High-Frequency and High-Resolution Neighborhood House Price Dynamics:

Identifying Spatio-Temporal Hotspots and At-Risk Areas for London, England

JACOB L. MACDONALD

UNIVERSITY OF LIVERPOOL

GEOGRAPHIC DATA SCIENCE LAB

BERKELEY INSTITUTE FOR DATA SCIENCE

"PREDICTING NEIGHBOURHOOD CHANGE USING BIG DATA AND MACHINE LEARNING"

JANUARY 9, 2020

Motivation & Research Question

- Many economic and social urban processes occur in real time and over a continuum of space. Housing transactions can serve as a strong barometer for tracking such processes.

- Sparse data at the most granular of scales is limited in its representability and power in modeling direct neighborhood price levels and changes.

- High frequency and high resolution data can, however, be used to classify neighborhood dynamics and identify (cluster) like-areas in terms of their transaction dynamics.

- Using housing transactions aggregated to neighborhood units, this work defines similar clusters of (urban sub-market) areas in London based on their spatial and temporal price evolutions and spill-overs.

Study Region & Data

- Capital city of London (UK). The density of transactions in this area allow for granularly detailed analysis of neighborhood dynamics.

- Data from the HM Land Registry – tracking housing transactions from 1995 onwards.
  - Referenced by date of sale and postcode.
  - Includes typology classification; indicator for new sale; indicator for freeholds.

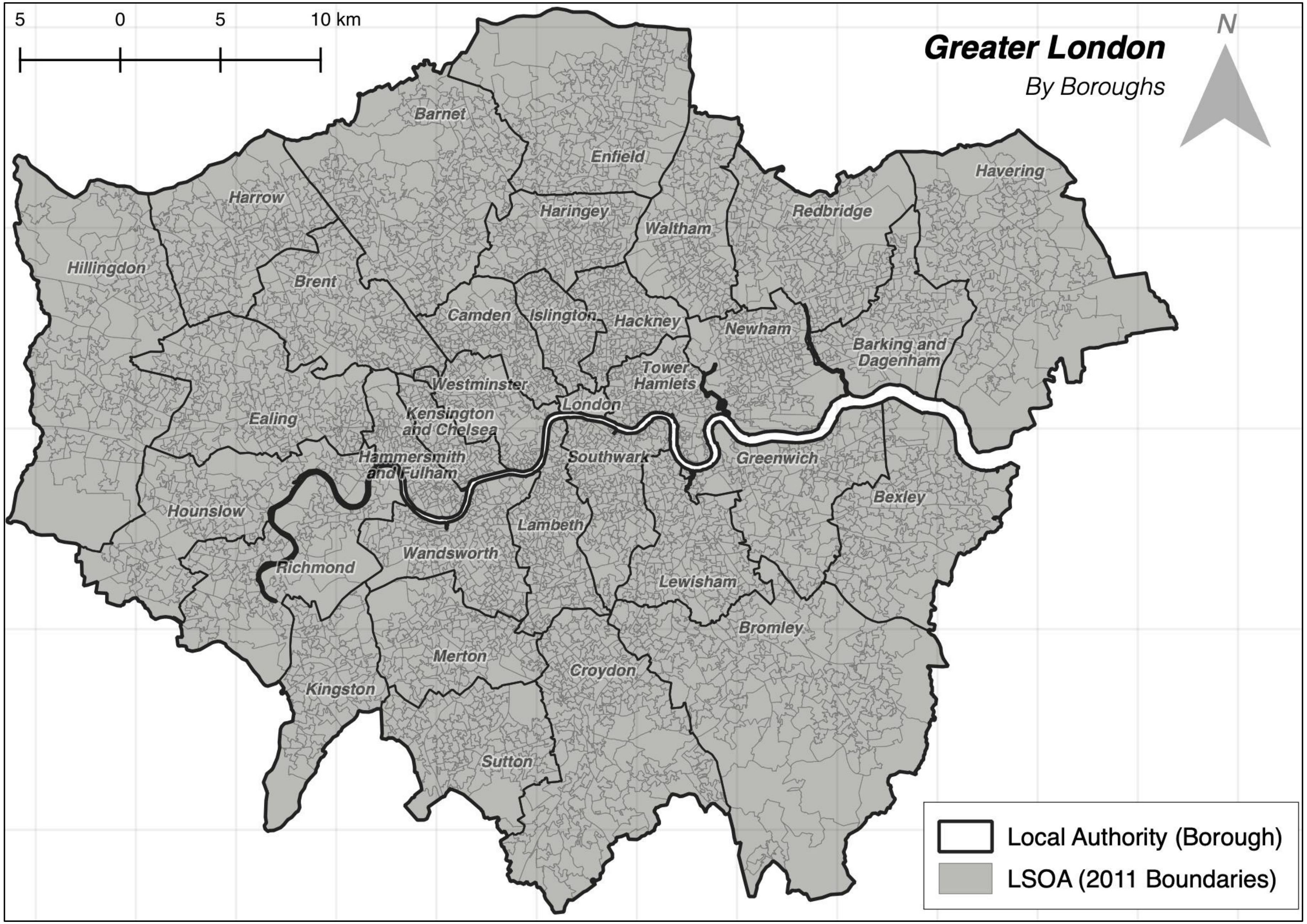
- Transactions geo-referenced to a hierarchy of spatial administrative units (LSOA and above) using the National Statistics Postcode Lookup.

