City Traffic Forecasting Using GPS Data

Yucheng Hu
Tsinghua Universtiy, Beijing, China

Joint work with

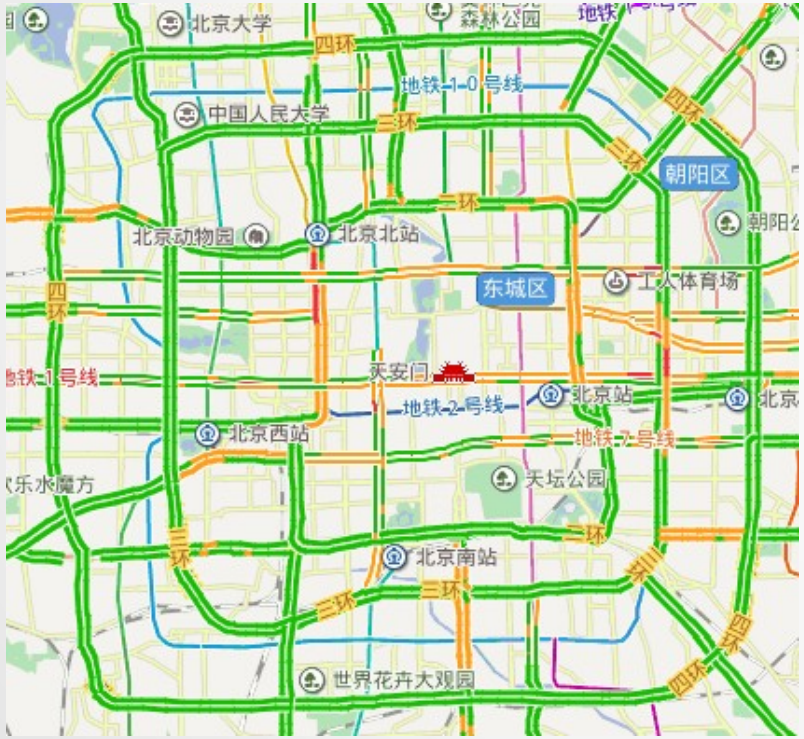
Weinan E Princeton University, Peking University
Tiejun Li Peking University
Hao Liu Beijing Transportation and Information Center
Haijun Yu Chinese Academy of Science
Haoshu Tian PrinceTechs

New Directions in Mathematical Approaches for Traffic Flow Management
Workshop II. Traffic Estimation
IPAM, UCLA, 2015.10.14

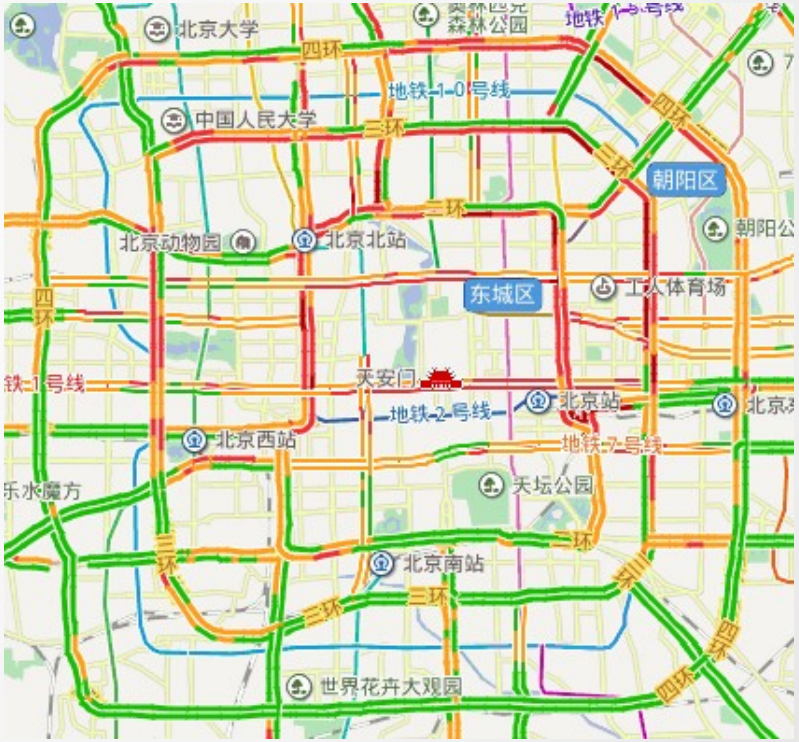

This is a “World Car-free Day”
650
FORLAND


This is NOT a parking lot
@法制晚报

The traffic problem in Beijing



12:00



18:00

source: Baidu Maps

The cost of traffic congestion

Cloth Food House Travel

Annual cost

$11.3 billion

Source: Peking University's National Development Research Institute

- **Money**
- **Fuel**
- **Happiness**
- **Lives**

Can mathematicians help?

Understanding traffic

Fluids
Particle system
Complex network
Big data

Taking actions

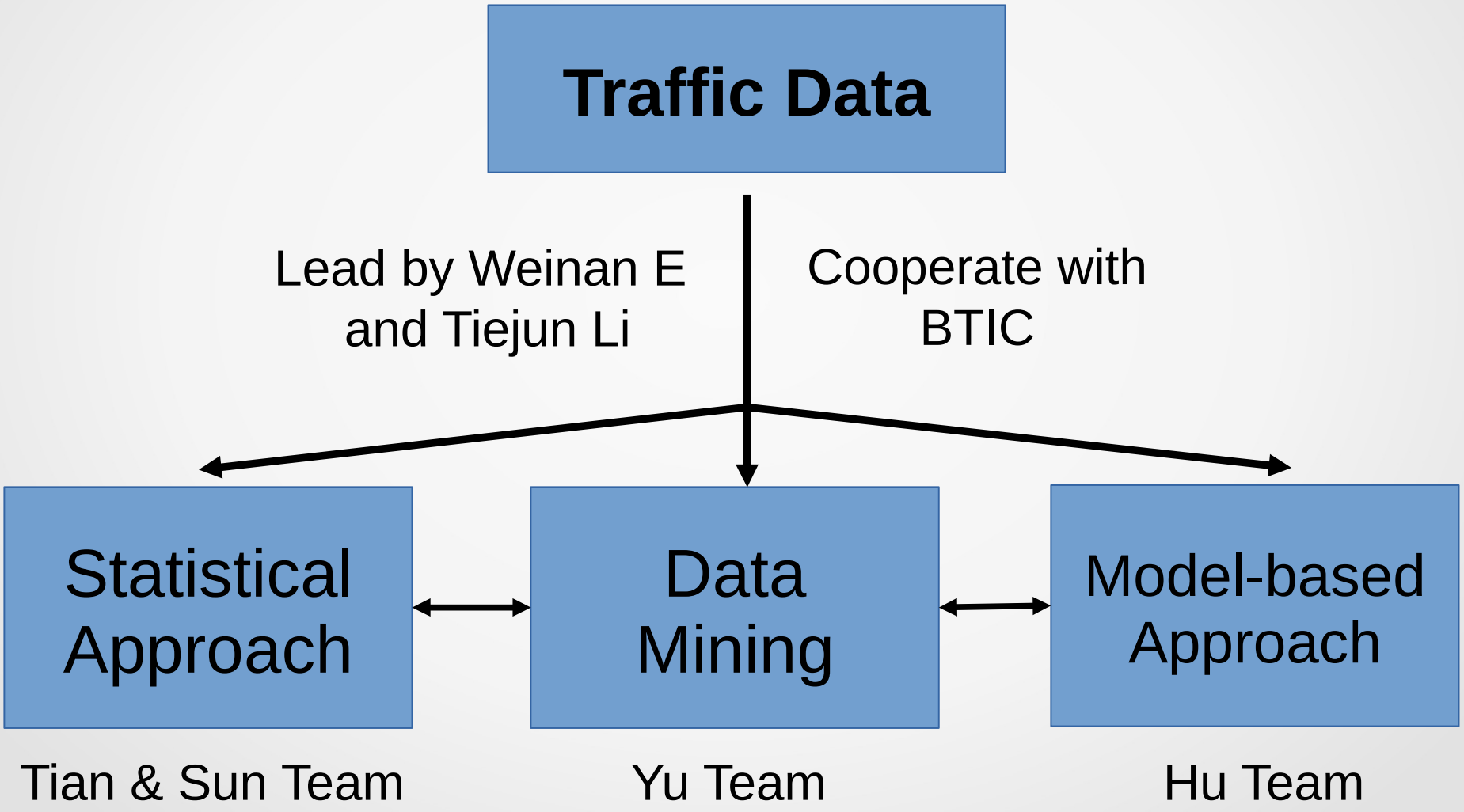
Route planning

Policy making

e.g. end-number license plate policy (1/5)

Source: Chinadaily

Big City, Big Data



Taxi GPS Data

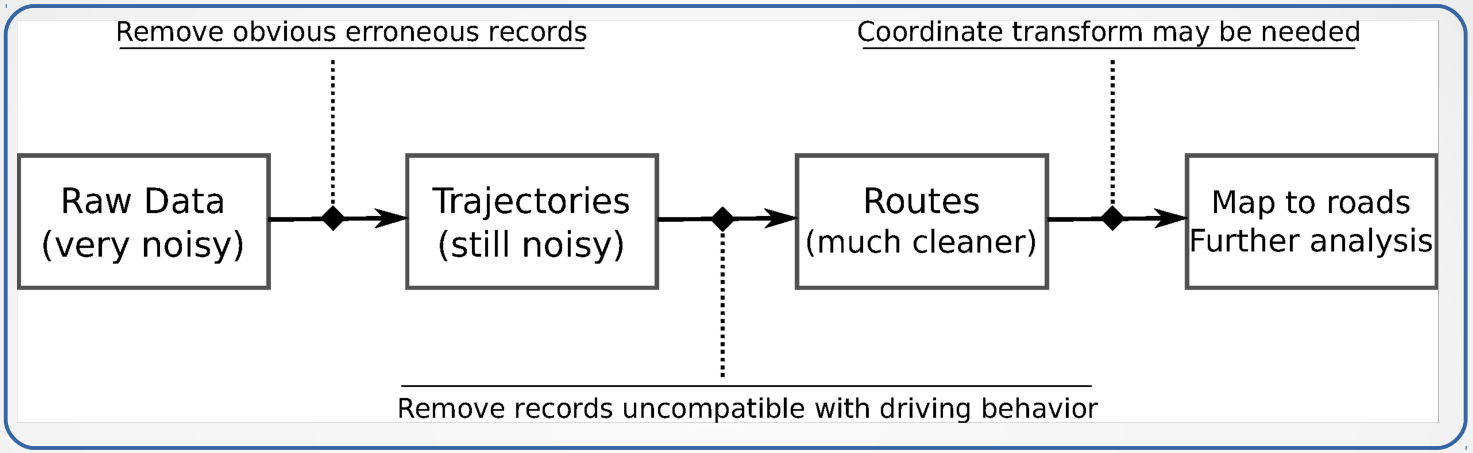
- From BTIC
- Near 70, 000 taxis
- Transmit signal once every minutes
- 4 months, 200G
- Real time in the future

