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

Book of Abstracts

Using Bayesian P-splines for modelling spatial point processes data

L. Altieri[1], F. Greco[1]
[1]Department of Statistical Sciences, University of Bologna, via delle Belle Arti 41, 40126, Bologna, Italy

Abstract. Two of the most discussed topics in recent point process analysis regard the choice of the model estimation method and of techniques for model evaluation and comparison. As for model estimation, a widespread approach consists in superimposing a grid over the observation window in order to discretize the model likelihood. Criticism highlights the relatively expensive computational times and the loss of information about the precise point locations. Starting from descriptive measures suited for point process data, we propose a strategy that allows to build a "sufficiently fine" grid resolution, i.e. a grid resolution that yields a negligible loss of information for the purpose of inference on the spatial distribution of the process intensity. In order to reduce the computational burden associated to fine grids, we propose to model the spatial effect by means of Bayesian P-splines. Model comparison is based on sampling from the posterior predictive distribution of the intensity and building suitable Bayesian p-values.

The use of spatial information in entropy measures

L. Altieri[1], D. Cocchi[1], G. Roli[1]
[1]Department of Statistical Sciences, University of Bologna, via delle Belle Arti 41, 40126, Bologna, Italy

Abstract. The concept of entropy has been firstly introduced in information theory and Shannon's formula of entropy (1948) became rapidly popular in many applied sciences. A relatively recent research field aims at accounting for space in entropy measures, as a generalization when the spatial location of the events ought to be accounted for. The main limitation of these developments is that all indices are computed conditional on a chosen distance. Space is always seen as exogenous, therefore the bivariate properties of entropy cannot be exploited. This work follows and extends the route for including spatial components in entropy measures. We start from the probabilistic properties of Shannon's entropy and investigate the characteristics of the quantity known as conditional entropy in order to include space as a second dimension. This way, we extend the proposal of univariate entropy measures to a bivariate context, in a setting where the probabilistic meaning of all components is well defined. As a direct consequence, we obtain an index satisfying the additivity property, as the global conditional entropy is a sum of partial entropy measures. Moreover, we are able to quantify the information brought by the inclusion of space.

Forecasting wind speed, wind direction and air pressure

D. Ambach[1], W. Schmid[1]
[1]Department of Statistics, European University Viadrina PO Box 1786, 15207 Frankfurt (Oder), Germany

Abstract. There are miscellaneous wind speed forecasting approaches developed in the course of time. However, this article considers an accurate multivariate model which predicts wind speed, wind direction and air pressure altogether. The wind direction and the air pressure are important to extent the forecasting accuracy of our wind speed forecasts. Moreover, accurate forecasting of the wind direction as good as possible can enhance the efficiency of a wind mill. The best forecast for the wind direction helps to bring the turbine into the predominant wind direction. We combine a multivariate seasonal time varying threshold autoregressive model with interactions (TVARX) with a threshold autoregressive conditional heteroscedastic (TARCHX) model. The model includes periodicity, conditional heteroscedasticity, interactions of different dependent variables and a complex autoregressive structure with non-linear impacts.

Adaptive spatial identification and oversampling of sites responding to prescribed profiles using administrative data

F. Andreis[1], M. Bonetti[1]
[1]Bocconi University, Milan, Italy

Abstract. Efficient and reliable monitoring systems are essential components of a functioning management of wildlife and agricultural ecosystems, as pointed out by many authors. Within this framework, survey sampling is usually the tool of choice for practical reasons, hence proposals on how to efficiently implement surveys for environmental monitoring purposes abound in the literature. In particular, during recent years, the topic of adaptive sampling has drawn the attention of the research community and has been studied and applied in a variety of instances in the environmental field. In this contribution, we propose a novel framework for sampling finite populations that uses administrative data as auxiliary information at the design level; our method, by exploiting the predictive capability of state-of-the-art machine learning methods, allows to produce samples in which units possessing some feature of interest are overrepresented, while retaining the possibility of making inference by suitably accounting for the induced imbalance. The potential of the method is demonstrated by means of an application based on an ISPRA air quality dataset, and future lines of research are discussed.

References

Andreis, F. and Bonetti, M. (2016) An adaptive strategy to oversample rare and clustered traits, Proceedings of the 13th International Meeting on *Computational Intelligence Methods for Bioinformatics and Biostatistics*, 1-3 September 2016, Stirling, UK.

Fattorini, L., Corona, P., Chirici, G. and Pagliarella, M.C. (2015) Design-based strategies for sampling spatial units from regular grids with applciations to forest surveys, land use, and land cover estimation. *Environmetrics* **26**, 216-228.

Grafström, A. (2012) Spatially correlated Poisson sampling, *Journal of Statistical Planning and Inference* **142**, 139-147.

ISPRA dataset, ID: 5053bebc-2478-4a16-85bb-d500adf5ba15. Qualità dell'aria nell'ambito delle procedure sullo scambio di informazioni previste dalle Decisioni 97 101 CE e 2001 752 CE.

Thompson, S.K. (2002) *Sampling*, 2nd ed., Wiley, New York.

Spatially balanced sampling of transects

R. Benedetti[1], R.L. Chambers[2], F. Piersimoni[3]
[1]Department of Economic Studies, University "G. d'Annunzio", Viale Pindaro 42, Pescara, Italy
[2]National Institute for Applied Statistics Research Australia (NIASRA), School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, NSW 2522, Australia
[3]Istat, Direzione centrale per la metodologia e il disegno dei processi statistici, Servizio metodi, qualità e metadati , Via Cesare Balbo 16, Rome, Italy

Abstract. Surveys are routinely used to gather data for environmental and ecological research. The units to be observed are often randomly selected from a finite population whose main feature is that it is made up of geo-referenced units. Consequently, the spatial distribution of these units represents basic information that can be used in sample design (Benedetti et al., 2015). Line or point transects are often used to estimate animal abundance, with enumerators recording the number of animals detected and the distance from each animal to the transect. The spread over the population of sampled transects can therefore have a considerable influence on the efficiency of the resulting estimates of abundance. Random sample selection criteria that ensure that selected transects are well spread in two or more dimensions can be used to define spatially balanced (Stevens and Olsen, 2004, Grafström, 2012, Grafström et al, 2012, Grafström and Tillé, 2013) samples of transects. Following Barabesi and Fattorini (2013), we use a model-based framework for estimation and demonstrate the superiority of an optimal spatially balanced sample of transects or points using a mix of design and model-based inference. Numerical comparisons based on real data are used to compare the relative efficiency, measured with respect to simple random sampling, of this optimal design with other approaches to spatially balanced sampling including the widely used Generalized Random Tessellation Stratified (GRTS) design (Stevens and Olsen, 2004, Barabesi and Franceschi, 2011).

