Land Value Estimation in Maricopa County, AZ: A Space-Time Local Regression Approach

Jeffrey P. Cohen*

David C. Lincoln Fellowship Symposium

January 29, 2021

professorjeffrey@gmail.com

*JPC Milo Consulting LLC; West Hartford, CT. The views expressed in this presentation are those of the author. He is not speaking, acting, or making representations on behalf of the University of Connecticut, its School of Business, Department of Finance, Connecticut Transportation Institute, or the Center for Real Estate and Urban Economic Studies; nor is he expressing institutional endorsement in relation to this research. The author, and not the University of Connecticut, is solely responsible for the entire content and for any errors.

Introduction/Background

Focus: How do we separate value of land from overall property values?

In practice:

Maricopa County in 2017 Assessments dataset: approx. 99% of residential property value assessments have approximately 20% allocated to land value

This can be very restrictive

Land Residual Approach (Davis et al., 2017):

Land value is the entire parcel value net of the depreciated construction costs

Land and structures are additive

Alternative: Option Value approach (Clapp, Cohen, Lindenthal, 2021)

Other alternatives:

Longhofer and Redfearn (2009):

-use vacant land and locally weighted regressions (LWR) to develop land value estimates

-spatial variation in implicit structure prices incorporated in a land valuation equation

Recent Innovations in LWR Estimation for Valuation:

-Cohen, Coughlin, Zabel (2020) – overall property value assessment, incorporate time in the kernel for LWR

Other Nonparametric Analysis of Land Value Estimation: Local Polynomial Regressions

Clapp (2004)

Cohen, Coughlin, Clapp (2017)

Build on These Other Approaches

Innovations:

-Somewhat greater tractability Relative to Clapp (2004) and Cohen, Coughlin, Clapp (2017): Use LWR instead of LPR

-Cleveland and Devlin (1988);

-McMillen and Redfearn (2010): hedonic applications

-I Incorporate a 3rd "Location" Dimension in LWR Estimation: Time

-Dependent Variable: Log of sale price per square foot of land area; attempts to address the "location value" problem

-Compare difference between my estimates and the Maricopa Assessor's land value estimates

Drawbacks:

-missing data on bedrooms and number of bathrooms

-estimated land price per square foot at the point of lat/long, rather than over the whole property

-Maricopa approach to assessing land value is based on 20% of total parcel assessment

Approach

Start with nonlinear version of hedonic model:

$ln(SP/LA)_{it} = f(Z_i, S_i, t_i) + \varepsilon_{it}$

SP/LA is sale price per square footage of land area

Z is vector of characteristics

S is lat and long

t is year of sale

We want only the part of [ln(SP/LA) – value of characteristics] that is dependent on location (lat, long, time)

Essentially, this requires estimation of:

$partres = ln(SP/LA) - Z\,\hat{\alpha}_R$

and then estimation of: $\quad$ $\text{Ln(Land value)} = E(partres \mid S)$

Note in the literature: Coefficients on Z and the nonlinear part of the model must be statistically independent; requires some adjustments ➔ Calculate Robinson (1983, 1986) Coefficients.

In other words, location value and improvement value must be statistically independent.

First, use LWR to determine sales price (*ln(SP/LA)**) and value of structure (Z^*) independent of space and time of sale:

- $ln(SP/LA)^* = ln(SP/LA) - E(ln(SP/LA) \mid S)$

- $Z^* = Z - E(Z \mid S)$.

Use Gaussian kernel, with kernel weights:

$$k_{ijt} = \exp\left(-\left(\frac{d_{ij}}{b}\right)^2\right) \cdot \exp\left(-\left(\frac{\tau_{ij}}{b}\right)^2\right)$$

Then, run separately for each NBHD, OLS regression of:

*ln(SP/LA)** on Z^*: $ln(SP/LA)^* = (Z^*)\alpha_R + \varepsilon$.

Call these parameter estimates: $\hat{\alpha}_R$. They are independent of nonlinear part of the model used below to calculate land value from *partres*

Then, can calculate for each property sale: $partres = ln(SP/LA) - Z\,\hat{\alpha}_R$

Take *partres* and use LWR to estimate for each property sale:

$$\text{Ln(Land value)} = \text{E}(\textit{partres} \mid \text{S})$$

Then, as described by Cohen, Coughlin and Zabel (2020) and more generally for OLS by Wooldridge (2016), land value per square foot of land area is given as:

$$\text{Land Value}_i = \widehat{\theta_\iota}\exp(\text{Ln(Land value)}),$$

where $\widehat{\theta_\iota} = \overline{\exp\,(\widehat{u_\iota})}$, and $(\widehat{u_\iota})$ is the vector of residuals from the LWR of *partres* on S for target point i.

These are not "location values" as in Cohen, Coughlin and Clapp (2017);

They are land value per square foot of land area

(benefits and drawbacks)

Data

Dependent variable: Log of sale price per square foot of land area

The variables in *Z* include:

-age at time of sale;

-number of bath fixtures;

-square footage;

-a dummy variable taking the value of 1 when the number of stories exceeds 1, and 0 otherwise;

-building quality (as denoted by r_iclass, where 0 is the lowest quality and 7 is the highest quality).

Tradeoff: Parsimony (avoiding the "curse of dimensionality" with LWR) and explanatory power

Using LWR captures many features that vary by location and time

Data Merging/Filtering:

-keep single family properties (most with pred_use between 100 and 186)

-Drop properties in bottom and top percentile of prices

-Useful deed types based on hedonic testing (discussions w/John Clapp):

The following 8 classes of deed types are retained in the dataset:

“WD, SD, JD, JC, AG, TD, D, 00”.

-Drop if sale price missing

-merge these estimates with the database, based on parcel id, with assessor assessed values for 2017, for parcels that have data on both the assessed values in that database and the land area per square foot in my database (so that I can calculate assessed land value per square foot of land area).