Five-Year Bin	Count	St. Dev.	Unadjusted Prices (£)			Inflation Adjusted Prices (2019 £)*		
			Avg.	Med.	Max.	Avg.	Med.	Max.
1995 - 2000	873,974	168,572	137,013	96,500	32,477,000	626,285	457,823	145,231,253
2000 - 2005	944,539	234,142	239,559	188,000	24,750,000	585,997	453,720	73,307,741
2005 - 2010	730,401	362,947	341,316	250,000	23,500,000	602,781	450,773	38,089,583
2010 - 2015	650,720	1,377,321	514,409	332,000	225,000,000	698,375	455,377	235,071,620
2015 - 2019	489,091	3,361,655	730,451	435,000	594,300,000	741,756	441,893	571,762,655

** Adjusted for inflation using monthly London Consumer Price Index (CPI) for durable housing goods (ONS; base year = 2015)*

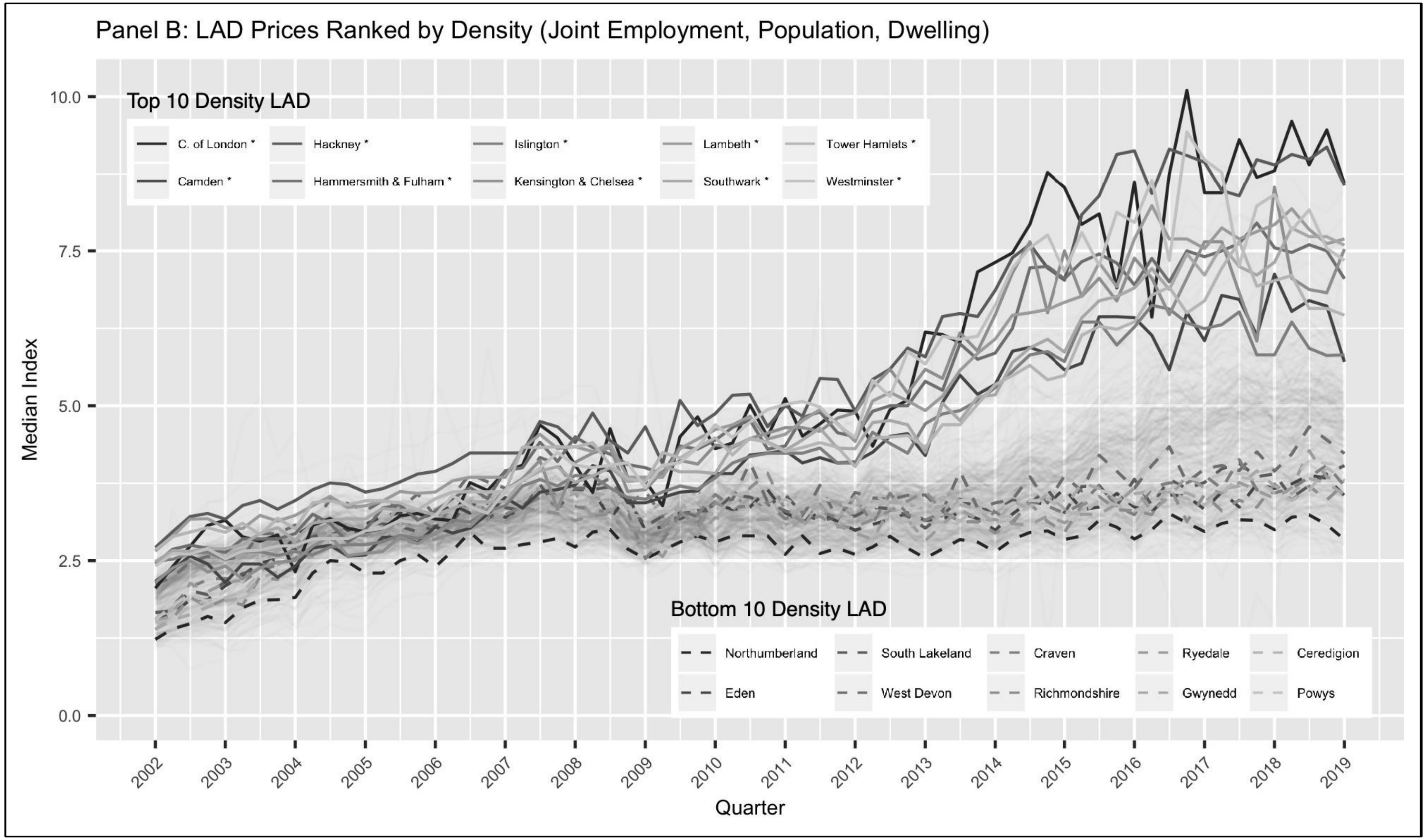
Study Region & Data

- Individual transactions are aggregated to neighborhood (LSOA) units using the median value of transactions in any given space-time bin.
- The base unit of analysis is LSOA neighborhoods (space) at monthly intervals (time) from 1995 to 2019.
- There is a trade-off that must occur between the number of transactions used to obtain the neighborhood price level, the size of the temporal bin, and the size of the spatial bin.
- While LSOA-months are appropriate for London, where transaction densities are the largest in the country, the sparsity of data cannot support this level of granularity outside the capital.



- N = 4,835 LSOA (Neighborhood) Units.
- Avg. size 0.33 km^2 with min. number of dwellings = 220 units ($4.60 \cdot 10^1$ per km^2).
- 33 Boroughs (Local Authorities).
- 3,175,800 transactions from 1995-2019.

London: Temporal Dynamics



London is a clear outlier in terms of transaction prices and growth.

High price growth in London is further correlated with areas of high income per capita and high densities of employment and population.

Even within the London boroughs however, there is large variation in prices and growth rates.

