Spatio-temporal Models for Optimizing Best Management Practices

Robert M. Sholl[1], Trevor J. Hefley[2]
Department of Statistics, Kansas State University

Background

Nonpoint source (NPS) pollution continues to threaten the safety and quality of Kansas water supplies[1]. The Kansas Department of Health and Environment (KDHE) and Kansas Department of Agriculture (KDA) work to subsidize the implementation of Best Management Practices (BMPs) to control NPS pollution introduction. However, as practice implementations become more involved and complex the allocation of funds transforms into a problem of optimizing outcome per dollar spent.

Objective

Develop a modeling framework for informing optimal practice implementation for NPS pollution reduction that:

1. Accounts for uncertainty
2. Maximizes computational efficiency
3. Produces easily interpretable results

Problem Statement

Presently, the allocation of funds is informed via basic quantitative techniques in coordination with expert knowledge. This creates significant burden on the subject matter experts and leaves large vulnerabilities for process breakdown should these experts vacate their positions.
Existing and common techniques for assessing ideal practices at specific locations for optimal NPS pollution reduction focus on large, deterministic systems and simulation models[2,3]. The popularity of these techniques is mostly attributed to their ability to fuse the fundamentally separated data sources for NPS pollution and BMP implementation. However, these methods are often cumbersome to implement, computationally inefficient, and lack inherent uncertainty quantification.

KDA/DOC CSIMS Data

BMP implementation data is recorded in extreme detail via the Cost-Share and Information Management System (CSIMS) database. The data is spatially organized at the 12-digit hydrologic unit code (HUC) (sub watershed) level and recorded temporally from 2004 to 2025.
This database has never been queried and tabled prior to this study and represents a serious challenge from a data cleaning standpoint. Due to the constantly changing nature of conservation practices, some practices have been aggregated to larger categories to avoid "single-use" practices creating problems during the modeling process.

KDHE Ambient Quality Data

NPS pollution tracking comes from a long-term surveillance data set provided by KDHE. The data is spatially organized at the 8-digit HUC (subbasin) level and recorded temporally from 1967 to 2024.
The data is comprehensive but incomplete; some pollutants were not tracked until much later in the data as compared to others. Pollutant loads are averaged across space and time to allow for fusion with CSIMS data. All data recorded prior to CSIMS tracking (2004) is considered as a historic reference.

Methods

To correct for spatial mis-alignment the dollars spent per practice in the CSIMS data is averaged per year, at the 8-digit HUC level. To consider temporal effects a 1-year lag is introduced. This assumes that practices influence pollutant loads long after implementation rather than immediately.
The data is then fit to three spatio-temporal models to assess performance within the study objectives and determine which will be implemented in a final dashboard product.

Multi-scale Geographically Weighted Regression

$$\boldsymbol{y}_{(s)} = \boldsymbol{X}\boldsymbol{\beta}_{(s)} + \boldsymbol{\epsilon}$$

Geographical Gaussian Process Regression

$$\boldsymbol{\eta} \sim GP\big(m(x), k(x, x')\big)$$

$$\boldsymbol{y}_{(s,t)} = \boldsymbol{\alpha}_{(s)} + \boldsymbol{\eta}_{(s)}\boldsymbol{\beta} + \boldsymbol{\phi}_{(s)}\boldsymbol{\tau} + \boldsymbol{\epsilon}$$

Gaussian Markov Random Field

$$\boldsymbol{\eta}_{v \in V} \equiv \boldsymbol{\gamma} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