-Drop repeat sales

-Focus on sales between 2015-2018 to be comparable with 2017 assessor’s values

-Drop properties with assessed land values less than $0.01 (fewer than 500 parcels)

-Left with 160,689 SF improved residential parcels that were sold 2015-18 and also have assessed value data

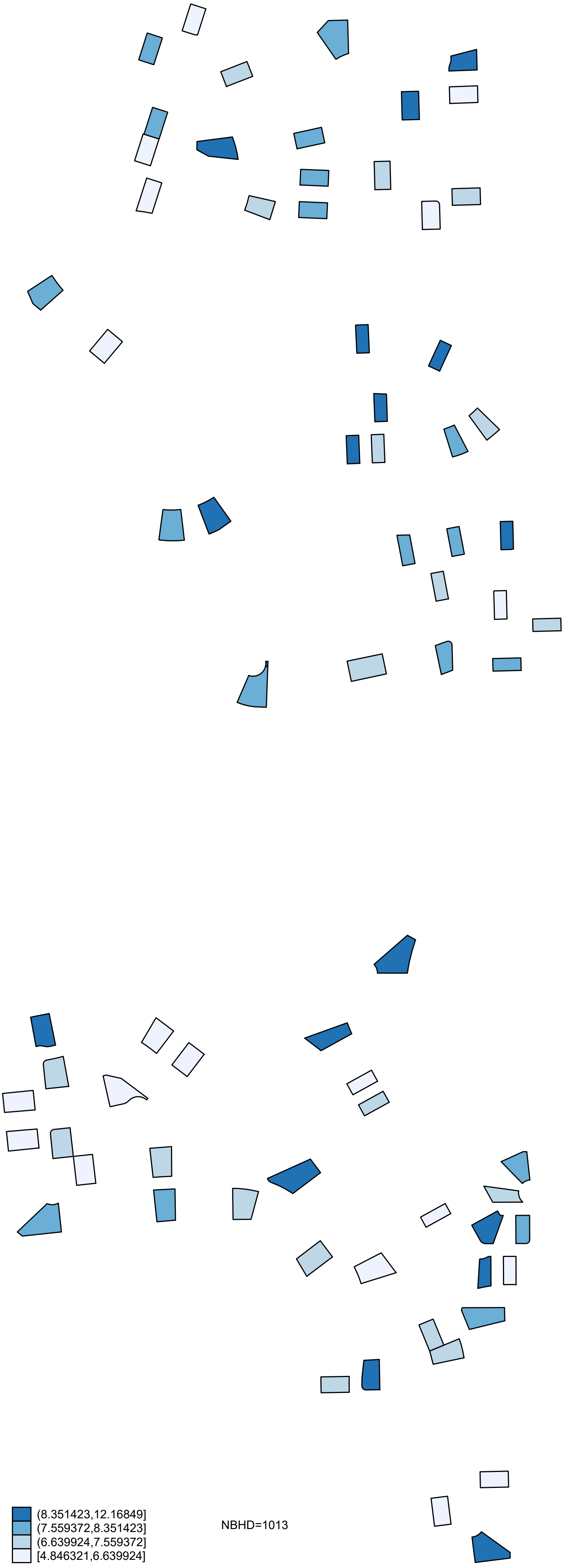
Summary Statistics

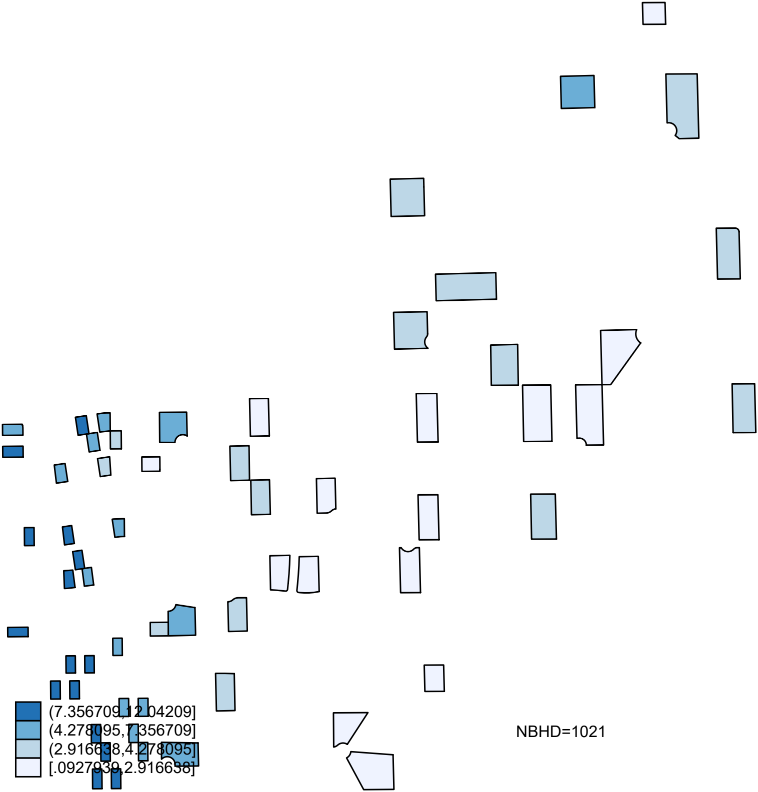
	mean	Sd	min	max
Sprice	288,146.7	160,277.6	29,500	1,310,526
Assess Land value	44,728.22	27,197.89	1	830,300
r_stories	1.210792	.4125602	1	4
Bathfix	8.000759	2.757856	2	45
r_imptotsqft	1,992.316	763.0406	252	9,966
r_iclass	3.360952	.5913903	0	6
Age	28.66132	20.06178	0	118
N	160,689			

