Summary Report of the Mercury Concentration Study in New Hampshire Estuary

January 9, 2018

The spatial dataset consists the mercury concentrations surveyed in the years 2000/2001 and 2003 at 97 locations in the largest estuary in New Hampshire; see Figure 1 for different measurement of mercury concentrations at different sampled locations. These surveys are part of a coastal monitoring program – National Coastal Assessment – that the US Environmental Protection Agency (EPA) has developed with the New Hampshire Department of Environmental Services.

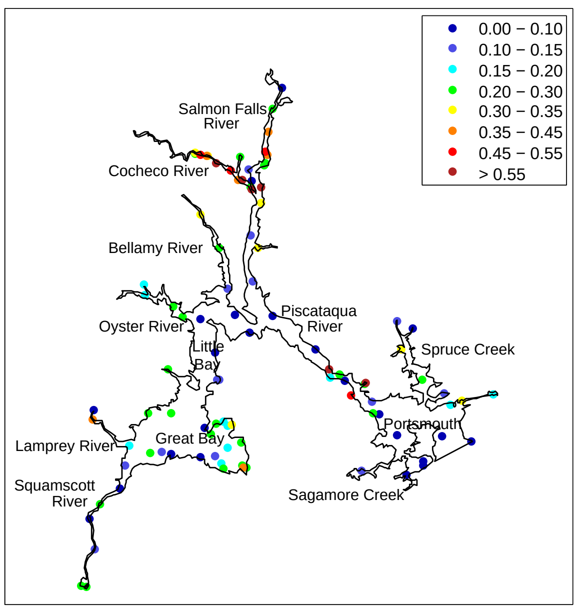


Figure 1: Estuaries in New Hampshire. Dots represent sample locations with different colors indicating different levels of mercury concentrations.

We consider a partial linear model with a linear term for the year effect:

$$\text{Concentration} = \beta_0 + \beta_1 \text{Year} + g(\text{Latitude}, \text{Longitude}). \tag{1}$$

To fit model (1), we consider the following five different methods:

- kriging (KRIG)
- thin-plate splines (TPS)

- geodesic low-rank thin-plate splines (GLTPS) by [2]
- finite element method (FEM) by [1]
- our proposed bivariate penalized splines over triangulations (BPST)

For KRIG, we use the Matérn covariance structure. For GLTPS, we set $k = 5$ as in [2], and the distance matrix is calculated using the code from: https://github.com/dill/gltps. For BPST and FEM, the smoothing or roughness parameter is selected by the GCV. Figure 2 shows the triangulation adopted by the BPST method.

