Inferring land use from mobile phone activity

Jameson L. Toole (MIT)
Michael Ulm (AIT)
Dietmar Bauer (AIT)
Marta C. Gonzalez (MIT)

UrbComp 2012
Beijing, China

THE BIG QUESTIONS

Can land use be predicted from mobile phone activity?

Can mobile phone data help measure temporal patterns of land use?

Are these patterns robust across cities?

SENSING URBAN SYSTEMS

Data

~~Scarcity~~

Abundance

Methods → Outputs

Traditional Data:

- Zoning Maps

Novel Data:

- Mobile phone activity

Statistical Physics

Machine learning

Geographic Boundaries (Ratti et al 2010)

Intra-city Mobility (Cho et al 2011)

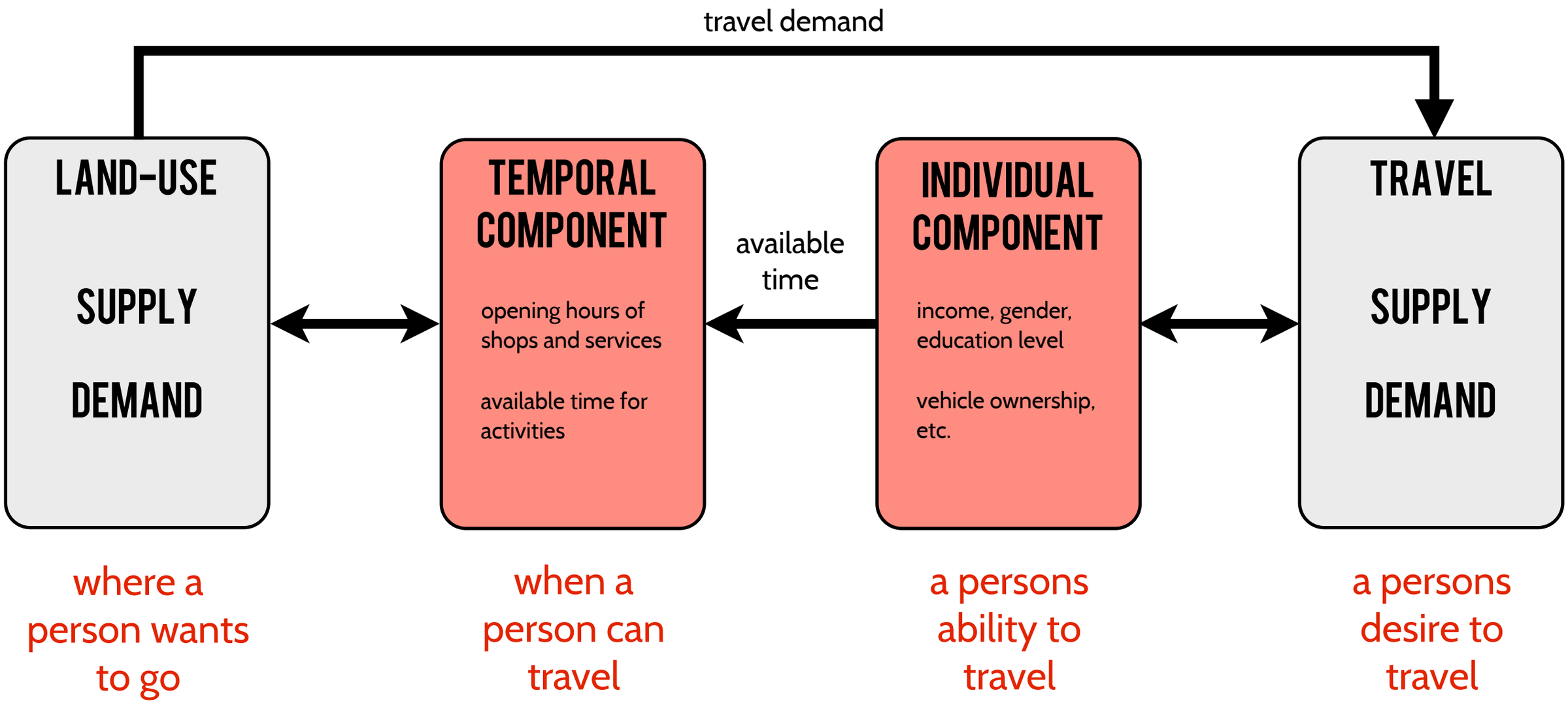
Inter-city Mobility

(Gonzalez et al 2008) (Wang et al 2009) (Simini 2012)

Unsupervised Classification

(Reades 2009) (Calabrese 2010) (Soto 2011)(Zheng 2012)

LAND USE AND TRAVEL BEHAVIOR



Static GIS data + New micro level data from mobile phones = Dynamic Land Use

adapted from: *K.T. Geurs, B. van Wee / Journal of Transport Geography 12 (2004) 127–140*

DATA

Zoning Data

- GIS shapefiles at the parcel level
- Provided by MASSDOT
- Aggregated to 5 zoning classification: Residential, Commercial, Industrial, Parks, Other

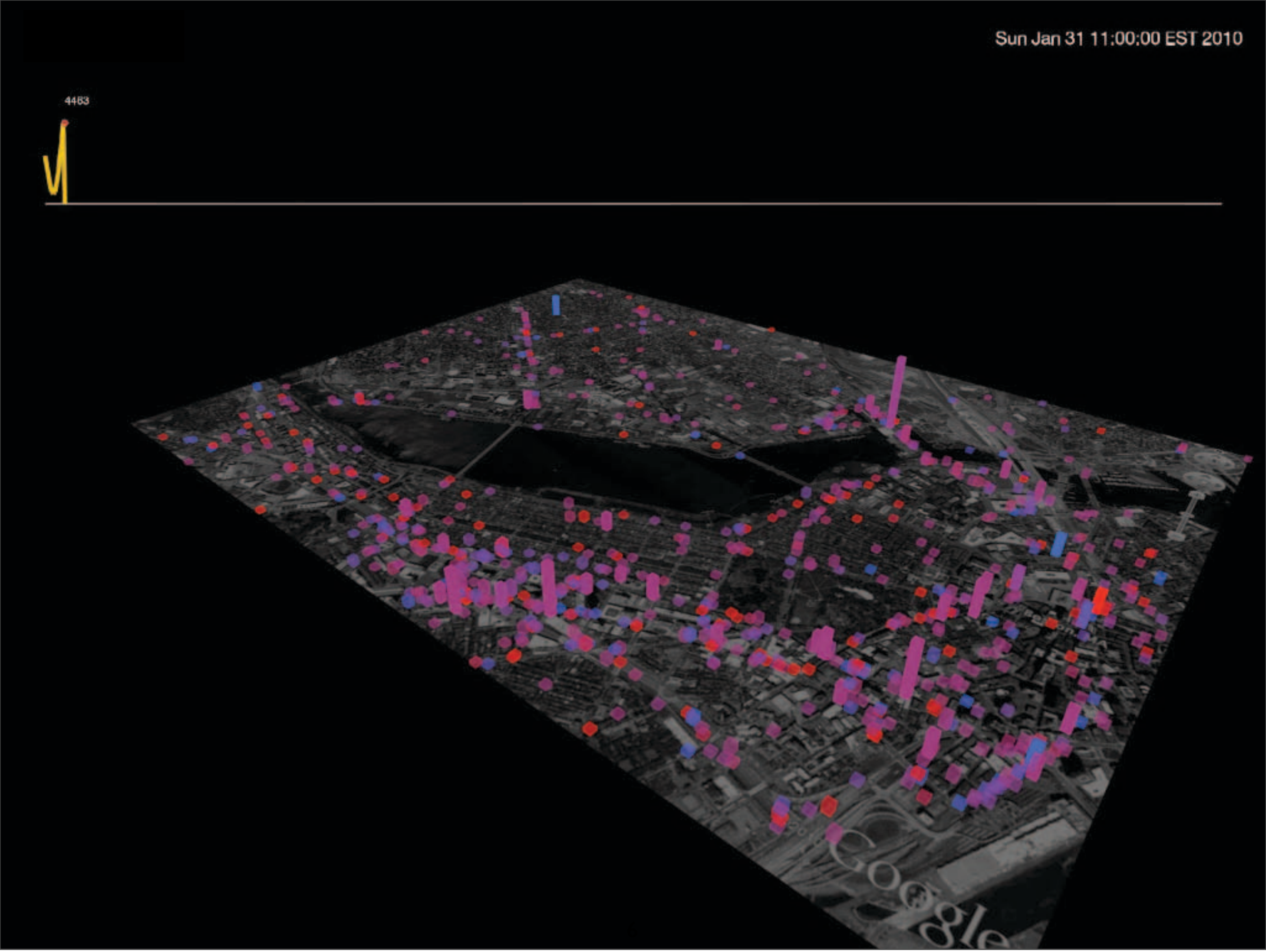
Mobile Phone

- ~600,000 Users in the Boston Metropolitan Area (Pop: ~3 million, Mkt Share: 30%-50%)
- 1 month of Call Detail Records (CDR) for voice calls and text messages.
- Each event geo-tagged with (Lat, Lon) triangulated from nearby towers