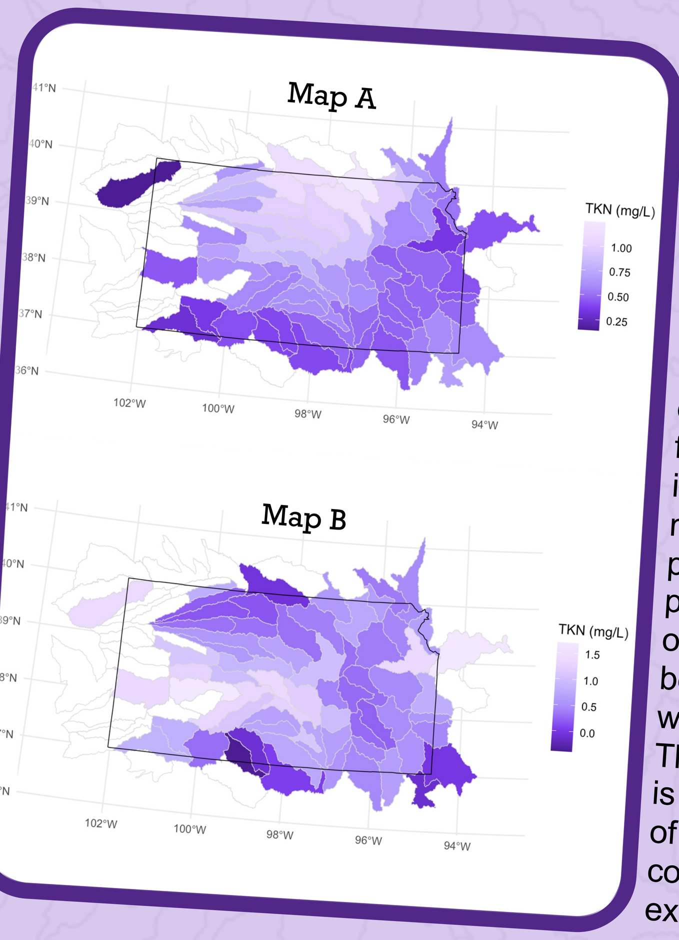
$$\boldsymbol{\Sigma}_{i,j}^{-1} \neq 0 \Leftrightarrow \{i, j\} \notin E$$

$$\boldsymbol{y}_{(s,t)} = \boldsymbol{\alpha}_{(s)} + \boldsymbol{\gamma}'_{(s)}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Results

Current

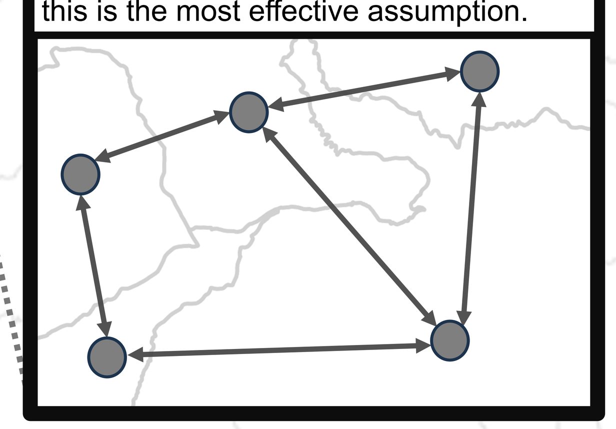
At present, the models can show point predictions across pre-specified space and time, as well as produce interval estimates.
Map A shows the expected distribution of TKN given that expenditure on relevant waste treatment practices is reduced to less than $100 on average for each HUC those practices have been implemented in previously.
Map B shows the same expected distribution but with the max reasonable expenditure on those practices.

