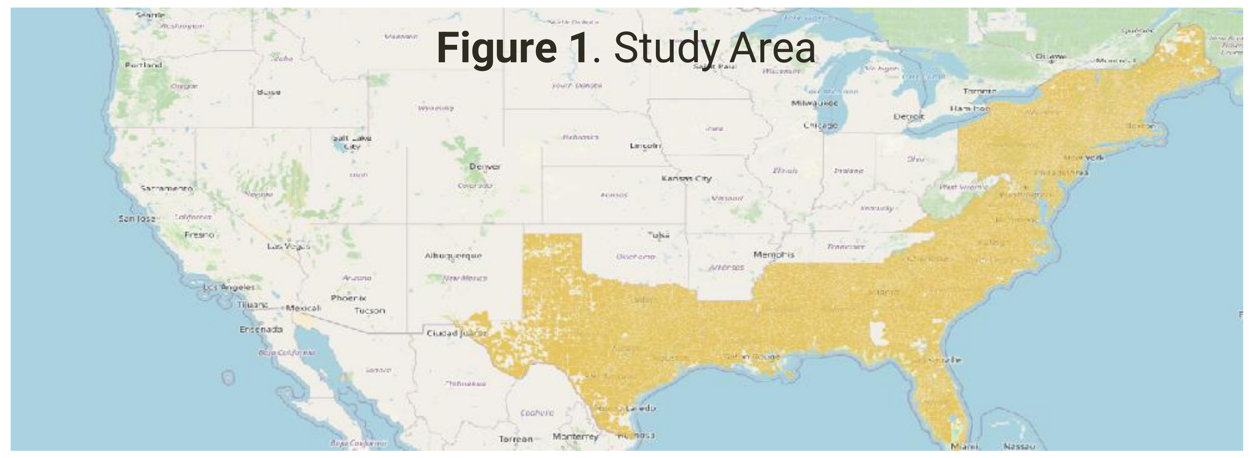


**Figure 1**. Study Area

# Forecasting and Analysis of Housing Market Response to Hurricane Based on Time Series Hawkes Process and Spatio-Temporal Analysis

**Harshavardhan Kamarthi (hkamarthi3@gatech.edu) & Huiying ("Fizzy") Fan (fizzyfan@gatech.edu)**

**Instructor: Dr. Aditya Prakash; CSE 8803 – IUC, Spring 2023**

Georgia Institute of Technology

## Motivation and Problem Definition

**Hurricane Events**
Hurricane events challenge urban housing and infrastructure and are common threat to east coast of the United States.

**Housing Market**
Housing price forecasting is a key building block for urban planning and help inform the public for earlier actions to reduce loss.

**Research Gap**
The predictive power of existing approaches is limited by reduced ability in understanding the complex interactions & spatio-temporal variables.

**Challenges in Modeling**

- Spatial and temporal autocorrelation
- Potentially high covariance
- Features of a discrete (abrupt drop) as well as a continuous variable (impact gradually dissipates)
- Time series of post-events response
- Risk perception vs. risk events
- Hierarchical structure (sociological clusters, etc.)

### *Problem Set-up*

We denote the multivariate time-series of monthly prices for $T$ months over $M$ zip-codes as $\mathbf{y}^{(1:T)} = \{\mathbf{y}^{(t)}\}_{t=1}^{T}$ where $\mathbf{y}^{(t)} \in \mathbb{R}^M$. Also let $\mathbf{y}_i^{(T)} = \{y_i^{(t)}\}_{t=1}^{T}$ be the univariate time-series of zip-code $i$. We characterize the shifts in prices of each of the time-series as *price-change* events $P_i^{(1:T)} = \{(p_1, t_1), (p_2, t_2), \ldots, (p_{N(i,T)}, t_{N(i,T)})\}$ , where $1 < t_j < T, \forall t_j \in P_i^{(1:T)}$ and $N(i,T)$ is the number of sudden price-shifts due to exogenous events such as hurricanes. Variables $p_j$ denote the type of price-change. We characterize these price-change events based on past works studying pricing data . We also observe the past exogenous events that influence price-changes as $E = \{(e_j, t_j)\}_{j=1}^{K}$.