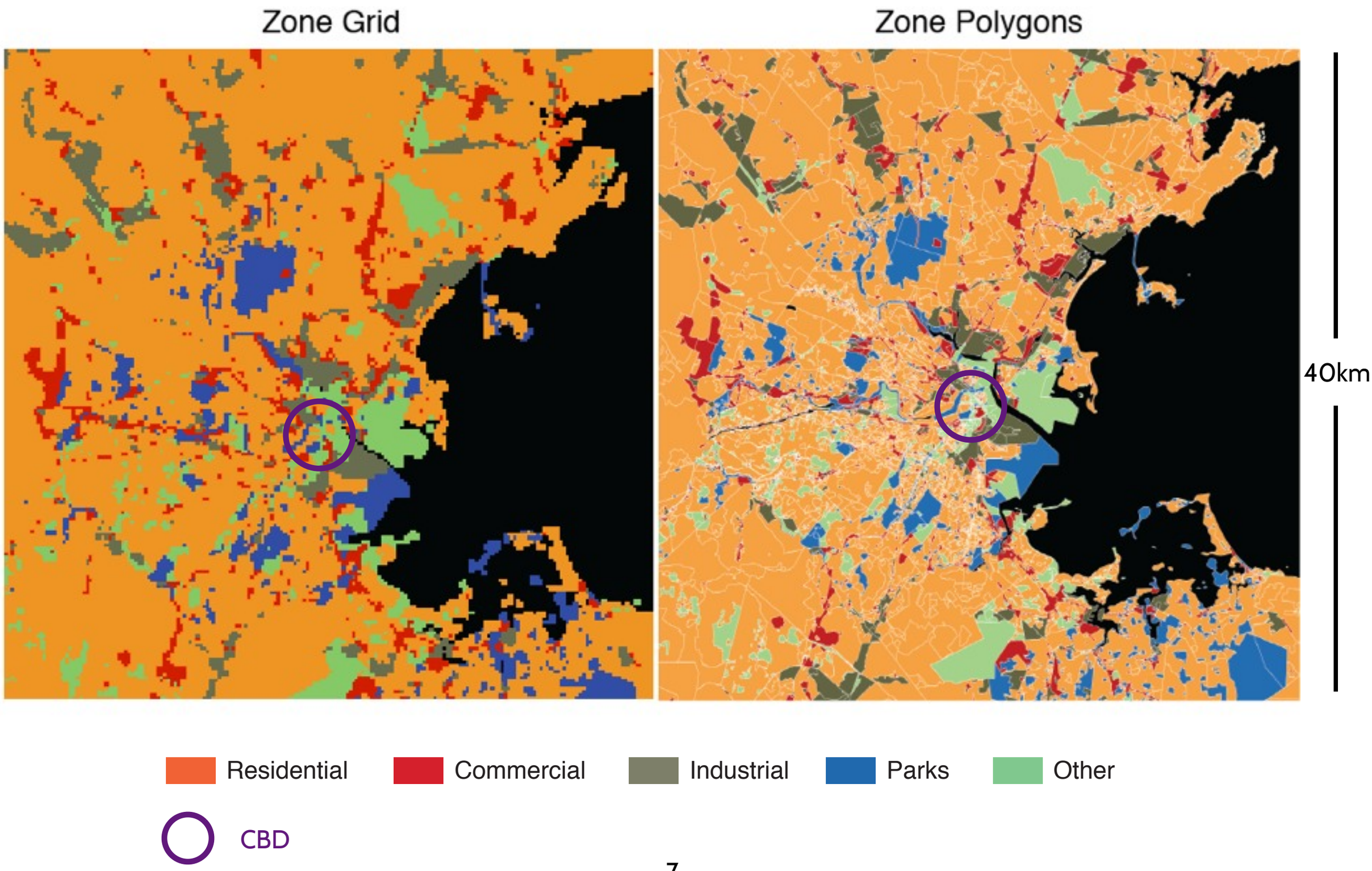
Grid

- Partition the region into 200m X 200m cells
- Each cell is labeled with the most prevalent zoning type within
- All mobile phone events are aggregated hourly into an average week