References

Barabesi, L. and Fattorini, L. (2013) Random versus stratified location of transects or points in distance sampling: theoretical results and practical considerations, *Environmental and Ecological Statistics* **20**, 215-236.

Barabesi, L. and Franceschi, S. (2011). Sampling properties of spatial total estimators under tessellation stratified designs, *Environmetrics* **22**, 271-278.

Benedetti, R., Piersimoni, F. and Postiglione, P. (2015) Sampling spatial units for agricultural surveys, in *Advances in Spatial Science Series*, Springer, New York.

Grafström, A. (2012) Spatially correlated Poisson sampling, *Journal of Statistical Planning and Inference* **142**, 139-147.

Grafström, A., Lundström, N.L.P. and Schelin, L. (2012) Spatially balanced sampling through the pivotal method, *Biometrics* **68**, 514-520.

Grafström, A. and Tillé, Y. (2013) Doubly balanced spatial sampling with spreading and restitution of auxiliary totals, *Environmetrics* **24**, 120-131.

Stevens, D.L. Jr and Olsen, A.R. (2004) Spatially balanced sampling of natural resources, *Journal of the American Statistical Association* **99**, 262-278.

Semiparametric M-quantile random effect model for indoor radon concentration

R. Borgoni[1], A. Carcagnì[1], N. Salvati[2]
[1]University of Milano-Bicocca, Milano, Italy
[2]University of Pisa, Pisa, Italy

Abstract. Indoor radon concentrations (IRC) depend on building characteristics as well as geological and geophysical factors that induce spatial regularity in IRC measures. However, these effects are far from being constant across the entire support of IRC (Borgoni, 2011). In this paper a M-quantile regression approach (Breckling and Chambers, 1988) is proposed to evaluate their impact at different level of the response variable. In particular, we extend the basic M-quantile model to include a spatial component in addition to other covariates. The spatial component is modelled by combining a random intercept (Chambers and Tzavidis, 2006) to catch the lithology effect on IRC with a semiparametric term, which is expected to grasp residual regularities across space (Pratesi et al., 2009). The flexible component is modeled via a thinplate bivariate spline of the geographical coordinates (longitude and latitude) of the sampling sites. Akin to Ruppert et al. (2003), we propose to treat the coefficients of the knots of the bivariate spline as a further random component in order to obtain smoother results. A robust maximum likelihood approach (Richardson and Welsh, 1995) has been adopted to estimate the model using the two-stage algorithm proposed by Tzavidis et al. (2015). The model is applied to a sample of IRC measures collected in two successive radon campaigns in Lombardy.

References

Borgoni, R. (2011) A quantile regression approach to evaluate factors influencing residential indoor radon concentration, *Environmental Modelling and Assessment* **16**, 239-250.

Breckling, J. and Chambers, R. (1988) M-quantiles, *Biometrika* **75**, 761-771.

Chambers, R. and Tzavidis, N. (2006) M-quantile models for small area estimation, *Biometrika* **93**, 255-268.

Pratesi, M., Ranalli, M.G. and Salvati N. (2009) Nonparametric M-quantile regression using penalised splines, *Journal of Nonparametric Statistics* **21**, 287-304.

Richardson, A.M. and Welsh, A.H. (1995) Robust estimation in the mixed linear model, *Biometrics* **51**, 1429-1439.

Ruppert, D., Wand, M.P. and Carrol, R.J. (2003) *Semiparametric Regression*, Cambridge Series in Statistical and Probabilistic Mathematics, CUP, Cambridge.

Tzavidis, N., Salvati, N., Schmid, T., Flouri, E. and Midouhas, E. (2015) Longitudinal analysis of the strengths and difficulties questionnaire scores of the millennium cohort study children in England using M-quantile random effects regression, *Journal of Royal Statistical Society* Series **A179**, 427-452.

Semiparametric statistical modeling for analysis of ecological processes

M. Brabec[1]
[1]Department of Nonlinear Modeling, Institute of Computer Science, The Czech Academy of Sciences, Praha, Czech Republic

Abstract. In this talk, we will describe a general approach to a broad class of ecological modeling problems based on modern semiparametric regression, additive models and their extensions with complexity penalties which can be motivated in a Bayesian way. Subsequently, we will illustrate the power of this approach on several examples of our previous work which has been done in collaboration with ecologists engaged in long-term forest composition studies. The illustrations will aim at somewhat nonstandard problems whose statistical treatments leads to formulations which are in the sprit of functional data analysis (FDA) and at spatial process modeling based on efficient computational approximations which will be useful for otherwise rather complicated spatial interaction.

Corals growth parameter estimation via hierarchical non-linear mixed-effects models

B. Cafarelli[1], C. Calculli[1], D. Cocchi[2]
[1]Department of Economics, University of Foggia, largo Papa Giovanni Paolo II 1, Foggia, Italy
[2]Department of Statistical Sciences, University of Bologna, via delle Belle Arti 41, Bologna, Italy

Abstract. The evaluation of demographic parameters in marine ecosystems is a crucial issue for conservation strategies and management purposes. Corals represent key species at this regard, especially in temperate waters. Their growth is satisfactorily described by the well-known von Bertalanffy function. Linearization techniques are currently employed for predicting growth, without taking care about the implicit assumption of limiting hypotheses and their lack of statistical accuracy. A more suitable approach based on hierarchical non-linear mixed-effects (HNLM) models is suggested in order to overcome any undue simplification in the non-linear relationship between coral growth and length. Such approach allows to model, at the same time, the influence of different sources of variability and the contribution of environmental factors on corals growth. The effectiveness of the proposal is assessed via a case study that analyzes the growth of solitary Mediterranean coral species, repeatedly measured at different sites.

