Spatiotemporal Estimation of Radon Exposure for Epidemiologic Risk Assessment, Regional Case Study

Kyle J. Sorensen, Richard L. Smith, Eric A. Whitsel, Jason M. Collins

Department of Statistics and Operations Research & Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina @ Chapel Hill

Introduction

Problem

- Radon levels are rising across North America, linked to trends in climate change
- Radon exposure is associated with lung cancer, strokes and other cardiovascular events
- Current estimates of radon exposure are limited, classified into three levels at low spatial resolution

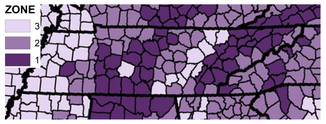

Figure 1: example of county-level resolution US EPA map of GRP radon zones in TN

Aim

- Create a spatial model for the geographic distribution of radon with some quantification of uncertainty
- Incorporate data accounting for geologic, atmospheric, and residential factors
- Provide improved, granular estimates of radon exposure

Solution

- Kriging and other spatial modeling techniques
- Spatial blocking cross-validation

Background

Radon

- Chemical element with symbol Rn, number 86
- Radioactive, colorless, odorless, and invisible gas
- Naturally occurring product of the decay of uranium

Exposure to radon

- Second leading cause of lung cancer
- Linked to strokes and other cardiovascular events
- Evidence of recent increases in North America

Radon and climate change - climate change may indirectly influence rises in radon exposure due to…

- Increased HVAC use
- Recycling of indoor air

Kriging

- Popular spatial modeling algorithm
- Model is a Gaussian process with…
  - Mean – function of covariates
  - Covariance – function of the spatial coordinates

Spatial dependence structure

- Nearby data is more similar than distant data
- Can cause artificially optimistic estimates of model performance

Spatial blocking cross-validation

- Folds from standard k-fold CV → geographically distinct regions
- Provides more realistic measure of model performance

Data

SRRS - EPA's State Residential Radon Survey

- Series of household-level short-term surveys
- 63,291 homes, 42 US states and six US territories
- Conducted between 1986 and 1992

GRP - USGS and EPA's Geologic Radon Potential

- Constructed from geologic, atmospheric and residential survey data
- Three levels:
  - "high" (estimated radon level > 4 picocuries per liter, or pCi/L) – zone 1
  - "moderate/variable" (2–4 pCi/L) – zone 2
  - "low" (< 2 pCi/L) – zone 3

Methodology

Data selection for example analysis

- Region encompassing Tennessee
- 4919 homes across 980 zip-codes
- Relatively high spatial variability in GRP

SRRS analysis

- Translated, log transformed radon exposure data
  - Better match to the assumption of a normally distributed response in the kriging algorithm
- Construction of heatmaps according to kriging estimates of radon exposure

SRRS+GRP analysis

- Formulation of z_i^*
  - Background radon level conditioned on observed radon level
  - Estimation of a variance parameter σ
  - Analogous translation and log transformation
- Construction of heatmaps according to kriging estimates of radon exposure using z_i^*

30-fold spatial blocking cross-validation

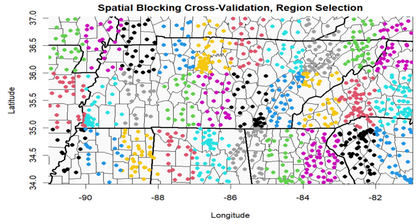

Figure 2: Visualization of region selection for spatial blocking cross-validation

