Supplementary material

SPATIOTEMPORAL PATTERNS AND PREDICTION OF MULTI-REGION HOUSE PRICES VIA FUNCTIONAL MIXED EFFECTS MODEL

Yilin CHEN, Haitao ZHENG

International Journal of Strategic Property Management
https://doi.org/10.3846/ijspm.2025.23639

1. Community information

Table S1 shows specific information for all communities in San Francisco, with Y representing participation in the study and N representing non-participation.

Table S1. Community list

ID	Community	Participation in research or not (Y/N)	ID	Community	Participation in research or not (Y/N)
1	Seacliff	Y	60	Peralta Heights	Y
2	Lake Street	Y	61	Holly Park	Y
3	Presidio National Park	N	62	Merced Manor	Y
4	Presidio Terrace	N	63	Balboa Terrace	Y
5	Inner Richmond	Y	64	Ingleside	Y
6	Sutro Heights	Y	65	Merced Heights	N
7	Lincoln Park / Ft. Miley	N	66	Outer Mission	Y
8	Outer Richmond	Y	67	Ingleside Terraces	Y
9	Golden Gate Park	N	68	Mt. Davidson Manor	Y
10	Presidio Heights	Y	69	Monterey Heights	Y
11	Laurel Heights / Jordan Park	Y	70	Westwood Highlands	Y
12	Lone Mountain	Y	71	Westwood Park	Y
13	Anza Vista	Y	72	Miraloma Park	Y
14	Cow Hollow	Y	73	McLaren Park	N
15	Union Street	Y	74	Sunnydale	Y
16	Nob Hill	Y	75	Visitacion Valley	Y
17	Marina	Y	76	India Basin	N
18	Telegraph Hill	Y	77	Northern Waterfront	Y
19	Downtown / Union Square	Y	78	Hunters Point	Y
20	Tenderloin	Y	79	Candlestick Point SRA	N
21	Civic Center	Y	80	Cayuga	Y
22	Hayes Valley	Y	81	Oceanview	Y
23	Alamo Square	Y	82	Apparel City	N
24	Panhandle	Y	83	Bernal Heights	Y
25	Haight Ashbury	Y	84	Noe Valley	Y
26	Lower Haight	Y	85	Produce Market	N
27	Mint Hill	N	86	Bayview	Y

ID	Community	Participation in research or not (Y/N)	ID	Community	Participation in research or not (Y/N)
28	Duboce Triangle	Y	87	Silver Terrace	Y
29	Cole Valley	Y	88	Bret Harte	Y
30	Rincon Hill	Y	89	Little Hollywood	Y
31	South Beach	Y	90	Excelsior	Y
32	South of Market	Y	91	Portola	Y
33	Showplace Square	N	92	University Mound	Y
34	Mission Bay	Y	93	St. Marys Park	Y
35	Yerba Buena Island	N	94	Mission Terrace	Y
36	Treasure Island	N	95	Sunnyside	Y
37	Mission Dolores	Y	96	Glen Park	Y
38	Castro	Y	97	Western Addition	Y
39	Outer Sunset	Y	98	Aquatic Park / Ft. Mason	Y
40	Parkside	Y	99	Fishermans Wharf	Y
41	Stonestown	Y	100	Cathedral Hill	Y
42	Parkmerced	N	101	Japantown	Y
43	Lakeshore	Y	102	Pacific Heights	Y
44	Golden Gate Heights	Y	103	Lower Pacific Heights	Y
45	Forest Hill	Y	104	Chinatown	Y
46	West Portal	Y	105	Polk Gulch	Y
47	Clarendon Heights	Y	106	North Beach	Y
48	Midtown Terrace	Y	107	Russian Hill	Y
49	Laguna Honda	Y	108	Financial District	Y
50	Lower Nob Hill	Y	109	Inner Sunset	Y
51	Upper Market	Y	110	Parnassus Heights	Y
52	Dolores Heights	Y	111	Forest Knolls	Y
53	Mission	Y	112	Buena Vista	Y
54	Potrero Hill	Y	113	Corona Heights	Y
55	Dogpatch	Y	114	Ashbury Heights	Y
56	Central Waterfront	Y	115	Eureka Valley	Y
57	Diamond Heights	Y	116	St. Francis Wood	Y
58	Crocker Amazon	Y	117	Sherwood Forest	Y
59	Fairmount	N			

2. Description of the comparison model

In our experiments, we use several existing spatiotemporal models for house price prediction to demonstrate the validity of our proposed model. Details of these comparative models are given below.