Extreme rainfall events analysis for understanding hydrologic change

V. Chiarello[1], E. Caporali[1], A. Petrucci[2]
[1]Dipartimento di Ingegneria Civile e Ambientale, Università degli Studi di Firenze, Firenze, Italy
[2]Dipartimento di Statistica, Università degli Studi di Firenze, Firenze, Italy

Abstract. Investigations on extreme hydrological events are often at the basis of environmental studies related to hydrologic cycle changes and more in general with climate change. Among the hydrological variables, extreme rainfall represents the paramount variable for the implications on flood risk assessment and territory protection measures definition. A regional frequency analysis for studying and understanding the annual maxima of rainfall depth, based on the index variable procedure, is implemented here in the catchment area of the Arno River in Tuscany. A geoadditive model for extremes assuming that the observations follow generalized extreme value distributions whose locations are spatially dependent is also implemented. In order to discuss the application of the two methods, the comparison of the maps of annual maxima of daily rainfall for different return periods is carried out.

A two-stage sampling strategy for assessing forest windthrow damages exploiting post-event Airborne Laser Scanning

G. Chirici[1], F. Bottalico[1], F. Giannetti[1], D. Travaglini[1], L. Fattorini[2], P. Corona[3]
[1]Università degli Studi di Firenze, Dipartimento di Gestione dei Sistemi Agrari, Alimentari e Forestali - GESAAF, Via San Bonaventura 13, Firenze, Italy
[2]Department of Economics and Statistics, P.za S. Francesco 8, 53100 Siena, Italy
[3]Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, Forestry Research Centre (CREA-SEL), Viale Santa Margherita 80, Arezzo, Italy

Abstract. In this study we present a two-stage sampling strategy exploiting a post-event single date Airborne Laser Scanning (ALS) data to assess windstorm damages in forest landscapes. ALS data are used for delineating damaged forest stands and for wall-to-wall evaluating the volume of fallen trees. Estimation of the total volume of fallen trees is then developed through a model assisted approach, where the ALS evaluation is used in the difference estimator as auxiliary variable. In the first stage a sample of the delineated forest stands is selected and the damages within are estimated in the second stage by means of line intercept sampling (LIS). The proposed method is able to produce both mapping of windthrown areas and the estimation of forest damages in terms of total volume of fallen trees with associated uncertainty. A case study is presented with reference to the huge windstorm that hit the Tuscany Region (Italy) hit by the 5th of March 2015 and caused extensive damages to trees in forest and urban areas. The relationship between ASL-based evaluations and field LIS estimates of fallen wood at stand level proved to be very close.

Citizen sciences and biodiversity monitoring

C. Coron[1]
[1]Université Paris Sud, Bâtiment 425, 15 Rue Georges Clémenceau, 91400 Orsay, France

Abstract. Citizen science programs have been largely developped for the last twenty years, notably in the eld of ecology and naturalism. In this work, we use two datasets resulting from such programs: a standardized dataset (provided by scientists) and an opportunistic dataset (provided by citizens). The rst one presents a small amount of data that are associated with a precise observation and reporting protocole, while the second one gathers a very large number of inhomogenous data. We aim at estimating relative abundances of several species on different regions of space and/or at dierent moments. We model data by Poisson variables whose intensity is unknown, and we first prove that combining these two types of datasets leads to a more precise estimation of these abundances, than provided by the sole standardized dataset. Second, taking into account habitat structuration of space and respective preferences of observers and observed individuals for different habitat types allows a more realistic and more proper application of our estimations.

Estimating animal home ranges and core areas using Brownian Bridge movement models: the effect of filtering low quality location data

D. De Angelis[1], S. Mancinelli[1], P. Ciucci[1]
[1]Department of Biology and Biotechnologies "C. Darwin", La Sapienza University of Rome, Rome, Italy

Abstract. The home range (HR) is amongst the most fundamental parameters when investigating space use patterns by animals. The recently proposed Brownian bridge movement models (BBMMs) are theoretically superior compared to other traditional HR estimators, as they alleviate the temporal and spatial autocorrelation that characterizes animal locations recorded at high frequency, such as those collected through GPS (Global Positioning System) devices. However, this estimator is still little used by biologists, probably due to the wide availability of commercial software solutions implementing traditional kernel HR estimators. In addition, no empirical test has ever been conducted to assess the effect of subsampling (i.e., filtering GPS locations of low quality) on BBMMs. By using GPS locations recorded on wide ranging species such as brown bears and wolves, the aims of our work are to: (i) illustrate the application of BBMM to typically highly correlated GPS data; (ii) implement an individual-based method to define core areas within the HR using BBMMs; and (iii) assess the effect on the HR estimates of removing GPS locations of lower accuracy according to increasing subsampling levels.

Sample-based spatial statistics: reconciling two literatures

M.M. Dickson[1], D. Giuliani[1], G. Espa[1]
[1]Department of Economics and Management, University of Trento, Trento, Italy

Abstract. A large body of spatial statistics for the analysis of spatial patterns of ecological and environmental data (from geostatistics to point pattern methods) generally does not take into account how data are collected, as it generically assumes simple random sampling (Ripley, 1988; Diggle, 2003, Cressie, 2015). However, in many real environmental applications data are collected according to more complex sampling designs, implying that the reliability of inferential results may not necessarily hold. In this work, we argue that a convenient approach to address this issue is to borrow from sampling theory in order to, on one hand, readjust estimation procedures when appropriate and, on the other hand, develop proper sampling designs to collect spatial data. As a worthwhile starting point, in the present work, we propose a general class of sample estimators of the main measures of spatial autocorrelation (such as the Moran's I and Geary's C indices) that allow accurate estimates and valid inference even when data are collected with an unequal probability sampling design. We investigate their statistical properties and we assess their performance, in terms of precision of the estimates and power of the test, through Monte Carlo simulation experiments. Moreover, we also focus on the problem of collecting sample data, which maintain the same spatial characteristics of population data. In particular, we study the development of a sampling strategy to select random samples that match the form and the level of spatial autocorrelation of the target variable in the reference population.

