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

Harshavardhan Kamarthi*
hkamarthi3@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Huiying ("Fizzy") Fan*
fizzyfan@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

KEYWORDS

time series analysis, spatio-temporal network, natural hazards, hurricane, housing market

ACM Reference Format:
Harshavardhan Kamarthi and Huiying ("Fizzy") Fan*. 2018. Forecasting and Analysis of Housing Market Response to Hurricane Based on Time Series Hawkes Process and Spatio-Temporal Analysis. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX).* ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

1.1 Background

Hurricane events challenge urban housing and infrastructure. In 2005, Hurricane Katrina resulted in a near-total evacuation and near-half damage to the housing units in New Orleans [32]. Scholars found a permanent property price drop in New York City after Hurricane Sandy, with a magnitude of 17-22% in areas directly influenced by the hurricane, and 8% in other flood-prone areas [24]. Under global climate change, analyses that support an enhanced understanding of the probability and severity of future hurricanes are becoming an increasingly popular and urgent research topic.

Among other subtopics, understanding and forecasting housing price changes in response to hurricane events are fundamental to the discussion. Housing price forecasting is a key building block for land use and transportation modeling [35]. Changes in housing prices are important indicators of community risks in the critical debate of adaptation. vs. migration [3, 33]. The ability to predict future changes in response to hurricane also help inform the public for earlier, precautionary actions to reduce loss [25].

The current research aims to tackle this problem with a modified Hawkes Process, to model both sequential time series and events data with deep-learning-based modification to Hawkes Process. We model spatiotemporal relations across zip codes. The analysis is expected to: (1). identify price changes in the future housing market; (2). differentiate types of price changes; (3). conduct end-to-end forecasting that leverages change points to adapt the parameters of the forecasting model. [1]

1.2 Research Questions

The proposed study can answer a number of important questions:

(1) What is the magnitude of the effect?
(2) When does the effect start to happen? Whether, and if so, when does the impact dissipate?
(3) Does the severity of the hurricane interfere with its influence on housing prices? Is there any pattern of tipping point or cascading failure?
(4) Does the prior risk perception interfere with storms' influence on housing prices?
(5) Do multiple events influence housing prices differently compared to a single event?

Housing market response to hurricanes started to attract attention in Urban and Land Economics since the 2000s [2, 13]. Despite the increasing abundance of data, the approaches to solve the problem remain similar. Previous analyses are primarily rooted in Economics, focusing on techniques like the hedonic pricing model and difference-in-difference model [17, 18, 26]. The hedonic pricing model estimates the value (or disvalue) of given environmental factor of interest by holding other internal and external factors constant. The difference in difference model reveals the estimated value of a factor by comparing the market outcome of two similar markets that differ by the factor of interest. With these approaches, some studies successfully estimated housing price change from historic events, and some provide important implications on confounding variables like coastal amenities and hurricane protective facilities. However, with its simple set up in statistical testing on single factor, the predictive power of such approaches is limited by their reduced ability in understanding the complex interactions among multiple factors, and with spatio-temporal variables. For this and other reasons, many of the aforementioned questions have only been inferred from a case by case observation, but not adequate to make any generic projections into the future.

A preliminary analysis of the data identifies a few challenges with the data that should need to be accounted for in the proposed methods:

- Spatial and temporal autocorrelation
- Many variables with potentially high covariance

*Both authors contributed equally

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

[1] **Website**: https://github.com/kage08/IUC_Project/blob/main/README.md
Code: https://github.com/kage08/IUC_Project

- Housing price change has features of a discrete event (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, county or state, etc.)
- Need a sensitivity analysis or reasoning to help explain and interpret the results

Despite the demand from real estate, transportation, natural hazards management and other disciplines to enhance the predictability of such models, there is a gap in the existing studies. The proposed research will contribute to the literature by introducing methods from other disciplines to integrate multiple agents of factors, disentangle their respective influences, develop a predictive model with complicated dataset, and ultimately attempt to answer the important questions identified at the beginning of the current section.