2.1. Single-output LSTM

Long Short-Term Memory (LSTM) is proposed based on the original Recurrent Neural Networks (RNN) (Hochreiter & Schmidhuber, 1997). It regulates the balance between memorization and forgetting by adding some multi-threshold gates. It can solve the problem of vanishing gradient during optimization in RNN. LSTM has been successfully applied to many sequence learning problems (Cho et al., 2014).

Single-output LSTM is used for single-objective, long-term, time-dependent learning and time series prediction. In this paper, the study uses Single-output LSTM for comparison modeling when a single point of a time series is used as the input set. The experiment is repeated for each community's data.

The following are some examples of using the Single-output LSTM for forecasting. Shi (2023) selected LSTM models for their study and explored and compared the suitability of these models using a dataset of second-hand house prices in Beijing. Usmani and Shamsi (2021) used RNN and LSTM Networks to forecast the National Stock Exchange and New York Stock Exchange stock market. Their study emphasized the adaptability of neural networks to various market environments. They obtained higher accuracy compared to traditional models.

2.2. Multi-output LSTM

Lee (2022) applied a multi-output LSTM model to predict multiple regions' house prices and transaction volumes. Data from a time series of a single point is generally limited to that location, ignoring some of the surrounding information. And multi-output LSTM can predict more accurately by exploiting the correlation of communities. It complements the problem that single-point datasets are insufficient to learn about relevant patterns.

In this paper, we investigate using multi-output LSTM for comparison experiments when time series data of the entire community is used as the input set. Sequential input values from each community are fed into an initial dense layer, and then LSTM is used as an encoder and decoder. LSTM autoencoder will implement feature learning and decompose auxiliary information such as spatial attributes. The extracted learning features are used for temporal modelling. Then, there is a dropout layer, which randomly sets previous neurons to zero by a certain percentage to prevent overprotection. Repeat the combination of the LSTM layer and the dropout layer twice. Finally, a multi-output dense layer is added to the architecture to output the predictions of the entire community. We train the multi-output LSTM 200 times with a batch size of 32.

2.3. CNN-LSTM

Ge (2019) proposed a model that jointly applies CNN and LSTM to the spatiotemporal prediction of house prices. It predicts the response of target spatial locations by considering the spatial dependence between communities. Finally, it is applied to predict community house price data in New York City and Beijing.

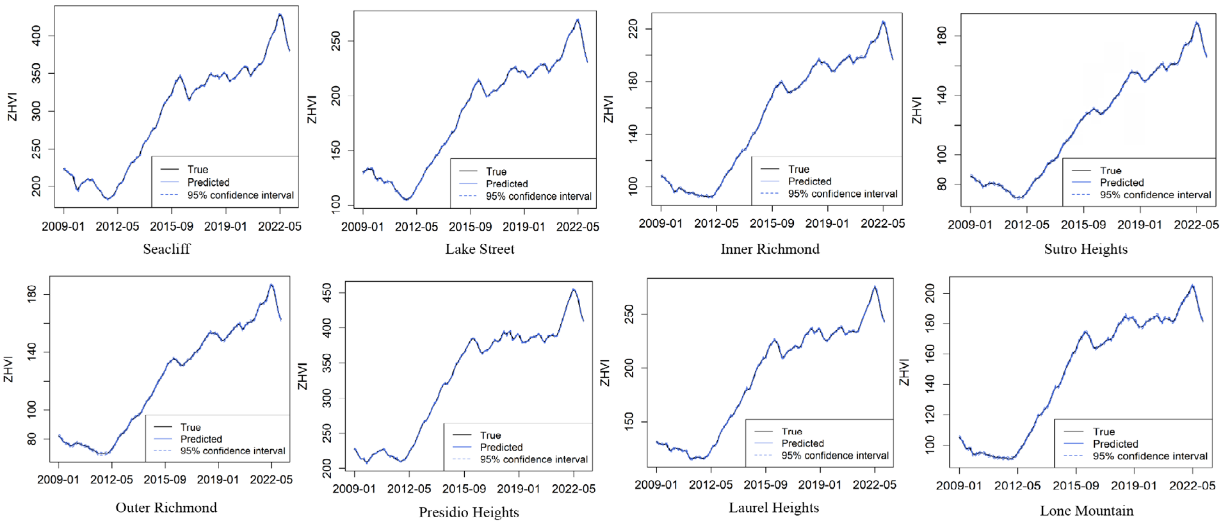
CNN-LSTM constructs the spatial structure in the following way. The structure is changed into a graph based on the community's spatial coordinates, and a graph convolutional neural network is utilized to indicate the spatial dependencies. Communities are represented as nodes. Distance weights between communities are calculated using Euclidean distances to connect weights topographically with weighted neighboring lines. Then, the feature vector matrix is fed into the convolutional layer. In the convolutional layer, the spatial dimensions are extracted using the peripheral nodes in the network. In the pooling layer, the dimensionality of the network will be reduced. The input two-dimensional matrix is compressed into a one-dimensional vector by convolution and pooling process. Finally, spatial correlations in the form of feature maps are obtained.