2	ZHTC	657954585156	[redacted]	2013-03-03 23:09:17	116.411102	40.001732	0	1	1	170	CZ
3	ZHTC	657954505750		2013-03-03 23:09:17	116.353783	39.980301	0	1	1	86	CZ
4	ZHTC	657953575558		2013-03-03 23:09:17	116.420181	39.903393	0	0	1	86	CZ
5	ZHTC	658055505656		2013-03-03 23:09:17	116.483383	39.985168	2	1	1	184	CZ
6	ZHTC	657757535752		2013-03-03 23:09:17	116.427353	39.939289	41	0	1	90	CZ
7	ZHTC	657757545457		2013-03-03 23:24:11	116.587601	40.077305	0	1	0	116	CZ
8	ZHTC	657753575151		2013-03-03 23:09:17	116.304268	39.909069	63	1	1	0	CZ
9	ZHTC	658057515456		2013-03-03 23:09:17	116.087349	39.947639	0	0	1	0	CZ

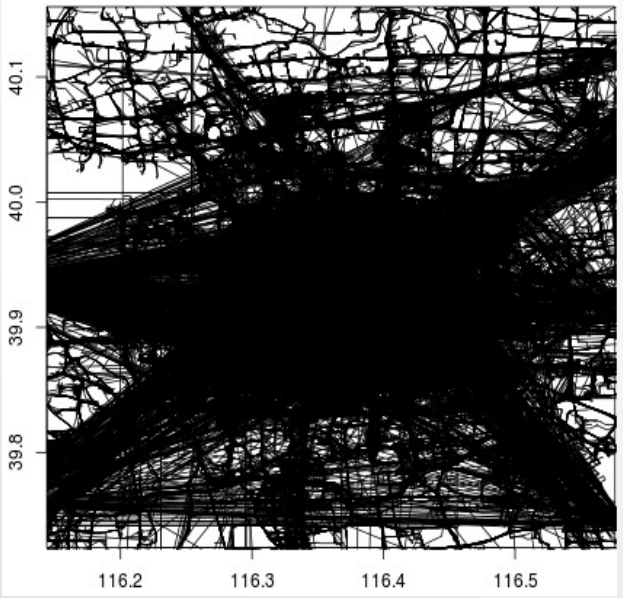
Sample data

Data Pre-processing

Trajectory-based denoising



Before



After

Traffic forecasting: overview

- **Pure Data**

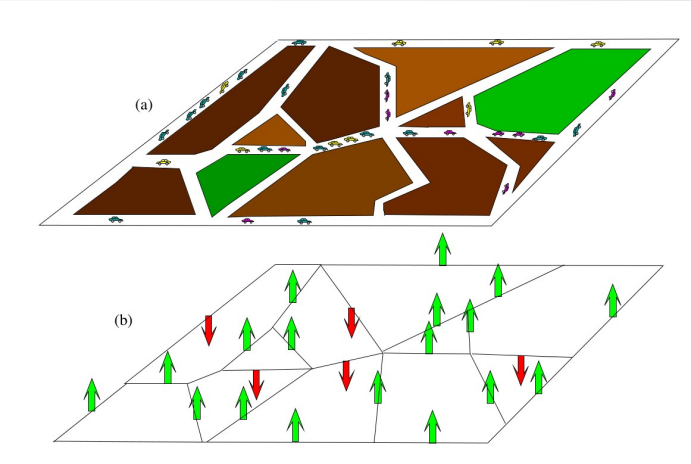
  Time-series

- **Road network + Data**

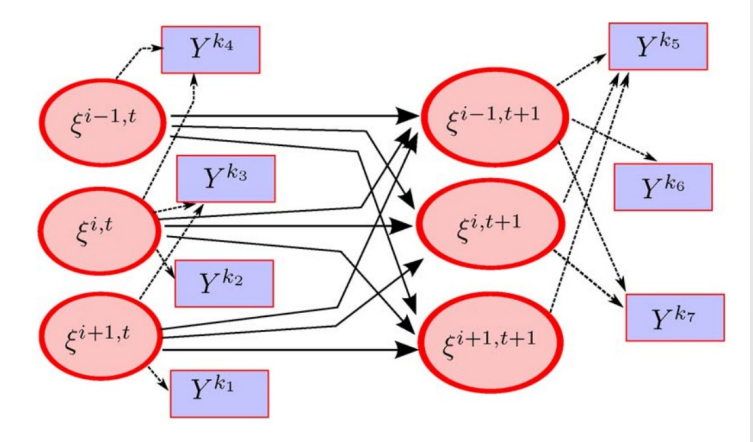
  Dynamic Bayesian network

- **Mechanism + Data**

  DYNASMART
  DynaMIT

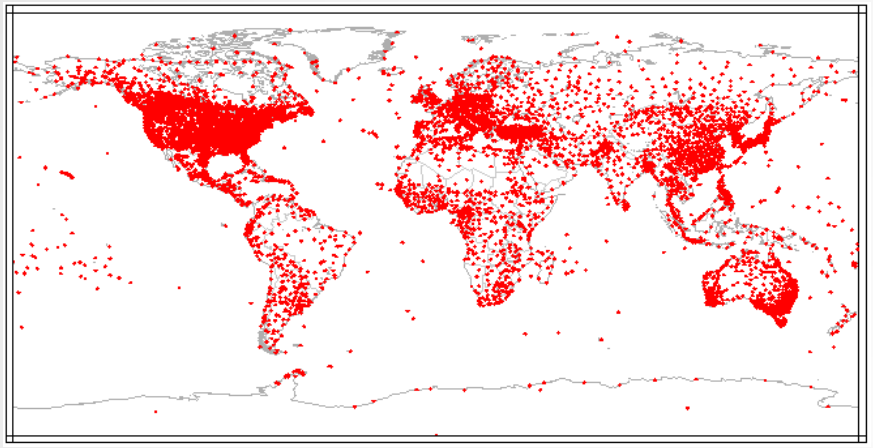


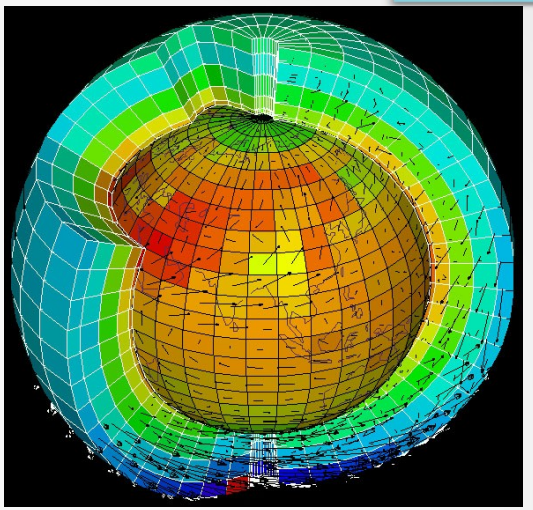
Furtlehner et.al, 2007



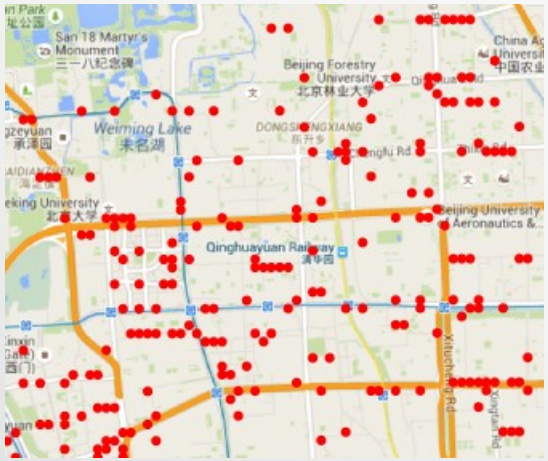
Hofleitner et.al, 2012

Numerical weather prediction

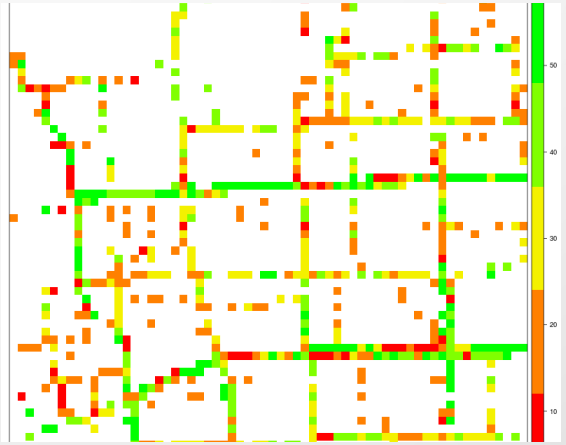




Data **Model**



Accidents
Weather
Social events

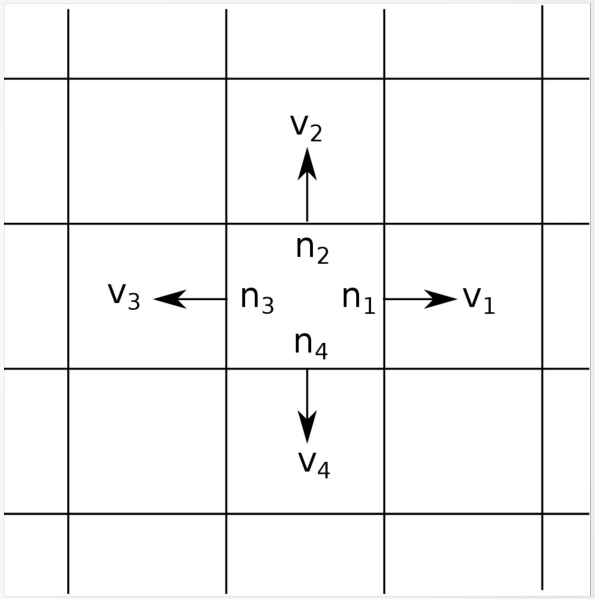


Modelling traffic

1. How many roads?
2. How many cars?
3. How do the cars move?

1. How many roads

- Partition Beijing into grids (100x100m)
- Count cars in each grid in every 10 mins
- Measure their average speed
- No car, no road
- More cars, more roads