2 LITERATURE SURVEY

House price modeling. The joint model for longitudinal and time-to-event data provides a possible solution to integrate continuous predictive information with time estimates to discrete events. The model is commonly used in medical studies to predict the probability of acute disease (e.g., heart attack) based on daily measurements (e.g., blood pressure, etc.). This type of model can reduce bias in parameter estimates by explicitly exploring the association between longitudinal (e.g., a regression between risk factor and multiple predictive variables) and time-to-event (e.g., exponential distribution of time to failure based on a parameter that indicates risk factor) models [29]. A Bayesian approach in the joint model can enhance the flexibility of the modeling process, for example, by allowing multiple random change points, incorporating hierarchical modeling, applying historical observations, etc. [1].

Modeling sequence of discrete events. Point processes like Hawkes process [15] and Poisson point process [12] have been used to probabilistically generate or model sequences of discrete events. These methods parameterize an intensity function $\lambda(t)$ that indicates the probability of occurrence of an event at infinitesimal time-period around time t. Many previous works have improved on the simple parametric function of $\lambda(t)$ using non-parametric models [19], recurrent neural networks [9, 22] and transformers [36] to capture complex temporal patterns as well as other multi-modal data sources like social media, audio, videos, health records, etc. [11, 28].

Spatio-temporal sequence modeling and forecasting. Modeling multivariate time-series from different sources such as geographical regions is an important widely studied machine learning problem [27]. unsupervised methods like Toeplitz inverse-Covariance Clustering (TICC) are effective in investigating the complicated interrelationships among multivariate time series data [16]. Graph neural networks(GNN) have also been used to capture spatial and other relation dependencies in many applications such as traffic prediction [6, 20, 34], general time-series forecasting [**?**], etc. GNN [14] refine representations of a node by using information from its neighbors using a message-passing scheme. These models input multivariate time-series from different regions and a network such as adjacency or mobility graph [7]. They use recurrent networks to capture temporal dependencies and graph neural networks to use spatial dependencies across time-series.

3 PROBLEM SETUP

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 [2, 18]. 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.

4 STUDY AREA AND DATASETS

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.

The analysis includes five main data sources, including property and rental value from Zillow Research, prior information on flood risk from FEMA, event updates from NOAA data, demographic control variables from U.S. Census, and location control variables from Open Street Map (OSM) (Table 1).

The data collection methods were derived in a different project and applied to an expanded study area for this work.

4.1 Property and Rental Value

We use the Zillow Home Value Index (ZHVI) and Zillow Rent Index (ZRI) to approximate monthly homeownership property price and home rental price, respectively, at the ZIP code level. ZHVI is a smoothed, seasonally adjusted approximation for the home-ownership housing price of an average-characteristic dwelling at the user-specified geographic level. ZRI is a smoothed approximation for a home rental price in a typical dwelling of a specific geographic level. Zillow provides both indices at the ZIP code level over monthly intervals. The earliest observation of ZHVI for units

Variable	Description
ZHVI	housing value index (buyer market)
ZRI	housing value index (rental market)
Storm	Categorical storm (no, low-, mid-, and high-severity)
FEMA 100	Percent of ZIP areas in FEMA 100-year floodplain
FEMA 500	Percent of ZIP areas in FEMA 500-year floodplain
Demographic	Census demographic variables
Location	Locational variables from OSM

Table 1: Summary of Data

in NC dates to April 1996, while that of ZRI is more recent, starting only in September 2010. For this study, we use monthly ZHVI estimates from January 2008 to December 2019, and ZRI from September 2010 to December 2019. Bulk data download is facilitated by the realEstAnalytics package in R.

Both ZHVI and ZRI exhibit very strong temporal autocorrelation. Temporal autocorrelation refers to the situation where two observations have similar values because of their proximity in time observed. In subsequent analysis, we correct this problem by transforming ZHVI and ZRI to their respective change rates. The optimal time lags are determined with a multi-temporal ANOVA test and t-test.