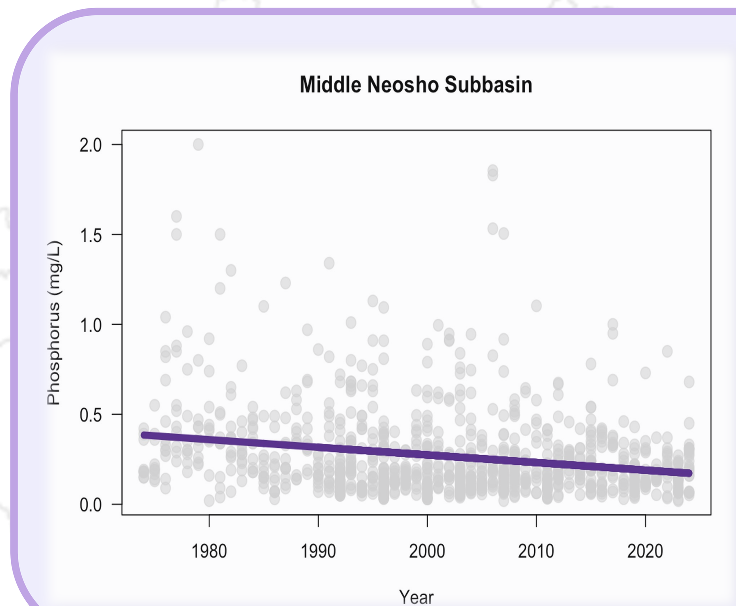


Expected

Due to the size and complexity of the data these models take significant time to run, assess, and tune. The expected result will be a strong answer on the best modeling option for implementation into a KDA dashboard tool for optimal funding of BMPs. The intent is to allow KDA decision makers to plug in the parameters for a proposed practice and determine if the outcome is sufficiently beneficial to follow through with funding.
The timeline for these results is controlled by the duration of model checking which continues to be an experimental process.

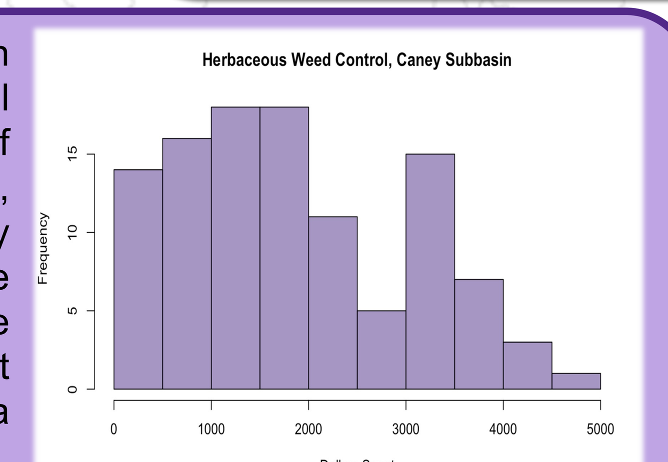
Spatial variability is considered between "queen" nearest neighbors, but further research would be needed to confirm if this is the most effective assumption.





Exploratory analyses have been performed at each spatial resolution to check assumptions, coordinate with subject matter experts, and build a repository for integration with a state government dashboard tool.
Currently, yearly aggregations occur at the HUC-8 level, inserting the assumption that averages are representative of the entire HUC. A possible expansion to this framework could consider the issue of year-over-year variance.

The use of dollars spent on practices as a predictor is useful to homogenize the units of measure for each practice, however more biologically significant units can be leveraged depending on the specific practice. While this isn't a feature of current results it is a feature of the model framework.



Acknowledgements

Kansas State University | Kansas Water Institute

References

[1] Kansas Department of Health and Environment, Division of Environment, Bureau of Environmental Field Services. (2019). *Kansas nonpoint source pollution management plan: 2019 update*.
[2] Arnold, J. G., Srinivasan, R., Muttiah, R. S. & Williams, J. R. (1998). LARGE AREA HYDROLOGIC MODELING AND ASSESSMENT PART I: MODEL DEVELOPMENT 1. , *34*. https://doi.org/10.1111/J.1752-1688.1998.TB05961.X
[3] Montefiore, L. R., & Nelson, N. G. (2022). *Can a simple water quality model effectively estimate runoff-driven nutrient loads to estuarine systems? A national-scale comparison of STEPLgrid and SPARROW.* Environmental Modelling & Software, 150, 105344.

Let's Collaborate!

rmsholl@ksu.edu
linkedin.com/in/robertmsholl