References

Cressie, N. (2015) *Statistics for Spatial Data*, revised edn, Wiley, New York.
Diggle, P.J. (2003) *Statistical Analysis of Spatial Point Patterns*, 2nd edn, Edward Arnold, London.
Ripley, B.D. (1988) *Statistical Inference for Spatial Processes*, Cambridge University Press, Cambridge.

Sampling and modelling spatial data with locational errors

M. M. Dickson[1], E. Taufer[1], M. Bee[1], D. Giuliani[1], G. Espa[1]
[1]Department of Economics and Management, University of Trento, Trento, Italy

Abstract. Geo-referenced registers are affected by locational errors, which could be an important source of bias in studies on spatial populations. These imprecisions could be caused by unintentional or intentional procedures, which are respectively due to technical problems in the geo-coding phase or to privacy protection issues. Consequently, units interested by locational errors are inaccurately positioned on the territory under study, result of placement of units to centroids of sub-regions or geo-masking procedures (Zimmermann, 2008, Allshouse et al., 2010, Jacquez, 2012). The two described types of locational errors can represent a relevant problem when conducting studies on natural resources data. In the present work, we show the effect of unprecise locations, both in sampling and in estimation. First, we propose a study, based on simulated and real environmental data, with the aim to understand the effects of locational errors in the selection of units for some spatial sampling designs, at different proportion of uncertain locations. Furthermore, we investigate the effect of not properly located units on spatial autoregressive models (see details in Cliff and Ord, 1981), commonly used in environmental context. We analyse both cases of partial or complete imprecision in the data, presenting also in this case a detailed discussion based on real and simulated data.

References

Allshouse, W.B., Fitch, M.K., Hampton, K.H., Gesink, D.C., Doherty, I.A., Leone, P.A., Serre, M.L. and Miller, W.C. (2010) Geomasking sensitive health data and privacy protection: An evaluation using an E911 database, *Geocarto International* **25**, 443-452.

Cliff, A.D. and Ord, J.K. (1981) *Spatial Processes: Models and Applications*, Taylor & Francis, London.

Jacquez, M.G. (2012) A research agenda: does geocoding positional error matter in health GIS studies?, *Spatial and Spatio-temporal Epidemiology* **3**, 7-16.

Zimmerman, D.L. (2008) Estimating the intensity of a spatial point process from locations coarsened by incomplete geocoding, *Biometrics* **64**, 262-270.

Estimation and testing procedures in Co-Regionalization Analysis with a Drift (CRAD): a timely summary, with applications using vegetation, terrestrial and atmospheric data

P. Dutilleul[1]
[1]McGill University, Montréal, Canada

Abstract. The abbreviation CRAD stands for "Co-Regionalization Analysis with a Drift", a statistical method of spatial data analysis performed in two phases: Phase 1, to estimate drifts in a multivariate spatial dataset, and Phase 2, to analyze correlations from the residuals, i.e., after removal of the estimated drifts. Pelletier et al. (2009a, 2009b) introduced CRAD as an alternative to the original "Co-Regionalization Analysis" (CRA), where the large-scale component of variability in the multivariate spatial dataset is considered random and modeled via a variogram (Goulard and Voltz, 1992, Goovaerts and Webster, 1994). In preliminary steps, Pelletier et al. (2004) had proposed a Generalized Least-Squares (GLS) estimation procedure for co-regionalization matrices in a Linear Model of Co-regionalization (LMC) and Larocque et al. (2007) had shown that even after this improvement, the uncertainty associated with the estimation of parameters in the LMC can still be very high in CRA. On the testing side, Dutilleul and Pelletier (2011) defined tests of significance for structural coefficients of correlation in CRA and CRAD applications. Recently, still in a spatial frame and as a further and novel extension to Dutilleul (1993) and Dutilleul et al. (2008), Dutilleul and Pelletier (unpublished) proposed a modified F-test of significance for the average R2 in "Redundancy Analysis" (RDA) (Rao, 1964, van den Wollenberg, 1977). (N.B.: The average R2 is the proportion of variance in the criterion variables that is reproducible linearly from the predictor variables in the RDA). In this talk, we will navigate in the CRAD world. First, I will give a broad description of the framework, with the main steps and their sequence. Then, I will explain with a good level of detail the procedures used for estimation and testing, e.g., by discussing the underlying assumptions. Thereafter, the focus will be on the most recent testing results. Finally, I will present a number of examples involving plant diversity, soil physicochemistry and greenhouse gas emission. A good part of the presented material is joint work with Bernard Pelletier (McGill University).

References

Dutilleul, P. (1993) Modifying the t-test for assessing the correlation between two spatial processes, *Biometrics* **49**, 305-314.

Dutilleul, P. and Pelletier, B. (2011) Tests of significance for structural correlations in the linear model of coregionalization, *Mathematical Geosciences* **43**, 819-846.

Dutilleul, P. and Pelletier, B. (2016) A valid parametric test of significance for the average R2 in redundancy analysis with spatial data. Unpublished manuscript.

Dutilleul, P., Pelletier, B. and Alpargu, G. (2008) Modified F-tests for assessing the multiple correlation between one spatial process and several others, *Journal of Statistical Planning and Inference* **138**, 1402-1415.

Goovaerts, P. and Webster, R. (1994) Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland European, *Journal of Soil Science* **45**, 79-95.

Goulard, M. and Voltz, M. (1992) Linear coregionalization model: tools for estimation and choice of crossvariogram matrix, *Mathematical Geology* **24**, 269-286.

Larocque, G., Dutilleul, P., Pelletier, B. and Fyles, J.W. (2007) Characterization and quantification of uncertainty in coregionalization analysis, *Mathematical Geology* **39**, 263-288.

Pelletier, B., Dutilleul, P., Larocque, G. and Fyles, J.W. (2004) Fitting the linear model of coregionalization by generalized least squares, *Mathematical Geology* **36**, 323-343.

Pelletier, B., Dutilleul, P., Larocque, G., and Fyles, J.W. (2009a) Coregionalization analysis with a drift for multiscale assessment of spatial relationships between ecological variables 1. Estimation of drift and random components, *Environmental and Ecological Statistics* **16**, 439-466.