4.2 Prior Flood Risk Information

Prior risk information is approximated with the FEMA Flood Hazard Layer, which classifies the East Coast of the U.S. into five flood categories, including regulatory floodway, 100-year floodplain, 500-year floodplain, 100-year floodplain in future condition, and areas with minimal risk. 100-year and 500-year floods refer to flood magnitudes that are seen, on average, once every 100 and 500 years. Each year, the magnitude of flood equivalent to 100- and 500-year flood is expected to happen at 1% and 0.2% chance, respectively. Floodplain is an area inundated under a flood event. Regulatory floodway contains preserved areas of land and watercourse to discharge floodwater in 100-year flood conditions. In FEMA Flood Hazard Layer, future conditions refer specifically to future land-use conditions under current zoning maps or comprehensive land-use plans [21].

4.3 Hurricane Events Data

In this study, we obtain hurricane events updates in GIS format from the National Hurricane Center (NHC) of the National Oceanic and Atmospheric Administration (NOAA) of the United States. We include all hurricane events recorded by NOAA in the Atlantic Basin Tropical Cyclone Zone (AL), totaling 204 events from August 2008 to December 2019, among which 56 storms have had an impact on inland states. For each event, we use hurricane wind swath polygons to approximate the impact area. The wind swath polygons delineate wind swaths with different wind speeds (34, 50, and 63 kts). In this study, we use the three wind speed levels to approximate the low, moderate, and high severity of storms. The hurricane data are downloaded with a script written in R [10]. For each ZIP unit in each month, the storm category is marked by the highest probability level wind swath that intersects the ZIP unit boundary in that particular month. If no wind swath polygon intersects the ZIP unit, we record a “0” for the storm variable.

4.4 Demographic Variables

We use the 5-year estimate American Community Survey data [5] for demographic control variables. We choose the 5-year ACS estimate over 1-year or 3-year estimates for two reasons. First, ZIP code level data are only available in 5-year estimates. Second, although 5-year data are less time-sensitive, they have better accuracy than 1-year or 3-year estimates (REF). The 5-year ACS is available for periods 2005-2009, 2006-2010, 2007-2011, 2012-2016, 2013-2017, 2014-2018, 2015-2019, 2016-2020, and 2017-2021. This study includes ACS data from 2012-2016 to control for variations among ZIP units. For these variables, variation across time is less important and the same value is applied to the same ZIP unit over time.

4.5 Location Variables

The location variables include access to beach amenities, availability of schools, hospitals, groceries, and other types of facilities, and distance to the urban center. The location variables are derived from Open Street Map (OSM). The OSM data is crowd-sourced geographic information, widely used by research and recognized for its comprehensive coverage and good accuracy.

5 EXPLORATORY DATA ANALYSIS

An exploratory data analysis was conducted, where changes in housing prices after 1-60 months of hurricane occurrences were compared among different categories of the hurricane. Figure 1 shows the estimated difference in magnitude and direction of change (an example of North Carolina, the complete result is attached in a separate file to this submission). As mentioned in the previous section, numbers indicate different categories of storms (0 mean no storm, 34 is a lower-severity storm, 50 mid-severity, and 64 high-severity). Colors of cells in the graph shows the difference in changes in housing value between two ZIP units that were hit by some category of storm versus ZIP units that did not experience any influence from hurricane. For example, the dark blue color in the 4th month of the 64-0 row in North Carolina (NC) suggests that areas that are hit by high-severity storm typically see a notably lower housing price increase (or higher housing price drop) compared to the control group that are not hit by the hurricane.

Here are a summary of findings from the exploratory data analysis:

- For almost all states, there is a pretty consistent "background" trend in the effect of hurricane on housing price (for example, housing price changes in ZIP areas that experience high-severity hurricane are always lower than the control in VA), and this trend seems to be linear with time. This could be related to confounding factors like proximity to the coastal area, access to amenities, and distance to urban center.
- For many states, there are sudden changes in housing market response, typically observed at 2-4 months after hurricane, 14-17 months, 27-28 months, and 38-42 months. These might be the price changes that are of interest to this research.