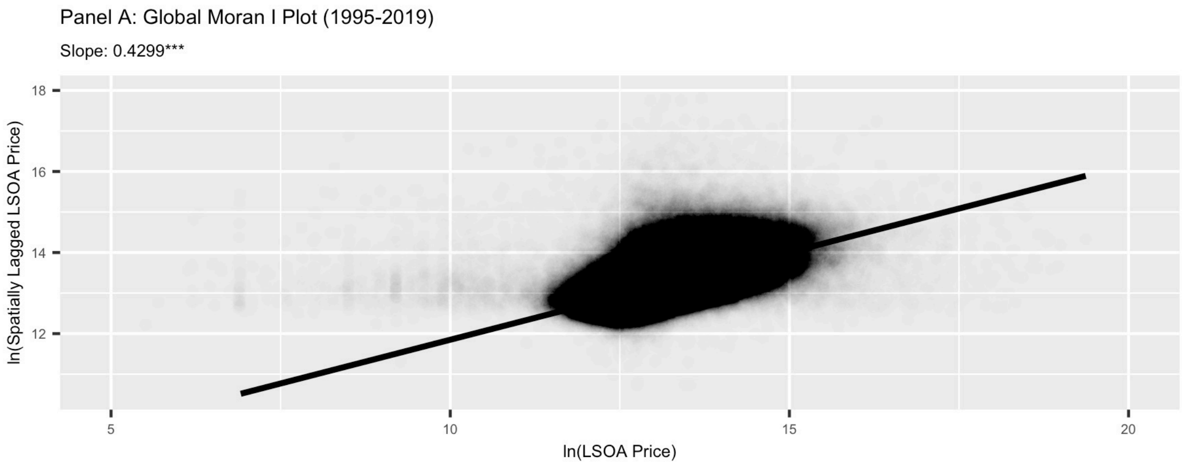
London: Spatial Dynamics

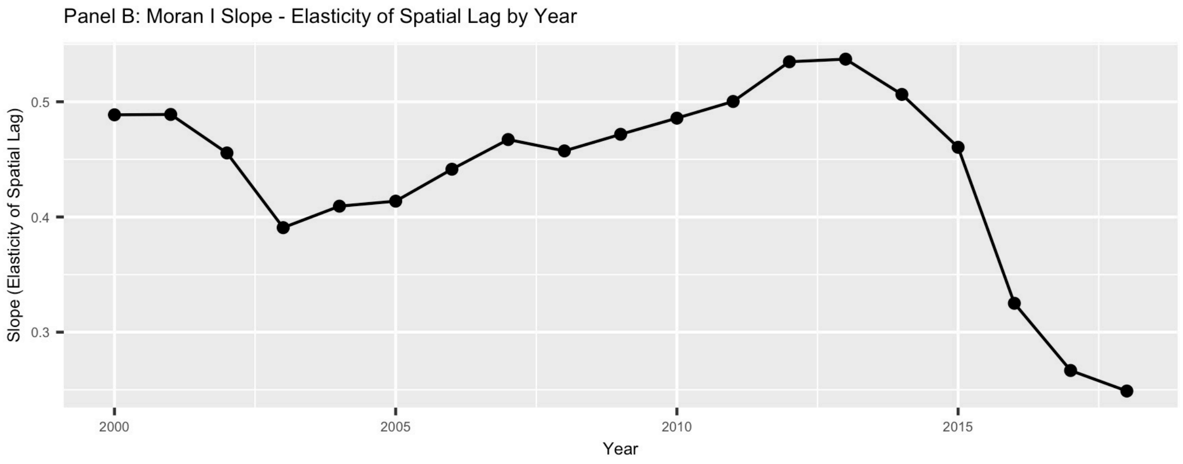
Neighborhoods have significant spatial spill-overs with high correlation between a unit's price and the spatial lag of contiguous tract prices.

Over time however, the strength of this spatial spill-over has been declining.

The decline in spatial correlation corresponds to the period of high growth in prices.

The growth across London is therefore not spatially equal in all areas – this high growth is driven by select high growth areas and isolated units.





Neighborhood Clustering Methodology

- Applying a (global) K-means cluster algorithm can characterize blocks of neighborhood units with relatively homogeneous prices, evolution, and spatial spill-overs.

- Unsupervised learning attributing a classification to areas of similar dynamics – maximizing homogeneity within each cluster (group) and heterogeneity between clusters.

- Clustering of neighborhoods is conditional on a set of input parameters which characterize each of the areas.

- Iterative sub-clustering breaks any potential large groupings down into sub-groups with marginally different characteristics.

Neighborhood Clustering Methodology

- Different sets of input parameters are used, each capturing the general idea on which we wish to cluster – broadly classified into three streams.

