Bayesian hierarchical modeling and analysis of spatial and spatio-temporal data
Proposed project for MISCADA

Advisor: Georgios P. Karagiannis

Academic year 2020-2021 @ Durham University

Description

The focus of this project is the study of statistical methods, and related computational tools that can be used for the analysis of spatial, or spatio-temporal data-sets, mainly in the Bayesian framework.

Data-sets with the colorful adjectives 'spatial' or 'temporal' contain data with labels indicating where or when (respectively) have been collected. Such data exist in problems in environmental, climatology, weather forecasting, epidemics, engineering, biology, etc... Spatial-temporal statistical methodologies utilizing such data-sets to build probabilistic models in order to perform predictions and inferences. Bayesian hierarchical modelling allows one to express and quantify uncertainties at different levels, as well as to perform more accurate predictions.

Project specific intended learning outcomes

By the end of this project, students will be able to design suitable Bayesian hierarchical models over space-time-stochastic domains, and apply these models for the analysis of real spatial-temporal data, as well as implement Bayesian computational tools for training these models. You will be exposed to the use of suitable software required to the practical implementation of the methodology.

Potential project directions

For example, some potential directions for the project/dissertation can be:

Big data methods: A satellite picture of a region may contain several 'pixels' (aka data), or be contaminated by noise. Processing all this information bears serious computational challenges.

Methods addressing high dimensionality : In several cases, the dimensionality of the problem is so large that makes traditional modeling or computational methods impractical. The challenge here is how to reduce dimensionality without significantly sacrificing accuracy.

Temporal data methods: In many cases the spatial data are time-dependent. The challenge is to capture the dynamic behaviours.

Multivariate data methods: Often when we collect data, (e.g., weather data), we collect several characteristics/variables (e.g., precipitation, pressure, etc...). Here, a challenge is how to model their dependencies.

Downscaling/upscaling methods What if we get data on one scale and we need to make inference on another one? Spatial data at different scales e.g. global, regional, local, etc.. are associated with different variabilities/dependencies.

Multi-resolution data methods: Often data from different sources are available. We may have weather data collected from different satellites (an old and an new); or from different sources a satellites and field stations.

Analysis of real data: Possible data-sets you can analyze as part of your projects can be: disease (like SARS, H1N1, COVID-19, etc...), CO_2 concentration, remote sensing satellite data, etc ...

or any other...

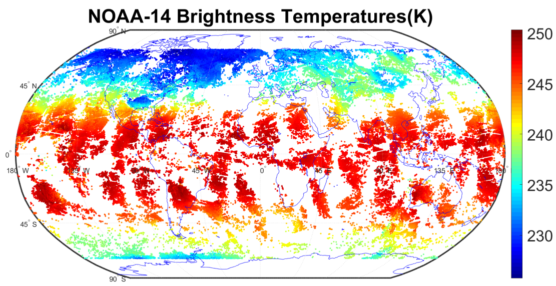
⋮

An example

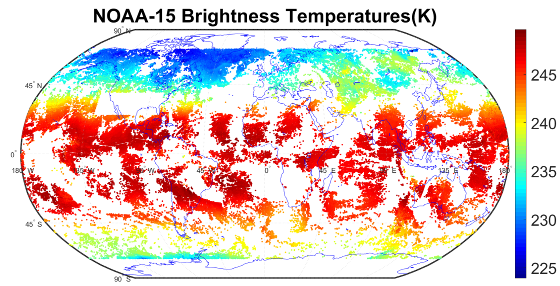
Fig 1 presents temperature data collected from an old, a-little-a-bit-rusty, satellite N-14 a long time ago. Fig 2 presents data collected from a new satellite N-15 which is in a better shape and hence more accurate.

A statistical challenge here is how to build a model properly combining big data sets from different sources at different accuracy, in order to learn what is going on at locations where measurements from the accurate source are not available (e.g. U.S.A, Oceania, etc...), as well as learn discrepancies between the different data sources, over time.

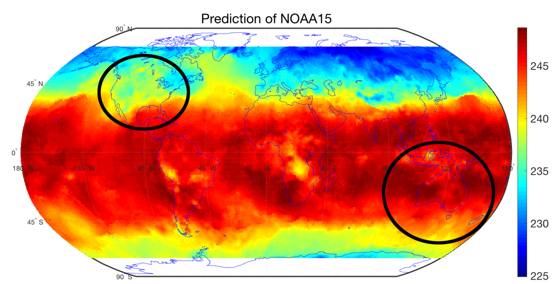
Fig 3 presents the predicted temperatures, at a specific time-step and over the whole area, produced by a Spatial-Temporal Statistics approach properly combining information available from different sources of different accuracy.



(1)



(2)



(3)

Borrowed from: Cheng, S., Konomi, B. A., Matthews, J. L., Karagiannis, G., & Kang, E. L. (2020). Hierarchical Bayesian Nearest Neighbor Co-Kriging Gaussian Process Models; An Application to Intersatellite Calibration. arXiv preprint arXiv:2004.01341.

Requirements

- Knowledge of Bayesian Statistics, and Regression models. ; Knowledge of R or Python

References

- Wikle, C. K. (2015). Modern perspectives on statistics for spatio-temporal data. Wiley Interdisciplinary Reviews: Computational Statistics, 7(1), 86-98. [LINK]

- Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). Hierarchical modeling and analysis for spatial data. CRC press.

Contact details

For further information, feel free to contact

Georgios Karagiannis (Office CM126b)
Email: georgios.karagiannis@durham.ac.uk
~~Telephone: +44 (0) 1913342718~~