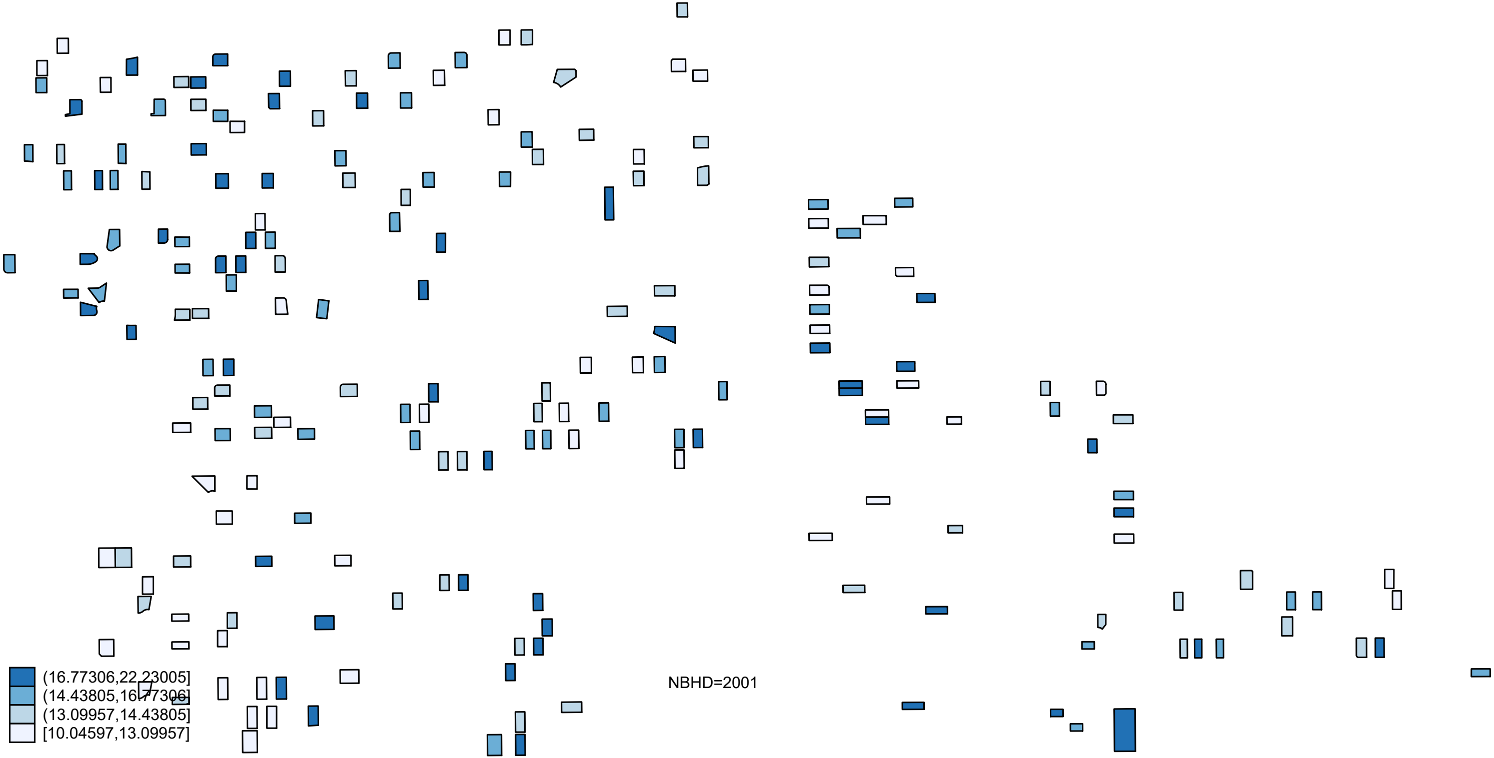
Land Value Per Square Foot of Land Area Results:

	Mean	StDev	Min	Max	N
Estimated	59.91424	13.27463	0.042866	6124.847	160,689
Assessor	5.63886	2.75078	0.000009	54.01522	160,689

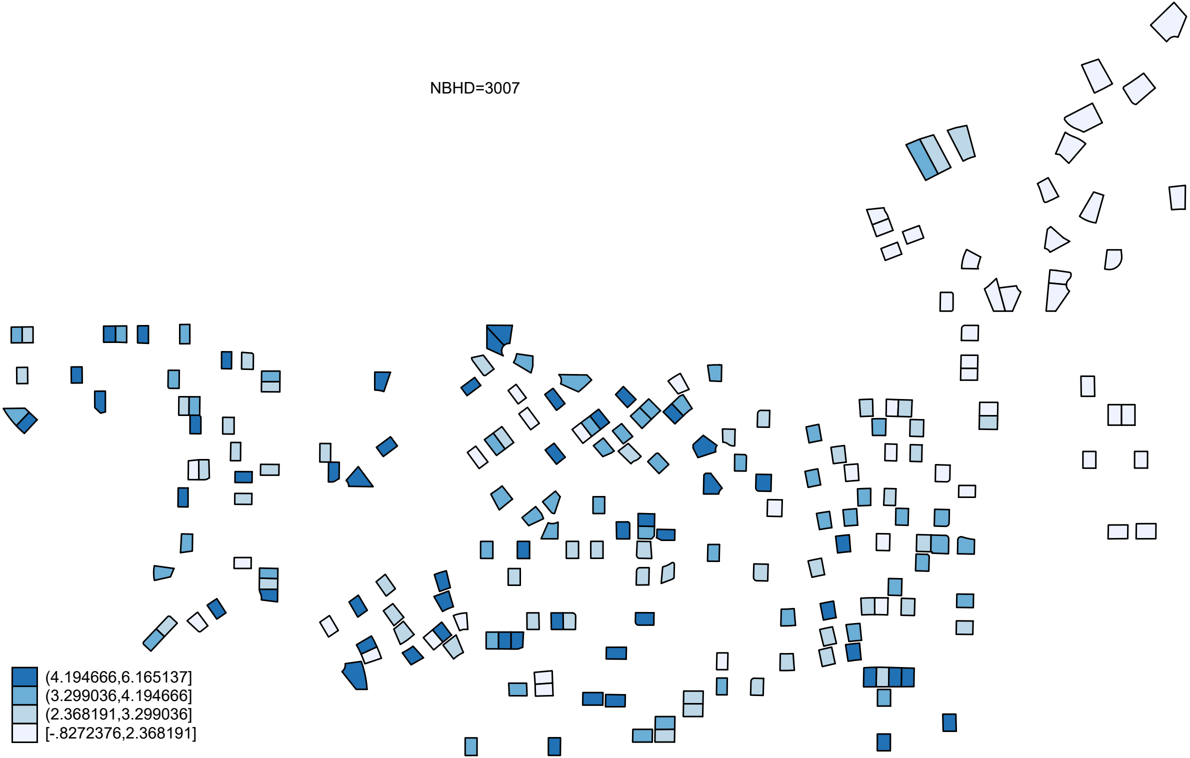
Some NBHD estimates look closer to the assessor's values than other NBHDs.