2. How many cars

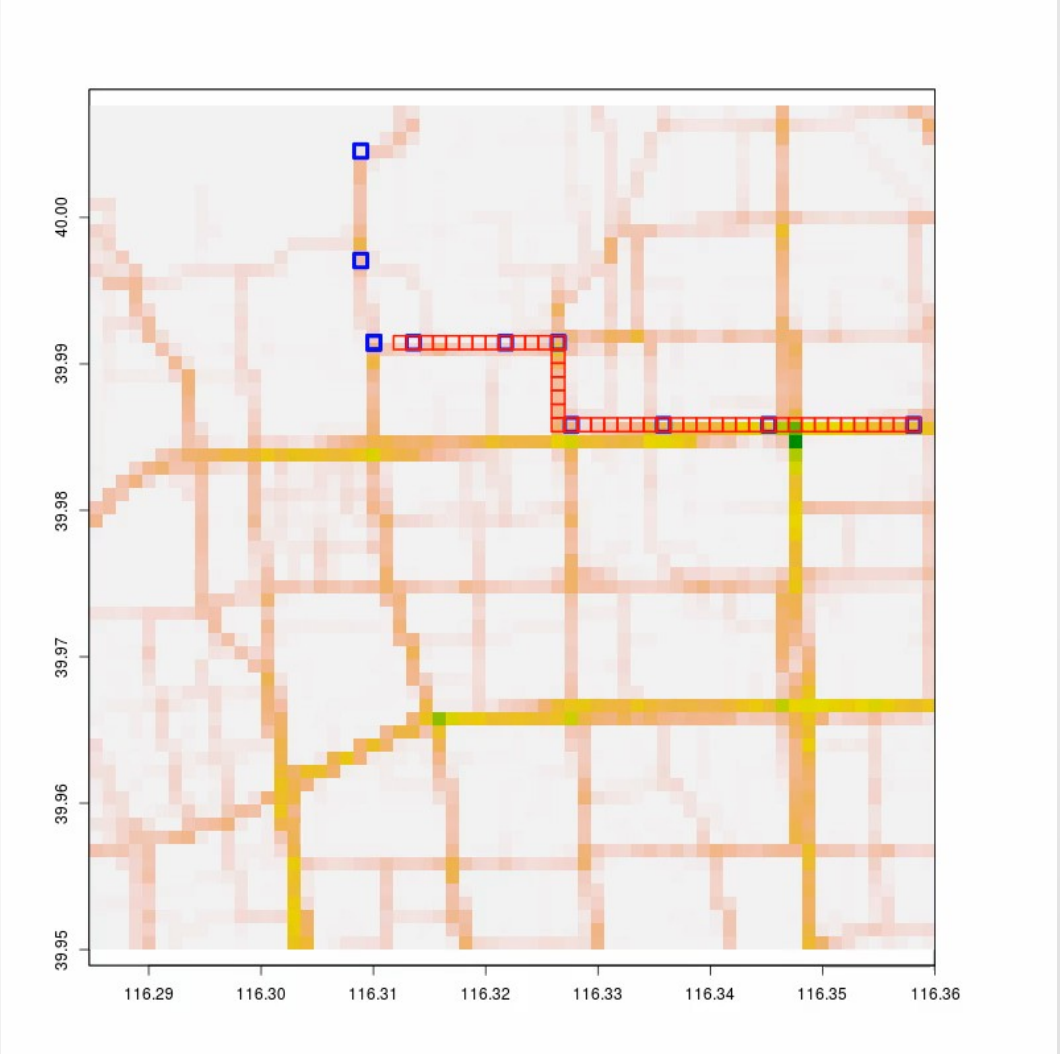
O/D Analysis

Current: historical path

Future: path sampling

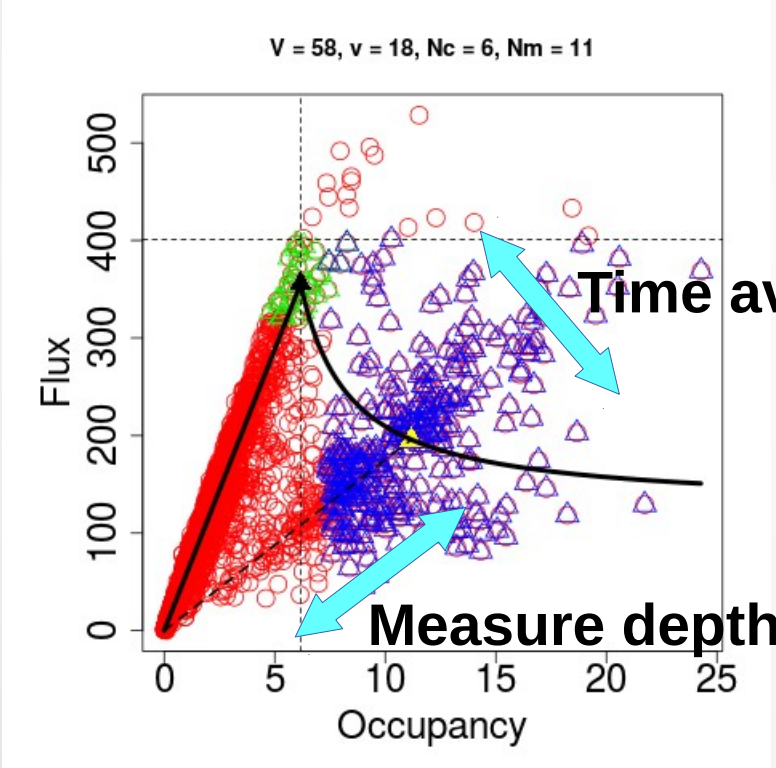
Taxi / Private car

More data is needed

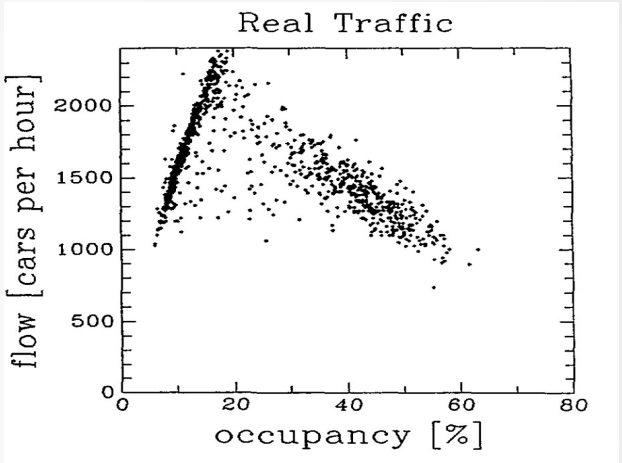


3. How do the car move

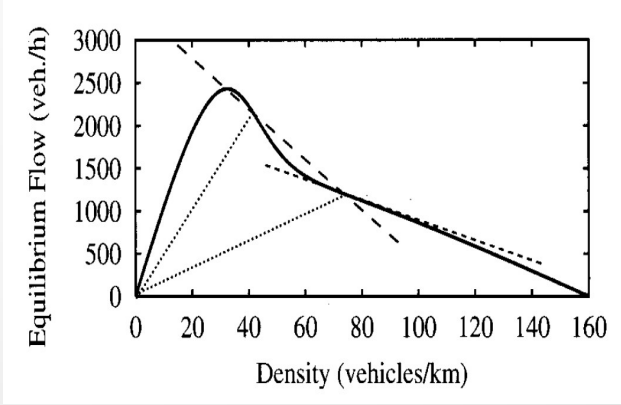
V(N) --- **constitutive law**



The fundamental diagram

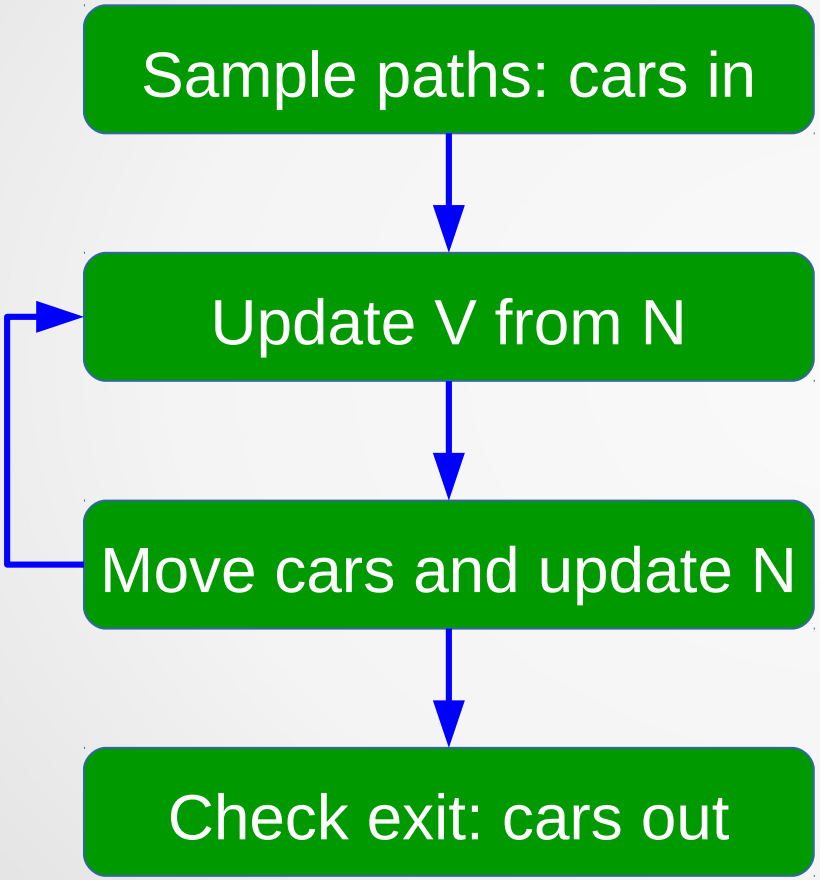


Nagel and Schreckenberg, 1992

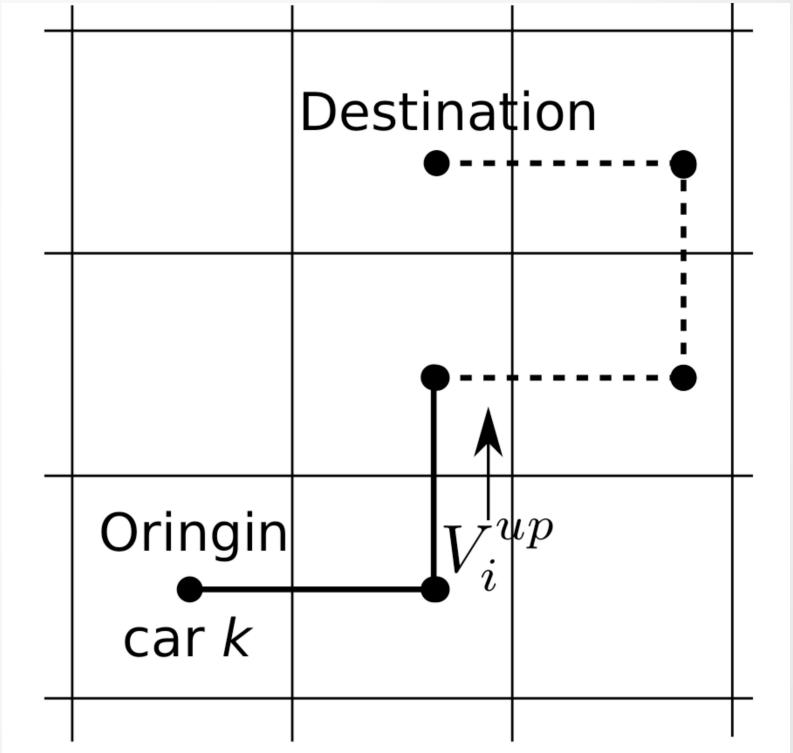


Helbing, 2001

Coarse-grained cellular automata model