- States behave very differently on hurricane. The housing market in northern states (ME, VT and NH) experience minimal influence from hurricane. New England states either see a "background" influence from hurricane and its confounding factors (CT, NY, NJ, and DE), or sees minimal influence except from certain months of sudden price changes (MA and RI). Southeastern states (NC, SC, GA, and FL) experience a potentially stronger influence from hurricane on housing price, as demonstrated in the sudden price drop at certain time points, as well as seemingly permanent price drop after the 38th month.

6 DETECTION AND FORECASTING HOUSING PRICE CHANGES

In this section, we briefly describe the class of models and methods we use for tackling the *price-change prediction* as well as forecasting tasks.

6.1 Preliminaries

Temporal Point Process. 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 λ_l can vary widely. For example, the widely used Hawkes process [15] has the form $\lambda(t) = \mu(t) + \sum_{t'<t} \phi(t - t')$, where $\mu(t)$ is the base intensity and function ϕ models the effect of past events on the probability of the next event's occurrence.

Neural models for TPP. There has been an increased interest in leveraging the representation power of the neural models to model more complex and flexible intensity functions [8, 11, 28]. For example, the Transformer Hawkes Process [36] encodes history past events till last event at t' into latent representations $\mathbf{h}(t')$ using a transformer [31]. The intensity function is then derived as

$$\lambda_l(t) = f_l\left(\alpha_l \frac{t - t'}{t'} + \mathbf{w}_l^T \mathbf{h}(t') + b_l\right) \quad (1)$$

where the first term models delay from the last event, the second term models the latent information from the transformer, and the last term is the bias component. f_l is the softplus function that constrains the intensity function within $[0, 1]$. Using the trained intensity functions, we can predict the next event occurrence via Monte Carlo sampling or other approximate integration methods.

6.2 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. The training data used for these models range from 2005 to the current year, while the forecasted results are for the period spanning from 2018 to 2023. We forecast for up to 12 months ahead as seen in Table 2.

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

Table 2: Average MAPE of forecasts of rental prices

We notice that simpler ARIMA models can't capture long-term relations due to changes in trends of time-series due to various economic and natural causes. Therefore, deep learning models capture these patterns better.

6.3 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. As the first step, chose these categories based on observations from past data. The problem of predicting future price fluctuation can be stated as a classification task and we indeed used classification models as baselines. To systematically capture the impact of current as well as past storm events and their long-term impacts as well as the history of prices, we use a Neural Hawkes Process to better capture the generative process that determines price fluctuation events.

6.4 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 $\arg\max_e f_{e,t}(\cdot)$.

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:

$$f_{e,t} = \text{softplus}\left(\alpha_e \frac{t - t'}{t'} + \mathbf{w}_1^T \Phi_1(\mathbf{y}_i^{1:t'}) + \mathbf{w}_2^T \Phi_2(\mathbf{e}_i^{1:t'}) + b_e\right) \quad (2)$$

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. Φ_1 is a transformer encoder that outputs a fixed-sized embedding that encodes price history. Similarly, Φ_2 is a transformer encoder that encodes past hurricane events. We use positional encoding similar to [36] 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 [36].

6.5 Classification Results

We compare our model with classification-based models that use RNN, Transformers, and Random Forest from past price data (Table 3). Our model is most accurate among all models.

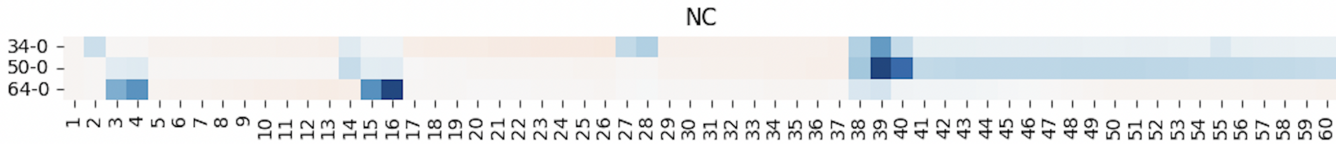

Figure 1: Results of the Preliminary Analysis (an example of the ZHVI housing market in North Carolina, rental market not included)

Model	Accuracy	F1
RNN	0.74	0.63
Transformer	0.81	0.75
RF	0.77	0.65
NHP	0.86	0.73

Table 3: Accuracy of event prediction

6.6 Using Demographic features

Effects of hurricane events on prices can be conditional on important demographic features like income and mobility of the population as well as features that indicate proximity to waterbodies. Therefore, we use census data collected from 2012 to 2016 for each of the zip codes as additional metadata along with sequence datasets for both classification and forecasting tasks.