Analysis and Conclusions

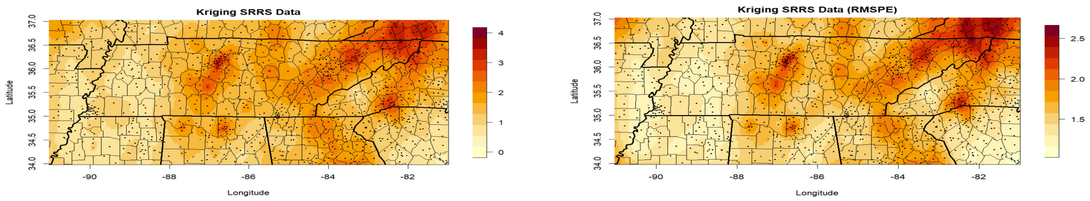

Figure 3: Maps of predicted radon exposure (pCi/L) for TN; predicted values and RMSPE based on SRRS data

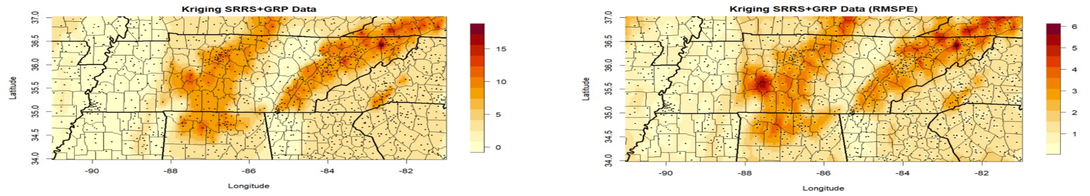

Figure 4: Maps of predicted radon exposure (pCi/L) for TN; predicted values and RMSPE based on SRRS+GRP data

Observations from Figures (2) and (3)

- Significant increase in range of estimated radon exposure (pCi/L)
  - Consistent with inclusion of variance parameter in SRRS+GRP analysis
  - Improved prediction of extreme SRRS values
- Increased range of RMSPE values
  - Relative error decreases in SRRS+GRP analysis

Spatial blocking cross-validation

- SRRS analysis:
  - Mean absolute error of 1.8 pCi/L
- SRRS+GRP analysis:
  - To be determined
  - Model shows signs of improved performance, especially in prediction of extreme values
- Will be used in validation of future models

Future Work and Recommendations

SRRS+RI analysis

- Can use US EPA's Radon Index (**RI**) data in place of GRP data (trichotomization of RI data)
  - Three levels → 15 levels
- Likely to capture additional variability in radon level due to
  - Aerial radioactivity
  - Geology type
  - Soil permeability
  - Architecture type

Bayesian hierarchical modeling

- Allows reversal of the conditioning in the SRRS+GRP analysis
- NIMBLE, an R package aiding in the construction of Bayesian hierarchical models, optimized for spatial data

Temporal component

- More recent radon measurement data exist, such as the US EPA's National Residential Radon Survey
- Forecasting of radon levels is desirable

References

[1] D.R. Roberts et al. "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure". In: Ecography 40.1 (2017), pp. 913–929. doi: 10.1111/ecog.02881.

[2] F.K.T. Stanley et al. "Radon exposure is rising steadily within the modern North American residential environment, and is increasingly uniform across seasons". In: Scientific Reports 9.18472 (2019). doi: 10.1111/ecog.02881.

[3] P. Symonds et al. "Home energy efficiency and radon: An observational study". In: Indoor Air 29.5 (2019), pp. 854–864. doi: 10.1111/ina.12575.

[4] S.H. Kim et al. "The prevalence of stroke according to indoor radon concentration in South Koreans: Nationwide cross section study". In: Medicine (Baltimore) 99.4 (2020). doi: 10.1097/MD.0000000000018859.

[5] W. Field. "Climate Change and Indoor Air Quality". In: Contractor Report (2010), pp. 1–15.

[6] R.L. Smith. Environmental Statistics. Lecture Notes, University of North Carolina at Chapel Hill. http://rls.sites.oasis.unc.edu/postscript/rs/envnotes.pdf, 2001.

[7] C.J. Paciorek et al. "Programming with models: writing statistical algorithms for general model structures with NIMBLE". In: Journal of Computational and Graphical Statistics 26 (2017), pp. 403–417. doi: 10.1080/10618600.2016.1172487.