disservices that trees may produce in the development of their functions if not managed in appropriate and integrated way. This contribution aims to demonstrate an approach to assess three disservices (aesthetic damage, failure hazard, pavement damage) by urban street trees at municipal level, starting from the municipal tree inventory. In the pilot study, from the population of street trees (N = 1769), a sample of 100 trees (i.e., around 5%) was extracted by stratified random sampling, where the strata were composed by tree species groups. In particular, we used a sampling scheme where the probability of a tree to be selected in the sample is higher for larger trees, under the hypothesis that the larger are the trees the greater are the caused disservices. In such a way, larger precision of the unbiased estimates of average or total values is expected for the considered population of urban trees. The results show a relatively low incidence of the selected disservices in our case study, albeit with high variability among species groups.

P Corona. CREA, Research Centre for Forestry and Wood, Arezzo, Italy, S Franceschi, C Pisani, University of Siena, Italy

A design-based critical look of k-NN spatial interpolation

Mapping the spatial pattern of natural and environmental resources is essential for monitoring and management. Many methods of spatial prediction have been developed allowing to construct wall-to-wall maps of the survey variable on the basis of a selected sample of locations or units. Among them, the k-nearest neighbour (k-NN) technique has been widely adopted for a long time. Many applications of the k-NN method are related to the coupling of field data and satellite data for forest and environmental attributes mapping, also owing to the increase in the availability of freely accessible remote sensed data for all locations or units in the population. Notwithstanding its popularity, the statistical properties of the k-NN predictors are still not clearly delineated. The only advanced investigations have been performed in a model-based framework in such a way that the properties of the resulting predictors strictly rely on the validity of the adopted super-population model. As consequence, this work focuses on the theoretically investigation of the asymptotic properties of k-NN predictors in a design-based framework, so that the statistical properties only stem from the probabilistic sampling schemes adopted to perform the surveys.

G Chirici, University of Florence, Italy

The evolution of National Forest Inventories in the era of big data and artificial intelligence

Remote sensing technologies evolved rapidly in the last 10 years because more data are available, with an unprecedented shortening of the revisiting time, and because more tools are available for processing these long multidimensional time series. At the same time forest decision makers at global, national, and local level augmented the amount of information needed to support policy objectives because of the new economic, environmental, and social values of forest ecosystems. National Forest Inventory (NFI) programmes need to rapidly evolve incorporating these new data sources and technologies. This contribution review most recent advancements in the use of remote sensing technologies to derive forest inventory or inventory-related information. More specifically we review the use of trend analysis of dense time series of optical high-resolution imagery for monitoring forest spatio-temporal trends with the use of cloud computing and artificial intelligence methods. Including methods on how this information can be incorporated in tradition NFI programmes offering the required spatial detail and accuracy across large areas.