For classification, we use a similar formulation as Equation 2 with an additional feed-forward network for using demographics features:

$$f_{e,t} = \text{softplus}\left(\alpha_e \frac{t - t'}{t'} + \mathbf{w}_1^T \Phi_1(\mathbf{y}_i^{1:t'}) + \mathbf{w}_2^T \Phi_2(\mathbf{e}_i^{1:t'}) + \mathbf{w}_3^T \Phi_3(\mathbf{z}_i) + b_e\right) \tag{3}$$

where zi is the census features of the region and Φ_3 is a feed-forward network with 3 hidden layers.

For the forecasting task, we use a similar model except we concatenate the embeddings from each data source and perfrom a linear aggregation to predict the forecast value k months ahead:

$$y_i^{(t'+k)} = \mathbf{W}^T \left(\Phi_1(\mathbf{y}_i^{1:t'}) \oplus \Phi_2(\mathbf{e}_i^{1:t'}) \oplus \Phi_3(\mathbf{z})\right) \tag{4}$$

where $\oplus$ is a concatenation operator. The overall structure of the model is in Figure 2.

We used 44 hurricane events for training and 12 hurricane events for testing. The test hurricane events chronologically occur after training events. The classification task is to predict the price change after 6 months from the event whereas the forecasting task involves prediction of price 3, 6 and 12 months from the event. We do 5-fold cross-validation to choose the optimal set of hyperparameters for the learning rate, batch size, and hidden unit size of feed-forward layers.

For forecasting, we compare our model with two sequential models: RNN and ARIMA that uses only past sequence data of prices and events as well as Linear Regression, Random Forest Regression ad 3 layer feed-forward neural network which uses the census data and last 6 months of price data as features. For classification, we compare with the same models except we train for classification. We also use logistic regression instead of Linear Regression.

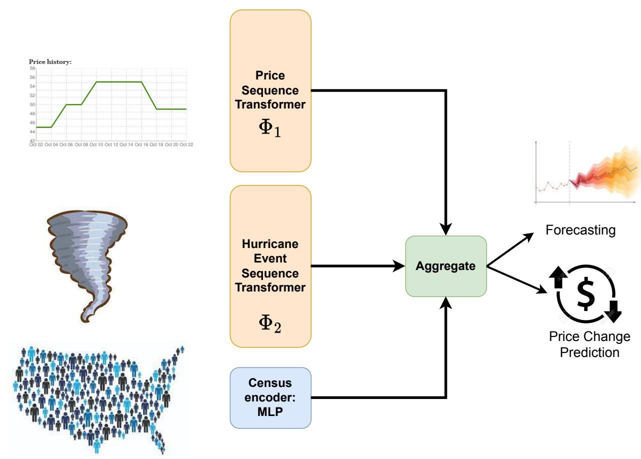

Figure 2: Model architecture

Table 4: MAPE of forecasts and F1 scores of classification for baselines and our model.

Model/Task	k=3	k=6	k=12	Avg	F1 (Classification)
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**

First, we observe in Table 4 that our model that combines sequential and static features outperforms others in forecasting and classification tasks. We also observe that simple MLP baseline that uses census data outperforms sequential models, showcasing the importance of demographic features in determining the nature of price dynamics in a region.

Exploring Demographics importance. In order to access the importance of each of the demographic feature, we used Integrated Gradient method[30] to capture the importance weights of inputs. We considered only the census features and normalized their absolute values for input between 0 to 1. We finally computed the average importance across all events in Figure 3.

The most important features detected by the model include "House Per Family member", "House Per Non-family member", features based on education and the ratio of houses per owner.