Pelletier, B., Dutilleul, P., Larocque, G. and Fyles, J.W. (2009b) Coregionalization analysis with a drift for multiscale assessment of spatial relationships between ecological variables 2. Estimation of correlations and coefficients of determination, *Environmental and Ecological Statistics* **16**, 467-494.

Rao, C.R. (1964) The use and interpretation of principal component analysis in applied research, *Sankhya* **A26**, 329-358.

van den Wollenberg, A.L. (1977) Redundancy analysis: An alternative to canonical correlation analysis, *Psychometrika* **42**, 207-219.

Chao and Zelterman estimators with observed heterogeneity, with application to a whale shark population

A. Farcomeni[1]
[1]La Sapienza - University of Rome, Deparment of Public Health and Infectious Diseases, Rome, Italy

Abstract. Chao estimator provides a lower bound for population size estimates under unobserved sources of heterogeneity without the need to specify a distribution for them. Zelterman estimator is robust to misspecification of the counting distribution. We describe generalized Zelterman and Chao estimators which include subject-occasion specific covariates, time-heterogeneity and behavioural responses to detection. As noted by Farcomeni and Scacciatelli (2013), there are no additional possible sources of (observed) heterogeneity. Taking into account known sources of heterogeneity within these estimators might lead to substantially lower mean squared errors. In fact, our generalized estimators have the same properties of the original ones, with possibly dramatic reductions in bias. For model choice we develop an ad-hoc biased empirical Focused Information Criterion. We illustrate with an original study about the population of Whale Sharks (Rhyncodon typus) in the South Ari atoll of Maldives, where time-heterogeneity and subjectoccasion specific covariates (e.g., length at each sighting) have an important effect.

Source of variability in Chlorophyll a fluorescence of Norway spruce (Picea abies (L.) H. Karst) trees - consequences for sampling

M. Ferretti[1], G. Bacaro[2], F. Cristofolini[3], A. Cristofori[3], S. Maccherini[4], E. Gottardini[3]
[1]TerraData environmetrics, Monterotondo M.mo, Grosseto, Italy
[2]Department of Life Sciences, Università degli Studi di Trieste, Trieste, Italy
[3]Research and Innovation Centre, Fondazione Edmund Mach (FEM), San Michele all'Adige, Trento, Italy
[4]Department of Life Sciences, Università degli Studi di Siena, Siena, Italy

Abstract. Forest health surveys in Europe are based on visual assessment of the reduction of the density of foliage on tree crown, termed crown transparency. This attribute has been criticized because its unclear relationship with tree physiological status and the need for more objective indicators of tree health has been frequently claimed. Chlorophyll a fluorescence of tree foliage permits to evaluate plant vitality in relation to different stressors and provides information on the status of photosynthetic apparatus of foliage of individual plants and stands. The question is therefore how to obtain reliable estimates of Chorophyll a fluorescence at tree level. Here we present the results of a pilot study carried out on Norway spruce trees in Trentino (N. Italy) under two different condition: juvenile trees growing into a forest nursery, and adult trees into a real forest. The aims were (i) to evaluate sources of variability that should be accounted for when designing a sampling strategy for Chorophyll a fluorescence and (ii) to estimate the required sampling effort.

Sampling an epiphytic lichen community with an adaptive design

S.A. Gattone[1], P. Giordani[2], T. Di Battista[1], F. Fortuna[1]
[1]University "G. d'Annunzio", Viale Pindaro 42, Pescara, Italy
[2]LUISS "Guido Carli" University, Rome, Italy

Abstract. Epiphytic lichens are sensitive organisms widely occurring in most terrestrial ecosystems. Obtaining robust information on the abundance and the distribution of these species at fine scale is a fundamental step for planning adequate conservation plans. However several lichen species are presently threatened, because their distribution is often spatially clustered on both large and local scales. The large-scale clustering is mainly due to the effect of habitat fragmentation, whereas on a local scale, both local dispersal and environmental conditions influence the spatial distribution and abundance of epiphytic lichen species. In this context, a suitable sampling design is given by adaptive cluster sampling useful to sample rare and clustered populations. In order to control the final sampling effort adaptive cluster sampling is combined with double sampling for stratification. This strategy consists in a two phase sampling. In the first phase the adaptive cluster sampling is used with a rapid assessment variable. The auxiliary information on the units selected in the first phase is then used for obtaining the strata. In the second phase, subsamples in each stratum are selected with proportional allocation. The novelty of this approach consists on the use of poststratification which may lead to gain in precision for the estimators. In this framework, a new type of estimator for the population mean which mimics the very simple estimator of stratified sampling is proposed. The sampling design has been used for estimating the population total of an epiphytic lichen community Lobarion pulmonariae in a forest area of the Northern Apennines (N-Italy), characterized by several species of conservation concern.

Efficient smoothing for spatial data collected over the globe

F. Greco[1], M. Ventrucci[1]
[1]Department of Statistical Sciences, University of Bologna, via delle Belle Arti 41, 40126, Bologna, Italy

Abstract. Penalized spline (P-spline) regression on a basis of B-splines is a stable and computationally efficient technique for smoothing. The key idea in P-splines is to model the smooth function as the sum of scaled B-spline basis functions, where each basis is defined over an arbitrarily large number of equally-spaced knots. Smoothness is imposed by a penalty on second order differences between the spline coefficients. This approach is frequently used to model spatial or spatio-temporal random effects in flexible additive regression models. In the spatial case, the basis is usually built as the tensor product of two marginal bases, for both latitude and longitude coordinates, taking knots on a regular grid (i.e. maintaining regular spacing in terms of Euclidean distances over latitude and longitude domain). However, when working with spatial data collected over wide areas, e.g. over the whole globe, a straightforward application of the P-spline method would require the definition of knots which are equally-spaced in terms of geodesic distances (i.e. maintaining regular spacing in terms of arc length distances over the globe). This work explores the relevance of geodesic grid systems in building efficient spatial models for global data. The example motivating our research is taken from a real case in climate studies, where the goal is to identify the Intertropical Convergence Zone using high-resolution remote sensing data.