Direct Price	Spatial Spill-Over	Temporal Evolution
Captures the direct "level" of a neighborhoods price and transactions. Measured using direct (inflation adjusted) prices and transaction density.	Captures the strength and similarity between a neighborhood's price and those of the contiguous neighborhoods nearby. Measured as the (inflation adjusted) 2nd order queen contiguous neighbor price average.	Captures both the short and long-run dynamics of a neighborhoods price evolution. Short-run effect measured by the average growth in the previous 6 months. Long-run effect measured by the parameter estimate of month impact on price growth over previous 6 years. $P_t^i = \beta_T^i T_t + \beta_{T^2}^i T^2{}_t + \varepsilon_t^i \quad \forall\, i, j; \quad t \in [t_{-72}, t_0]$

Neighborhood Clustering Methodology

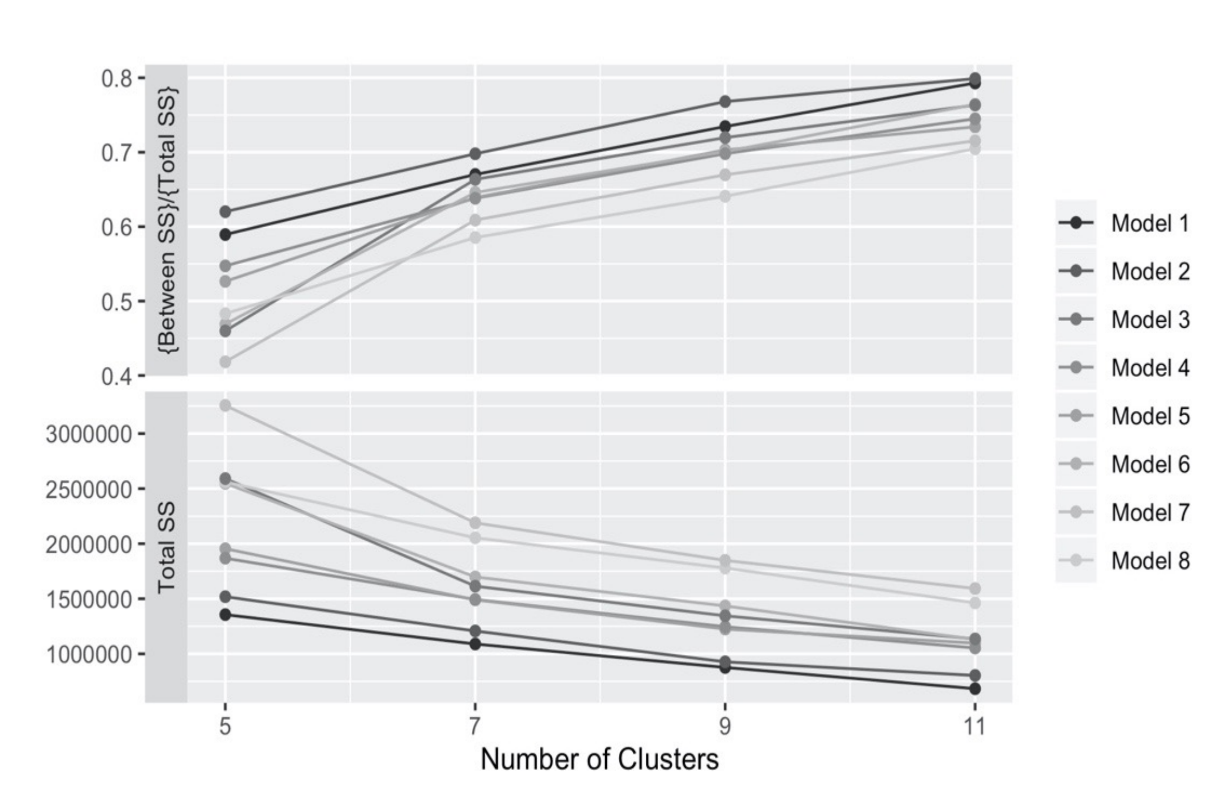
- The clustering is applied to a spatio-temporal longitudinal database of neighborhood (LSOA) units at monthly intervals.

- The clustering input variables (spatial spill-overs and temporal parameter estimates) are generated on a rolling basis for each LSOA at every month.

- Clustering on global (all years) and on inflation adjusted prices removes the significant temporal effect which would almost exclusively define the clusters groups if not.