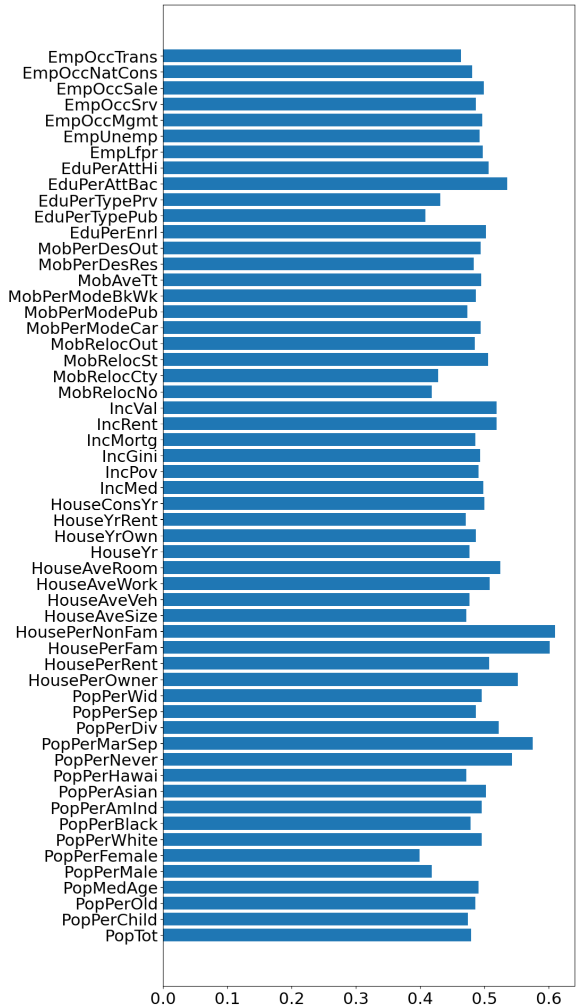

Figure 3: Importance weights of demographic features captured from census data

6.7 Forecasting leveraging Spatial-relations

In previous sections, we forecasted and performed price change predictions on a group of zip codes that were affected by an hurricane event. 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.

While we forecast for about 290 zip codes simultaneously, different regions may have different dynamics and factors affecting the price changes, not all of which are explicitly captured by exogenous variables. Forecasting for individual zip codes separately is infeasible due to the small dataset size. Therefore, we leverage geographical similarity to cluster regions together and forecast for each of them separately.

We test with the following methods of clustering:

- Cluster by state
- 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) [4] or Spectral clustering (SC) [23]. 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 use the training data from 2008 to 2016 and evaluate from 2017 to 2021. For tuning the hyperparameters we perform 5-fold cross validation on training data and report the two best configurations in the results reported in Table 5.

Table 5: Average MAPE scores of forecasting using different spatial aggregation methods

Cluster/Horizon	3	6	12
All Regions	0.17±0.07	0.22±0.11	0.25±0.14
State	0.19±0.08	0.21±0.12	0.24±0.17
LCD (k=3)	0.11±0.04	0.12±0.05	**0.15±0.03**
LCD (k=5)	0.13±0.06	0.15±0.06	0.17±0.04
SC(k=3, n=4)	**0.09±0.03**	**0.11±0.06**	**0.15±0.03**
SC(k=5, n=8)	0.10±0.04	0.13±0.06	0.18±0.04

As reported in Table 5, Spectral clustering with $k = 3$ and $n = 4$ provided the best results on test data followed by Louvain's Community Detection method. This shows that effectively using geographical similarity to learn from regions experiencing similar price dynamics is effective in increasing forecasting performance.

7 CONCLUSION

We collect and analyze the housing price data from 20 states over a 12-year period identifying useful statistical indicators to quantify the effect of hurricanes on housing prices. Based on these observations, we design a Neural Hawkes process-based method for predicting future price changes and forecasting future house prices leveraging past events, price data and census features. We also explore leveraging geographical similarity across zip codes and observe improved forecasting performance.