Grid provides privacy


Sun Jan 31 11:00:00 EST 2010
4463
Google

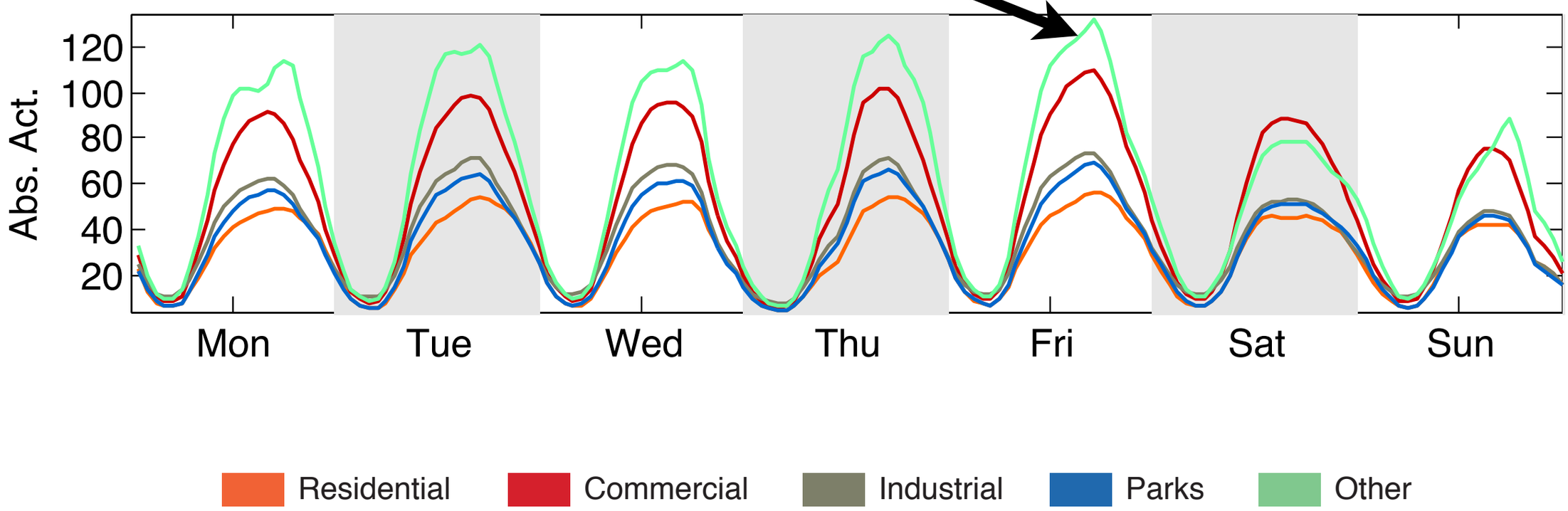
ZONING DATA


Zone Grid
Zone Polygons
40km
Residential
Commercial
Industrial
Parks
Other
CBD

ZONED USE & DYNAMIC POPULATION

Results: Absolute Activity

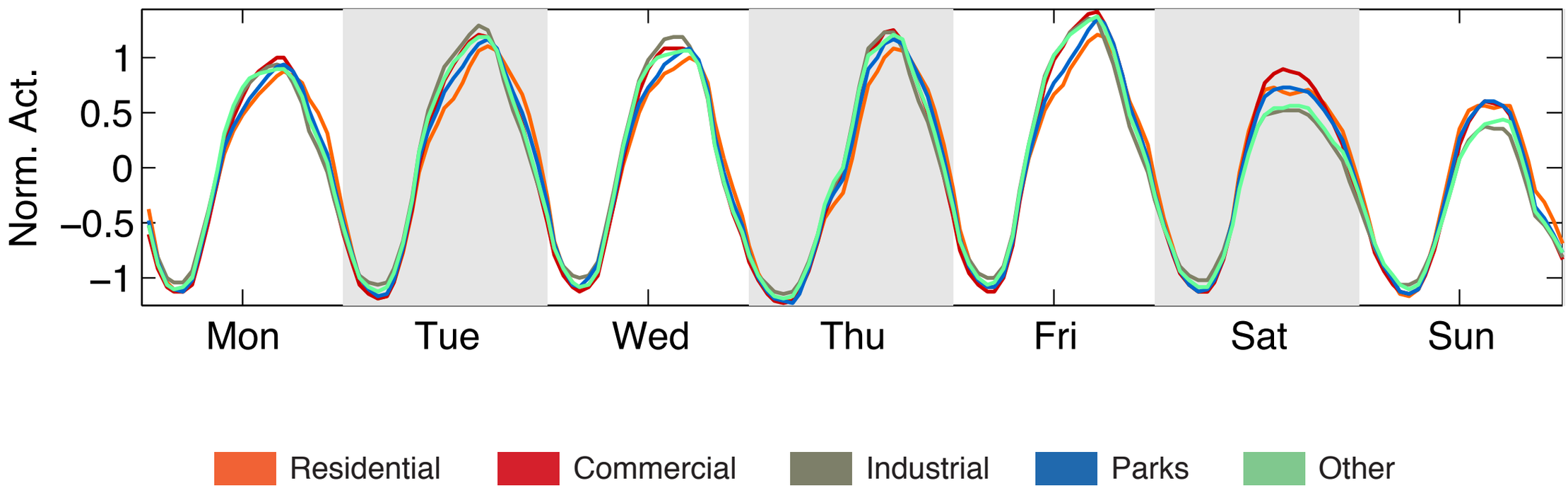
Downtown Boston is mostly zoned as OTHER thus more activity is related to more people.



ZONED USE & DYNAMIC POPULATION

Results: Normalized Activity

$$a_{ij}^{norm}(t) = \frac{a_{ij}^{abs}(t) - \mu_{a_{ij}^{abs}}}{\sigma_{a_{ij}^{abs}}}$$

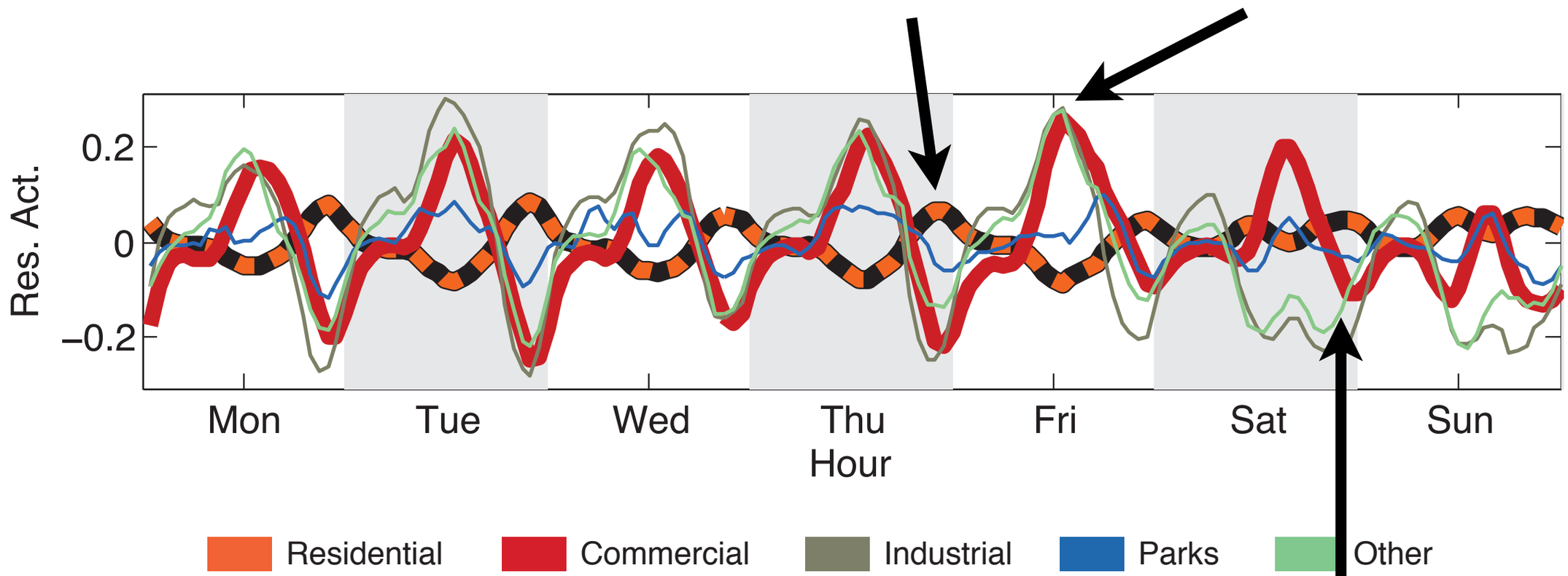


Similarities are related to how individuals use mobile phones, not how they move.

ZONED USE & DYNAMIC POPULATION

Results: Residual Activity $$a_{ij}^{res}(t) = a_{ij}^{norm}(t) - \bar{a}^{norm}(t)$$

Inverse relationship between RESIDENTIAL and COMMERCIAL



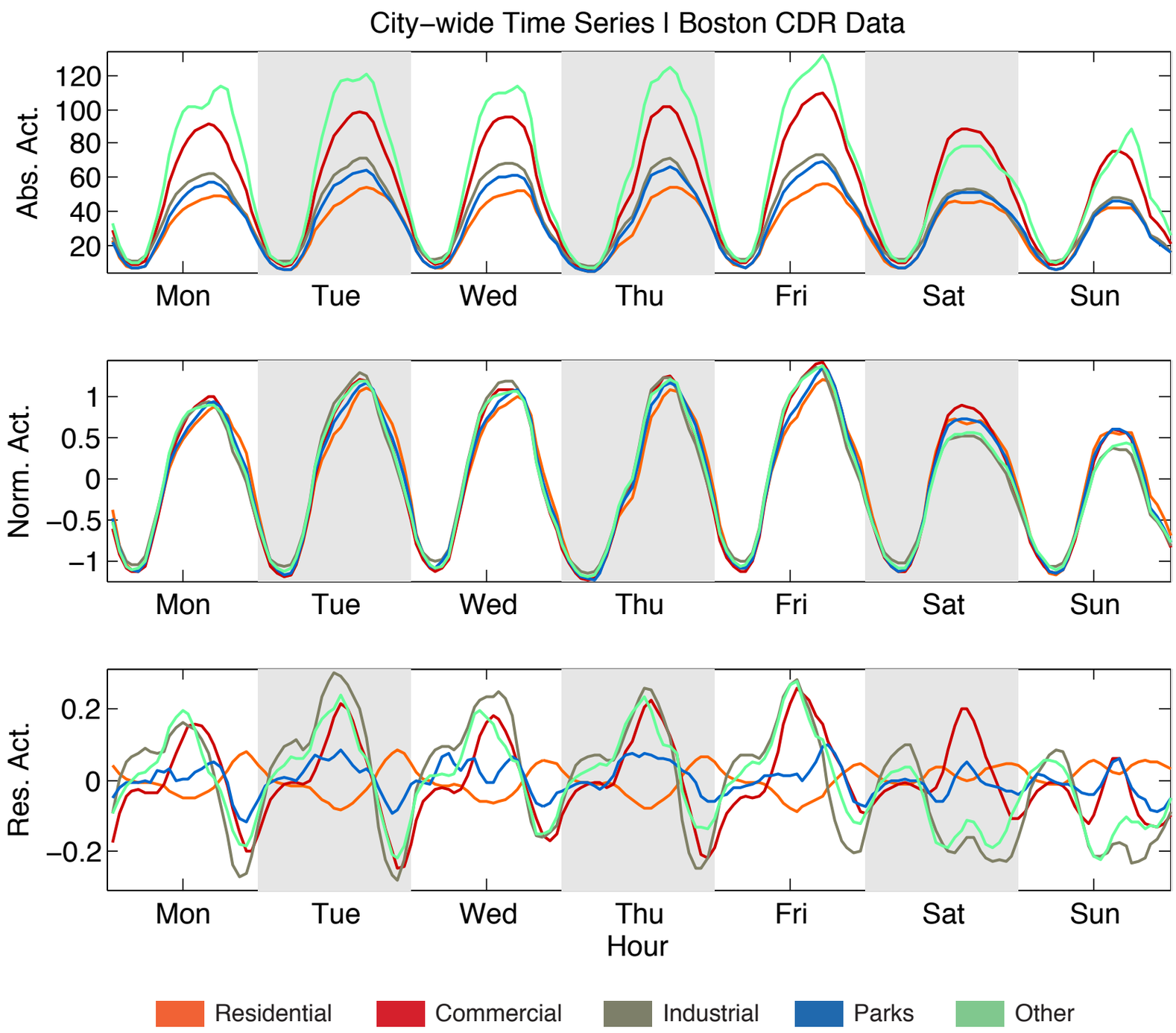
Weekends are quantitatively different than weekdays.