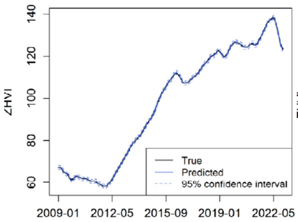
After that, the vectors are fed into the LSTM layer. The LSTM layer first selectively forgets factors such as historical house price data from different communities. The LSTM unit decides to store new information to update the state. Finally, LSTM determines the output values and provides them to the fully connected layer. The CNN-LSTM output is decoded using the fully connected layer to get the final prediction.

Before constructing the CNN-LSTM model, several hyperparameters need to be preset. These include basic parameters such as the number of CNN layers, the number of LSTM layers, the number of fully connected layers, and the number of nodes per layer of the data. The optimal hyperparameters are also determined using a randomized search method with 5-fold cross-validation. The basic structure of the model was finalized. CNN uses a two-layer structure. The first layer is a convolutional layer, and the second layer is a pooling layer. We use an LSTM layer containing 300 nodes and a fully connected layer containing 100 nodes. The above structural configuration obtained the best prediction performance in this experiment.

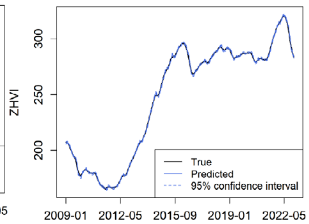
3. All community fitting curves

Below are the fitted curves for all 101 communities in San Francisco that participated in the study.