mesh size: 256x256



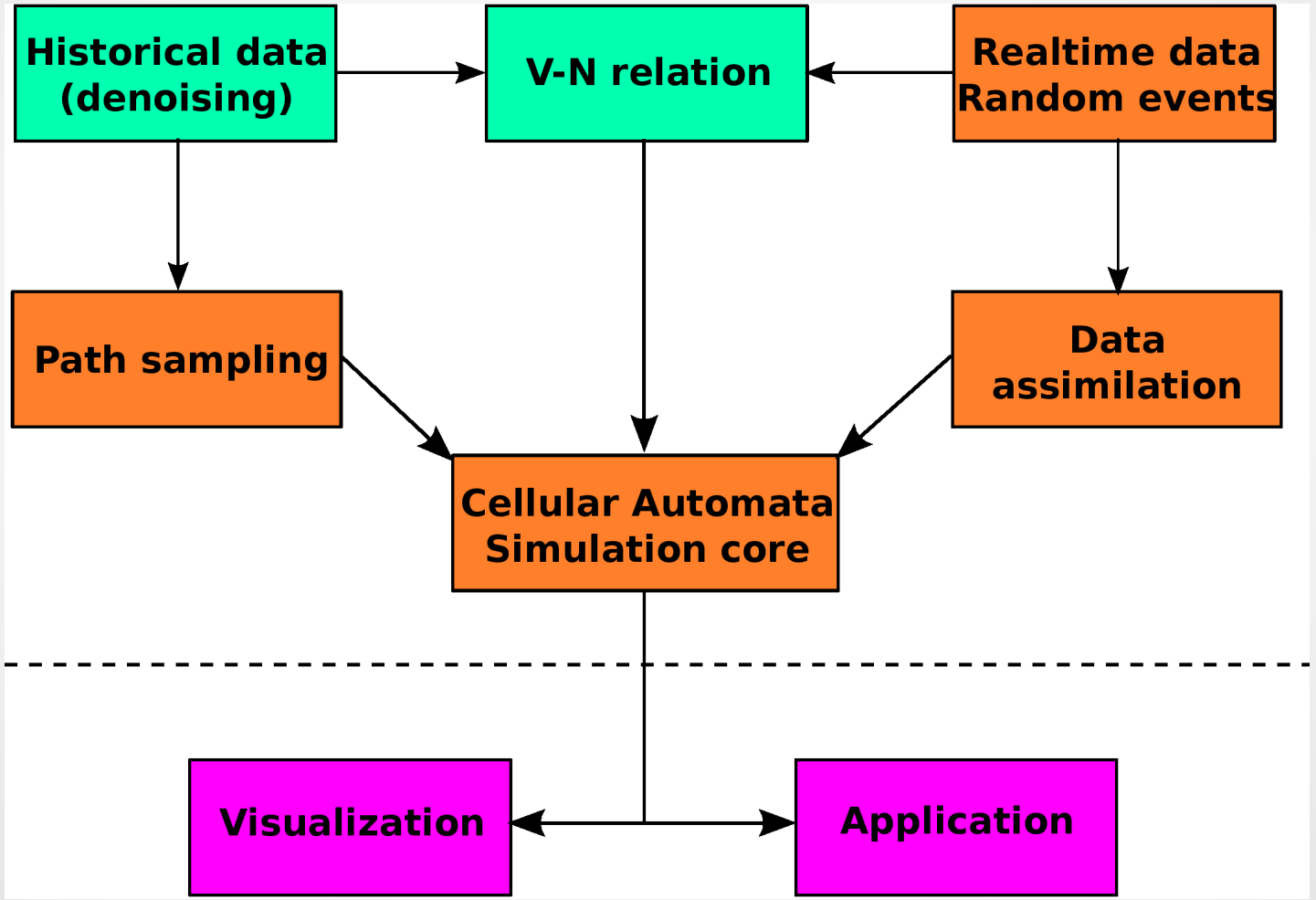
Features

- Powered by big data
- Model inside
- Requires no road information
- Similar work-flow with NWP

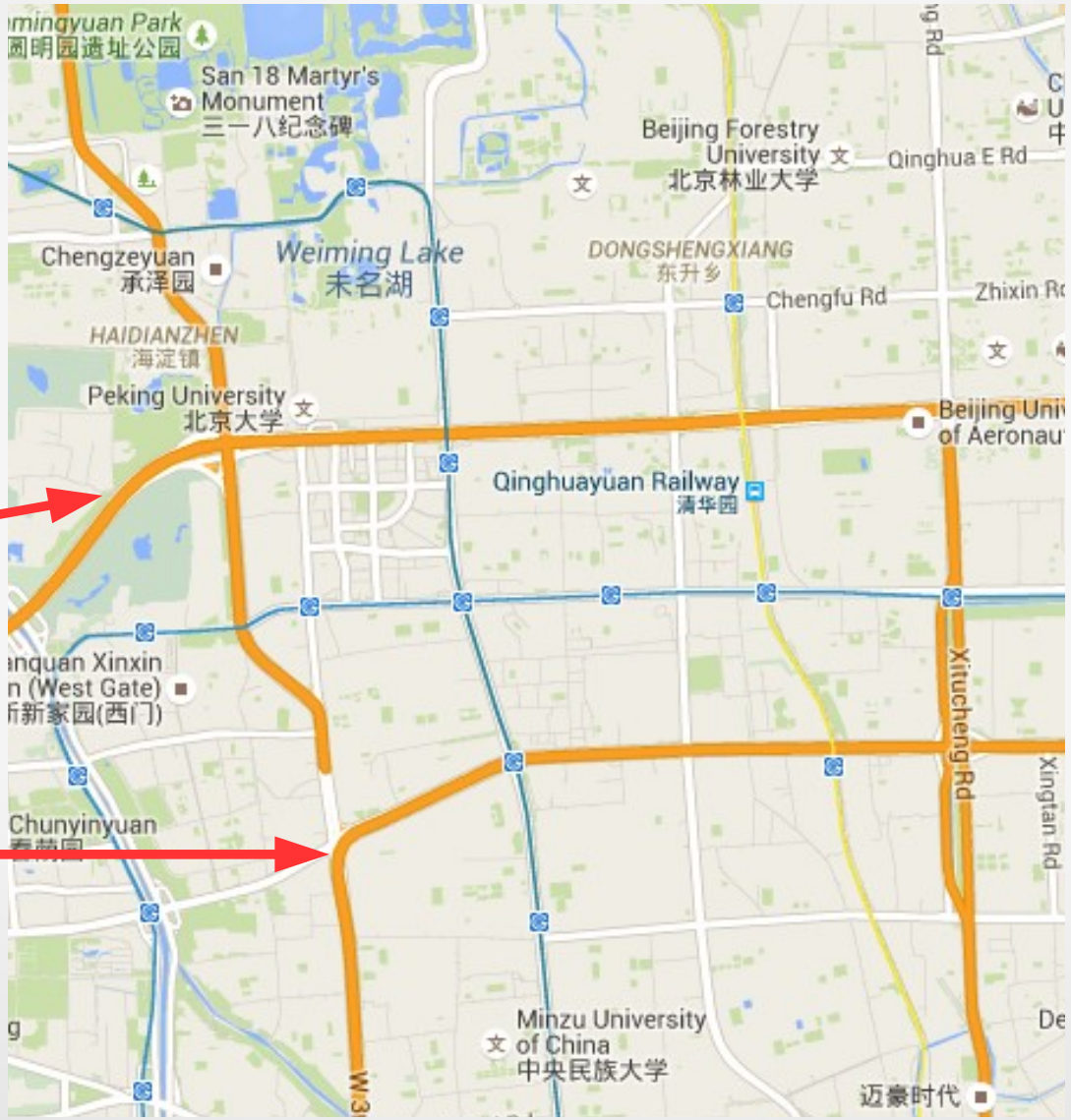
Difficulties

- Heavy computational load
- Choice of V-N relation
- The effect of private cars
- Incorporate real-time data

Work-flow


Historical data (denoising)
V-N relation
Realtime data Random events
Path sampling
Data assimilation
Cellular Automata Simulation core
Visualization
Application