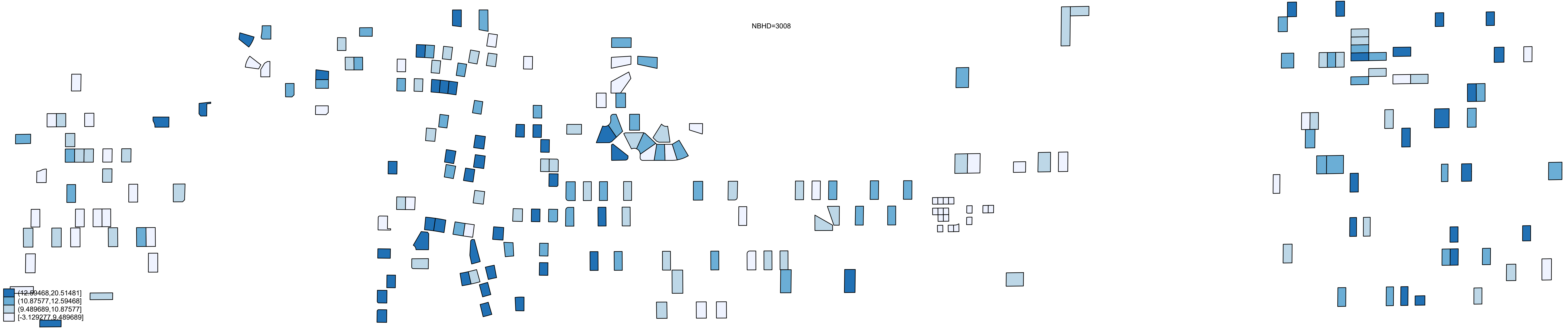

(8.351423,12.16849]
(7.559372,8.351423]
(6.639924,7.559372]
[4.846321,6.639924]
NBHD=1013

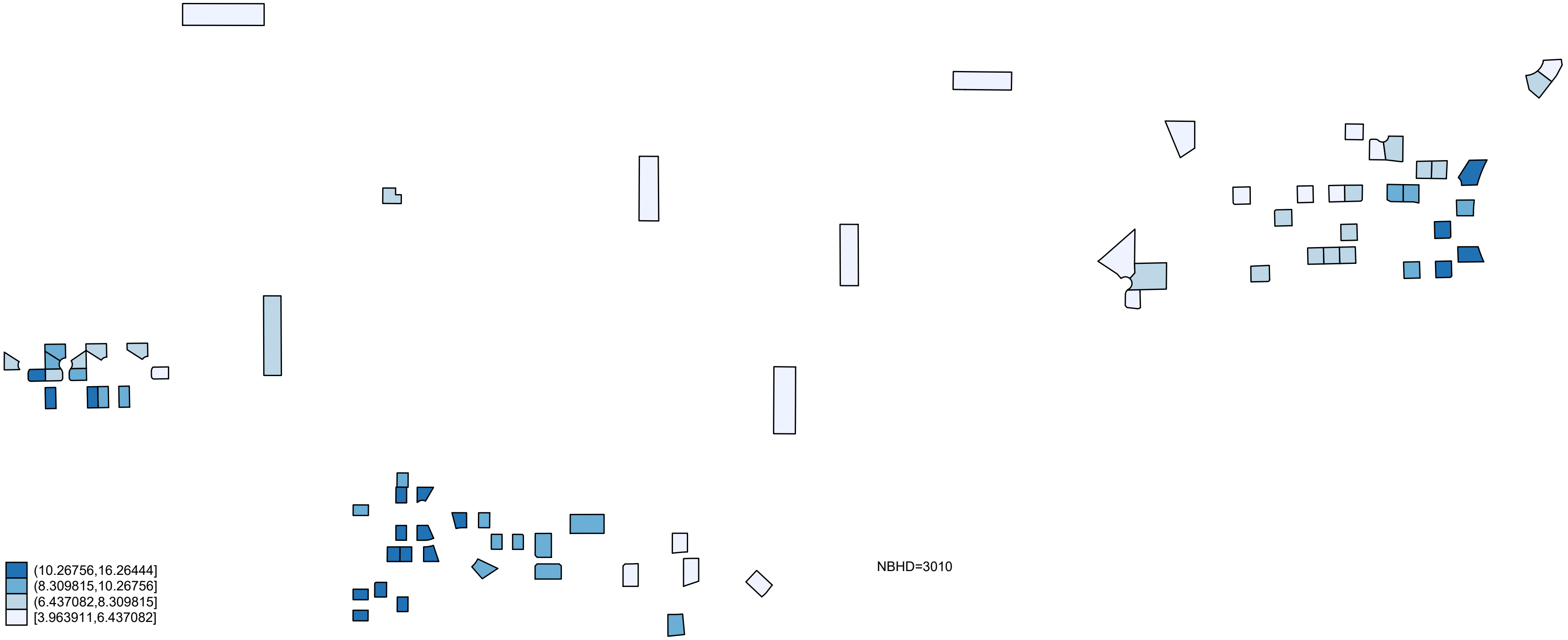

(7.356709,12.04209]
(4.278095,7.356709]