ZONED USE & DYNAMIC POPULATION

static population

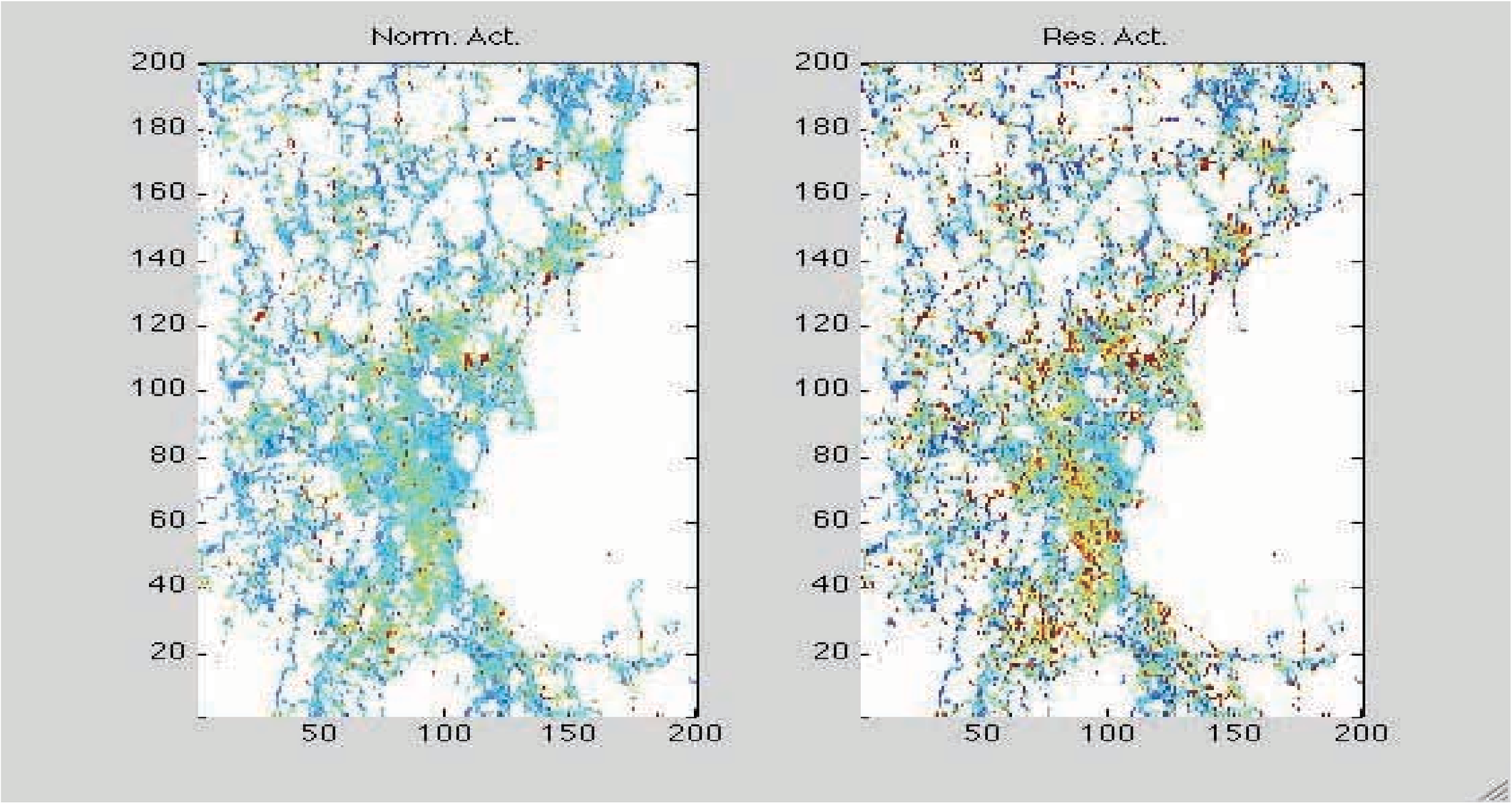
phone use

movement



ABSOLUTE ACTIVITY

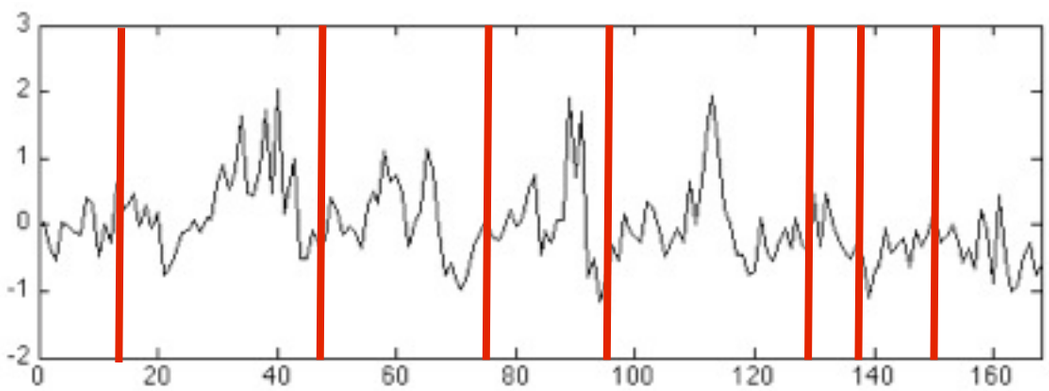
Results:

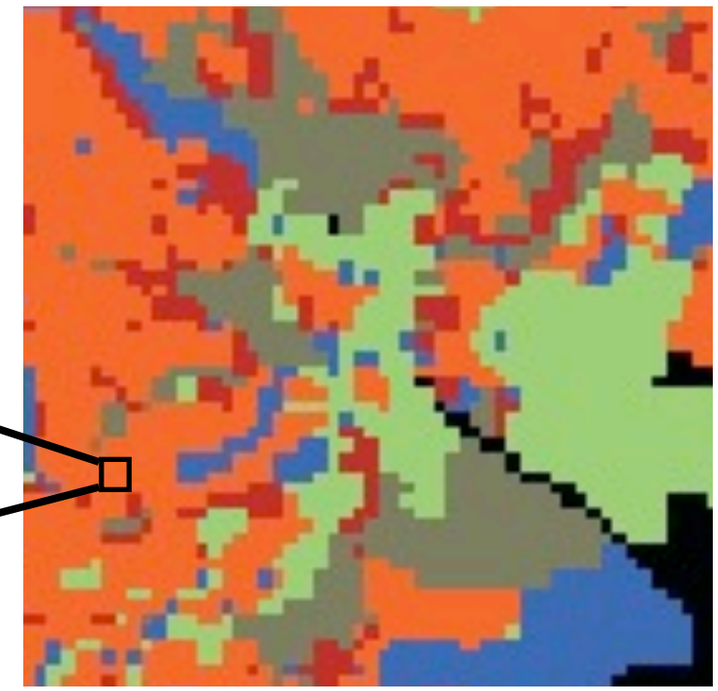


Low Activity  High Activity

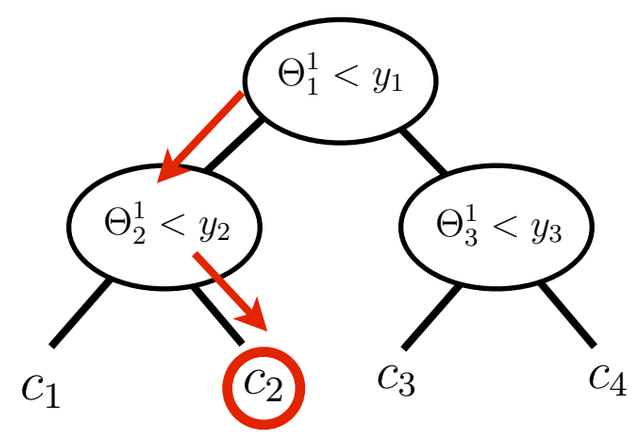
INFERRING LAND USE

Methods - Supervised Learning: Random Decision Tree





$$h(\mathbf{x}, \Theta_k) = \mathbf{c}$$

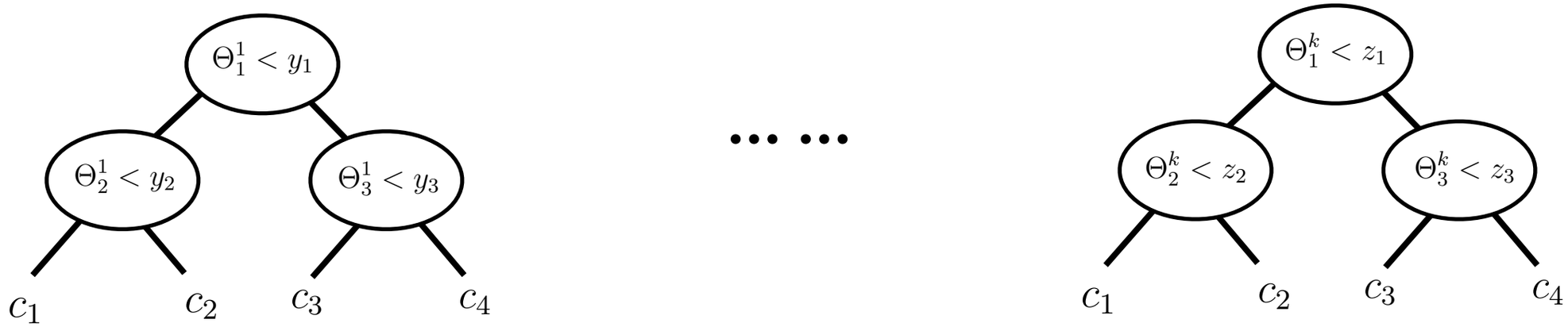


INFERRING LAND USE

Methods - Supervised Learning: Random Forest

Random Forest: (Breiman 2001) $\{h(\mathbf{x}, \Theta_k),\ k = 1, \ldots\}$

$\mathbf{x}$ is a $T \times 1$ vector
Θ_k is a $m \times 1$ vector, $(m < T)$



INFERRING LAND USE

Methods - Supervised Learning: Random Forest

Random Forest: (Breiman 2001) $\{h(\mathbf{x}, \Theta_k),\ k = 1, \ldots\}$

$\mathbf{x}$ is a $T \times 1$ vector
Θ_k is a $m \times 1$ vector, $(m < T)$

$$\{h(\mathbf{x}, \Theta_k)\} = \{\mathbf{c}_2\, \mathbf{c}_2\ \ \mathbf{c}_3\ \mathbf{c}_2 \cdots\cdots \mathbf{c}_1\ \mathbf{c}_2\ \ \mathbf{c}_1\, \mathbf{c}_4\}$$

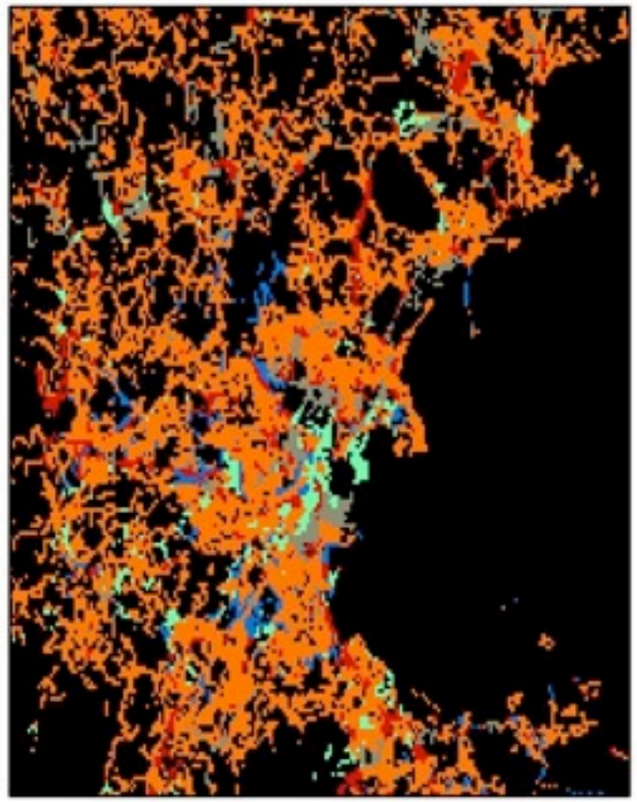
$$\hat{\mathbf{c}} = \text{mode}(\{h(\mathbf{x}, \Theta_k)\}) \qquad perf = \frac{\sum_i^N I(\hat{c}_i = j)}{N}$$

INFERRING LAND USE

Results:

Prediction

Second pass



Actual

Thresh	Total Accuracy	Weights					Confusion Matrix Res	Com	Ind	Prk	Oth
500	0.538	(0.60	0.10	0.10	0.10	0.10)	0.62	0.21	0.15	0.01	0.01
							0.30	0.48	0.19	0.00	0.02
							0.33	0.27	0.38	0.00	0.02
		Res	Com	Ind	Prk	Oth	0.52	0.26	0.18	0.02	0.02
	Share:	0.74	0.09	0.08	0.04	0.05	0.37	0.28	0.25	0.00	0.10

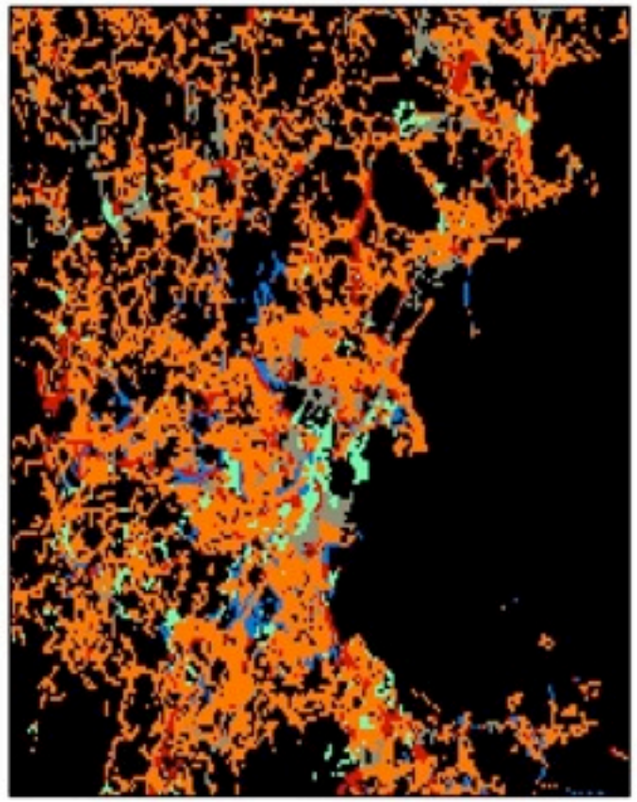
Res Com Ind Prk Oth

INFERRING LAND USE

Results:

Prediction

Second pass



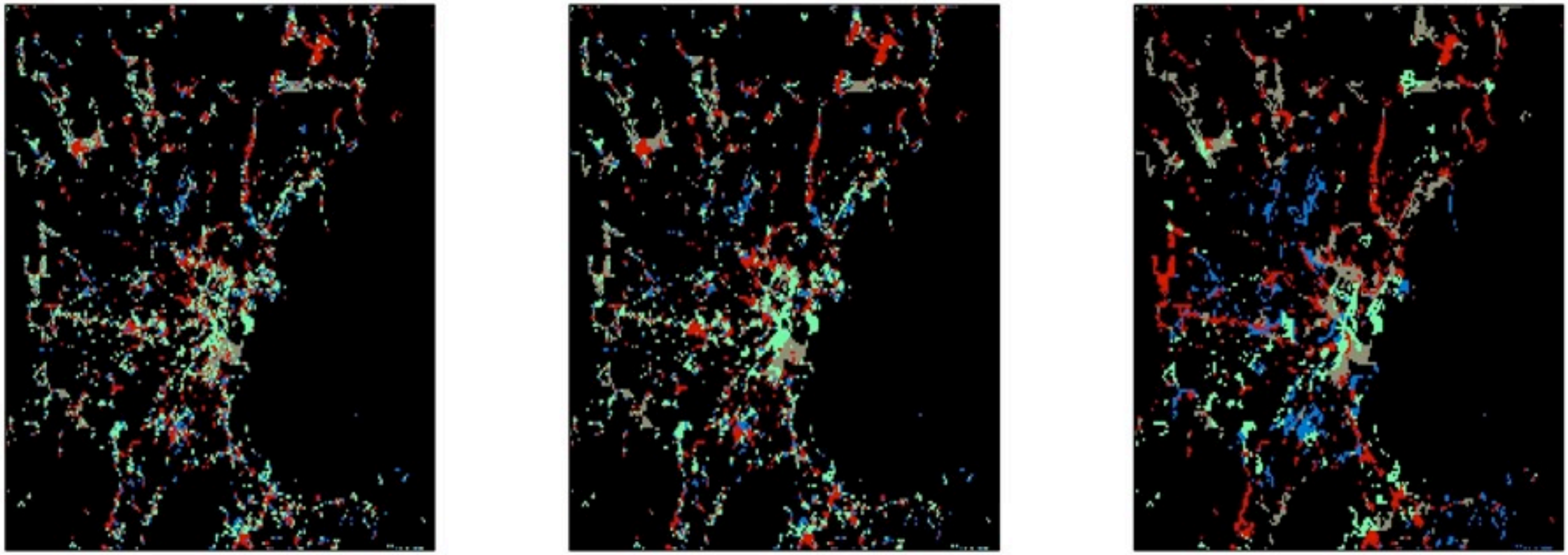
Actual

Thresh	Total Accuracy	Weights					Confusion Matrix Res	Com	Ind	Prk	Oth
500	0.538	(0.60	0.10	0.10	0.10	0.10)	0.62 +	0.21 +	0.15 +	0.01 +	0.01 = 1
							0.30	0.48	0.19	0.00	0.02
							0.33	0.27	0.38	0.00	0.02
		Res	Com	Ind	Prk	Oth	0.52	0.26	0.18	0.02	0.02
	Share:	0.74	0.09	0.08	0.04	0.05	0.37	0.28	0.25	0.00	0.10

