Estimating means in Spatio-Temporal Datasets – Discussion minutes

The discussion started with a brief overview of the different components that are estimated as part of the mean structure for spatio-temporal datasets. These include

- Spatial trends and/or Temporal trends
- Regression component
- Seasonality
- Autocorrelation
- Mean components based on pde's driven by the underlying science or physical processes

One of the frequently encountered type of spatio-temporal dataset constitutes of regular observations in time at numerous locations. Air quality data is one example of such rectangular datasets. Various non-parametric methods like polynomial regression, splines etc were discussed that estimate the temporal trend site-by-site. Some of these methods determine the smoothness automatically. Estimation of trends independently at the sites assumes space-time separability and may not be suitable for datasets of non-separable origin. To circumvent this restriction, some of the above mentioned methods use a family of similar curves for estimating the site-by-site trend and thereby incorporate the spatial dependence between the sites. These methods usually leave a set of temporally uncorrelated residuals from which the spatial covariance can be estimated. However, it was agreed that removing the trend does not necessarily remove non-stationarity and one should still test for it among the residuals.

One of the alternative approaches discussed uses multivariate dynamic linear models that can capture the trend and autocorrelation jointly. Functional regression approach that models all the locations at the same time-point jointly can be suitable for data coming monthly or annually. A related problem of predicting the mean at a new location with unobserved covariates may require simultaneous modeling of the response and the covariates.

More complicated spatio-temporal datasets were discussed like those coming from measurements of hurricane forecasting errors or temperatures at different locations and levels inside the ocean. The mean components for many of these datasets satisfy physical constraints which needs to be accounted for.

There was also discussion about how to identify which variation can be explained by the mean and which one can be attributed to the variance. In the end the general consensus reached was that one should try to put as much science into the trend as possible. All knowledge about the covariates, physical processes guiding the responses should be exploited for estimating the mean. The remaining unexplained variation can be modeled as part of the variance.