Semiparametric regression with errors in covariates for the analysis of sensor data on soil-plant variability

G. Jona Lasinio[1], A. Pollice[2], R. Rossi[3], M. Amato[4], T. Kneib[5]
[1]La Sapienza Università di Roma, Rome, Italy
[2]Università degli Studi di Bari "Aldo Moro", Bari, Italy
[3]Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Rome, Italy
[4]Università della Basilicata, Potenza, Italy
[5]Georg-August-Universität, Göttingen, Germany

Abstract. Alfalfa is one of the most important forage legumes in the world. Tools for the rapid assessment of soil-plant variability are crucial for the implementation of tactical and strategic management. We have analyzed the relationship between alfalfa NDVI repeated measurements and multi-depth soil electrical resistivity (ER) used as a proxy of soil variability. Observations of both quantities were made using sensors at a very refined spatial resolution. NDVI data were obtained at 4 sampling occasions with point locations changing over time and the number of samples ranged from 110k to 250k. ER data where obtained in one occasion at three different depths. Modeling these data involves the solving of several issues mostly linked to the spatial and temporal misalignment and the large data size. In this work we are going to present a model based approach to all of the above. NDVI and ER data were interpolated (upscaled) on a 2574 points square grid covering the field. To account for the uncertainty on the covariate (ER) we adopted a measurement error type approach and to analyze the relation between NDVI and ER a semiparametric approach is illustrated. Estimates are obtained in the Bayesian framework using BayesX implementation.

Covariances between environment and animals: disentangling the effects of observation and process in spatially explicit studies

V. La Morgia[1], S. Focardi[2]
[1]Department of Wildlife, Italian National Institute for Environmental Protection and Research (ISPRA), Bologna, Italy
[2]Institute for Complex System, National Research Council (ISC-CNR), Firenze, Italy

Abstract. Many wildlife conservation and management programs rely on the assumption that correlative evidences between environmental features and animal presence (or abundance) may suggest some kind of cause-effect relationships. Such correlative analyses often fail to disentangle the environmental effects on our capacity to detect the species of interest, from the effects of such factors on the processes underlying animal distribution. In this paper we review two methods, density surface and occupancy modelling, that actually allow to explicitly model both observation and state processes, and we illustrate these procedures with a number of different examples. These include the application of Density Surface Modelling techniques to study the spatial interaction between deer species, and the use of occupancy modelling to assess the actual relationship between the presence of elusive mammal species (mountain hares and snow voles) and habitat features. Spatially-explicit studies may significantly contribute to understand ecological processes and the impact of anthropogenic stressors on animal distribution, but the assessment and monitoring of the effects of environmental variables should rely on recently established methodologies, correcting for animal detectability.

Recent experiences using the Italian Land Use Inventory to assess land use changes in Italy

M. Marchetti[1], L. Sallustio[1], P. Corona[2], M.C. Pagliarella[3], M. Munafò[4], B. Lasserre[1], L. Fattorini[3]
[1]Dipartimento di Bioscienze e Territorio, University of Molise, Contrada Fonte Lappone, Pesche, Isernia, Italy
[2]Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, Forestry Research Centre (CREA-SEL), Viale Santa Margherita 80, Arezzo, Italy
[3]Dipartimento di Economia e Statistica, University of Siena, Piazza San Francesco 8, Siena, Italy
[4]Italian National Institute for Environmental Protection and Research (ISPRA), Via V. Brancati 48, Rome, Italy

Abstract. The Land Use and Land Cover Changes (LUC) can be analyzed using both cartographic and inventory approaches. The latter, in particular, provide estimates of the accuracy of the sampling strategy adopted , allowing objective and scientifically sound comparisons of the estimates at different times. The possibility to assess the statistical accuracy and the possibility of frequent updates suggest the inventory approach as a valid and reliable instrument for the LUC assessment. The evaluation of LUC and the selection of a reliable and accurate approach usable as a standard for a large series of experiences plays a primary role as a support for future land use planning. Based on a tessellation stratified sampling (TSS) the Italian Land Use Inventory (IUTI) has been further implemented during last years to better understand and characterize LUC, their causes and possible effects on ecosystem services. Particularly, the principal implementations and experiments aimed at: a) analyzing recent LUC in Italy (particularly forest expansion and urban growth); b) proposing the integration and comparison of land use and land cover analysis as a quick and effective instrument to provide essential information to support land use planning; c) proposing the integration of inventory and cartographic approaches to assess the impact of LUC on ecosystem services provision; d) proposing and testing a two-phase strategy for LUC estimation based on a one-per-stratum stratified sampling (OPSS) to reduce costs and time but minimizing the precision lost.

How to reduce survey costs in evaluating the growing trends in ICP forest plots

M. Marchi[1], C. Ferrara[2], G. Bertini[1], G. Fabbio[1], S. Fares[2], L. Salvati[2]
[1]Council for Agricultural Research and Economics, Forestry Research Centre (CREA-SEL), Viale S. Margherita 80, Arezzo, Italy
[2]Council for Agricultural Research and Economics, Research Centre of Plant-Soil Interactions (CREA-RPS), Via della Navicella 2-4, Rome, Italy

Abstract. High-quality spatial data and long-term time series are the basis of any research activities dealing with natural resources analysis. However, measurements over space and/or time can rarely be implemented due to both survey costs and the investigated phenomena structure complexity. As a result, an adequate sampling design is fundamental to allow a robust statistical analysis to be representative of a relevant target variables set. The growth-trend can be considered a good proxy of forest health and, in this work, the sampling strategy of CONECOFOR ICP plots has been evaluated to define a more efficient and cost-effective procedure. Two different sub-sampling strategies have been implemented based on a measure of (i) the dominant layer, and (ii) a random sub-sample of sub-squared plots. A result comparison shows that the growth trends of the dominant layer are strictly related to auto-ecology species as well as spatial structure and age of the forest. Consequently, on the one hand, sampling effort could not be effectively reduced; on the other hand, the sub-sampling method seems to be a good and flexible procedure, influenced only by the horizontal distribution of trees and their coefficient of variation.

