Proposal Title

Master in Statistics Mathematics

Alejandro M. Ouslan

Supervisors:
Dr. Raul E. Macchiavelli
Dra. Damaris Santana
Dr. Julio C. Hernandez
Dr. Roberto Rivera Santiago

University of Puerto Rico, Mayaguez

September 18, 2025

Abstract

This research looks to compare the preformance of Spatial regressions using a predefined weights matrix and semi parametric regessions with a spatila smoother

1 Proposal Keywords

Spatial simulation, Spatial Regressions, GAMs, Tensor Products

2 Introduction

3 Background and Motivation

- Provide background information and context of the research problem.
- What is the research scope?
- What is the question or problem you are trying to answer or solve?
- Why is this important?
- How is it done today, and what are the current limitations and challenges?

4 Systematic Literature Review

- What is the state of the art in this area of research?
- Provide proper references.
- It is recommended that PRISMA guidelines be followed (`https://www.prisma-statement.org/`).

5 Aims and Objectives

- What are you trying to accomplish?
- How are you going to do the work?
- What is new in your approach? How is your approach going to be different from others?
- Why do you think it will be successful?
- Describe the novelty in your approach or how are you going to improve on current approaches.

6 Research Plan and Methodology

Starting for a simple regression of ordinary least squares (OLS) can be defined as follows:

$$y_i = \alpha + \sum_{i=1}^{p} x_i \beta_i + \epsilon \tag{1}$$

where y is our expicatory variable and x is the independent variables. If we wanted to study the corrolation that a location i has on its neighbors, the classic approach would be to define a relationship matrix W which details how each location relates to all other locations. This classic Spatial regressions are an extension of the normal regresion with with the addition of a spatialy term, which can be defined as follows:

$$y = X\beta + \rho W X + \epsilon \tag{2}$$

expanding the matrixes we get the following:

$$y_{it} = \alpha + \sum_{i=1}^{p} x_{it} \beta_i + \rho \sum_{j=1}^{N} w_{ij} x_{jt} + \epsilon \tag{3}$$

This spatial can be addapted spatialy control for either your dependent term

$$y = \alpha + X\beta + \rho WY + \epsilon \tag{4}$$

Or even the error term

$$\begin{aligned} y &= \alpha + X\beta + u \\ u &= \gamma Wu + \epsilon \end{aligned} \tag{5}$$

Continuing from the SDM we can express the model a non linear equation as follows:

$$y = \alpha + \sum_{i=1}^{p} x_{it}\beta_i + f(C_i) + \epsilon \tag{6}$$

where C is the centroid of the observations and $f(C)$ is a function given the centroid of the individuals.

The overall hypothesis is whether the semiparametric methode preforms better on average given that we do not have reason to believe what is W. We expect to find what are to mesure the effects of picking a wrong W and wheather we can miticate them with using a spetial smoother.

7 Prototype Design and Implementation

The Strategy for this resarch is to simulate data in the folowing format:

$$y \sim \alpha + \sum_{i=1}^{p} x_{it}\beta_i + \rho \sum_{j=1}^{N} w_{ij}x_{jt} + \epsilon \tag{7}$$

Where $\epsilon \sim N(0, \sigma^2$ and W is defined in multiple ways this could be show in the following examples:

8 Success and Impact

The SDM generaly carries the problem that is very influenced on how W is defined and there is now systematic methedology to pick an apropiet W. This task is mostly delegated to domain experties. Given that the semiparametric model proposed does not depend on W this research intends to look if this model has a better preformance than the SDM model. Preformance of this model is primarily defined as:

$$\frac{\sum_{n=1}^{N}(y_n - \hat{y}_{nSDM})^2}{N}; \quad \frac{(y - \hat{y}_{GAM})^2}{N} \tag{8}$$

where N is the number of simulations, y is the predetermine outcome variable that we picked before the simulations are run. In addition we would also use look at the performance of β which can be shown in the simal maner:

$$\frac{(\beta - \hat{\beta}_{SDM})^2}{N}; \quad \frac{(\beta - \hat{\beta}_{GAM})^2}{N} \tag{9}$$