Spatial Interpolation for Missing Temperature Data

Yanrong Mu, Ben Lee, and Murali Haran

Department of Statistics, The Pennsylvania State University, and Network for Sustainable Climate Risk Management

Introduction

- **Goal:** Interpolate temperatures across spatial locations
- **Challenges:**
  - Temperature exhibits spatial dependence
  - Other variables affect temperature (elevation, distance from sea)
  - Large data sets pose computational difficulty for prediction methods
- **Our approach:**
  - Fit a model to temperature data, accounting for the effect of elevation
  - Use mean squared prediction error (MSPE) to evaluate methods

Why is This of Interest

- Modern data sets in atmospheric science are often spatial
- Not only interested in "how much", but also "how much is where"
- We may want to interpolate information across spatial observations
- Demonstrate spatial interpolation method for temperature data

Data and Methodology

- 14727 weather stations in the U.S. and part of Canada
- Data: average of highest temperature across 12 months

Methodology:

- **Use kriging based on Gaussian process**
  - Kriging: interpolate missing spatial data based on the observed values
  - Standard kriging: better for small data sets used "geoR" package in R
  - Lattice kriging: better for large data sets used "LatticeKrig" package in R
- **Model comparison**
  - Split data into 90% "training" and 10% "test" data
  - Fit model on "training" data
  - Interpolate/ predict "test" locations
  - Use mean squared prediction error (MSPE) to evaluate models

$$MSPE = \sum_{i \in location} (true\ value_i - prediction_i)^2$$

- If MSPE is smaller, then the model is better

Model Building

- **Exploratory Data Analysis (EDA)**
  - Plot empirical variogram
  - Fit variogram using different models
  - Decide whether to include nugget
- **Model Fitting**
  - Fit different models (exponential and spherical) based on EDA
  - I used "geoR" package in Rstudio
  - Use MSPE to validate different models (exponential and spherical)

Covariance Model	Exponential	Spherical
MSPE	4.214176	4.209564

- **Check the Effect of Elevation on Temperature**
  - Model form (linear regression on elevation):
    - y_i: temperature at location i; x_i: covariate (elevation) at location i
    - $y_i = x_i * \beta + \varepsilon_i$, where β is parameter vector, ε_i is error at location i
    - $\varepsilon_1, \ldots, \varepsilon_n$ is spatially dependent (from a Gaussian process)
  - Decide whether to include covariate

MSPE	Without Covariate	With Covariate
Exponential	4.214176	2.443368
Spherical	4.209564	2.439944

Standard Kriging VS Lattice Kriging

- Standard kriging is infeasible for large data sets, eg. over 10,000 data points
- Standard kriging produces smaller MSPE than lattice kriging

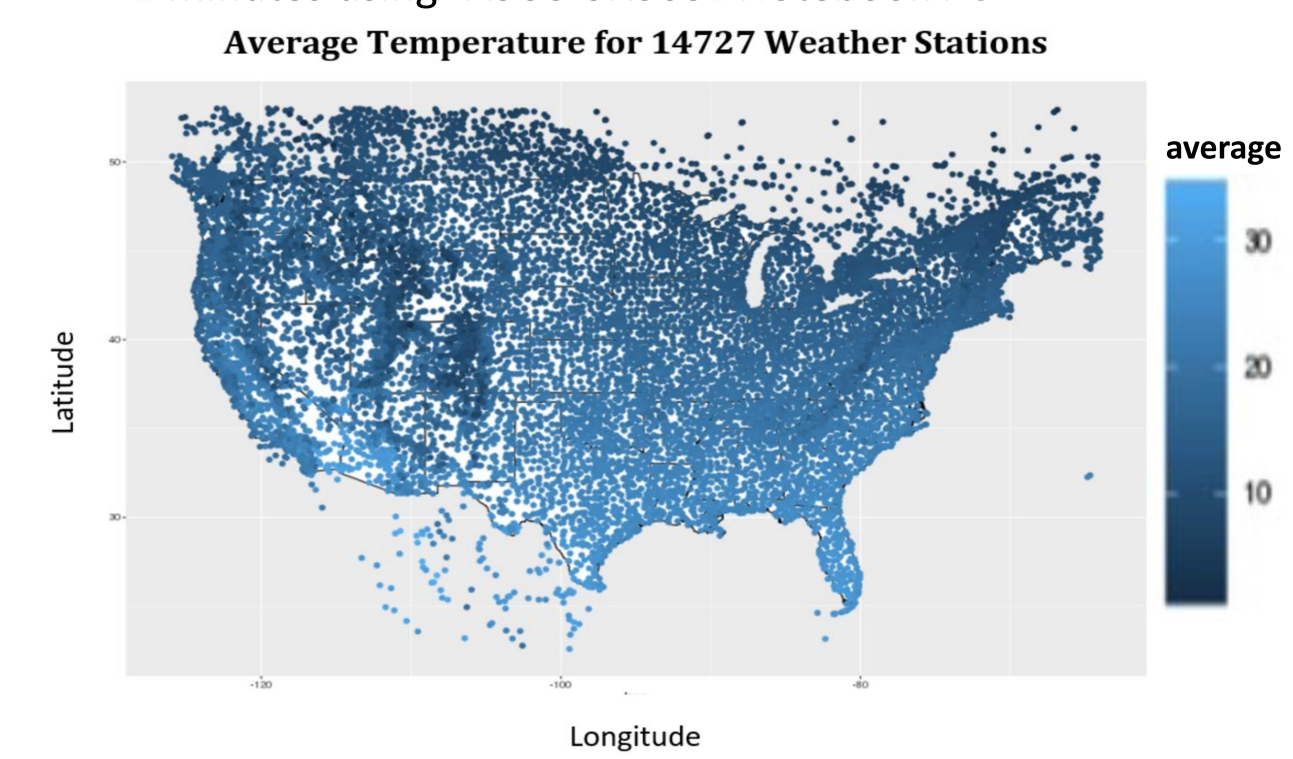
	Lattice Kriging (5000 data)	Standard Kriging (5000 data)
MSPE	2.93251	2.380298