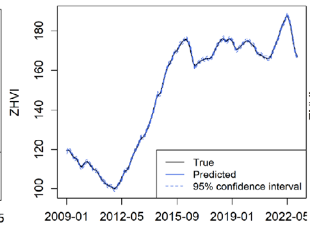




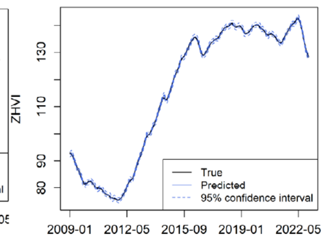
Anza Vista



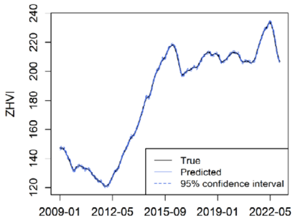
Cow Hollow



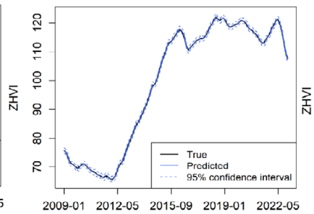
Union Street



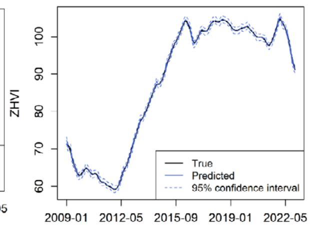
Nob Hill



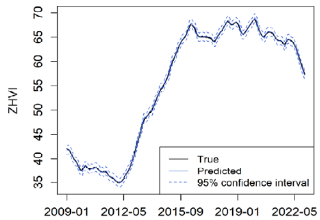
Marina District



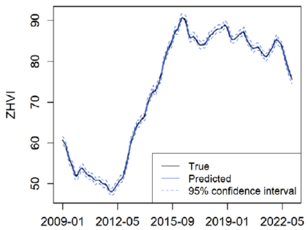
Telegraph Hill



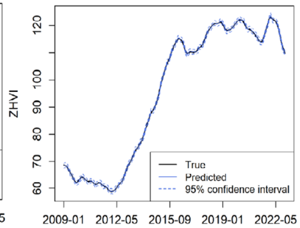
Downtown



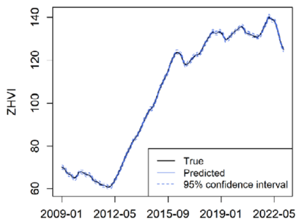
Tenderloin



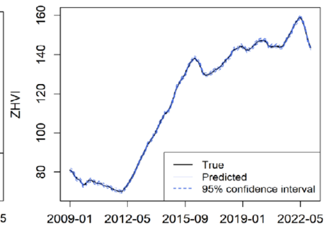
Civic Centerf



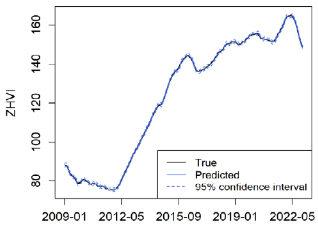
Hayes Valley



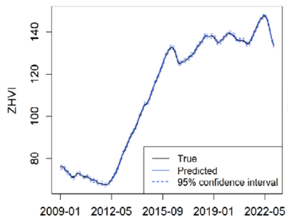
Alamo Square



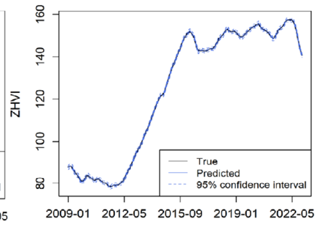
Panhandle



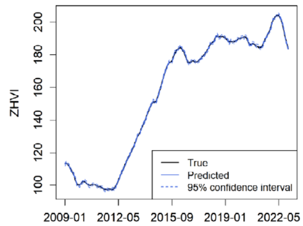
Haight



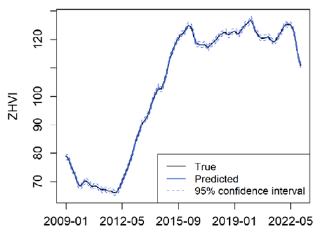
Lower Haight



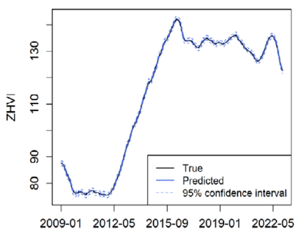
Duboce Triangle



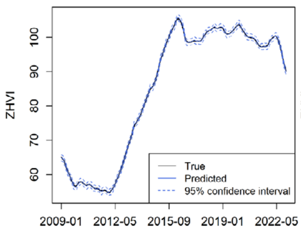
Cole Valley



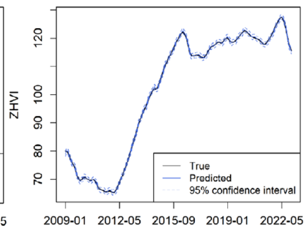