(2.916638,4.278095]
[.0927939,2.916638]
NBHD=1021

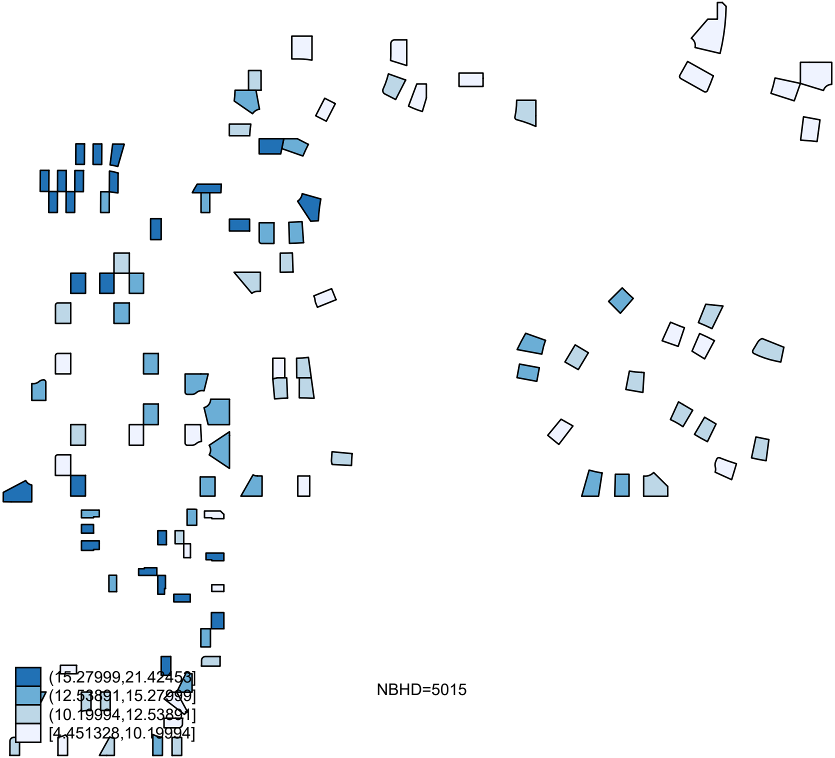


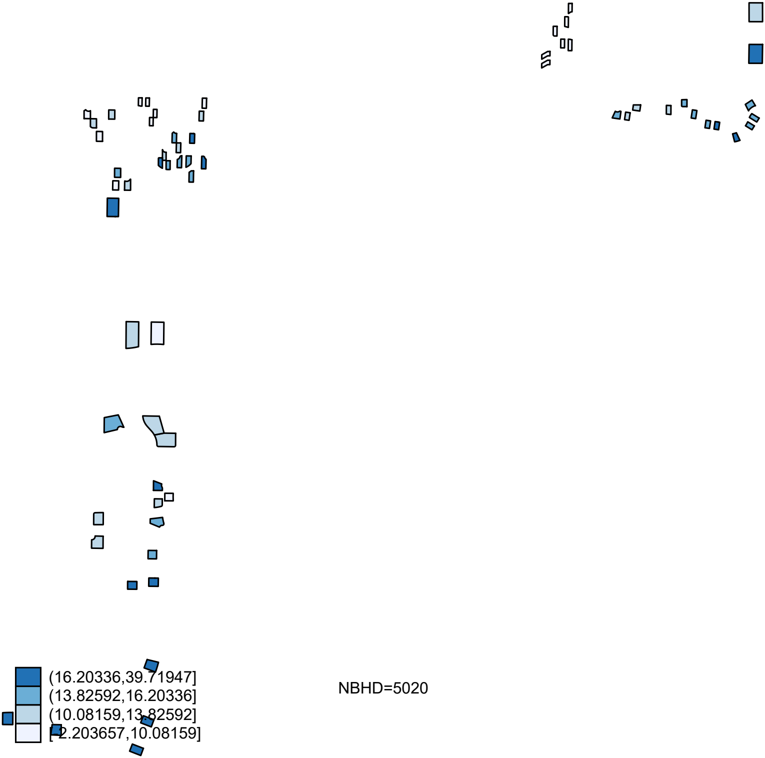
(16.77306,22.23005]
(14.43805,16.77306]
(13.09957,14.43805]
[10.04597,13.09957]
NBHD=2001