As part of future work, we would conduct a more detailed sensitivity analysis of the model to specific important geographical and demographic features to better understand the economic mechanisms underlying the housing market response to hurricanes.

REFERENCES

[1] Maha Alsefri, Maria Sudell, Marta García-Fiñana, and Ruwanthi Kolamunnage-Dona. 2020. Bayesian joint modelling of longitudinal and time to event data: A methodological review. *BMC Medical Research Methodology* 20 (4 2020), 1–17. Issue 1. https://doi.org/10.1186/S12874-020-00976-2/TABLES/5

[2] Okmyung Bin, Thomas W. Crawford, Jamie B. Kruse, and Craig E. Landry. 2008. Viewscapes and Flood Hazard: Coastal Housing Market Response to Amenities and Risk. *Land Economics* 84 (8 2008), 434–448. Issue 3. https://doi.org/10.3368/LE.84.3.434

[3] Richard Black, Stephen R.G. Bennett, Sandy M. Thomas, and John R. Beddington. 2011. Migration as adaptation. *Nature 2011 478:7370* 478 (10 2011), 447–449. Issue 7370. https://doi.org/10.1038/478477a

[4] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008, 10 (2008), P10008.

[5] US Census Bureau. 2018. Understanding and Using American Community Survey Data: What All Data Users Need to Know. (2018).

[6] Weiqi Chen, Ling Chen, Yu Xie, Wei Cao, Yusong Gao, and Xiaojie Feng. 2020. Multi-range attentive bicomponent graph convolutional network for traffic forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 3529–3536.

