# Research Proposal

## Problem background

The spatial interaction model (SIM) is a common model to analyze and simulate various types of flow data. Among them, the gravity model is the most representative, widely used in the study of flow data (like population migration flow). However, those traditional SIMs usually suffer from a significant effect of spatial structure which will cause parameter mis-calibration and significant spatial autocorrelation.(Hu and Pooler, 2002) Some modifications of traditional SIMa have been brought up. Among them, the competing destinations model (CDM) has received growing attention. The model recommends individuals use a hierarchical destination selection process, where the first stage chooses a broad area as the destination range and then the second step selects particular locations from this broad region.(Fotheringham, 1983)

Parameter calibration in SIMs is of great importance, so is in CDMs. The critical argument in SIMs (and CDMs) is distance-decay parameter. Researchers proposed different methods to find optimal distance-decay parameters in CDMs. One popular idea is to transform the gravity models into an optimization problem for solving a linear system by applying a functional transformation to it. And then an error term, representing the difference between theoretical flow and the actual flow, is being introduced and used as the optimization object for the linear programming model.

There are multiple methods that have been studied to seek for the best or optimals results for linear models. Here I am interested in applying the generalized additive model (GAM) to solve the optimization problem so that I could find the optimal distance-decay parameters in CDMs.

## Objective

Building on the problem background, this study will focus on how to build CDM in Python and apply GAM to implement parameter calibration in CDM. Plus, I will conduct a small empirical test of CDM to explore the spatial interaction among different counties in the U.S for the period 2015-2019.

## Data

In this study, the county-to-county migration flows of the United States are used.This dataset covers the decade 2015–2019 and includes migratory flows among counties in the U.S. Additionally, population data will be applied in this project to measure a county's attractiveness; However, there are additional variables that may be used to assess or influence

a destination's attractiveness, such as income, crime statistics, gross domestic product (GDP), and so on. If time allows, this research may be interested in various ways to measure it.

The data mentioned above are all tables with CSV or Excel formats, from the U.S. Census Bureau websites, which could be accessed through the super links above in this proposal. Additionally, cartographic boundaries of counties will be utilized in this project in order to obtain spatial information of counties so that we could calculate the distances among different counties and build the spatial interaction model

Data tidying is necessary in this project, which mainly includes reading data, spatial and aspatial join data, defining projection, distances calculation, etc..

**Python tools**

Spatial Interaction Modeling (SpInt) module will be used in this project. As the name indicated, this module could be used to implement the traditional spatial interaction models and also support the parameter calibration in gravity models by using an entropy maximizing (EM) framework or the equivalent information minimizing (IM) framework. pyGAM may also be referred to in this project, which is for GAM in Python.

Additionally, I will write my own codes in this project because the package so far does not support parameter calibration by using GAM, which is also the main purpose of this project.

**Envisioned challenges**

Obviously, how to combine GAM with parameter calibration would be a big challenge since few studies focus on this. How to speed up the calculation would also be challenging.

**Exploratory data analysis**

So far, I just made a basic plot to show what the county boundary data with population information looks like. This part will be updated later.

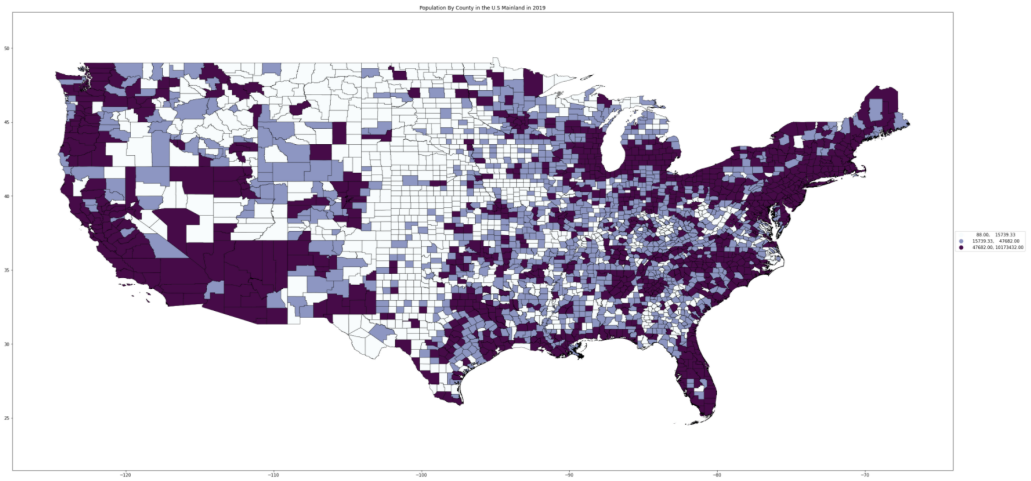


Figure 1. Population By County in the U.S Mainland in 2019

## References

Hu, P., Pooler, J., An empirical test of the competing destinations model. *J Geograph Syst* 4, 301–323 (2002).

Fotheringham AS, A new set of spatial interaction models: The theory of competing destinations. Environment and Planning A 15:15–36 (1983a).