We introduce the *price-change prediction* problem as follows: given past prices of all regions $\mathbf{y}^{(1:T)}$, hurricane events $E$ and past price-changes $P = \{P_i^{(1:T)}\}_{i=1}^{M}$, we generate a model $M$ that, given the data till week $T$ predicts the time-stamp and the type of future price-change $(p_{N(i,T)+1}, t_{N(i,T)+1})$ for each of the zip-codes $i$. Note that a price-change event can occur after some delay from the exogenous event or the event may not significantly impact the price time-series. We also forecast the future $k$ values of the price-change $\mathbf{y}^{(T:T+k)}$ based on the predicted future price-change event.

**Table 2**. MAPE of Forecasts

| Model / MAPE | Avg. | 1 | 6 | 12 |
|---|---|---|---|---|
| RNN | 0.106 | 0.043 | 0.117 | 0.125 |
| ARIMA | 0.209 | 0.169 | 0.195 | 0.318 |

## Data and Methods

### 1. Study Area and Data

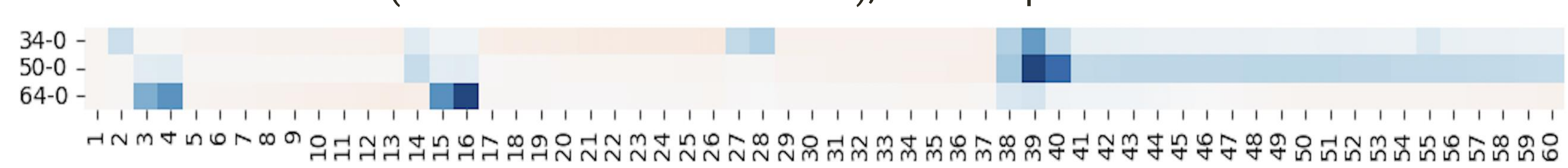
The study uses a spatiotemporal dataset from the eastern coast of the United States (including ME, VT, NH, MA, RI, CT, NY, NJ, DE, MD, PA, VA, NC, SC, GA, FL, AL, MS, LA, and TX) for a 12-year period from August 2008 to December 2019 (before the pandemic). The analysis is conducted at a granularity of the ZIP code unit level on monthly basis. In total, the analysis covers 13,889 ZIP units on the east coast and spans over a period of 137 months.

This study uses five main categories of data (totaling 0.99GB on disk, property value, property value reference, hurricane events, demographic variable and location variable (Table 1).

**Table 1**. Summary of Data

| Data | Source | Size, Format | Temporal | Spatial |
|---|---|---|---|---|
| Property Value | Zillow Research | 1, continuous | Monthly | ZIP code |
| Property Ref. | Zillow Research | 3, continuous | Monthly | N/A |
| Hurricane | NOAA | 1, categorical | Monthly | ZIP code |
| Demographic | Temporal | 30, varied | N/A | ZIP code |
| Location | OpenStreetMap | 4, continuous | N/A | ZIP code |

**Figure 2**. Exploratory data analysis of housing value difference between storm categories over time (month after hurricane event), an example of North Carolina




### 2. Preliminaries

Temporal point process (TPP) is a probability distribution over non-periodic sequences of events over a time-interval. TPP generates a sequence of events $\{(e_i, t_i)\}_{i=1}^{K}$ where $\{t_i\}_i$ are the time of event occurrences and $\{e_i\}_i$ are usually categorical variables of the type of events. TPPs are characterized by intensity function $\{\lambda_l(t)\}_{l=1}^{L}$ for each of the $L$ types of events where, intuitively, $\lambda_l(t)\Delta t$ is approximately the probability of occurrence of event $l$ in the interval $[t, t + \Delta t]$.

