Experiments on the Earth: Smoothing and Spatial Statistics for Geophysical Applications Using fields

1. Introduction

fields is a collection of programs for curve and function fitting with an emphasis on spatial data and is written in the R language. The package contains support for smoothing splines and spatial process estimators (Kriging). At the heart of the package is a unified philosophy for fitting smooth curves and surfaces from noisy data. The structure and functionality in fields make it an excellent tool for the analysis of the types of data found in many geophysical applications as well as in many other areas.

2. Splines and Kriging

Given noisy observations of a function \( f \), i.e., \( Y = f(X) + \epsilon \), the goal is to construct a prediction of \( f \) at any location, \( x \). In the estimation of \( f \), two questions arise:

- How does one build in constraints on the smoothness of the estimate of \( f \)?
- How does one characterize the uncertainty of an estimated function?

The penalized least-squares criterion is essentially a cost function accounting for both the fit to the data and the roughness of the estimated function.

\[
\sum (Y_i - f(x_i))^2 + \lambda \int (f''(x))^2 \, dx
\]

Splines are the solution to the minimization of this cost function. The roughness penalty controls the smoothness of the solution through the smoothing parameter \( \lambda \). Splines, process estimators, or Kriging, assume that the unknown function is a realization of a Gaussian process, i.e., \( f \) has mean \( \mu \( f \( \) \) \) and covariance function \( \Sigma \). The Kriging estimate is the conditional mean of \( f \) given the observations. The choice of covariance function controls the smoothness of the estimate.

UNIFYING these ideas, note that:

- Splines \( \rightarrow \) Kriging:
  - The fit to the data defines log density of \( f \):
  - the roughness penalty defines log density of \( f \):
- Kriging \( \rightarrow \) splines:
  - Posterior logLikelihood(\( f \), Prior(\( f \))):
  - Posterior mode is minimizing logLikelihood(\( f \), log Prior(\( f \))):

EXAMPLE: Monthly Colorado temperatures:
- 351 stations.
- Daily maximum and minimum for the month of July.
- These temperatures were plotted against elevation and an estimate of \( f \) produced. Can you tell which approach, splines or Kriging, was used?

3. Large Data Sets

PREDICTION for large spatial fields requires a computationally efficient approach that often begins with compactly supported covariance functions, either through direct specification of the covariance or via a tapering approach. The compact support of these covariance functions minimizes the number of required matrix operations and the size of the covariance matrices are sparse. Taking advantage of this sparse structure can lead to potentially large gains in terms of both speed and storage. The class sparse (sparse matrices), embedded in the fields package, exploits sparseness and allows for working with large and even very large data sets.

4. Multivariate Data

MULTIVARIATE spatial models can also be easily constructed in the fields framework. Consider the bivariate case where there are two measurements at each spatial location:

\[
Y_i = X_i \beta + Z_i \gamma + \epsilon_i
\]

where \( X \) includes spatial trends, \( Z \) includes additional covariates, \( \beta \) and \( \gamma \) are random effects with \( \text{E}(\beta) = 0 \).

\[
\Sigma = \begin{bmatrix}
\Sigma_{\beta} & \Sigma_{\beta\gamma} \\
\Sigma_{\gamma\beta} & \Sigma_{\gamma}
\end{bmatrix}
\]

K is a spatial covariance, and \( b \) is random intercept. This model allows sufficient flexibility to account for different scales between the variables and different degrees of smoothing.

EXAMPLE: North Carolina average winter temperature and precipitation:
- 66 stations.
- Elevation included as additional covariate.
- Trends with latitude/elevation for temperature and with longitude for precipitation.
- More spatial smoothing for precipitation.

Figure 1: Colorado stations and regression of temperature on elevation.

Figure 2: Illustration of covariance tapering.

Figure 3: Sparse covariance matrix and its Cholesky decomposition.

Figure 4: Image plot of predictions of US precipitation anomalies.

Figure 5: North Carolina stations.

Figure 6: Estimated \( \Sigma_{\beta} \) (top) and \( \Sigma_{\gamma} \) (bottom).

Figure 7: Predictions for average winter temperature (top) and precipitation (bottom).

2007 Joint Statistical Meetings, Salt Lake City, UT. This research is supported in part by the Geophysical Statistics Project at the National Center for Atmospheric Research and the National Science Foundation through grants DMS-0355474, ATM-0534173 (first author), and DMS-0621118 (second author).

2015 Joint Statistical Meetings, San Francisco, CA. This work was supported by the National Science Foundation through grants DMS-1246975 second author.)