Res Com Ind Prk Oth

INFERRING LAND USE

Results:



Prediction Second pass Actual

Thresh	Total Accuracy	Weights					Confusion Matrix Res	Prk	Ind	Con	Oth
500	0.396	(N/A	0.30	0.30	0.20	0.20)	N/A	N/A	N/A	N/A	N/A
							N/A	0.50	0.19	0.11	0.19
							N/A	0.27	0.37	0.12	0.24
		Res	Com	Ind	Prk	Oth	N/A	0.31	0.18	0.29	0.21
	Share:	0.00	0.33	0.31	0.16	0.20	N/A	0.26	0.24	0.15	0.34

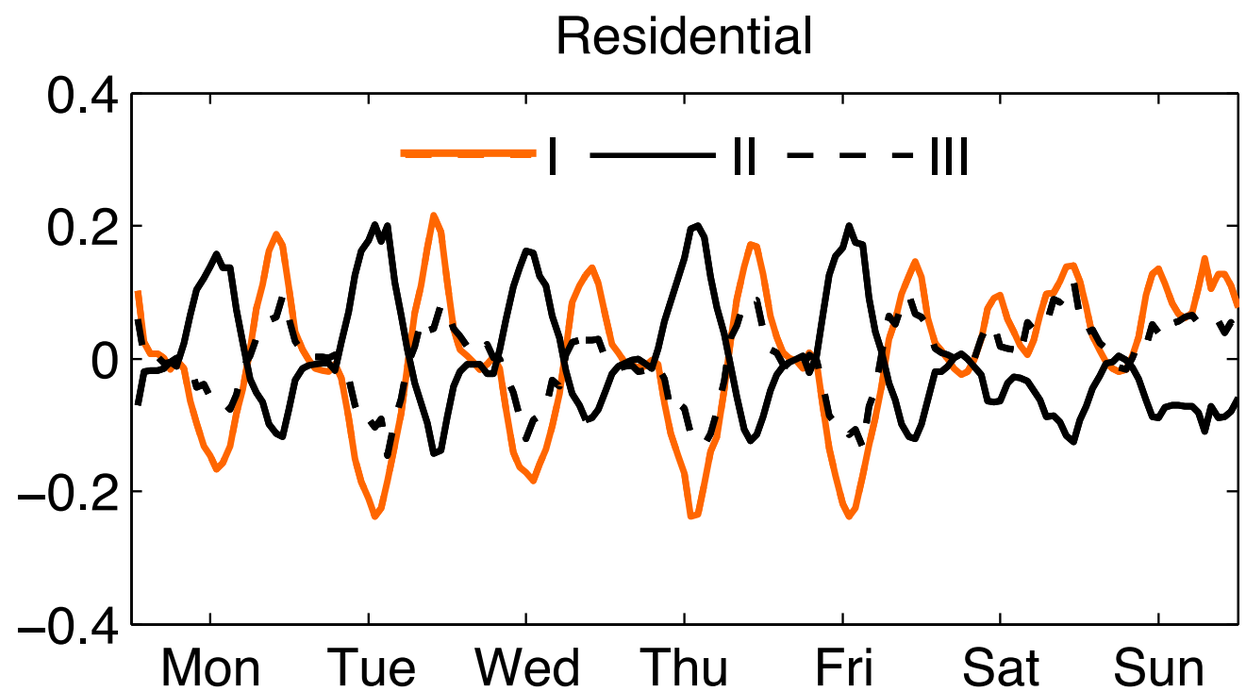
INFERRING LAND USE

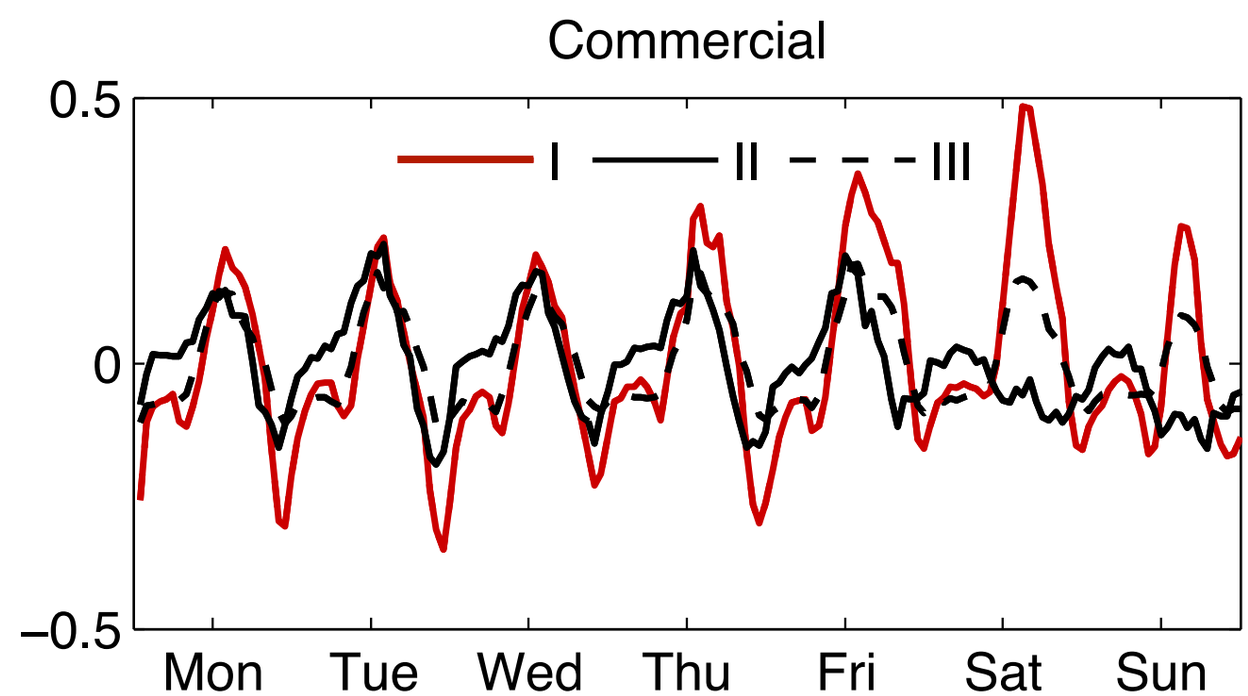
CLASSIFICATION ERROR ANALYSIS

Group I — Cells correctly predicted to be a given use.

Group II — Cells of a given use incorrectly predicted to be a different use.

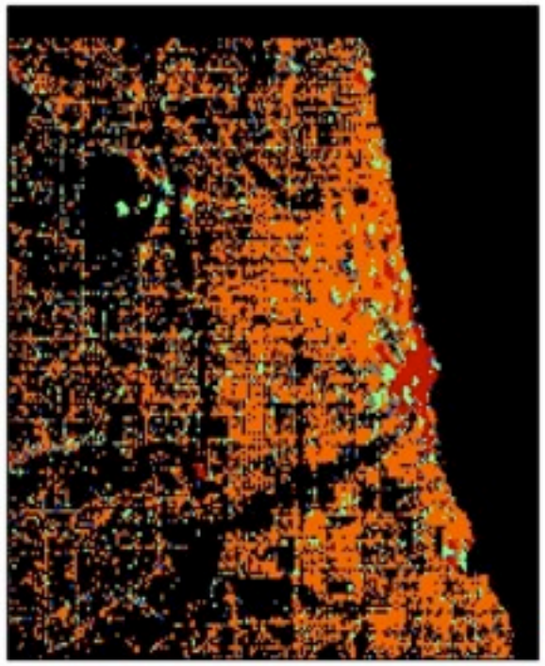
Group III — Cells of a different use incorrectly predicted to be a given use.