The form of the function $\lambda_l$ can vary widely. For example, the widely used Hawkes process has the form $\lambda(t) = \mu(t) + \sum_{t' < t} \phi(t - t')$, where $\mu(t)$ is the base intensity and function $\phi$ models the effect of past events on the probability of the next event's occurrence

### 3. Forecasting Prices

We investigate the housing time-series data and perform univariate forecasting to analyze the complexity of the problem. To accomplish this, we use two types of models, namely, the Autoregressive Integrated Moving Average (ARIMA) model and the Deep Learning-based Recurrent Neural Network (RNN) model, and forecast the rental prices at all zip codes. Average MAPE of forecasts of rental prices is shown in Table 2. Deep learning models capture these patterns better.

### 4. Price Change Predictions

The goal of this study is to measure the impact of hurricane events to the price changes. We have access to past hurricane events and the categories of the flood. We measure the type of price change **10 months** from the event.

We categorize the kinds of price changes into following events: 1) $\geq 30\%$ increase 2) $10 - 30\%$ increase 3) $< 10\%$ price change 4) $10 - 30\%$ decrease 5) $\geq 30\%$ price decrease.

### 5. Model Overview

For each type of event $e$, we learn an intensity function $f_{e,t}$ that uses the history of past events till $t$ and other metadata to predict the future occurrence of the event. The next event is determined as $\mathrm{argmax}_e f_{e,t}(\cdot)$.

The functional form of $f_{e,t}$ can have multiple components that model multiple data sources. We use the following form as a first-cut iteration:

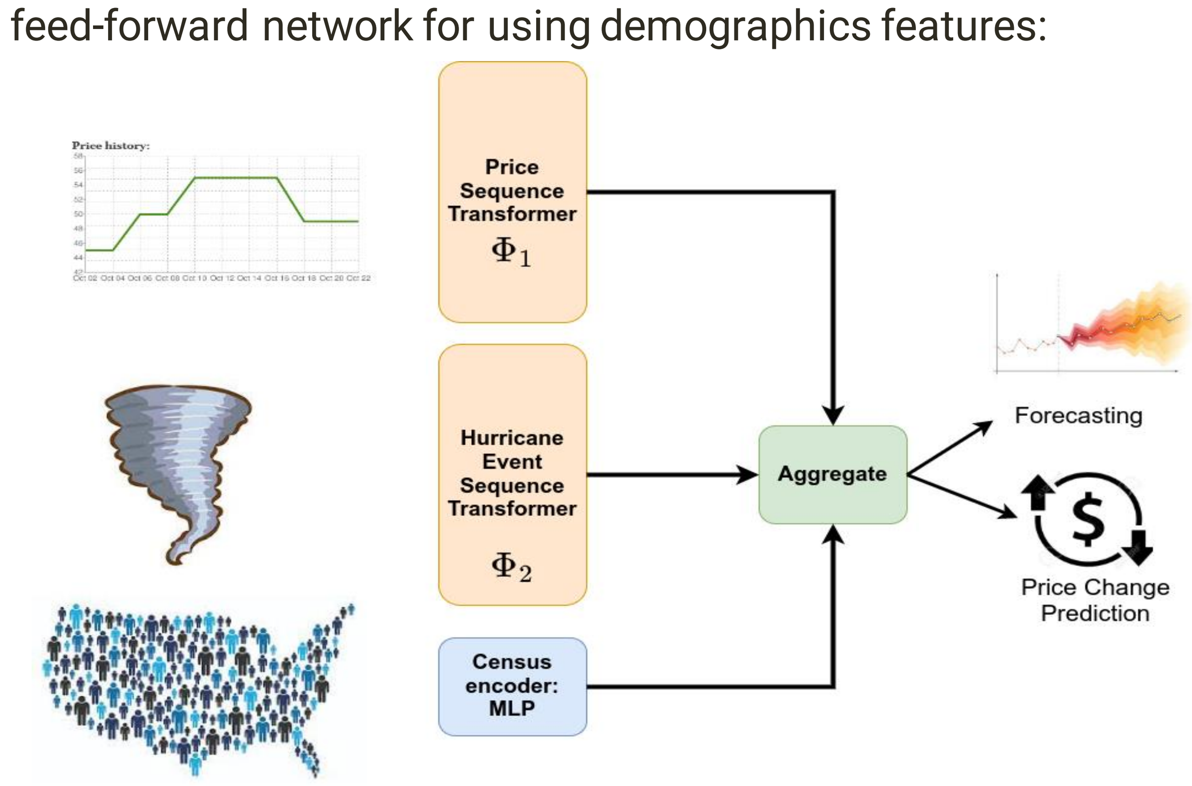
$$f_{e,t} = \text{softplus}\left(\alpha_e \frac{t - t'}{t'} + \mathbf{w}_1^T \Phi_1(\mathbf{y}_i^{1:t'}) + \mathbf{w}_2^T \Phi_2(\mathbf{e}_i^{1:t'}) + b_e\right)$$

