Accounting for Spatial Anonymization in Household Surveys

Umut Altay[1], John Paige[2], Andrea Ingeborg Riebler[3] and Geir-Arne Fuglstad[4]

[1] *Norwegian University of Science and Technology, Norway, umut.altay@ntnu.no*
[2] *Norwegian University of Science and Technology, Norway, john.paige@ntnu.no*
[3] *Norwegian University of Science and Technology, Norway, andrea.riebler@ntnu.no*
[4] *Norwegian University of Science and Technology, Norway, geir-arne.fuglstad@ntnu.no*

Household surveys are a key source of data for the estimation of demographic and health indicators in low- and middle-income countries. These surveys consist of spatial data with coordinates for observed household clusters. The coordinates of the household clusters have been anonymised to protect privacy. Each coordinate has been randomly displaced according to a known jittering distribution. In this talk, we describe a computationally fast approach to account for the uncertainty in the measurement locations by combining the stochastic partial differential equation (SPDE) approach for spatial modelling and the Template Model Builder (TMB) method for inference. This allows us to investigate the effect of jittering on parameter inference and prediction. Our main example is the Demographic and Health Surveys (DHS) survey in Kenya from 2014. Through a simulation study designed based on this survey, we quantify the effect of jittering on prediction for different magnitudes of jittering and discuss how large jittering could be made before it has an adverse effect.

Modelling extreme short-term precipitation with the blended generalised extreme value distribution

Silius M. Vandeskog[1], Sara Martino[2] and Daniela Castro-Camilo[3]

[1] *University of Science and Technology, Norway, silius.m.vandeskog@ntnu.no*
[2] *University of Science and Technology, Norway*
[3] *University of Glasgow, UK*

Short-term extreme precipitation can cause flash floods, large economic losses and immense damage to infrastructure. In order to prepare for future extreme precipitation, a spatial Bayesian hierarchical model is applied for estimating large return levels of short-term precipitation in Norway.

A modified version of the generalised extreme value (GEV) distribution, called the blended GEV (bGEV) distribution, is used as a model for the yearly maxima of short-term precipitation. Inference with the GEV distribution is known to be difficult, partially because its support depends on its parameters. This is not a problem with the bGEV distribution, which has the right tail of a Fréchet distribution and the left tail of a Gumbel distribution, resulting in a constant support that allows for simpler inference. Fast inference is performed using integrated nested Laplace approximations (INLA).

We propose a new model for block maxima that borrows strength from the peaks over threshold methodology by linking the scale parameter of the bGEV distribution to the standard deviation of observations larger than some threshold. The new model is found to outperform the standard block maxima model when fitted to the yearly maxima of short-term precipitation in Norway. Model fit is evaluated using the scaled threshold weighted continuous ranked probability score (StwCRPS), with a weight function that only focuses on large quantiles.

Ensemble updating of categorical state vectors

Margrethe Kvale Loe[1] and Håkon Tjelmeland[2]

[1] *NTNU, Norway, margrethe.loe@ntnu.no*
[2] *NTNU, Norway, haakon.tjelmeland@ntnu.no*

Ensemble filtering methods represent a class of approximate, Monte Carlo-based algorithms for solving the stochastic filtering problem. Instead of computing the sequence of forecast and filtering distributions explicitly, ensemble filtering methods use a set of samples, an ensemble, to represent them empirically. The main challenge in ensemble filtering methods is the updating of a prior (forecast) ensemble to a posterior ensemble. This needs to be done every time new data become available. Generally, the updating of the ensemble cannot be carried out exactly, and approximations are required. In the present study, we consider one such approximate updating method appropriate for state vectors whose elements are categorical variables. The method is based on a Bayesian and generalised view of the well-known ensemble Kalman filter (EnKF). In the EnKF, Gaussian approximations to the forecast and filtering distributions are introduced, and the forecast ensemble is updated with a linear shift. Given that the Gaussian approximation to the forecast distribution is correct, the EnKF linear update corresponds to conditional simulation from a Gaussian distribution with mean and covariance such that the posterior samples marginally are distributed according to the Gaussian approximation to the filtering distribution. In the proposed approach for categorical vectors, the Gaussian approximation to the forecast distribution is replaced with a Markov chain model, and instead of a linear update, we characterise, for each forecast ensemble member, a class of decomposable graphical models for simulating a corresponding posterior ensemble member. To make the update robust against the assumptions of the assumed forecast and filtering distributions, an optimality criterion is formulated. The proposed framework is Bayesian in the sense that the parameters of the assumed forecast distribution are treated as random. Results from a simulation example where each variable of the state vector can take three different values are presented.

Flexible Spatial Covariance Structures for 3D Gaussian Random Fields

Martin Outzen Berild and Geir-Arne Fuglstad

Department of Mathematics,
Norwegian University of Science and Technology, Norway
(martin.o.berild@ntnu.no, geir-arne.fuglstad@ntnu.no)

Numerical solutions for complex dynamical systems such as the oceans are computationally demanding and the computed forecasts will not be an exact match for the future. However, we can use the numerical solution to construct a statistical model that matches some of the properties observed in the numerical solution. We will use a spatial model consisting of a spatially varying mean together with a 3D Gaussian random field (GRF). The spatial model can then be updated at a given time point based on local measurements, for example, from an autonomous underwater vehicle (AUV).

We use the stochastic partial differential equation (SPDE) approach to achieve a computationally efficient description of the GRFs through sparse matrices. The SPDE approach makes it easy to introduce non-stationarity in the dependence structure in such a way that positive-definite precision matrices are guarenteed. Furthermore, the spatially varying anisotropy can be intuitively specified through a vector field controlled by parameters.

The approach is applied to model the ocean mass outside of Trondheim, where the parameters are found by fitting the model to simulated data from a complex ocean model (SINMOD). Then an evaluation of the predictive power of the model is obtained by strictly proper scoring rules on collected data from underwater vehicles from the same water mass.

Geostatistical modelling combining point observations and nested areal observations - A case study of annual runoff predictions in the Voss area

Thea Roksvåg[1,3], Ingelin Steinsland[1], Kolbjørn Engeland[2]

[1] *Department of Mathematical Sciences, NTNU, Norway*
[2] *The Norwegian Water Resources and Energy Directorate (NVE)*
[3] *Norwegian Computing Center*

In this study, annual runoff is estimated by using a Bayesian geostatistical model for interpolation of hydrological data of different spatial support. That is, streamflow observations from catchments (areal data), and precipitation and evaporation data (point data). The model contains one climatic spatial effect that is common for all years under study, and one year specific spatial effect. Hence, the framework enables a quantification of the spatial variability that is due to long-term weather patterns and processes. This can contribute to a better understanding of biases and uncertainties in environmental modeling. By using integrated nested Laplace approximations (INLA) and the stochastic partial differential equation approach (SPDE) to spatial modeling, the two field model is computationally feasible and fast. The suggested model is tested by predicting 10 years of annual runoff around Voss in Norway and through a simulation study. We find that on average we benefit from combining point and areal data compared to using only one of the data types, and that the interaction between nested areal data and point data gives a spatial model that takes us beyond smoothing. Another finding is that when climatic effects dominate over annual effects, systematic under- and overestimation of runoff over time can be expected. On the other hand, a dominating climatic spatial effect implies that short records of runoff from an otherwise ungauged catchment can lead to large improvements in the predictability of runoff.