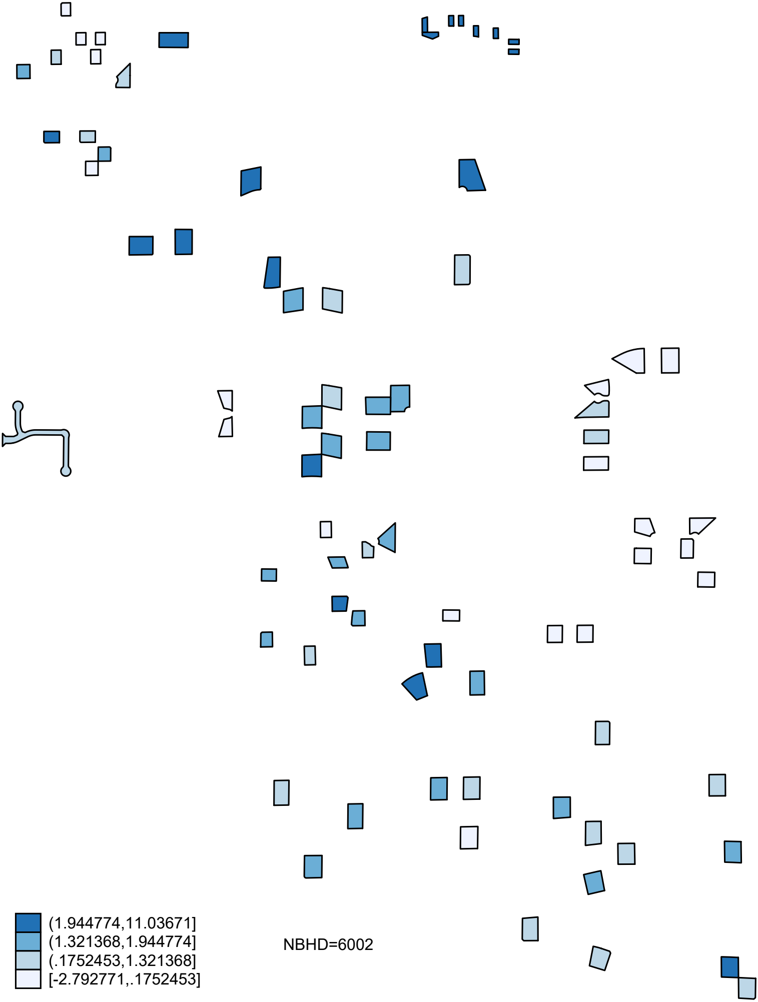

NBHD=3007
(4.194666,6.165137]
(3.299036,4.194666]
(2.368191,3.299036]
[-.8272376,2.368191]

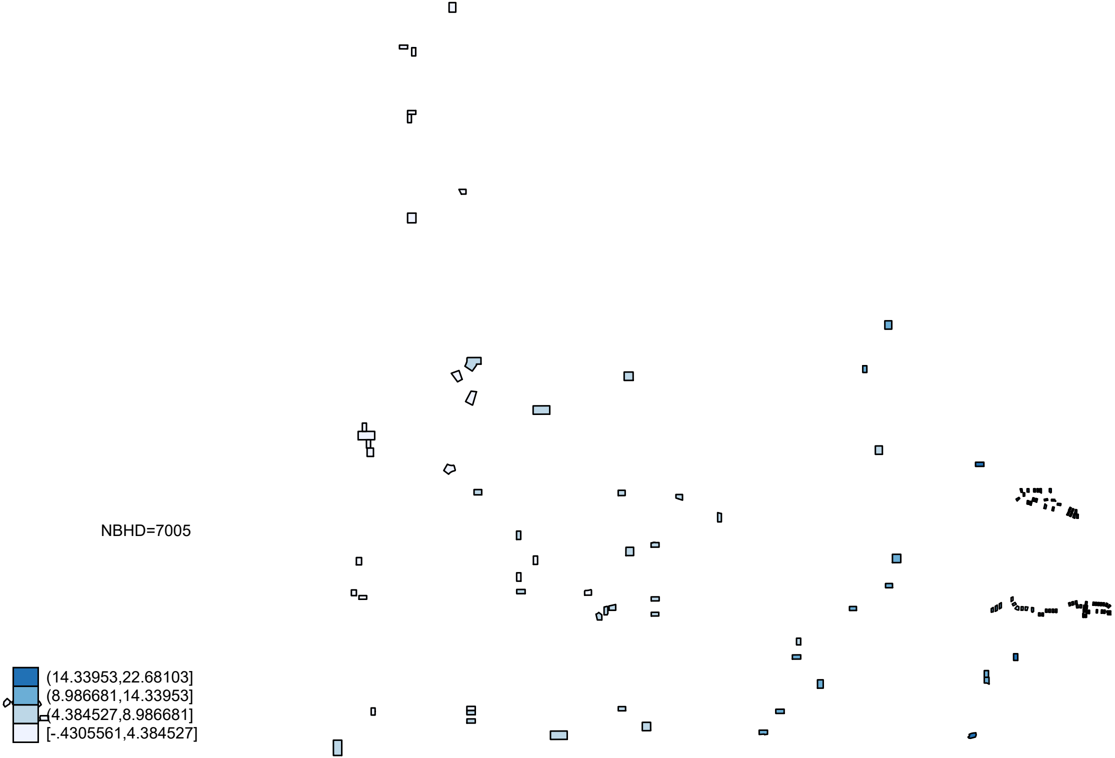

NBHD=3008
(12.59468,20.51481]
(10.87577,12.59468]
(9.489689,10.87577]
[-3.129277,9.489689]


(10.26756,16.26444]
(8.309815,10.26756]
(6.437082,8.309815]
[3.963911,6.437082]
NBHD=3010


(15.27999,21.42453]
(12.53891,15.27999]
(10.19994,12.53891]
[4.451328,10.19994]
NBHD=5015


(16.20336,39.71947]
(13.82592,16.20336]
(10.08159,13.82592]
[2.203657,10.08159]
NBHD=5020

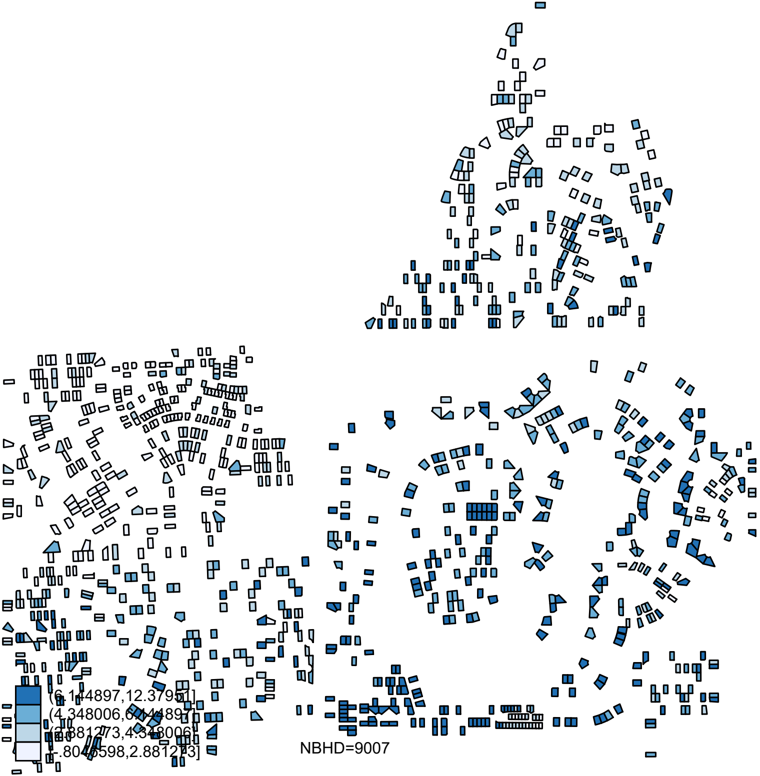

(1.944774,11.03671]
(1.321368,1.944774]
(.1752453,1.321368]
[-2.792771,.1752453]
NBHD=6002


NBHD=7005
(14.33953,22.68103]
(8.986681,14.33953]
(4.384527,8.986681]
[-.4305561,4.384527]