Rincon Hill



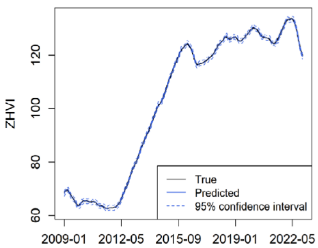
South Beach



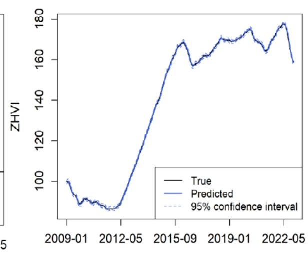
South of Market



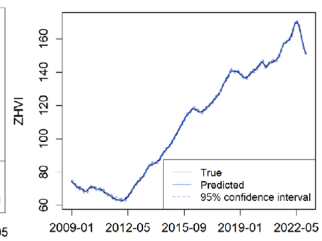
Mission Bay



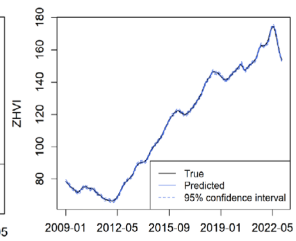
Mission Dolores



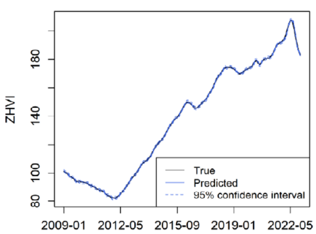
Castro



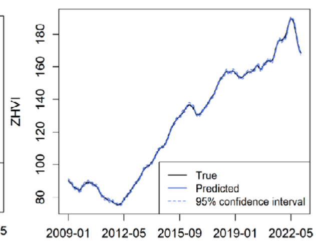
Outer Sunset



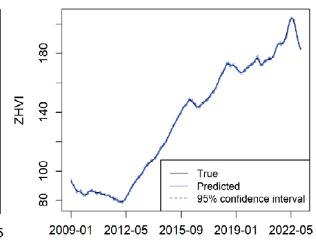
Parkside



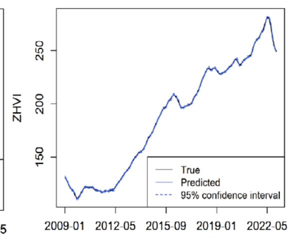
Stonestown



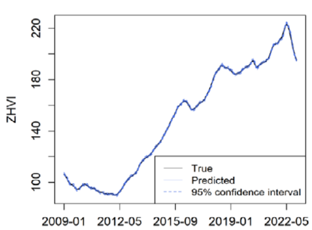
Lakeshore



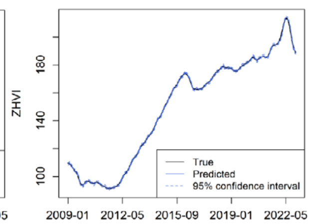
Golden Gate Heights



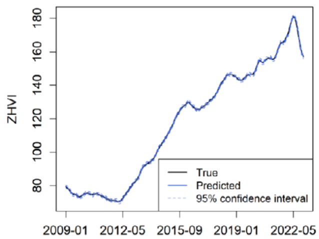
Forest Hill



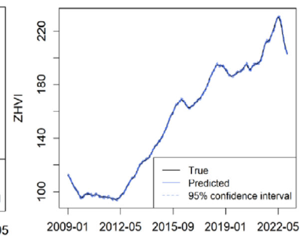
West Portalff



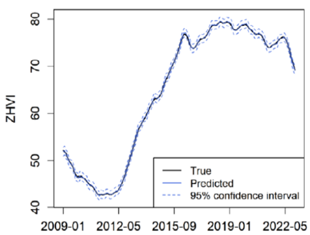
Clarendon Heights



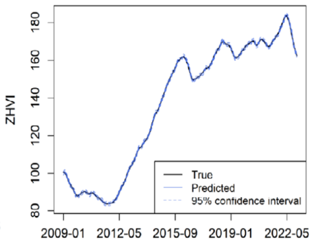
Midtown Terrace



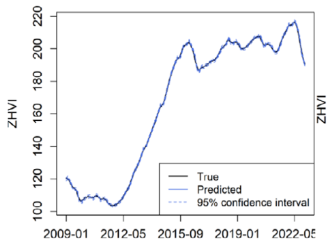
Laguna Honda



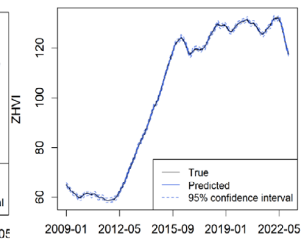
Lower Nob Hill



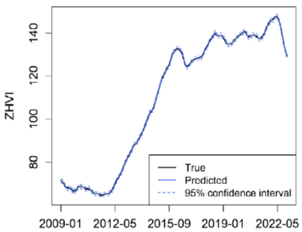