where $t'$ is the last hurricane event time-stamp. $\mathbf{y}_i^{1:t'}$ is the history of prices for a zip code till time $t$. $\mathbf{e}_i^{1:t'}$ is the history of hurricane events. $\Phi_1$ is a transformer encoder that outputs a fixed-sized embedding that encodes price history. Similarly, $\Phi_2$ is a transformer encoder that encodes past hurricane events. We use positional encoding similar to to provide time-stamp information.

Therefore, the intensity function encodes delay from previous hurricanes as well as history of prices and events. The model is trained on the log-likelihood score from training data with Monte-Carlo samples used for approximating over integral similar to.

### 6. Demographic Features

Effects of hurricane events on prices can be conditional on demographic features like income and mobility of the population as well as features that indicate proximity to waterbodies. We also use census data collected from 2012 to 2016 for each of the zip codes. We use similar formulation with an additional feed-forward network for using demographics features:




*MAPE of forecasts and F1 scores of classification for baselines and our model.*

| Model/Task | k=3 | k=6 | k=12 | Avg | F1 (class.) |
|---|---|---|---|---|---|
| Linear/Logistic | 0.29 | 0.33 | 0.51 | 0.38 | 0.52 |
| RF | 0.06 | 0.14 | 0.16 | 0.12 | 0.65 |
| MLP | 0.04 | 0.10 | 0.10 | 0.08 | 0.69 |
| ARIMA | 0.17 | 0.20 | 0.32 | 0.23 | NA |
| RNN | 0.04 | 0.12 | 0.13 | 0.10 | 0.63 |
| Ours | 0.04 | 0.08 | 0.10 | 0.07 | 0.82 |

### 7. Forecasting Leveraging Spatial-relations

In this section, we explore a more general problem of forecasting prices at any given time (with or without hurricanes) using past price data as well as census data. We leverage geographical similarity to cluster regions and forecast for each of them.

We test two methods of clustering, by state and by spectral clustering: Use Latitude-Longitude information to first construct an adjacency graph based on $k$ nearest neighbors. Then we cluster with either Louvain's Community Detection Algorithm (LCD) or Spectral clustering (SC) . We tweak the parameters $k$ and $n$ (no. of clusters) to get best result.

For each cluster, we train a single model using past prices, hurricane events, and census data to predict prices for horizons of 3, 6 and 12 months.

We see that effectively using geographical similarity via Spectral Clustering to learn from regions experiencing similar price dynamics is effective in increasing forecasting performance.

## Discussion & Conclusion

**Effective Modeling Scheme**
We derived an effective modeling scheme that can achieve very high predictive and classification accuracy for complicated multi-dimensional dataset.

**Accurate Housing Price Prediction**
Our analysis contributes to the literature with an unprecedentedly accurate prediction of hurricane influence on future housing values.

**Hurricane Influence**
The model with hurricane influence has significantly higher prediction accuracy, implying that hurricanes are highly relevant in housing price fluctuations.

**Difference by Demographic**
Demographic variables related to household type (percentage of family/married household, etc.) strongly influence the prediction of housing value.

**Future Works**
More detailed post-analysis of the model output and a sensitivity analysis will likely provide further insights to shed light on the economic mechanism of the housing market response to hurricanes.