Results: 2D small

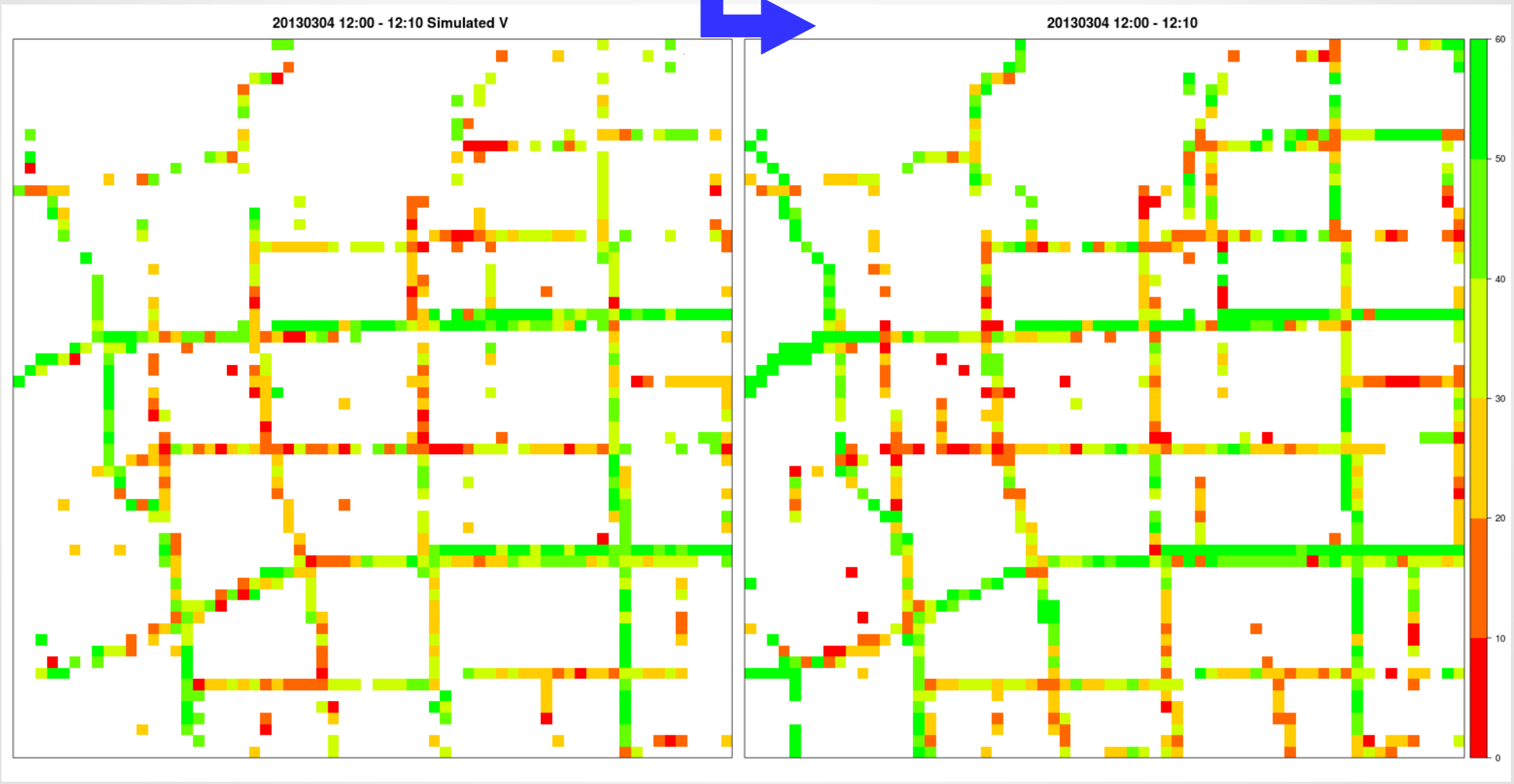


4th ring NW

3rd ring NW

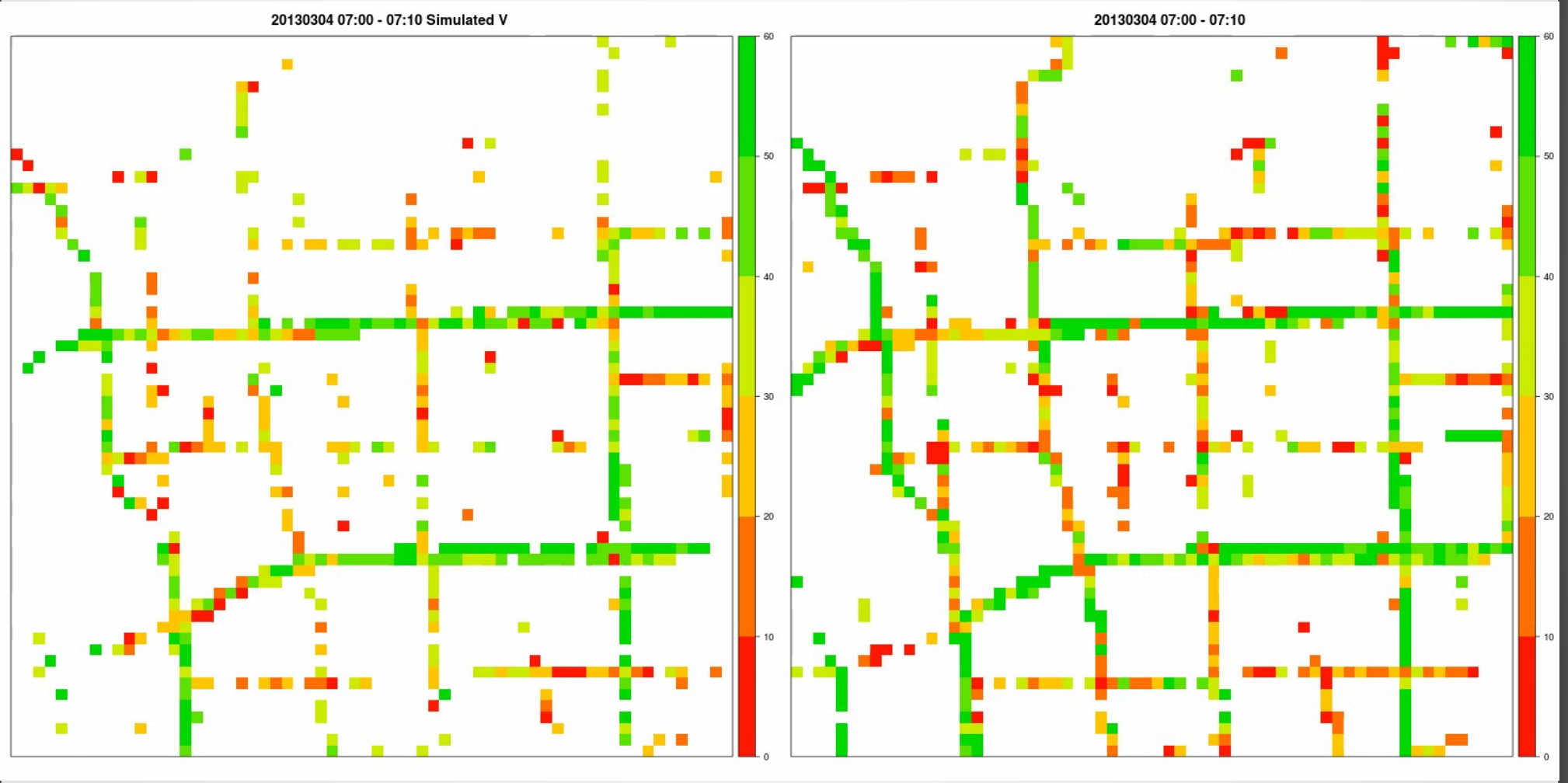
mesh size 64x64

Results: 2D small



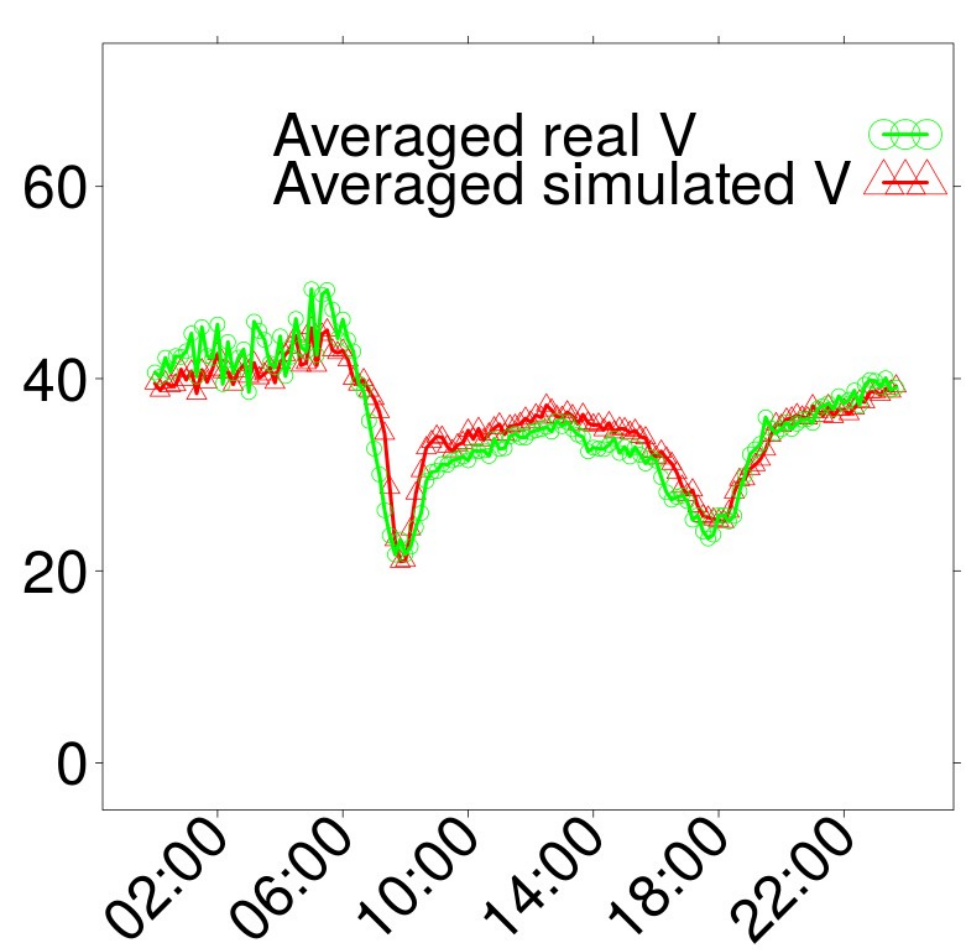
Simulation **Real data**

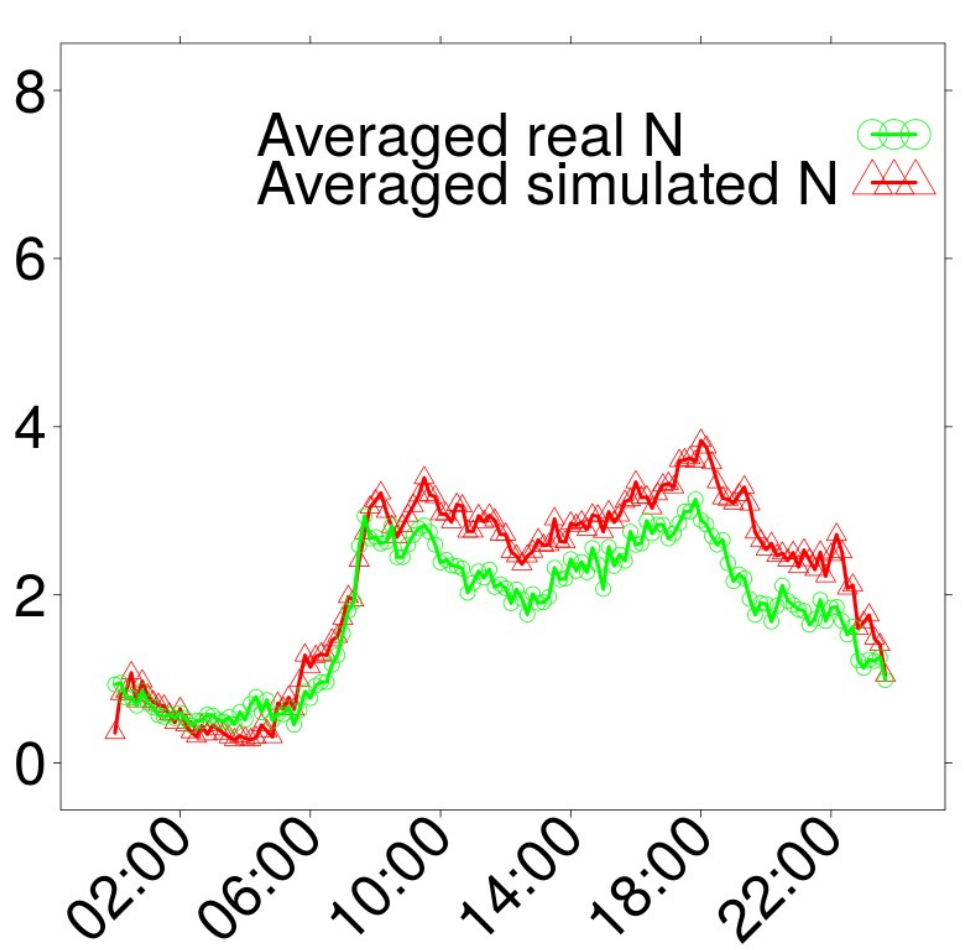
Results: 2D small



Simulation **Real data**

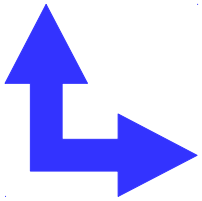