Upper Market



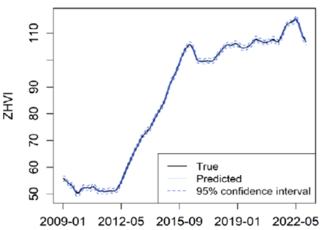
Dolores Heights



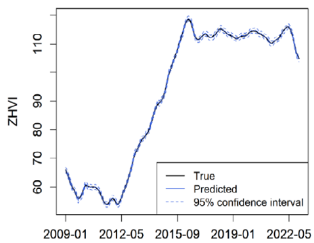
Mission



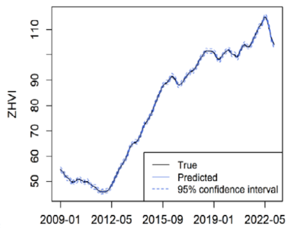
Potrero Hill



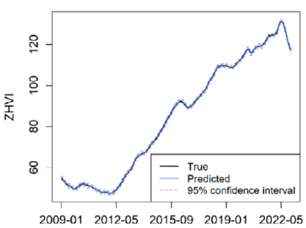
Dogpatch



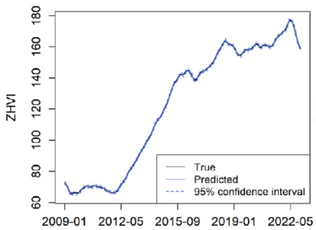
Central Waterfront



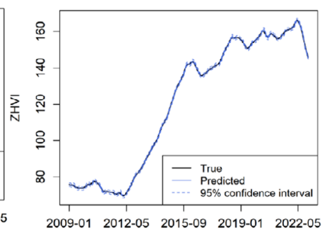
Diamond Heights



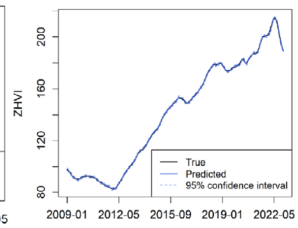
Crocker Amazon



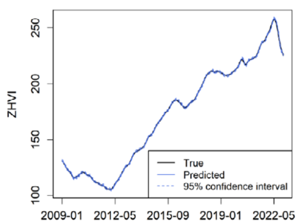
Peralta Heights



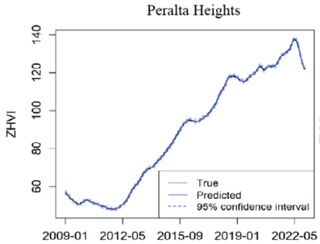
Holly Park



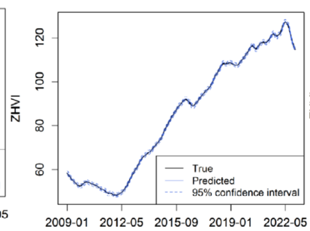
Merced Manor



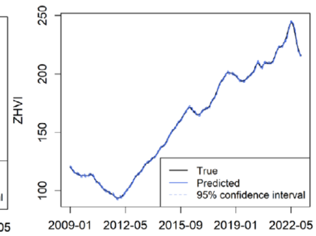
Balboa Terrace



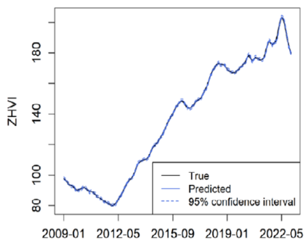
Ingleside



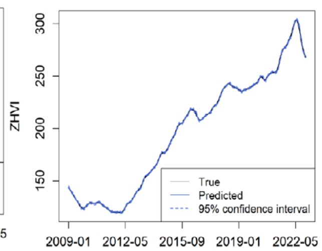
Outer Mission



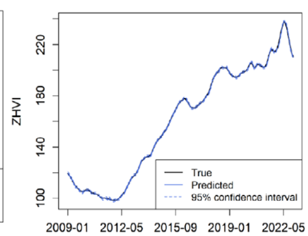
Ingleside Terrace



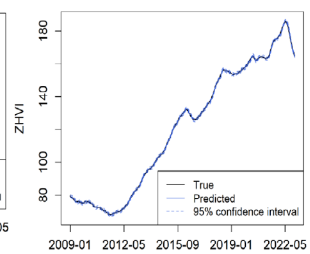
Mount Davidson Manor



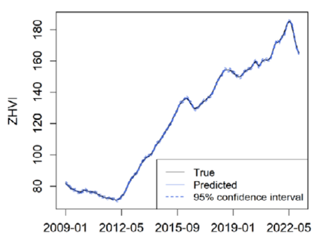
Monterey Heights



