Statistical Modeling for Complex Observation Data: Applications to Intelligent Transport System

Supervisor: [Prajamitra Bhuyan]

Research Group: [Statistics and Data Science]

Project description:

As the world becomes increasingly dominated by large-scale data-oriented complex systems, such as the Internet of Things, national transportation and cyber systems, the demand for automatic monitoring and mitigation for extreme and high-consequence events to enhance the safety, security and economic wellbeing of the smart cities is pressing. The effect of such events manifests in a number of ways, and novel statistical methodology and machine learning techniques are required to provide resilience to these systems. However, if the data used to train an algorithm contains biases against certain characteristics, then the resulting estimate will be biased. Using causal methods, we are aiming to ensure algorithmic fairness by taking into account different structural biases and compensating for them effectively. The `potential' outcome framework has been extensively studied in many causal inference settings.

The aim of this project to develop modeling framework to incorporate pre-intervention and post-intervention information based on the propensity score and outcome regression model. These methods are intended to assist in analysing the causal effect of various interventions on an intelligent transport system. In such applications, practitioners are interested to foresee the effect of an intervention on extreme quantiles of the outcome of interest.

There is an ample amount of literature on the estimation of quantile treatment effect in various application areas. However, the existing methods are not suitable for this purpose. In this project, we consider addressing this issue and propose a novel estimation method utilizing tools of causal inference and extreme value theory. It is also worth noting that we are obliged to assume that the outcomes of one unit are not affected by the treatment assignment of any other units. One possible direction of further studies could be using an improved design and defining several types of treatment effects and developing associated estimation methodologies for the setting where there may be clustered interference.

Further information:

[How to apply]

[Entry requirements]

[Fees and funding]

Efficient spatio-temporal modelling and interpolation of Global Sea Surface Height

Supervisor: Arthur Guillaumin

Research Group: Statistics and Data Science

Project description:

Big environmental data combined with modern data science offer a key opportunity to better understand our environment and address climate change. Yet environmental data poses specific challenges due to spatio-temporal dependence: 1. Modelling. Most models of spatio-temporal dependence are homogenous, which in general is not an accurate description of real-world phenomena. 2. Estimation. More complex parametric models of covariance are often not amenable to estimation via exact likelihood due to computational inefficiency and lack of robustness to model misspecification.

In this project, you will develop new parametric covariance models and estimation methods for the analysis of Sea Surface Height (SSH). The surface height of the oceans is monitored by passing satellites on a global scale. The modelling of SSH is vital to a better understanding of the global climate and to making more accurate interpolation via kriging [1].

Firstly, your research will focus on estimating the parameters of a spatial covariance model of Sea Surface Height. You will pursue recent developments in quasi-likelihood estimation [2, 3, 4, 5, 6] for spatio-temporal data to propose a parametric estimation method that is both computationally and statistically efficient. The idea behind quasi-likelihood estimation is to maximize a computationally efficient approximation to the exact likelihood.

Secondly, you will develop more advanced parametric models of covariance that can incorporate some additional physical phenomena that drive SSH. Possible directions for this part of the project range from relaxing the assumption of spatial homogeneity [7], to modelling temporal dependence and seasonal patterns [8]. These methodological developments can also have an impact in other application areas such as econometrics.

This project will be in collaboration with C. Wortham and J. Early, NorthWest Research Associates, Seattle, USA, who have been granted funding by the NASA to develop mapping software for SSH.

References

1. Stein, M. L. (1999). Interpolation of Spatial Data: Some Theory for Kriging. 1st ed., Springer Series in Statistics, Springer.

2. Sykulski A. M., Olhede S. C., Guillaumin A. P., Lilly J. M., Early J. J (2019). The Debiased Whittle Likelihood. Biometrika 106(2), 251–266.
3. Guillaumin A. P., Sykulski A. M., Olhede S. C., Simmons F. V. (2022). The Debiased Spatial Whittle Likelihood. Journal of the Royal Statistical Society: Series B (under review).
4. Guinness J. (2018). Permutation and Grouping Methods for Sharpening Gaussian Process Approximations. Technometrics 1706, 1-15.
5. Guinness J, Fuentes M. (2017). Circulant Embedding of Approximate Covariances for Inference from Gaussian Data on Large Lattices. Journal of Computational and Graphical Statistics 26(1), 88-97.
6. Fuentes M. (2007). Approximate likelihood for large irregularly spaced spatial data. Journal of the American Statistical Association 102(477), 321-331.
7. Matsuda, Y., Yajima, Y. (2018). Locally stationary spatio-temporal processes. Japanese Journal of Statistics and Data Science 1, 41–57.
8. Napolitano A. (2016). Cyclostationarity: New trends and applications. Signal Processing 120, 385-408.

Further information:

How to apply

Entry requirements

Fees and funding

Bayesian spatial modelling for biodiversity

Supervisor: **[Silvia Liverani]**

Research Group: **[Statistics and Data Science]**

Project description:

Many models assume that observations are obtained independently of each other. However, distance between observations can be a source of correlation, which needs to be accounted for in any model. For example, pollution has a spatial smooth pattern and measurements close in space are likely to be very similar. Spatial models therefore have to take into account the spatial autocorrelation in datasets in order to separate the general trend (usually depending on some covariates) from the purely spatial random variation.

This project will focus on developing and applying Bayesian spatial and spatio-temporal modelling techniques to enhance our understanding of the association between, and predict, plant species that are at risk of extinction and areas in need of protection in the face of climate change, changing land use (especially agriculture) and pollution. The pollutants of interest are nitrogen and phosphate-based fertilizers. We will leverage spatial distribution data for the entire British flora, studying changing trends in distribution and land use over 60 years. We will also include data for a range of different measures of genomic diversity (e.g. genome size and polyploidy) together with climate and soil data to uncover the role of biological and abiotic factors in predicting species at risk of extinction, and landscapes at increased risk of biodiversity loss under differing land use and climate change scenarios. The research will be undertaken in collaboration with Dr Ilia Leitch, Senior Research Leader at the Royal Botanic Gardens, Kew.

One statistical challenge that arises is that the data are available at different resolutions, and advanced methods are required to model misaligned spatial and spatio-temporal data. We will leverage recent work by Dr Silvia Liverani on Bayesian methods for misaligned areal data, and extend them to meet the needs of this research challenge in the study and understanding of biodiversity.

Further information:

How to apply

Entry requirements

Fees and funding