Spatio-temporal geostatistics of functional data: some preliminary results in analysis of temperature vertical profile

F. Ndongo[1], A. Fassò[1], F. Finazzi[1]
[1]University of Bergamo, Bergamo, Italy

Abstract. It is not trivial to capture the spatio-temporal dependence due to the interaction between space and time encapsulated in a spatio-temporal data. The concept of functional data brought by Functional Data Analysis (FDA) allows to consider a vector of finite points as a single entity treated as function or curves. Exploiting that concept into a spatio-temporal framework does not necessary carry out the spatio-temporal complexity of the phenomenon under study. Since the function is built along one dimension and data are considered a single entity along this dimension, the information along this dimension will be hide and it will carry out a marginal description of the phenomenon along the other dimension. For instance, if we use the temporal dimension to build the functions, we will have spatially correlated functions and spatial statistics is performed losing any information about time. The same order of problem occurs when we use spatial dimension for building the functions, we build temporally correlated functions and time series analysis is performed. Therefore, it is clear that we have to build spatio-temporally correlated functions in other to take into account the spatio-temporal complexity of the phenomenon under study. For this purpose, here we consider a third dimension along which we build the functions in order to compute analysis along the other two. As a motivating example, we will consider some preliminary results in the analysis of atmospheric vertical profiles. In particular, we consider the global space-time series of radiosonde observations related to atmospheric profiles of Essential Climate Variable (ECV) such as temperature given by RAOB network (www.raob.com).

Spatially-balanced sampling vs unbalanced sampling for assessing forest change: evidence in favour of spatial balance

M.C. Pagliarella[1], P. Corona[2], L. Fattorini[3]
[1]University of Perugia, Perugia, Italy
[2]Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, Forestry Research Centre (CREA-SEL), Viale Santa Margherita 80, Arezzo, Italy
[3]Dipartimento di Economia e Statistica, University of Siena, Piazza San Francesco 8, Siena, Italy

Abstract. After Kyoto protocol, the monitoring of forest cover at large scale plays a prominent role. Effective forest change assessment is one of the main target. With this aim, we investigated two alternative strategies. We carried out a comparison between a spatially balanced sampling, in which spatial units are selected all over the study area in order to obtain an even spread of the sampled units, against an unbalanced sampling where the spatial units are mostly selected where change is present. Estimates are accomplished with and without using auxiliary information, respectively by means of the Horvitz-Thompson and difference estimators. In order to check the validity of the strategies proposed and to assess the accuracy achieved, simulation experiments are carried out on real data regarding two areas located in the Central Italy where high and low change occurred between 2000 and 2012. The results are in favor of balanced schemes

Observations of storms flooding along a natural sandy beach

L. Parlagreco[1], G. Aretusi[2], L. Fontanella[3], L. Ippoliti[3], P. Valentini[3], S. Devoti[1], L. Liberti[1]
[1]Istituto Superiore per la Protezione e Ricerca Ambientale, Roma, Italy
[2]Torre del Cerrano, Area Marina Protetta, Italy
[3]University G. d'Annunzio, Chieti-Pescara, Italy

Abstract. In the last decade, the scientific community has paid increasing attention to the development of vulnerability assessment methods for coastal communities and infrastructure to flooding phenomenon. In this work, we discuss the use of optical remote sensing observations to derive the wave field on the shore and to estimate the impact of beach flooding during storms. We describe the use of short time series (ten minutes per hour) of video data collected at 2Hz from a videomonitoring station, known as timestack (Chickadel et al., 2003, Stockdon et al., 2006, Holmann and Stanley, 2007, Flores et al., 2016), and discuss the associated statistical issues of estimating the single storm impact on a beach. The study refers to the natural sandy coast of Torre del Cerrano (Abruzzo).

References

Chickadel, C., Holman, R. and Freilich, M. (2003) An optical technique for the measurement of longshore currents, *Journal of Geophysical Research* **108**.

Stockdon, H., Holman, R., Howd, P. and Sallenger, H. (2006) Empirical parameterization of setup, swash, and runup, *Coastal Engineering* **53**.

Holman, R. and Stanley, J. (2007) The history and technical capabilities of Argus, *Coastal Engineering* 54.

Flores, R., Catalán, P. and Haller, M. (2016) Estimating surfzone wave transformation and wave setup from remote sensing data, *Coastal Engineering* **114**.

Identifying spatial regimes in environmental and ecological models

P. Postiglione[1], R. Benedetti[1], A. Cartone[1]
[1]Department of Economic Studies, University "G. d'Annunzio", Viale Pindaro 42, Pescara, Italy

Abstract. This paper is aimed at examining the presence of spatial pattern of parameters when modeling geocoded environmental and ecological data. The empirical evidence often suggests that we are very far from being able to accept the assumption that the parameters vary in space smoothly and continuously as is expected from the typical solution suggested by Geographically Weighted Regression (GWR). The existence of natural boundaries and of areas of similar ecology, delineated by homogenous topography, temperature, and land cover, gives rise to the adoption of a more realistic hypothesis concerning spatial regime regressions that allow the model coefficients to vary between discrete spatial subsets of the data. We observed that the fit of several models is not satisfactory if, in the empirical analysis, we consider the entire data set as one sample. The phenomenon should be analyzed considering zones as belonging to different sub-samples with similar ecological status. The existence of this heterogeneity affects spatially distributed data, and often is neglected in the analysis. We review and compare the performance on real data sets of four different approaches that have been introduced in literature. The findings reveal that, computationally intensive, purely spatial algorithms (Postiglione et al., 2010, 2013) accomplish better results than multipurpose solutions as the EM algorithm (Dempster et al., 1977) used to identify finite mixture regressions (Turner, 2000) and binary recursive partitioning based on parametric models (Zeleis et al., 2008).

References

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood fromincomplete data via the EM algorithm, *Journal of the Royal Statistical Society* **B39**, 1-22.

Postiglione, P., Andreano, M.S. and Benedetti, R. (2013) Using constrained optimization for the identification of convergence clubs, *Computational Economics* **42**, 151-174.

Postiglione, P., Benedetti, R. and Lafratta, G. (2010) A regression tree algorithm for the identification of convergence clubs, *Computational Statistics and Data Analysis* **54**, 2776-2785.