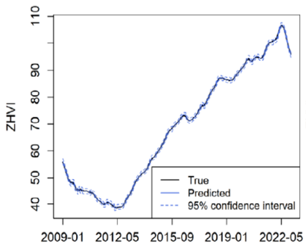
Westwood Highlands



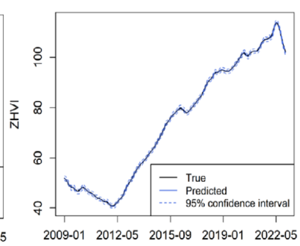
Westwood Park



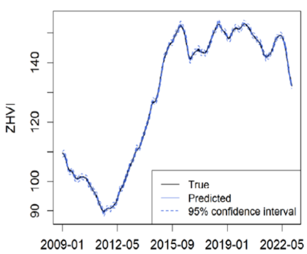
Miraloma Park



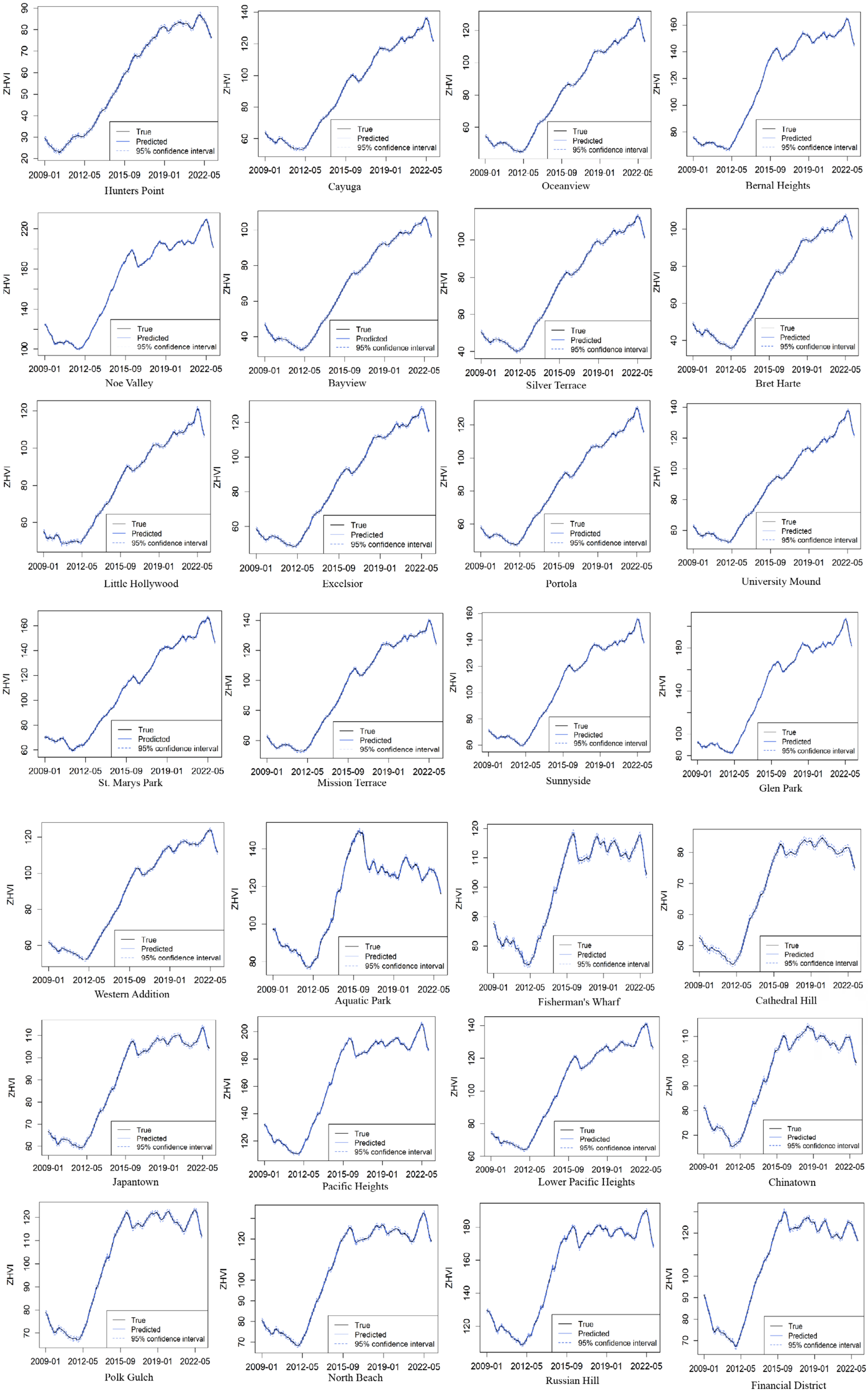
Sunnydale

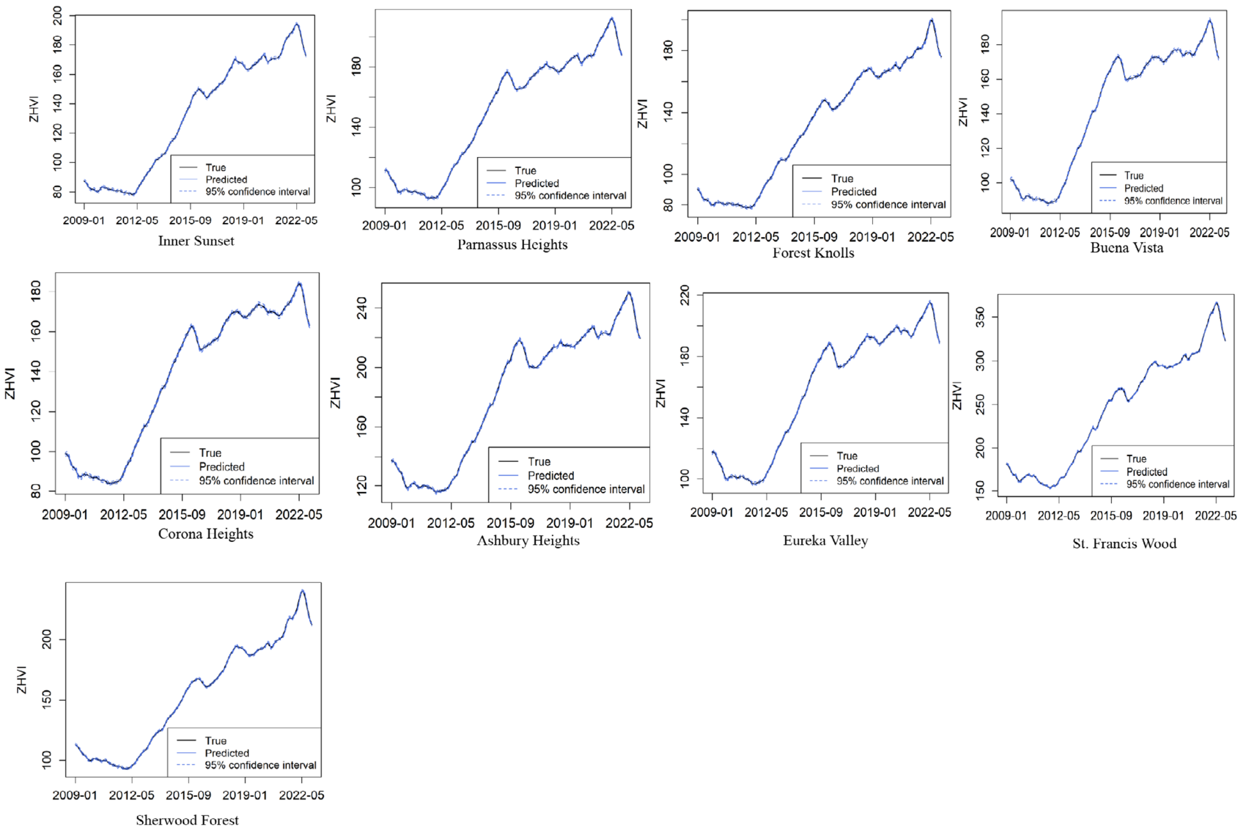


Visitacion Valley



North Waterfront


ZHVI
True
Predicted
95% confidence interval
2009-01 2012-05 2015-09 2019-01 2022-05
Hunters Point
Cayuga
Oceanview
Bernal Heights
Noe Valley
Bayview
Silver Terrace
Bret Harte
Little Hollywood
Excelsior
Portola
University Mound
St. Marys Park
Mission Terrace
Sunnyside
Glen Park
Western Addition
Aquatic Park
Fisherman's Wharf
Cathedral Hill
Japantown
Pacific Heights
Lower Pacific Heights
Chinatown
Polk Gulch
North Beach
Russian Hill
Financial District



References

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). *Learning phrase representations using RNN encoder–decoder for statistical machine translation*. arXiv. https://doi.org/10.48550/arXiv.1406.1078

Ge, C. (2019, June 10–13). *A LSTM and graph CNN combined network for community house price forecasting* [Conference presentation]. 2019 20th IEEE International Conference on Mobile Data Management (MDM), Hong Kong, China. https://doi.org/10.1109/MDM.2019.00-15

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Lee, C. (2022). Forecasting spatially correlated targets: Simultaneous prediction of housing market activity across multiple areas. *International Journal of Strategic Property Management*, *26*(2), 119–126. https://doi.org/10.3846/ijspm.2022.16786

Shi, S. (2023). Comparison of real estate price prediction based on LSTM and LGBM. *Highlights in Science, Engineering and Technology*, *49*, 294–301. https://doi.org/10.54097/hset.v49i.8521

Usmani, S., & Shamsi, J. A. (2021). News sensitive stock market prediction: Literature review and suggestions. *PeerJ Computer Science*, *7*, Article e490. https://doi.org/10.7717/peerj-cs.490