ROBUSTNESS ACROSS CITIES

Prediction

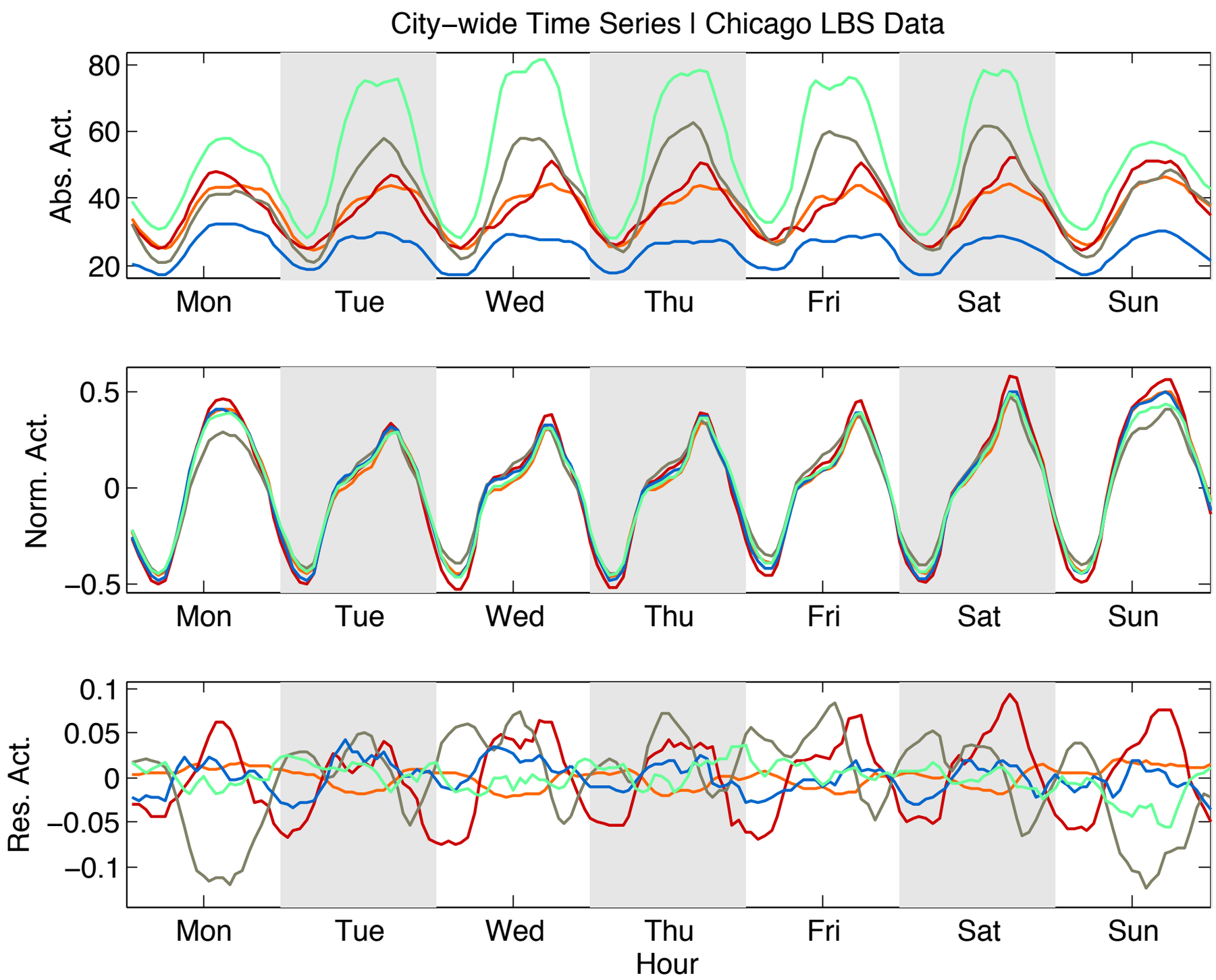


Second pass

Actual

Thresh	Avg. Error	Cutoffs					Cross Tabulation				
							Res	Com	Ind	Par	Oth
100	0.468	(0.50	0.12	0.12	0.12	0.12)	0.66	0.09	0.08	0.05	0.12
							0.53	0.15	0.10	0.05	0.16
							0.52	0.10	0.19	0.05	0.14
		Res	Com	Ind	Prk	Oth	0.59	0.10	0.11	0.06	0.13
	Share:	0.63	0.09	0.09	0.09	0.10	0.55	0.12	0.10	0.06	0.17

ROBUSTNESS ACROSS CITIES



INFERRING LAND USE

Summary:

- Introduced a way to reconcile and normalize traditional and new data sources
- Used temporal activity patterns to identify land uses
- Performed error analysis to determine why classifications failed
- Similar results obtained for different cities

Implications:

- Mobile phone activity can be radically different depending on zoned land use.
- Traditional zoning maps should be augmented to account for real-time use.
- Dynamic land use patterns seem relatively consistent despite differences in history and regulatory classification