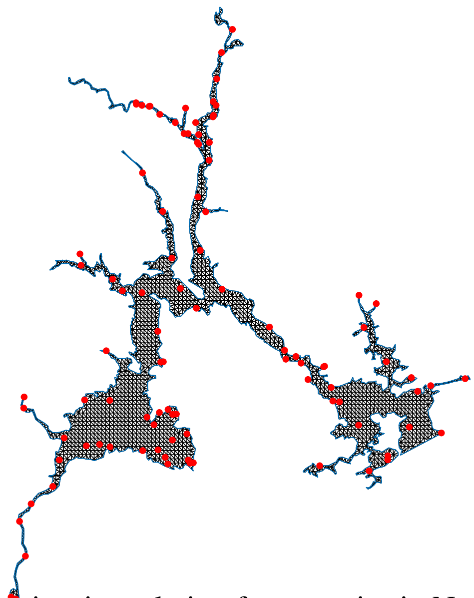


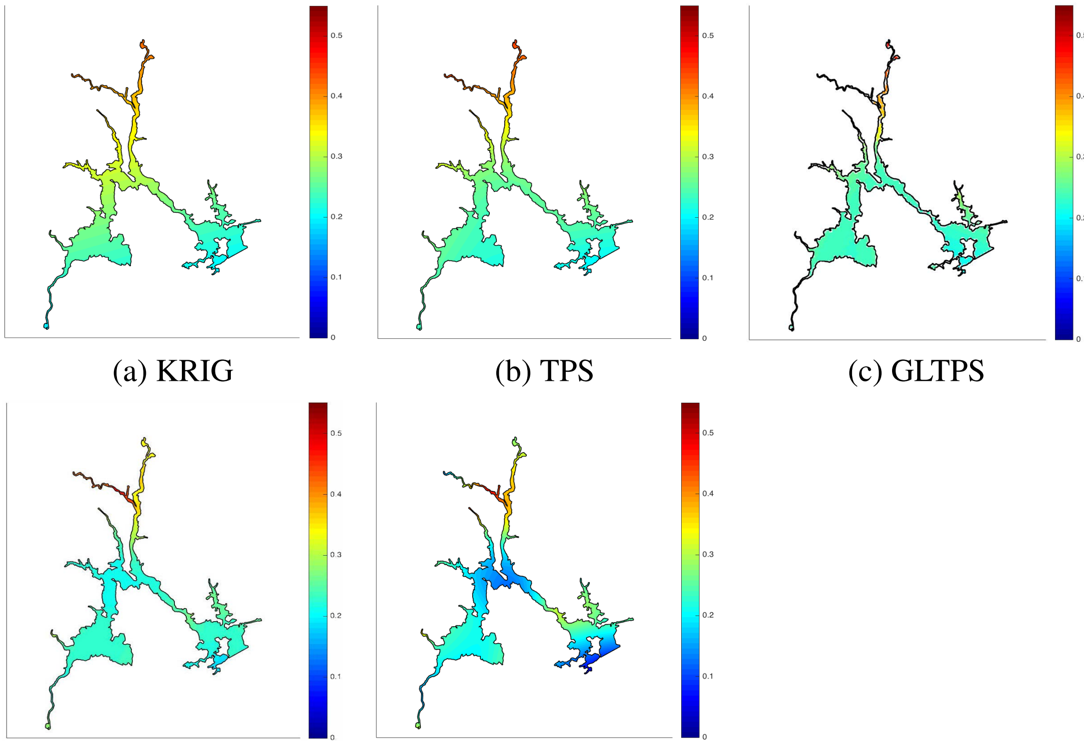
Figure 2: Domain triangulation for estuaries in New Hampshire.

Table 1 summarizes the coefficient estimation results based on different methods. All the methods agree that "year" has a very significant effect.

Table 1: Estimated coefficients with the standard errors.

	KRIG	TPS	GLTPS	FEM	BPST
Year	0.0511	0.0485	0.0465	0.0391	0.0253
s.e.	0.02	0.02	0.02	0.02	0.02

Prediction maps at 20×20 m resolution level using different methods are shown in Figure 3. The GLTPS procedure is a little computation-intensive to make such a high-resolution prediction, so we decrease its resolution to 150×150 m. The spatial distribution in Figure 1 shows generally higher values in the Salmon Falls River and Cocheco River, and lower values in the Piscataqua River and the Portsmouth area, and some localized low spots the Great Bay. Figure 3 illustrates the overspill from the Northern part to the middle area when an ordinary spatial smoothing (such as KRIG and TPS) is used, as it smoothes across the Salmon Falls River and Cocheco River with high concentration levels in the northern part. This problem is mitigated for GLTPS and FEM. The BPST smoother does not show signs of leakage in the Piscataqua River and the Portsmouth area of the estuaries, as other methods do.



(a) KRIG (b) TPS (c) GLTPS

(d) FEM (e) BPST

Figure 3: Prediction maps of mercury concentrations over the estuaries in New Hampshire.

To evaluate different methods, we estimate the out-of-sample prediction errors of each method using the 10-fold cross-validation. We randomly split all the observations into 10 roughly equal-sized parts. For each $k = 1, \ldots, 10$, we leave out part k, fit the model to the other four parts (combined), and then obtain predictions for the left-out kth part. Table 2 summarizes the mean squared prediction errors (MSPE) of the mercury concentrations based on different methods. The MSPEs are similar, and the BPST gives the lowest MSPE.

Table 2: Estimation errors and 10-fold cross-validation prediction errors for mercury concentrations.

Method	KRIG	TPS	GLTPS	FEM	BPST
RMSE	0.1008	0.1379	0.1367	0.1393	0.1225
RMSPE	0.1498	0.1501	0.1505	0.1436	0.1407

References

[1] Ramsay, T. (2002), "Spline smoothing over difficult regions," *Journal of the Royal Statistical Society, Series B*, 64, 307–319.

[2] Wang, H. and Ranalli, M. G. (2007), "Low-rank smoothing splines on complicated domains," *Biometrics*, 63, 209–217.