Results: 2D small

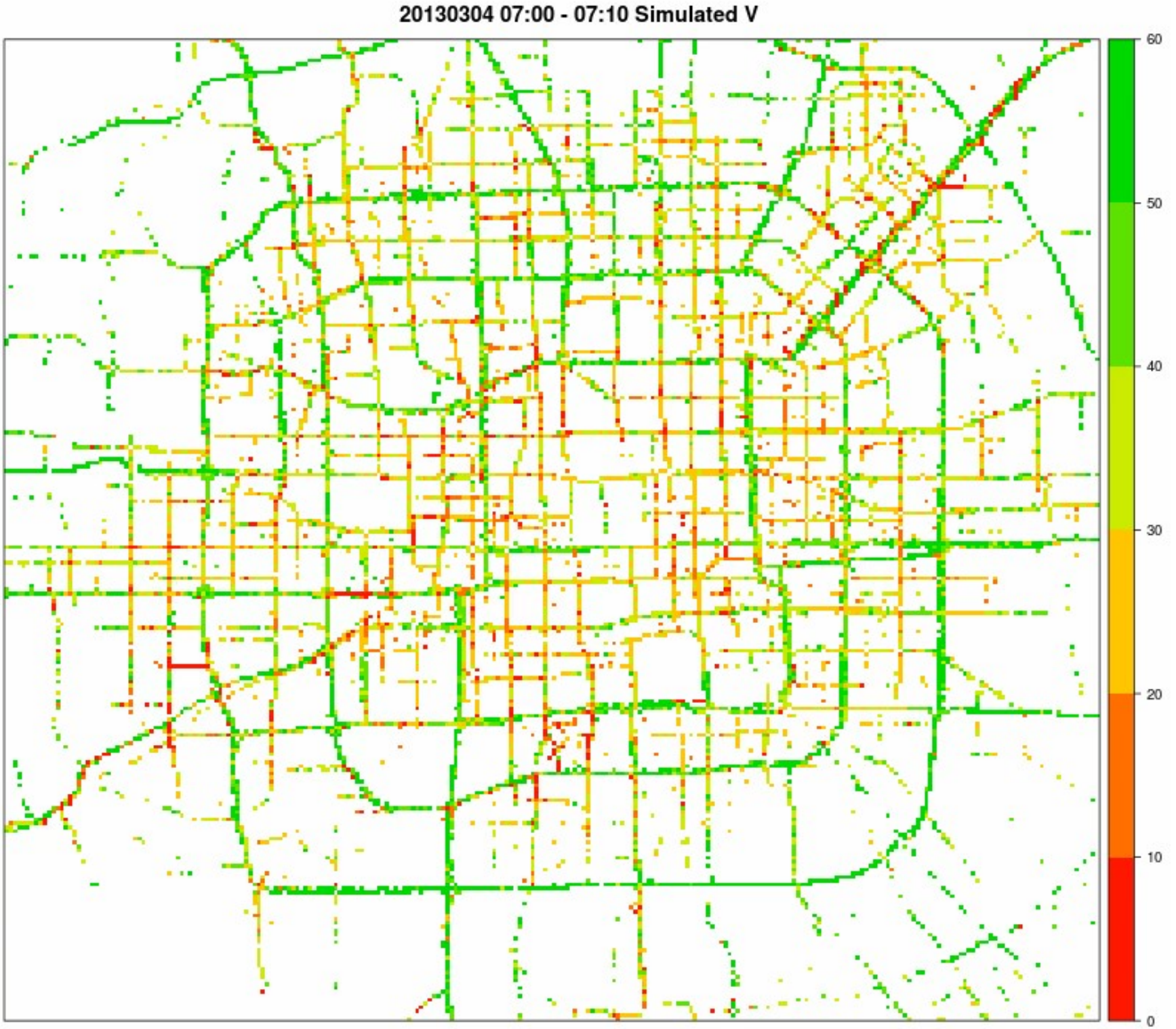




2D large

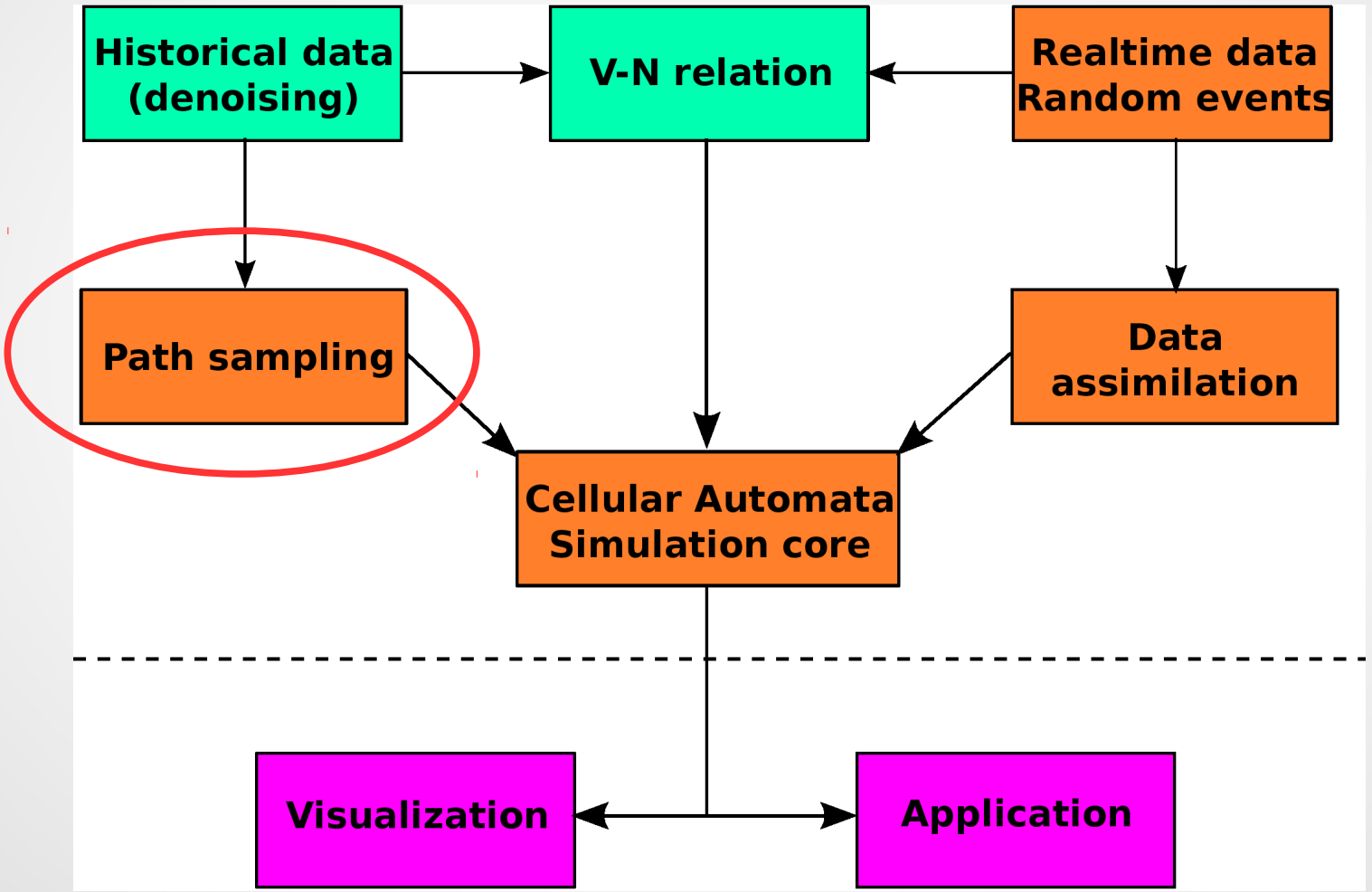
256x256





Sampling Taxi Origin-Destination Demands in Large Metropolitan Environment

O-D sampling


Historical data (denoising)
V-N relation
Realtime data Random events
Path sampling
Data assimilation
Cellular Automata Simulation core
Visualization
Application