Model & Cluster Selection

	Direct	Spatial	Temporal
Model 1	Prices (2019 Adj.)	2nd Order Queen Contiguity Spatial Lag (2019 Adj.)	β_T; β_{T^2}
Model 2	Prices (2019 Adj.)	2nd Order Queen Contiguity Spatial Lag (2019 Adj.)	Average Growth T-6 Months; β_T; β_{T^2}
Model 3	Prices (2019 Adj.)	2nd Order Queen Contiguity Spatial Lag (2019 Adj.)	Average Growth T-6 Months; β_T; β_{T^2}; Mean Square Error
Model 4	Prices (2019 Adj.)	2nd Order Queen Contiguity Spatial Lag (2019 Adj.)	β_T; β_{T^2}; Mean Square Error
Model 5	Prices (2019 Adj.); Transactions/ km^2	2nd Order Queen Contiguity Spatial Lag (2019 Adj.)	β_T; β_{T^2}
Model 6	Prices (2019 Adj.); Transactions/ km^2	2nd Order Queen Contiguity Spatial Lag (2019 Adj.)	Average Growth T-6 Months; β_T; β_{T^2}
Model 7	Prices (2019 Adj.); Transactions/ km^2	2nd Order Queen Contiguity Spatial Lag (2019 Adj.)	Average Growth T-6 Months; β_T; β_{T^2}; Mean Square Error
Model 8	Prices (2019 Adj.); Transactions/ km^2	2nd Order Queen Contiguity Spatial Lag (2019 Adj.)	β_T; β_{T^2}; Mean Square Error