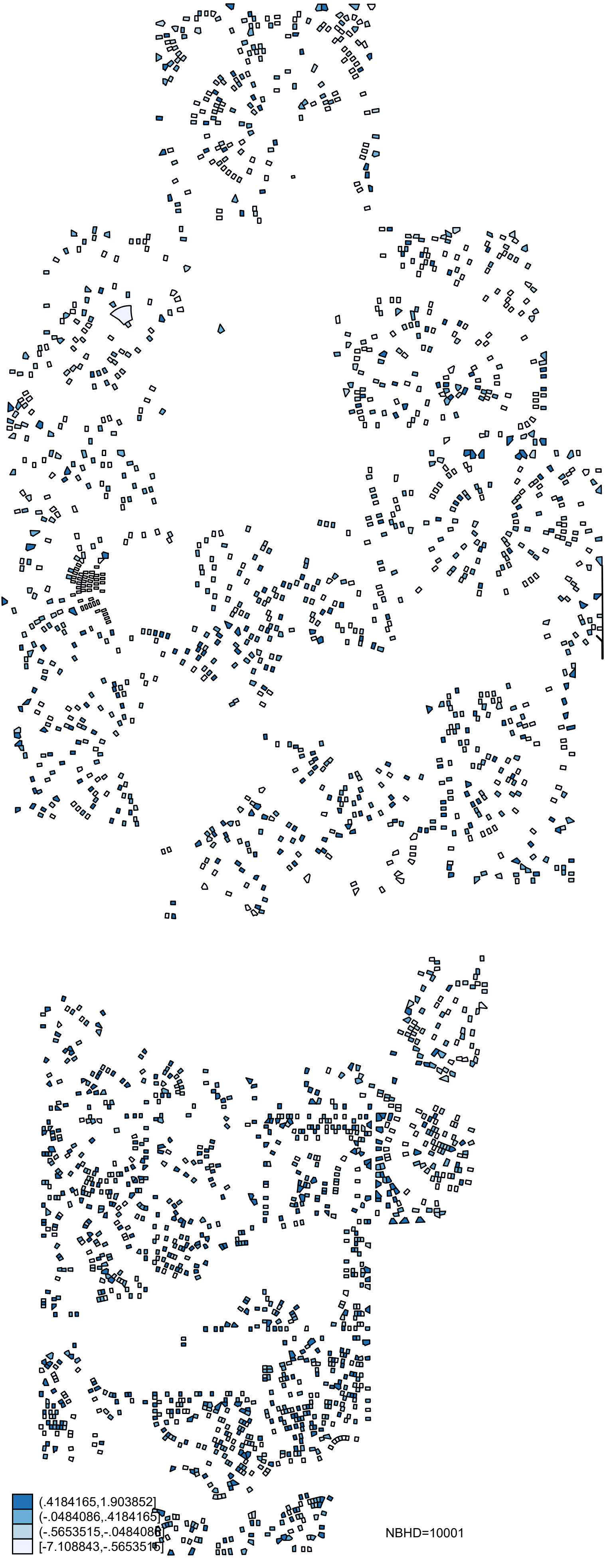

(17.95982,18.32381]
(9.539215,17.95982]
(8.047572,9.539215]
[.3746424,8.047572]

NBHD=8006

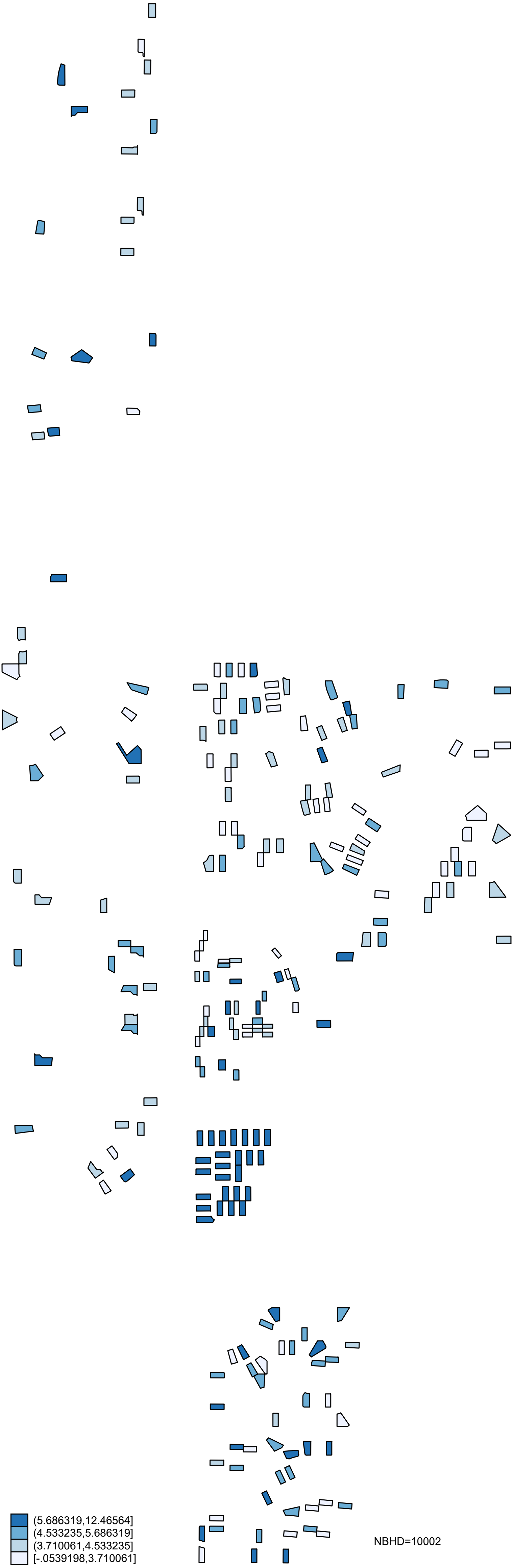

(13.7613,17.00467]
(11.1835,13.7613]
(7.028782,11.1835]
[1.607134,7.028782]
No data
NBHD=9006