O-D data

a

Raw data

Pre-process

Routes

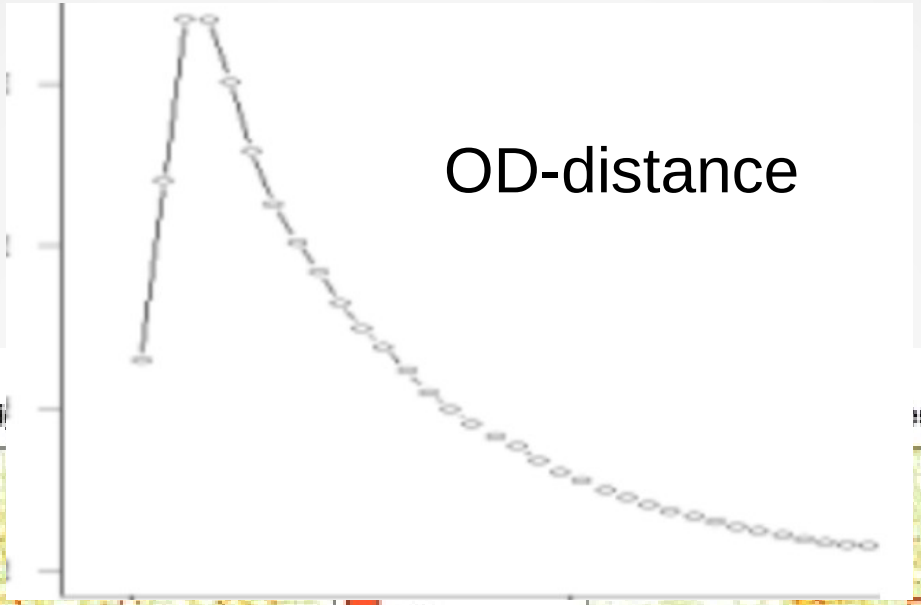
46	2013-03-05 11:56:58	116.40073	39.90125	0
47	2013-03-05 12:00:16	116.40066	39.90478	0
48	2013-03-05 12:02:28	116.40066	39.90482	1
49	2013-03-05 12:04:19	116.40062	39.91098	1
50	2013-03-05 12:05:46	116.40282	39.91281	1
51	2013-03-05 12:06:46	116.40441	39.91402	1
52	2013-03-05 12:08:30	116.40879	39.91419	1
53	2013-03-05 12:13:37	116.41022	39.91418	1
54	2013-03-05 12:14:20	116.41113	39.91414	1
55	2013-03-05 12:14:58	116.4115	39.91389	0
56	2013-03-05 12:15:15	116.41149	39.91384	0
57	2013-03-05 12:15:16	116.41149	39.91383	0
58	2013-03-05 12:16:13	116.41172	39.91365	0

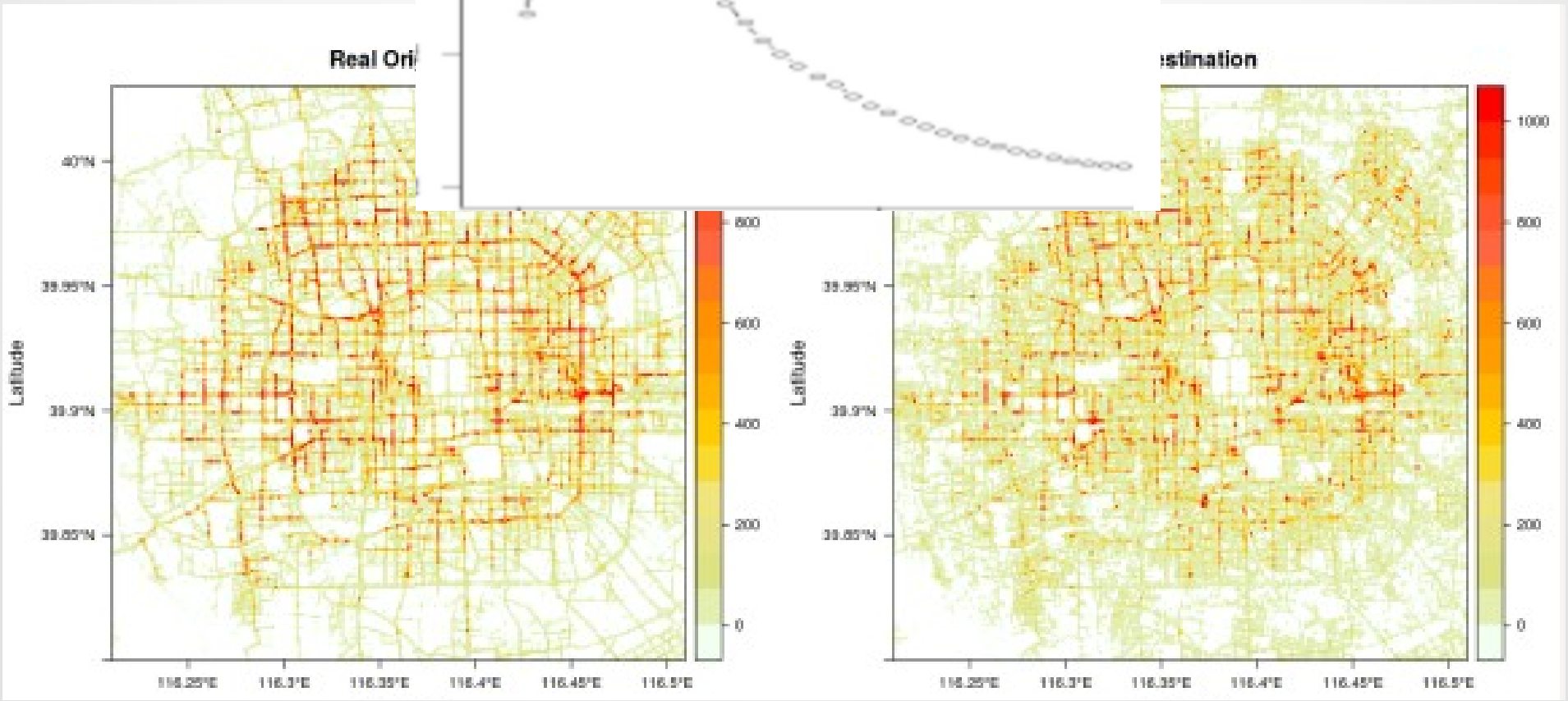
Origin

Destination

b

O-D statistical





O-D sampling

For some time period 8:00-9:00, given

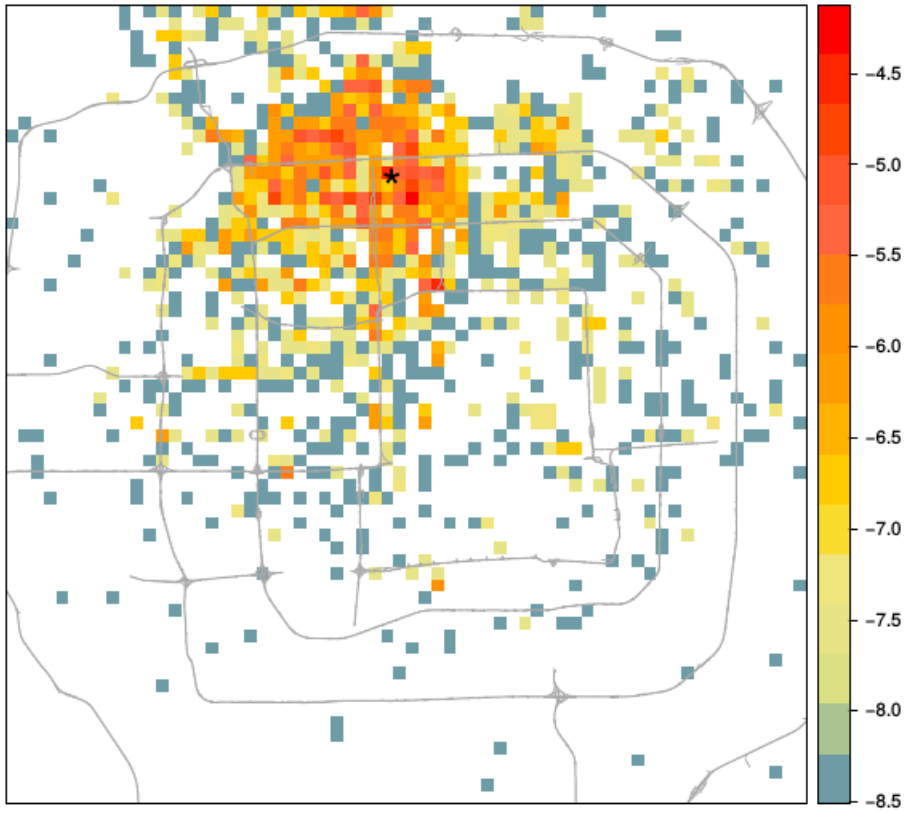
- N: the total number of OD;
- $P_O(\mathbf{x})$: the probability density of origins;
- $P_D(\mathbf{y})$: the probability density of destinations;
- $P_{dist}(r)$: the probability density of OD-distance.

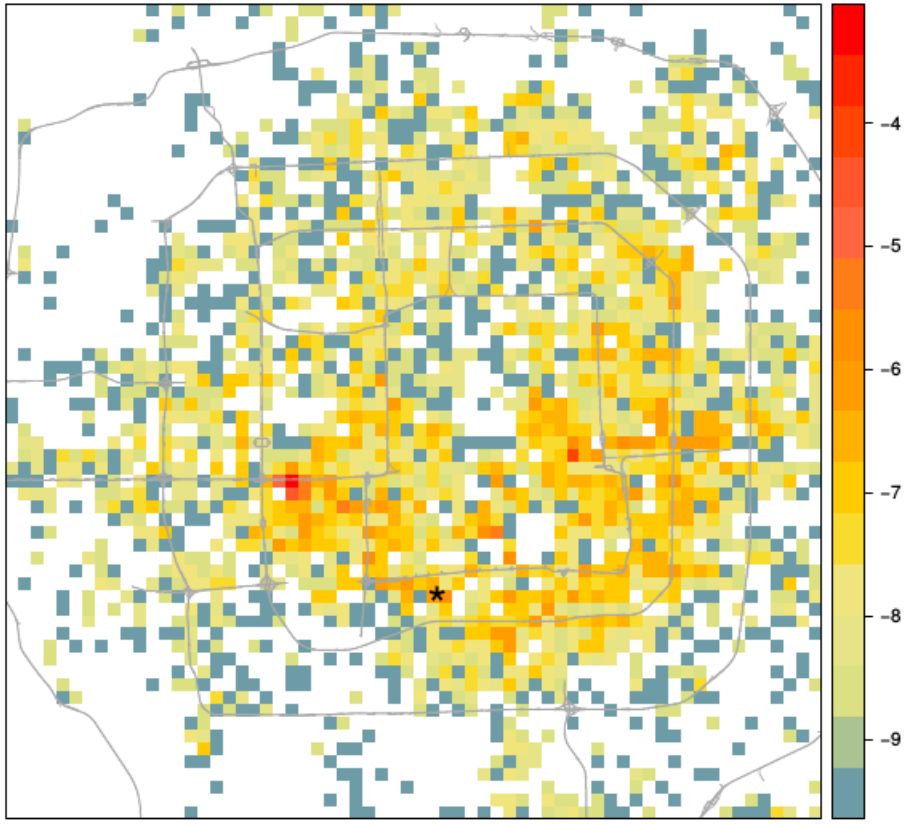
Generate OD samples that fit the above quantities.