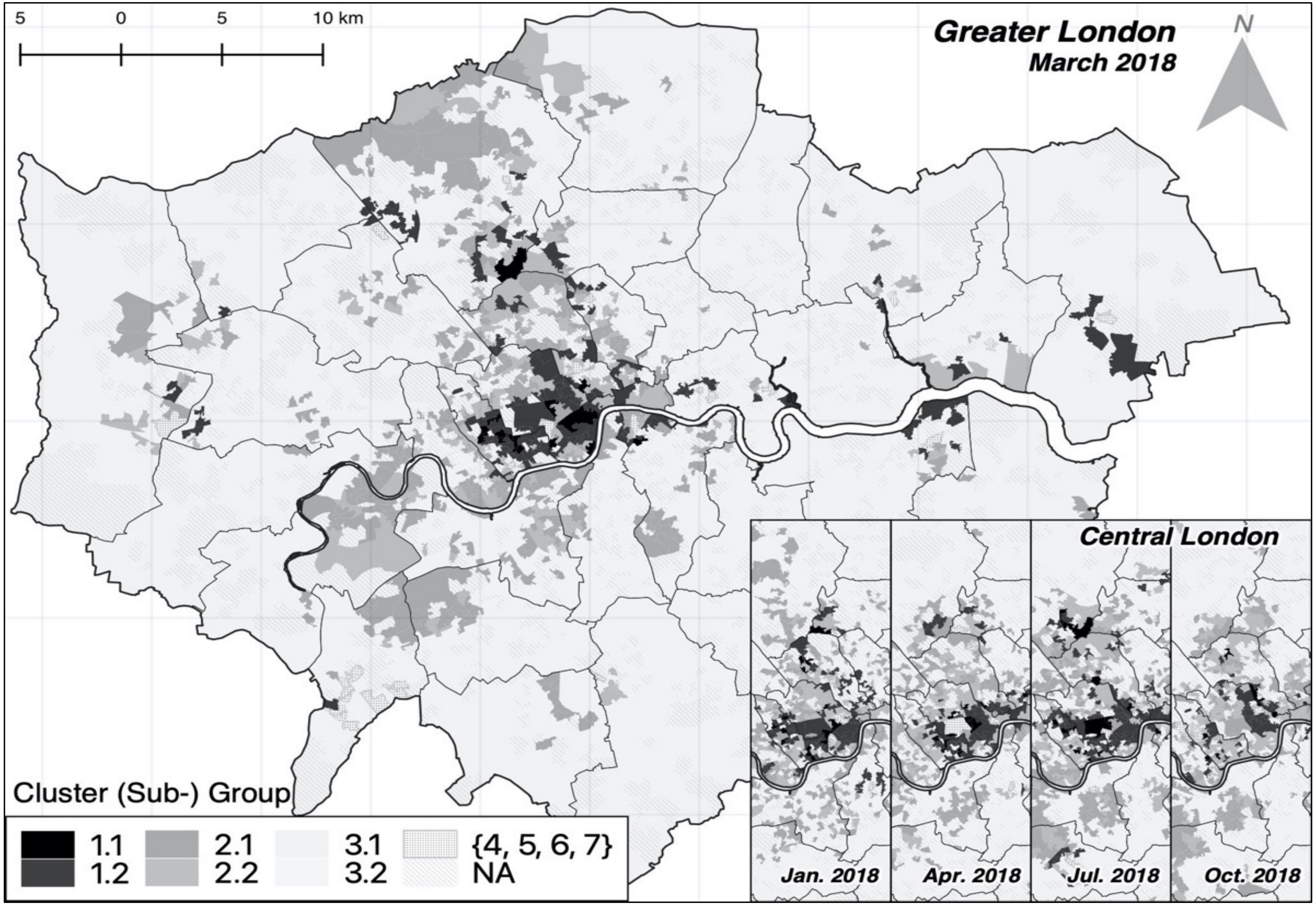


- Preferred model is the most parsimonious, **Model 1**, with 7 cluster groups.
  - Lowest within-group SS and highest between-group explanatory power

Cluster Group Dynamics

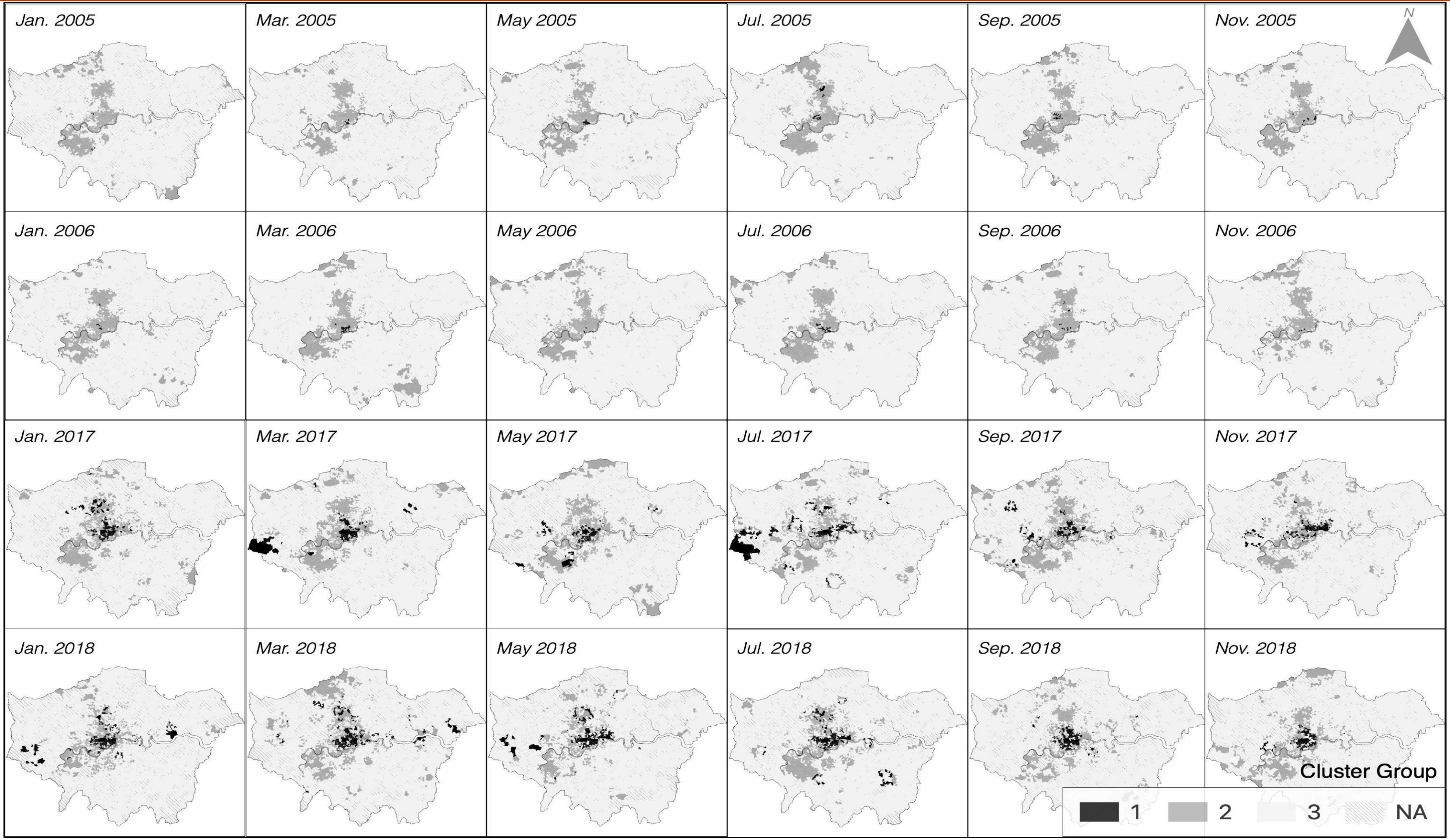
Cluster Group		Global N	Prices (2019 Adj.)	Spatial Lag (2019 Adj.)	β_T	β_{T^2}
1	1.1	2,498	5,275,559	2,167,247	3,576	182
	1.2	10,532	1,268,897	2,896,450	19,269	-101
2	2.1	87,979	1,099,612	901,229	615	44.6
	2.2	34,477	1,034,324	1,426,841	9,005	-56.5
3	3.1	430,816	366,849	402,687	1,015	5.3
	3.2	258,244	611,007	598,915	2,526	-3.7
4		607	17,528,299	1,451,392	-210,182	3,040
5		367	1,004,352	14,157,599	16,029	-70.2
6		43	94,286,123	1,358,529	-511,230	9,153
7		10	1,173,276	725,891	-12,956,880	92,070

