World Academy of Science, Engineering and Technology
International Journal of Biological and Ecological Engineering
Vol:15, No:12, 2021

Open Science Index, Biological and Ecological Engineering Vol:15, No:12, 2021 publications.waset.org/abstracts/137277.pdf

Imputation of Incomplete Large-Scale Monitoring Count Data via Penalized Estimation

Authors : Mohamed Dakki, Genevieve Robin, Marie Suet, Abdeljebbar Qninba, Mohamed A. El Agbani, Asmâa Ouassou, Rhimou El Hamoumi, Hichem Azafzaf, Sami Rebah, Claudia Feltrup-Azafzaf, Nafouel Hamouda, Wed a.L. Ibrahim, Hosni H. Asran, Amr A. Elhady, Haitham Ibrahim, Khaled Etayeb, Essam Bouras, Almokhtar Saied, Ashrof Glidan, Bakar M. Habib, Mohamed S. Sayoud, Nadjiba Bendjedda, Laura Dami, Clemence Deschamps, Elie Gaget, Jean-Yves Mondain-Monval, Pierre Defos Du Rau

Abstract : In biodiversity monitoring, large datasets are becoming more and more widely available and are increasingly used globally to estimate species trends and con- servation status. These large-scale datasets challenge existing statistical analysis methods, many of which are not adapted to their size, incompleteness and heterogeneity. The development of scalable methods to impute missing data in incomplete large-scale monitoring datasets is crucial to balance sampling in time or space and thus better inform conservation policies. We developed a new method based on penalized Poisson models to impute and analyse incomplete monitoring data in a large-scale framework. The method al- lows parameterization of (a) space and time factors, (b) the main effects of predic- tor covariates, as well as (c) space–time interactions. It also benefits from robust statistical and computational capability in large-scale settings. The method was tested extensively on both simulated and real-life waterbird data, with the findings revealing that it outperforms six existing methods in terms of missing data imputation errors. Applying the method to 16 waterbird species, we estimated their long-term trends for the first time at the entire North African scale, a region where monitoring data suffer from many gaps in space and time series. This new approach opens promising perspectives to increase the accuracy of species-abundance trend estimations. We made it freely available in the r package ‘lori’ (https://CRAN.R-project.org/package=lori) and recommend its use for large- scale count data, particularly in citizen science monitoring programmes.

Keywords : biodiversity monitoring, high-dimensional statistics, incomplete count data, missing data imputation, waterbird trends in North-Africa

Conference Title : ICSE 2021 : International Conference on Statistical Ecology
Conference Location : Vienna, Austria
Conference Dates : December 27-28, 2021