25th International Congress on Modelling and Simulation, Darwin, NT, Australia, 9 to 14 July 2023
mssanz.org.au/modsim2023

Sophisticated tools for spatio-temporal data exploration

Sevvandi Kandanaarachchi [a], Petra Kuhnert [b], Andrew Zammit-Mangion [c] and Christopher K. Wikle [d]

[a] *CSIRO Data61, Melbourne, Australia*
[b] *CSIRO Data61, Brisbane, Australia*
[c] *National Institute for Applied Statistics Research Australia, School of Mathematics and Applied Statistics, University of Wollongong, Australia*
[d] *Department of Statistics, University of Missouri, Columbia, United States*
Email: sevvandi.kandanaarachchi@data61.csiro.au

Abstract: Spatio-temporal data underpin many critical processes such as weather, crop production, wildfire spread and epidemiological and disease function. Models of these processes can reveal changing characteristics in both space and time and can help inform decision-makers. A recent example is during the pandemic years, spatio-temporal models were used to inform public policy. While there are many spatio-temporal modelling methods and packages, tools specifically designed for exploratory data analysis are somewhat lacking. Exploratory data analysis is a vital step in the end-to-end process of statistical and machine learning modelling. A lack of tools for exploratory spatio-temporal data analysis may lead to researchers starting the modelling process prematurely and make suboptimal modelling choices. We aim to fill this gap by contributing ***stxplore*** – an R package equipped with useful functionality designed for spatio-temporal data exploration.

All functions in ***stxplore*** are designed to provide visually useful outputs. Furthermore, all computations can be performed using either data frames or *stars* objects in the R framework. Data frames are traditional, general purpose data structures in R, used for tabular data, while ***s****tars* objects cater for geospatial data. These object classes are defined in the R package *stars*, which has gained popularity within the research community, and are a newer addition to the R geospatial package ecosystem. The package ***stxplore*** can work with either of these objects, i.e. the functions in ***stxplore*** can take either data frames or *stars* objects as input.

The functionality of ***stxplore*** can be grouped into three main categories: 1) simple exploration, 2) covariance calculation, 3) PCA-like spatio-temporal computation. Simple exploration functions allow the visualisation of spatial snapshots through time or temporal snapshots at any given location. Consider a spatio-temporal dataset containing daily maximum temperature data at different locations. We can view spatial maps for any given time point, or temporal fluctuations for any given spatial location. Furthermore, ***stxplore*** can be used to visualise spatial means averaged over time, or temporal means averaged over space. Another useful plot provided by ***stxplore*** is the Hovmöller plot which can be used to visualise the spatio-temporal evolution of the process along one spatial dimension, by averaging over the other spatial dimensions. For example, Hovmöller plots can be used to visualise temperatures by latitude over time, by averaging over longitude. In a similar way, ridgeline plots that consist of multiple distribution plots can be organised by longitude/latitude or by time. With ridgeline plots we can therefore analyse how the distribution of the data changes over time or over space.

Covariance calculations can be used to understand the underlying correlation structure of the data. Empirical spatial covariances and space-time variograms are useful tools to visualise the effect of distance on the quantity of interest. Empirical orthogonal functions and spatio-temporal canonical correlation analysis, included in ***stxplore***, involve computations that are similar to PCA. Empirical orthogonal functions are often used as a dimension reduction technique for spatio-temporal data, while spatio-temporal canonical correlation analysis is done to uncover the correlations between two spatio-temporal datasets. Covariance calculations and PCA-like calculations are often under-utilised methods in spatio-temporal data exploration, mainly due to lack of visualisation tools and associated documentation. However, these methods can help us draw useful insights about the data and inform us of better modelling choices. The package website https://sevvandi.github.io/stxplore/ contains examples with R code. In this presentation we showcase ***stxplore*** using aerosol data captured over Australia during the 2019/20 bushfires.

REFERENCES

Wikle, C.K., Zammit-Mangion, A., Cressie, N., 2019. Spatio-Temporal Statistics with R. Chapman. Hall/CRC Press.

Keywords: *Spatio-temporal data, exploratory data analysis, data visualisation*