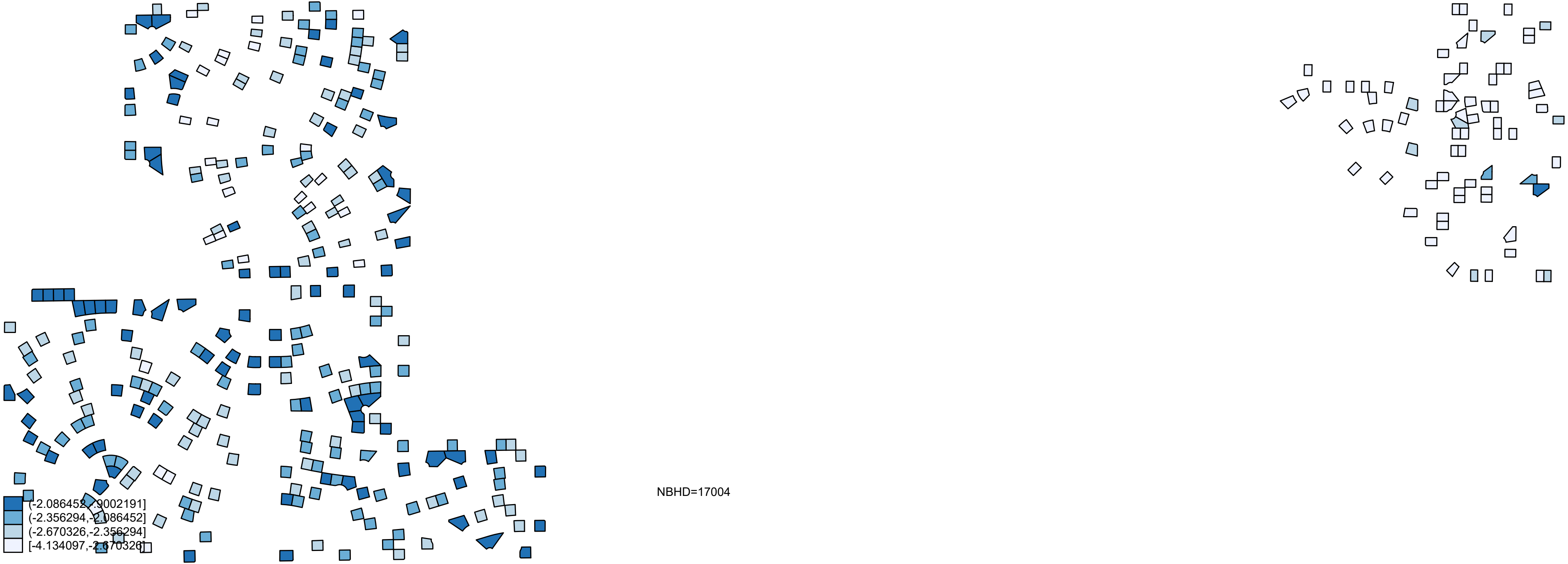
(6.144897,12.37951]
(4.348006,6.144897]
(2.881273,4.348006]
[-.8046598,2.881273]
NBHD=9007



(.4184165,1.903852]
(-.0484086,.4184165]
(-.5653515,-.0484086]
[-7.108843,-.5653515]
NBHD=10001


(5.686319,12.46564]
(4.533235,5.686319]
(3.710061,4.533235]
[-.0539198,3.710061]
NBHD=10002


(2.022412,4.075459]
(1.741514,2.022412]
(1.452409,1.741514]
[.492213,1.452409]
No data
NBHD=11007


(-2.086452,-.9002191]
(-2.356294,-2.086452]
(-2.670326,-2.356294]
[-4.134097,-2.670326]

NBHD=17004

Conclusions

-Previous studies using the Clapp (2004) and Cohen Coughlin Clapp (2017) methods mostly limited to development of land price indexes

-Here I incorporate time of sale as an additional dimension

-Adjust the models to obtain land value per square foot of land area

-My estimates likely high; assessors' estimates roughly uniform percent of entire property value. "True" values likely somewhere in between

-Potential causes for discrepancy:

> -Data on Bedrooms and number of bathrooms unavailable
>
> -Land may be worth more at the centroid of the property than at the fringe of the parcel – but my estimate of land value per sq foot of land area is at the lat/long/time point
>
> -Perhaps earlier literature using Local Polynomial Regressions is preferable to LWR?

-In some neighborhoods of Maricopa County, my approach may be a useful alternative to assessors' current approach

Potential Future Work:

-Consider another city with smaller parcels that also have data on number of bedrooms and bathrooms

-Why does this approach work well in some neighborhoods but yield greater variability in others?

-Compare with land leverage model of Davis and Palumbo (2008):

-Cohen, Coughlin and Clapp (2017): to do this comparison, consider $partres = ln(SP/LA) - Z\ \hat{\alpha}_R$

-Add back in $Z\ \hat{\alpha}_R$ and then subtract depreciated construction costs C_{it} , so that: $partres\_c = ln(SP/LA) - C_{it}$

-Then: $\text{Ln(Land value)} = E(partres\_c \mid S)$

-How do land values compare with these two approaches?