CEI Projects - June, 2015

Noel Cressie, Andrew Zammit Mangion, and Sandy Burden

The Centre for Environmental Informatics has projects available in the following areas. The details of specific projects can be tailored to the strengths and goals of interested students.

1 Multivariate spatial covariance models

Multivariate geostatistics is based on modelling all covariances between all possible combinations of two or more variables and their locations in a continuously indexed domain. Multivariate spatial covariance models need to be built with care, since any covariance matrix that is derived from such a model has to be nonnegative-definite. This project explores the conditional approach to model construction and the fitting of such models to spatial datasets.

2 Predicting exceedances

An exceedance region or hotspot can be defined as the set of locations where the response process of interest exceeds some specified threshold. They could be identified as the the area for which the predicted response of a spatial or spatio-temporal process exceeds a given threshold. In this project, various loss functions that can be used to predict exceedances will be defined and explored. An important application is to estimation of near-surface carbon dioxide sources and sinks. These require computationally efficient modelling techniques that effectively model the tails of the distribution when there is spatial variation in the uncertainty of the predictions.

3 Statistical comparison of global atmospheric chemistry models

Atmospheric chemistry models are an important and widely used tool for understanding and interpreting the behaviour of trace gases in the atmosphere, including climate- and pollution-relevant species. The Southern Hemisphere Model Intercomparison project (SHMIP) evaluated the fit of four chemistry-climate and chemical transport models against in-situ, column, and aircraft observations for gases such as carbon monoxide (CO) and formaldehyde (CH2O). The aim of this project is to use the model's output and observations to compare the fit of the models for CO and other gases using spatial and spatio-temporal statistical model assessment techniques that are available in the R software.

4 Statistics for remote sensing data

Traditional statistical methods do not necessarily scale well with data size and model complexity. One area where this problem is particularly apparent is when estimating and predicting with geostatistical spatio-temporal models from satellite remote sensing data. The project's aim is to assess current state-of-the-art software available in R for this purpose, highlight the key limitations in the software, and propose (not implement) possible ways in which these may be remedied based on current literature. Suggested reference: http://cran.r-project.org/web/views/SpatioTemporal.html

5 Statistics and parallel computation

Statistics has been relatively slow in the uptake of parallel computation, despite its prevalent use in other areas of data science. In this project, the use of parallelism in Markov chain Monte Carlo methods will be investigated, both with the use of multiple chains and with "in-chain" parallelisation. The student will make use of CEI parallel computation resources and existing code using parallel packages available in R. Suggested reference: Kontoghiorghes, E. J. (Ed.). (2005). Handbook of Parallel Computing and Statistics. CRC Press.

6 SAR models for large datasets

The spatial autoregressive (SAR) model is widely used in spatial econometrics to model Gaussian processes on a discrete spatial lattice but, for very large datasets, fitting it becomes computationally prohibitive. An alternative is the Spatial Random Effects (SRE) model and an additive measurement-error model, calibrated to the SAR model of interest using the Generalised Moran operator. In this project the specification of the spatial weights matrix that is used in these models will be explored using 2011 Australian Census data.

7 Spatial Census data

Census data summaries are ubiquitous in statistical models for understanding social, environmental, and economic processes. When the data include spatial location information, the Census summaries form a spatial lattice over the region of interest, and inference can be made that accounts for spatial variation in the process of interest. However, for many statistical models inference also depends on the number and location of areas that are used in the analysis. In this project the use of alternative areas (or zones) is explored using hierarchical spatial statistical models.