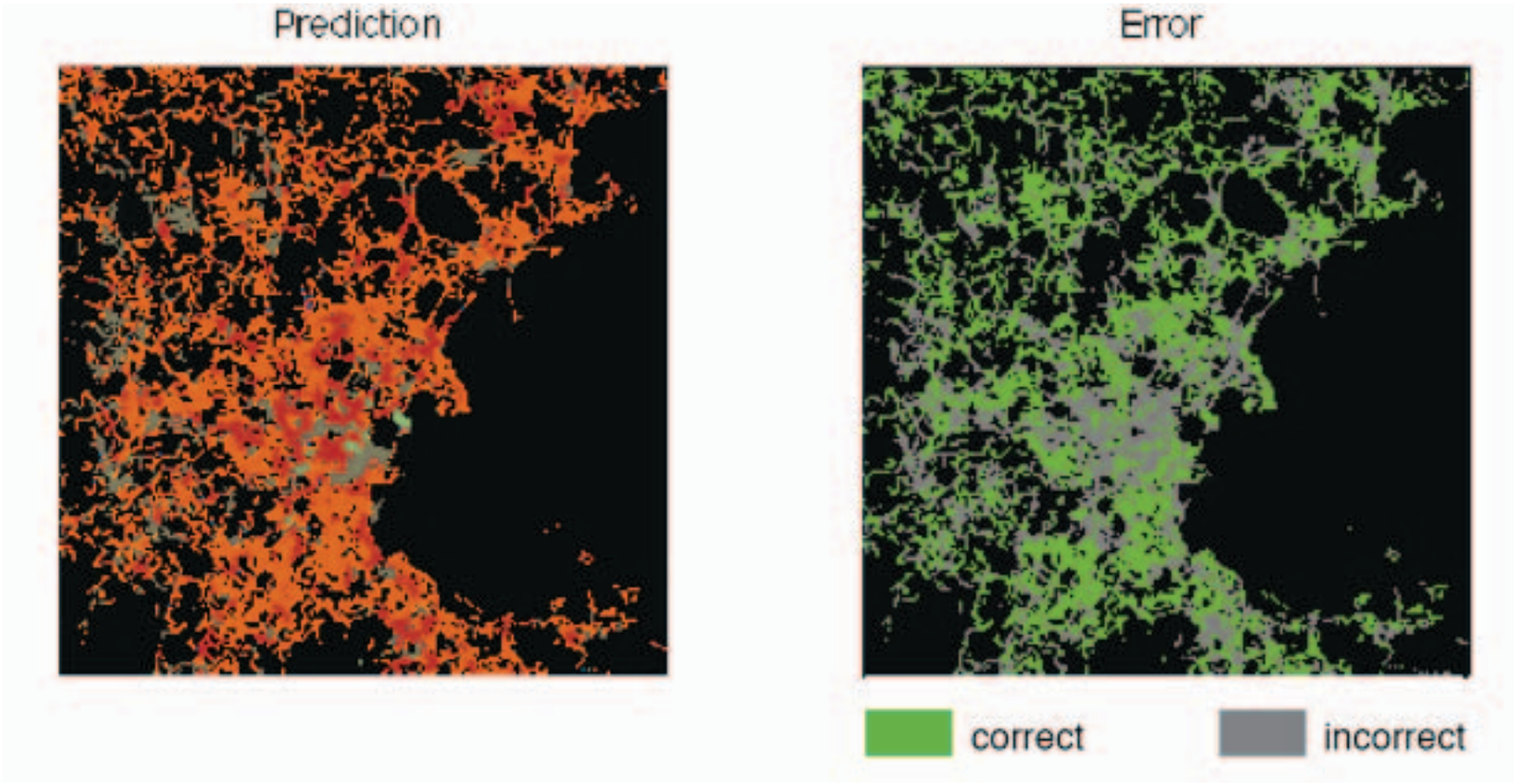
FUTURE CONTRIBUTIONS

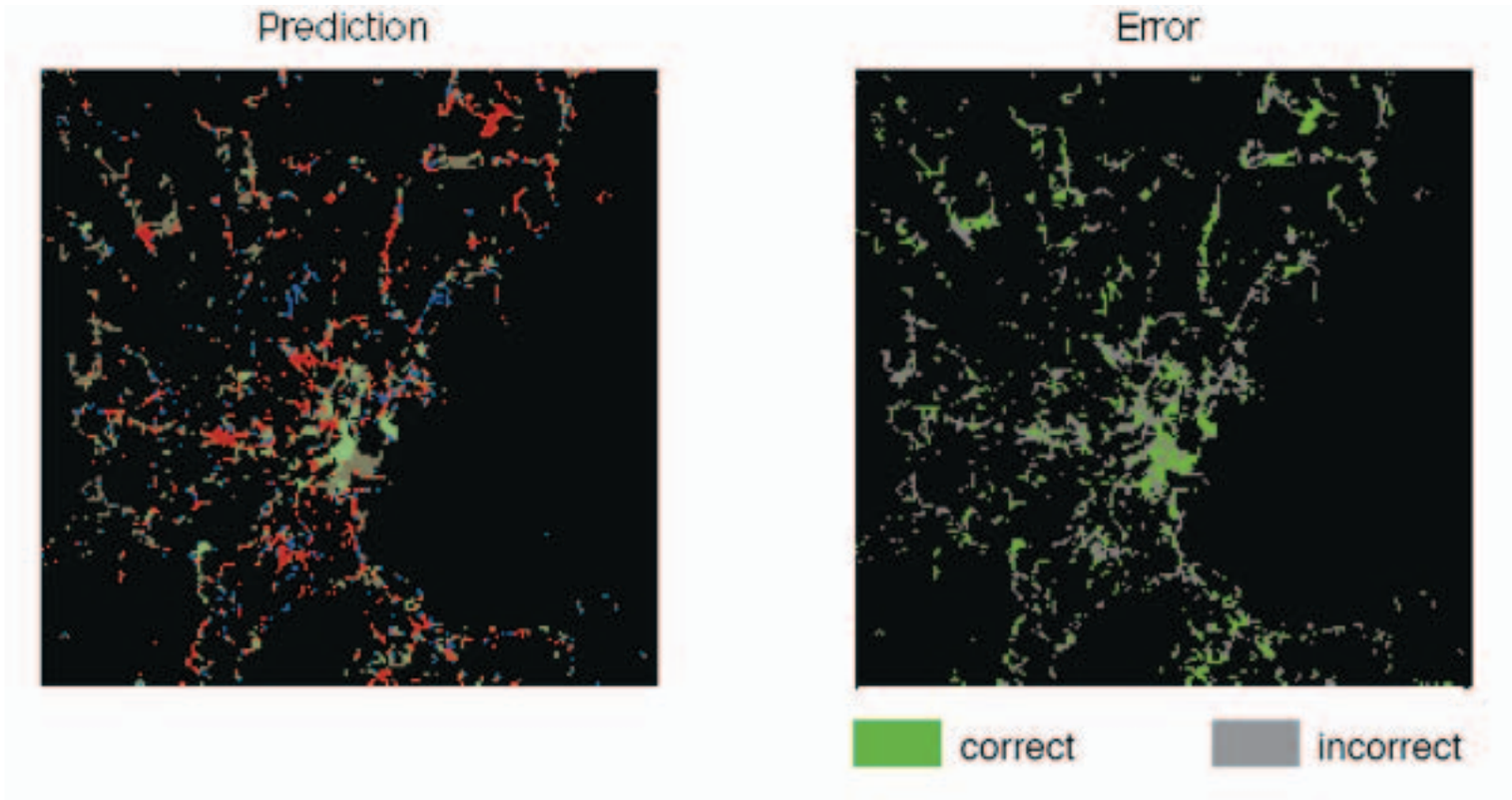
- Use more sophisticated feature spaces for zoning classification.
- Introduce new data sources.
- Measure urban system dynamics in multiple cities around the globe to identify similarities and differences in behavior.
- Apply normalization metrics to other data sources.
- Combine our understanding of temporal land use with measurements of human mobility to develop better models of intra-city travel behavior.

THANK YOU!

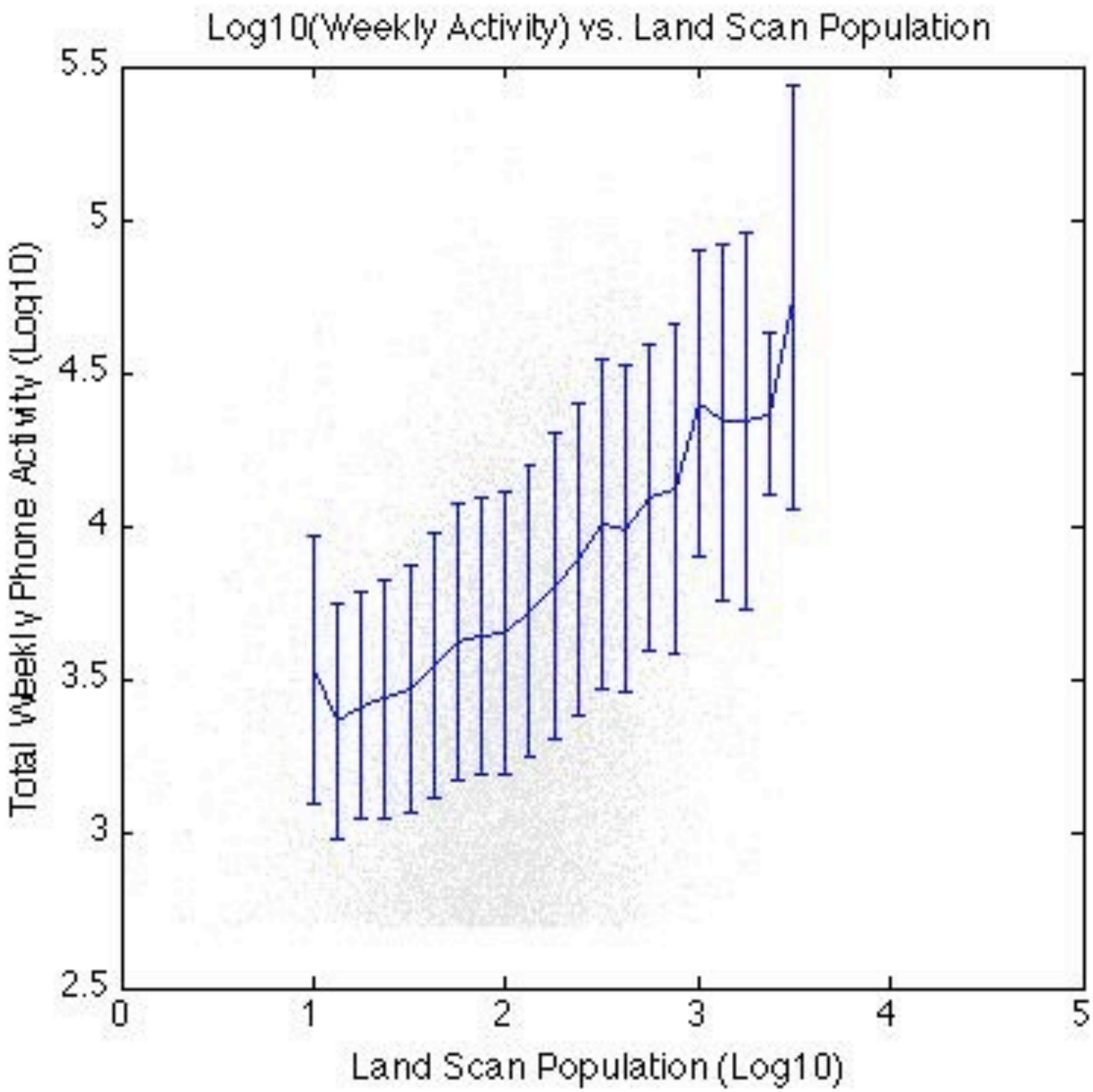
QUESTIONS?

ERRORS


Prediction
Error
correct
incorrect


Prediction
Error
correct
incorrect

CORRELATION WITH STATIC POPULATION



Classification Error Analysis

Group I Cells correctly predicted to be a given use.

Group II Cells of a given use incorrectly predicted to be a different use.

Group III Cells of a different use incorrectly predicted to be a given use.