[7] Songgaojun Deng, Shusen Wang, Huzefa Rangwala, Lijing Wang, and Yue Ning. 2020. Cola-GNN: Cross-location attention based graph neural networks for long-term ILI prediction. In *Proceedings of the 29th ACM international conference on information & knowledge management*. 245–254.
[8] Prathamesh Deshpande, Kamlesh Marathe, Abir De, and Sunita Sarawagi. 2021. Long horizon forecasting with temporal point processes. In *Proceedings of the 14th ACM international conference on web search and data mining*. 571–579.
[9] Nan Du, Hanjun Dai, Rakshit Trivedi, Utkarsh Upadhyay, Manuel Gomez-Rodriguez, and Le Song. 2016. Recurrent marked temporal point processes: Embedding event history to vector. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1555–1564.
[10] R Core Team (Firm. 2019. A language and environment for statistical computing. (2019). https://www.R-project.org/.
[11] Vinayak Gupta, Srikanta Bedathur, Sourangshu Bhattacharya, and Abir De. 2021. Learning temporal point processes with intermittent observations. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 3790–3798.
[12] Peter Guttorp and Thordis L Thorarinsdottir. 2012. What happened to discrete chaos, the quenouille process, and the sharp markov property? some history of stochastic point processes. *International Statistical Review* 80, 2 (2012), 253–268.
[13] Daniel G. Hallstrom and V. Kerry Smith. 2005. Market responses to hurricanes. *Journal of Environmental Economics and Management* 50 (11 2005), 541–561. Issue 3. https://doi.org/10.1016/J.JEEM.2005.05.002
[14] William L Hamilton, Rex Ying, and Jure Leskovec. 2017. Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584* (2017).
[15] Alan G Hawkes. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 1 (1971), 83–90.
[16] Vadim Kapp, Marvin Carl May, Gisela Lanza, and Thorsten Wuest. 2020. Pattern Recognition in Multivariate Time Series: Towards an Automated Event Detection Method for Smart Manufacturing Systems. *Journal of Manufacturing and Materials Processing 2020, Vol. 4, Page 88* 4 (9 2020), 88. Issue 3. https://doi.org/10.3390/JMMP4030088
[17] Seung Kyum Kim and James K. Hammitt. 2022. Hurricane risk perceptions and housing market responses: the pricing effects of risk-perception factors and hurricane characteristics. *Natural Hazards* 114 (12 2022), 3743–3761. Issue 3. https://doi.org/10.1007/S11069-022-05541-2/TABLES/6
[18] Seung Kyum Kim and Richard B. Peiser. 2020. The implication of the increase in storm frequency and intensity to coastal housing markets. *Journal of Flood Risk Management* 13 (9 2020), e12626. Issue 3. https://doi.org/10.1111/JFR3.12626
[19] Shuang Li, Shuai Xiao, Shixiang Zhu, Nan Du, Yao Xie, and Le Song. 2018. Learning temporal point processes via reinforcement learning. *Advances in neural information processing systems* 31 (2018).
[20] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. 2017. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926* (2017).
[21] Anthony S. Lowe. 2003. The federal emergency management agency's multi-hazard flood map modernization and the national map. (2003), 1133–1135.
[22] Hongyuan Mei and Jason M Eisner. 2017. The neural hawkes process: A neurally self-modulating multivariate point process. *Advances in neural information processing systems* 30 (2017).
[23] Andrew Ng, Michael Jordan, and Yair Weiss. 2001. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems* 14 (2001).
[24] Francesc Ortega and Süleyman Taşpınar. 2018. Rising sea levels and sinking property values: Hurricane Sandy and New York's housing market. *Journal of Urban Economics* 106 (7 2018), 81–100. https://doi.org/10.1016/J.JUE.2018.06.005
[25] Katerina Papagiannaki, Vassiliki Kotroni, Kostas Lagouvardos, and Giorgos Papagiannakis. 2019. How awareness and confidence affect flood-risk precautionary behavior of Greek citizens: The role of perceptual and emotional mechanisms. *Natural Hazards and Earth System Sciences* 19 (7 2019), 1329–1346. Issue 7. https://doi.org/10.5194/NHESS-19-1329-2019
[26] Adam B. Pollack and Robert K. Kaufmann. 2022. Increasing storm risk, structural defense, and house prices in the Florida Keys. *Ecological Economics* 194 (4 2022), 107350. https://doi.org/10.1016/J.ECOLECON.2022.107350
[27] Zahraa Al Sahili and Mariette Awad. 2023. Spatio-Temporal Graph Neural Networks: A Survey. *arXiv preprint arXiv:2301.10569* (2023).
[28] Oleksandr Shchur, Ali Caner Türkmen, Tim Januschowski, and Stephan Günnemann. 2021. Neural temporal point processes: A review. *arXiv preprint arXiv:2104.03528* (2021).
[29] Maria Sudell, Ruwanthi Kolamunnage-Dona, and Catrin Tudur-Smith. 2016. Joint models for longitudinal and time-to-event data: A review of reporting quality with a view to meta-analysis. *BMC Medical Research Methodology* 16 (12 2016), 1–11. Issue 1. https://doi.org/10.1186/S12874-016-0272-6/TABLES/5
[30] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. 2017. Axiomatic attribution for deep networks. In *International conference on machine learning*. PMLR, 3319–3328.
[31] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
[32] Jacob Vigdor. 2008. The Economic Aftermath of Hurricane Katrina. *Journal of Economic Perspectives* 22 (9 2008), 135–54. Issue 4. https://doi.org/10.1257/JEP.22.4.135
[33] Kira Vinke, Jonas Bergmann, Julia Blocher, Himani Upadhyay, and Roman Hoffmann. 2020. Migration as Adaptation? *Migration Studies* 8 (12 2020), 626–634. Issue 4. https://doi.org/10.1093/MIGRATION/MNAA029
[34] Bing Yu, Haoteng Yin, and Zhanxing Zhu. 2017. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875* (2017).
[35] Chengxiang Zhuge, Chunfu Shao, Jian Gao, Chunjiao Dong, and Hui Zhang. 2016. Agent-based joint model of residential location choice and real estate price for land use and transport model. *Computers, Environment and Urban Systems* 57 (5 2016), 93–105. https://doi.org/10.1016/J.COMPENVURBSYS.2016.02.001
[36] Simiao Zuo, Haoming Jiang, Zichong Li, Tuo Zhao, and Hongyuan Zha. 2020. Transformer hawkes process. In *International conference on machine learning*. PMLR, 11692–11702.