Turner, T.R. (2000) Estimating the rate of spread of a viral infection of potato plants via mixtures of regressions, *Applied Statistics* **49**, 371-384.

Zeileis, A., Hothorn, T. and Hornik, K. (2008) Model-based recursive partitioning, *Journal of Computational and Graphical Statistics* **17**, 492-514.

A two-phase sampling strategy for estimating tree species diversity from large scale forest inventories

N. Puletti[1], M. Marcheselli[2], C. Pisani[2], P. Corona[1]
[1]Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, Forestry Research Centre (CREA-SEL), Viale Santa Margherita 80, Arezzo, Italy
[2]Department of Economics and Statistics, P.za S. Francesco 8, 53100 Siena, Italy

Abstract. Estimating forest biodiversity is an important issue for environmental management. A two-phase strategy is here proposed. In the first phase, Tessellation Stratified Sampling is performed by partitioning the study area into a grid of quadrats and randomly selecting a point in each quadrat. The first-phase points are subsequently classified as forest or non-forest using aerial imageries and, in the second phase, a sample of points is selected from those classified as forest by means of Simple Random Sampling Without Replacement. The second-phase points constitute the centers of circular plots which are visited on the ground to quantify tree species and abundance. This strategy is applied to estimate tree abundance and biodiversity indicators in Casentino forests (Central Italy).

Sensing the dierence: measuring Rao's Q diversity from satellite images in an open source space

D. Rocchini[1], M. Marcantonio[1,2], C. Ricotta[3]
[1]Fondazione Edmund Mach, Research and Innovation Centre, Department of Biodiversity and Molecular Ecology, Via E. Mach 1, S. Michele all'Adige, Trento, Italy
[2]Technische Universität Berlin, Department of Landscape Architecture and Environmental Planning, Geoinformation in Environmental Planning Lab, Berlin, Germany
[3]Department of Environmental Biology, University of Rome "La Sapienza", Rome, Italy

Abstract. Measuring biodiversity is a key issue in ecology to protect and manage ecosystems in a more efficient manner. However, estimating biodiversity from field observations might present difficulties related to costs and time needed. Moreover, a continuous data update for biodiversity monitoring purposes might be difficult. From this point of view, remote sensing represents a powerful tool since it allows to cover wide areas in a relatively low amount of time. One of the most commonly used indices of biodiversity is Shannon's entropy H, which is strictly related to environmental heterogeneity, and thus to species diversity. However, Shannon's entropy might show drawbacks once applied to remote sensing data, since it considers relative abundances but it does not explicitly account for distances among pixels' numerical values. In this study we propose the use of Rao's Q applied to remotely sensed data, providing a straightforward R-package function to calculate it in 2D systems.

Urban development and wildfires: a panel data analysis on Italian provinces

L. Secondi[1], T. Laureti[1], A. Tomao[1], L.D. Mancini[1], D. Giuliarelli[1], P. Corona[2]
[1]University of Tuscia, Viterbo, Italy
[2]Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, Forestry Research Centre (CREA-SEL), Viale Santa Margherita 80, Arezzo, Italy

Abstract. Wildfires are major disturbances in Mediterranean ecosystems worldwide. Various causes including biophysical and anthropic drivers have been proved to affect the probability of fire occurrence. Furthermore, the vast majority of wildfires are human-induced, particularly in the Countries around the Mediterranean basin. Despite its importance, the influence of human factors on the spatial and temporal patterns of wildfire needs to be better understood. The aim of this study is to explore the relationships between socio-economic variables and wildfire occurrence for a panel of Italian provinces. A rich dataset is constructed by integrating wildfire occurrence data with information on the socio-economic and demographic characteristics taken from various official sources. Fixed and random effects panel data models are used in order to properly examine the relationships between urbanization and wildfires. We evaluate and discuss the relationships among wildfire occurrence and human development activities including land-use intensity, economic intensity, population density, spatial distribution of infrastructure and public service facilities.

A generalized regression estimator for spatial population

A. Vagheggini[1], D. Cocchi[1]
[1]Department of Statistical Sciences, University of Bologna, Bologna, Italy

Abstract. Generalized Regression (GREG) estimator is a powerful tool which, in a design-based finite population framework, allows to correct the estimates of synthetic population quantities through the use of auxiliary variables known for all the elements before sampling. In this work we propose a modified version of the GREG estimator, which, by employing the Geographically Weighted Regression (GWR) model, is able to take into account the spatial relationship between locations as well as use auxiliary variables. Theoretical and simulated results show that the estimator has good properties and, at the same time, outperforms the usual Horwitz-Thompson estimator.

Is vantage point count a reliable method for monitoring roe deer populations?

M. Zaccaroni[1], F. Dell'Agnello[1], G. Ponti[1], F. Riga[2], L. Fattorini[3]
[1]University of Florence, Department of Biology, Via Madonna del Piano 6, Florence, Italy
[2]ISPRA - Institute for Environmental Protection and Research, Via V. Brancati 48, Rome, Italy
[3]University of Siena, Department of Economics and Statistics, Piazza S. Francesco 8, Siena, Italy

Abstract. Vantage point counts carried out on all open areas of a study region could be used for monitoring roe deer populations as relative abundance indexes. From a practical point of view, to cover all the open areas of a study region may be too demanding in terms of operators, time and organization. In most cases vantage point counts are performed on a portion of the open areas usually selected by means of purposive criteria Thus, no objective conclusion about the accuracy and the precision of these values as estimates of the corresponding relative abundance indexes can be drawn. This fact precludes the use of a rigorous statistical testing to compare estimates achieved at different times for detecting changes. On the other hand, if the areas were selected by means of probabilistic sampling schemes and statistically sound estimators were adopted, then estimation of the sampling errors, construction of confidence intervals and change assessment are possible, together with a post hoc power analysis for evaluating the probability of failing to detect a true change. The aim of this study is to propose a sampling strategy that allows to perform all these statistical steps and to evaluate its performance on a real situation in Tuscany (Central Italy) in which all the open areas were surveyed. Results discourage the use of vantage point counts for monitoring purposes.