

Guest editorial: Advances in contemporary environmental statistics

Ashis SenGupta · Richard L. Smith

Received: 20 June 2014 / Published online: 14 February 2015
© Springer Science+Business Media New York 2015

From March 4–6, 2013, the first SAVI workshop on Environmental Statistics was held at the Statistical and Applied Mathematical Sciences Institute (SAMSI), in Research Triangle Park, North Carolina. SAVI is an abbreviation for Science Across Virtual Institutes, an international initiative for scientific collaboration sponsored by the US National Science Foundation and various international counterparts. As one of the outcomes of this initiative, two of the mathematical sciences institutes in the US (SAMSI and Institute for Computational and Experimental Research in Mathematics (ICERM)) teamed up with four mathematics and statistics institutes in India, with co-sponsorship from the Department of Science and Technology of the Indian Government. Activities funded have included joint workshops and research visits in both directions.

The workshop on Environmental Statistics was planned with 14 invited speakers, seven from the US and seven from India. In the event, one Indian participant was obliged to withdraw at the last minute and was replaced by a visitor from the U.K. In addition, participants had the opportunity to present posters. This special volume comprises of seven papers from this workshop. The papers represent a good cross-section of the work that was presented in the workshop.

Simmons et al. present an approach to benchmark dose estimation using Bayesian model averaging. Benchmark dose estimation is a key step in establishing safe stan-

A. SenGupta (✉)
University of California, Riverside, CA, USA
e-mail: eestashis@gmail.com

A. SenGupta
Indian Statistical Institute, Kolkata, WB, India

R. L. Smith
Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, NC, USA
e-mail: rls@samsi.info

dards for toxic materials, by estimating the proportion of affected individuals corresponding to a series of doses of a toxic material. However, the results of this procedure are highly sensitive to the choice of risk model assumed for the analysis. Bayesian model averaging (BMA) is a technique that allows numerous risk models to be tried with posterior estimates consisting of suitably weighted averages over all the models. The method is applied to assess the carcinogenic properties of goldenseal root powder. The paper follows on from a talk given by co-author Walter Piegorsch at the workshop, in which he overviewed a number of different methods of benchmark dose estimation.

Biswas, Hwang and Angers apply BMA in a different context: to assess the effects of different air pollutants to clinic visits for respiratory diseases in Taiwan. Their final model assesses the joint effects of four pollutants—SO₂, NO₂, ozone and particulate matter (PM₁₀)—together with temperature as a meteorological confounder.

A different type of pollution problem is considered in the paper by Chattopadhyay, Mondal and Biswas. They are concerned with using a large number of remotely sensed atmospheric measurements to assess soil moisture in an experiment conducted in Iowa. To achieve a suitable classification, they propose K-means clustering combined with either principal component analysis (PCA) or independent component analysis (ICA). The distinction is that whereas PCA breaks the data into uncorrelated components, ICA makes the stronger assumption that components are independent. The authors present a survey of current computational methods for ICA and demonstrate why the method is superior to PCA in the proposed context.

In a paper originally presented as a poster at the workshop, Vsevolozhskaya and colleagues consider a multiple testing problem in functional data analysis. The problem is motivated by a geologic carbon sequestration experiment in which the objective is areal detection of CO₂ leakage. The authors consider two methods (originally due to Fisher and to Šidák) for combining p values in a multiple testing experiment, and discuss the properties of the resulting combined test when the individual experiments are correlated, as they inevitably are in the case of sampling at different locations of a spatial field. The principal methodological contribution of the paper is a resampling-based procedure for determining the overall test.

Another paper addressing a problem in spatial statistics is that by Fang and Chan, who discuss the fitting of generalized additive models when the errors are spatially correlated. They propose maximum likelihood and restricted maximum likelihood methods of estimation and present theoretical results for asymptotic posterior normality, for model selection, and for testing temporal independence in a spatio-temporal context. As an application, they discuss how the abundance of Alaska plaice depends on various covariates in the eastern Bering Sea.

SenGupta, Kulkarni and Hubale address the problem of predictive inference for future observations for environmental events, following a Weibull distribution, based on a new ‘Pivot’ quantity. Compared to the existing methods, the proposed prediction interval has excellent performance even for small sample sizes. Environmental applications of their procedure are made to intervals between earthquakes, to median rainfall measurements, and to wind energy. Such predictive inference as above can substantially benefit the planning of future strategies in several such areas.

The final paper is by Auffhammer, Li, Wright and Yoo, and discusses a problem in paleoclimate estimation, closely related to the talk given by Bo Li at the workshop.

In this paper, they compare two regression approaches (reverse regression and inverse regression) for determining the relationship between a tree-ring series and directly measured temperatures, with the objective of using the tree-ring series to reconstruct temperatures from times before the era of modern meteorological measurements. The paper presents a number of reasons why the inverse regression approach is to be preferred, especially in the presence of temporal correlation, though it also hints at reasons why a third approach based on Bayesian hierarchical models, though much more intensive computationally, may in fact be best of the three.

Overall, these papers present a wide range of environmental statistics but the common theme of all of them is the combination of rigorous methodological development with challenging environmental applications. The editors acknowledge the contributions of all seven papers, as well as the others that were presented at the workshop, as a worthy summary of contemporary environmental statistics.

The Guest Editors acknowledge with thanks the encouragements and supports received from National Science Foundation, USA and Department of Science and Technology, Government of India, in particular from the officials, Professors Bimal K. Roy, Director, Indian Statistical Institute, India and Govindan Rangarajan, Indian Institute of Science, Bangalore, India. They also wish to express their deep appreciations, for the efforts and cooperations extended in preparing the scholarly reports, to the Referees: Drs. Rahul Bhattacharya, Atanu Biswas, Asis Chattopadhyay, Kalyan Das, Tanujit Dey, Paramjit Gill, Apratim Guha, Steward Huang, Snehalata Huzurbazar, Myron Katzoff, Sungsu Kim, Hemangi V. Kulkarni, Somesh Kumar, Amitava Mukherjee, Manju Pandey and Baris Sürücü.