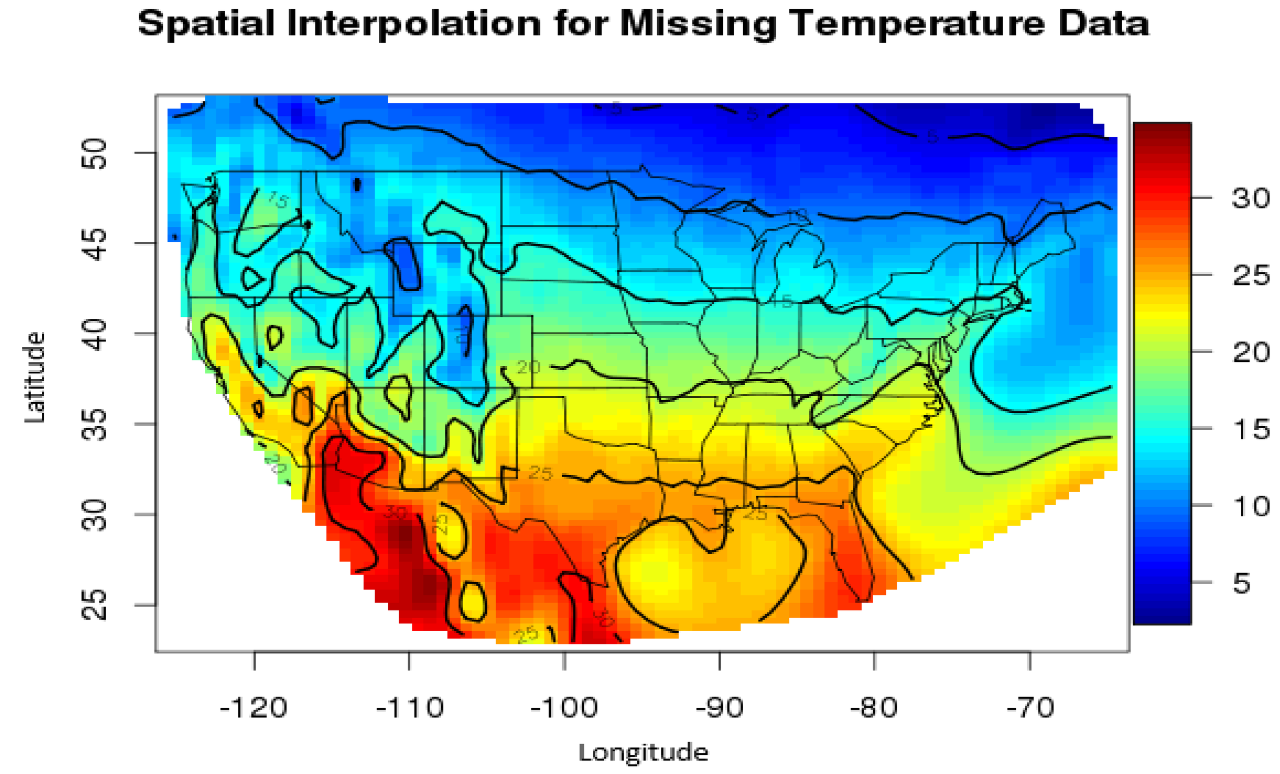
Results

- Exponential and spherical covariance model perform similarly based on MSPE
- Elevation is a significant spatial covariate (including elevation, reduce MSPE)
- Standard kriging is more reliable than lattice kriging
- Lattice kriging is necessary for large data sets; standard kriging is infeasible

Interpolation Using Lattice Kriging

- **Lattice kriging runs quickly with almost 15,000 data points**
  - ~ 2 minutes using "ASUS UX303L Notebook PC"





Maps were made using "ggplot2" and "LatticeKrig" packages in R

Conclusions and Future Work

Conclusions:

- Exponential and spherical covariance model perform similarly
- Elevation should be considered as covariate of temperature
- Standard kriging is more reliable when it is feasible
- Lattice kriging runs more quickly and necessary for large data sets

Future Work:

- Consider more than one covariate
- Extend applications to precipitation data (eg. semicontinuous)
- Work with non Gaussian data

Contact:
Yanrong Mu
The Pennsylvania State University,
Network for Sustainable Climate Risk Management,
Department of Statistics
Email: ywm5117@psu.edu

Acknowledgements:

- This work was supported by the National Science Foundation through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507.
- Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References:

- Diggle, P.J. & Ribeiro Jr, P.J. Model Based Geostatistics Springer, New York, 2007
  Paulo J. Ribeiro Jr & Peter J. Diggle geoR: a package for geostatistical analysis R-NEWS, 2001
- H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.
- Nychka D, Hammerling D, Sain S and Lenssen N (2016). "LatticeKrig: Multiresolution Kriging Based on Markov Random Fields." R package version 5.5
- Oliver Schabenberger, Carol A. Gotway, 2005. *Statistical Methods for Spatial Data Analysis*. Washington, DC: CHAPMAN & HALL/CRC
- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.