- Cluster groups 1, 2 and 3 are each sub-classified into two sub-clusters each.
- These cluster groups have a sufficiently large grouping able to support the sub-classification into marginally different within-groups.


5
0
5
10 km
Greater London
March 2018
N
Central London
Jan. 2018
Apr. 2018
Jul. 2018
Oct. 2018
Cluster (Sub-) Group
1.1
1.2
2.1
2.2
3.1
3.2
{4, 5, 6, 7}
NA

Cluster Group Classification

- **Cluster 1**: Ultra-high-priced neighborhoods concentrated together in the city center. High prices, high growth and high-priced neighbors.
- **Cluster 2**: Moderately priced neighborhoods peripheral to the core city center. The million-pound dwellings located along the Thames.
- **Cluster 3**: Ambient level price neighborhoods with moderate growth. Sub-grouped by the intensity of prices and growth.
- **Cluster 4, 6, 7**: Price outliers and declining neighborhoods.
- **Cluster 5**: Isolated neighborhood surrounding by ultra-high-priced areas.


Jan. 2005
Mar. 2005
May 2005
Jul. 2005
Sep. 2005
Nov. 2005
N
Jan. 2006
Mar. 2006
May 2006
Jul. 2006
Sep. 2006
Nov. 2006
Jan. 2017
Mar. 2017
May 2017
Jul. 2017
Sep. 2017
Nov. 2017
Jan. 2018
Mar. 2018
May 2018
Jul. 2018
Sep. 2018
Nov. 2018
Cluster Group
1
2
3
NA

Discussions

- The emergence of ultra-high-priced and high growth neighborhoods in the capital have only begun over the past decade.
- The rolling nature of clustering – at every month – captures the ebbs and flows of neighborhood dynamics as they evolve and change over time.
- This high-frequency neighborhood classification captures the dynamic nature of sub-urban housing markets.
- Through these changes we can track which neighborhoods have consistent categorization, or which neighborhoods persistently remain the same.