Conditional probability

$$p_D(\mathbf{y}) = \int_{\mathbf{x}} p_O(\mathbf{x}) p(\mathbf{y}|\mathbf{x}) d\mathbf{x}$$

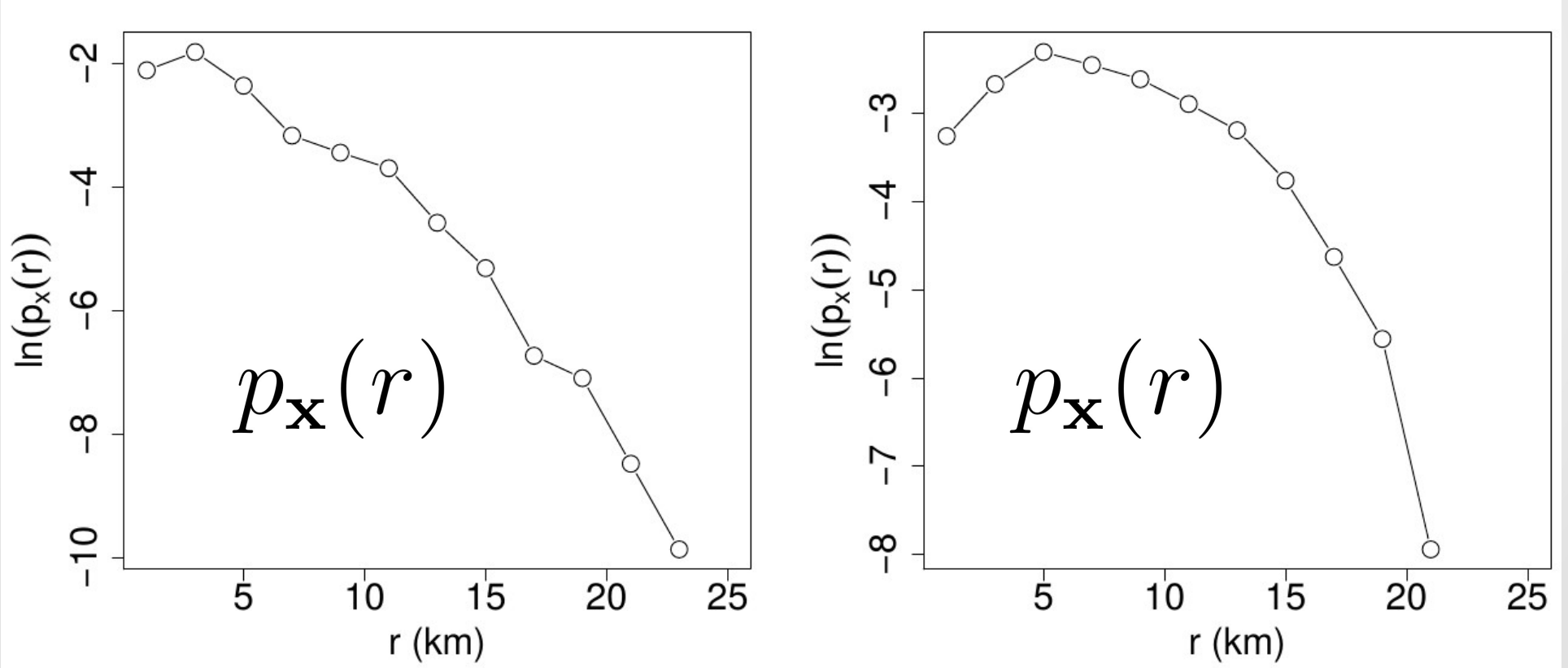
OD matrix: 256^4





O-D modeling

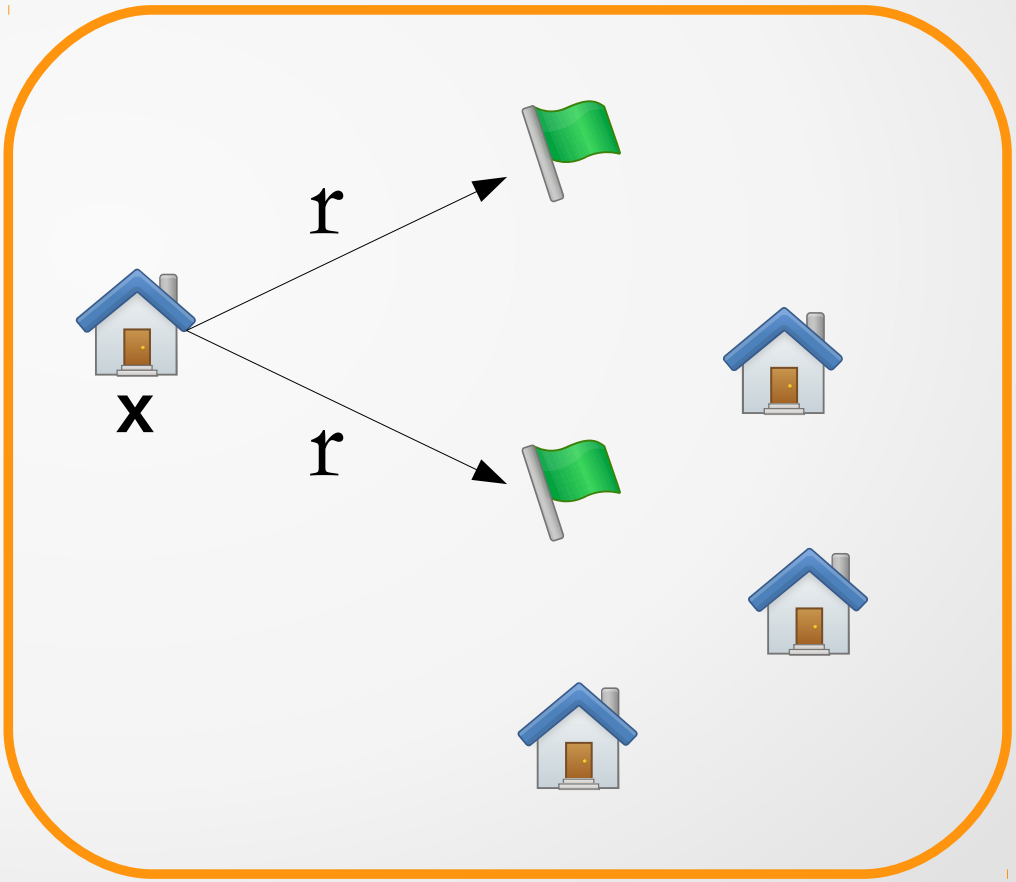
$$p(\mathbf{y}|\mathbf{x}) = p(r|\mathbf{x})p(\mathbf{y}|r,\mathbf{x})$$



Attractivity

$$p(\mathbf{y}|r,\mathbf{x}) = A^r_{\mathbf{x}} q(\mathbf{y})$$

$$q(\mathbf{y}) = p_D(\mathbf{y})$$

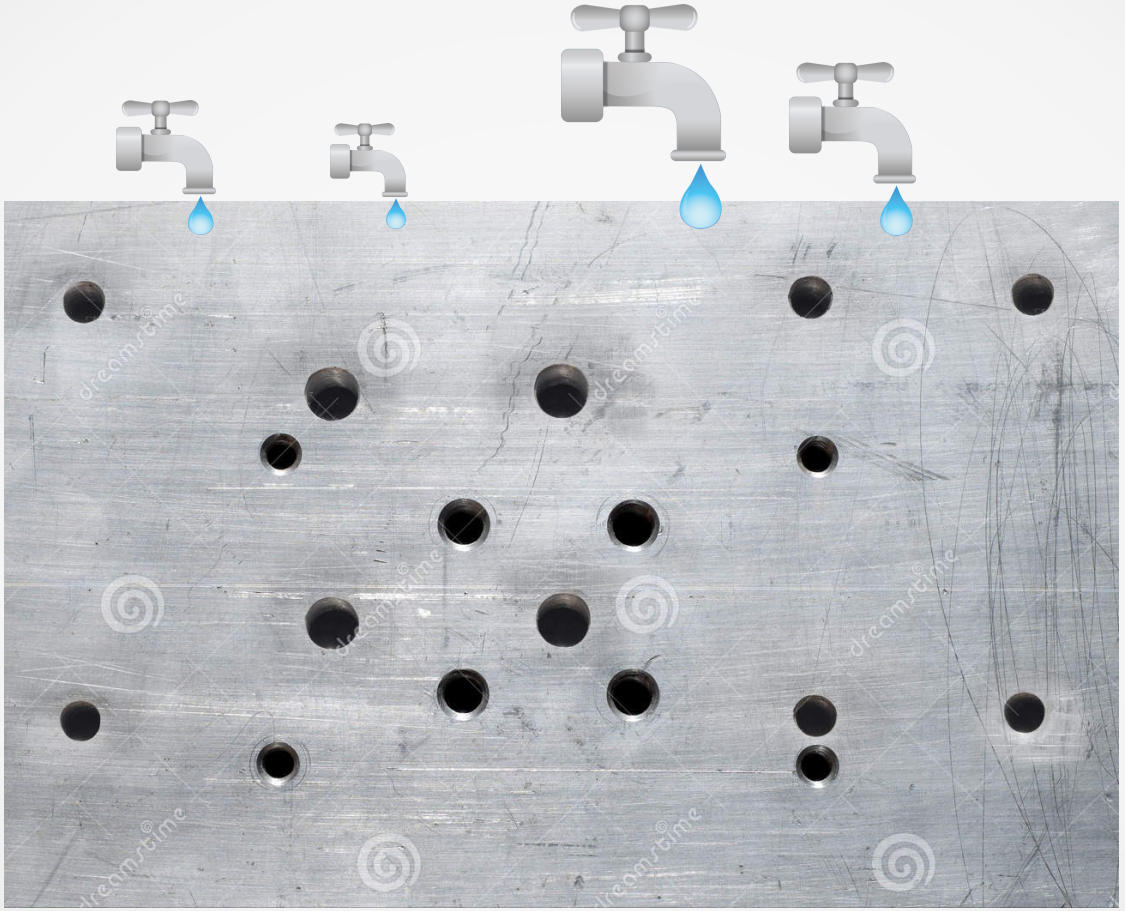


Radiation-absorbtion

$p_O(\mathbf{x})$

$q(\mathbf{y})$

$p_D(\mathbf{y})$



Consistency relation

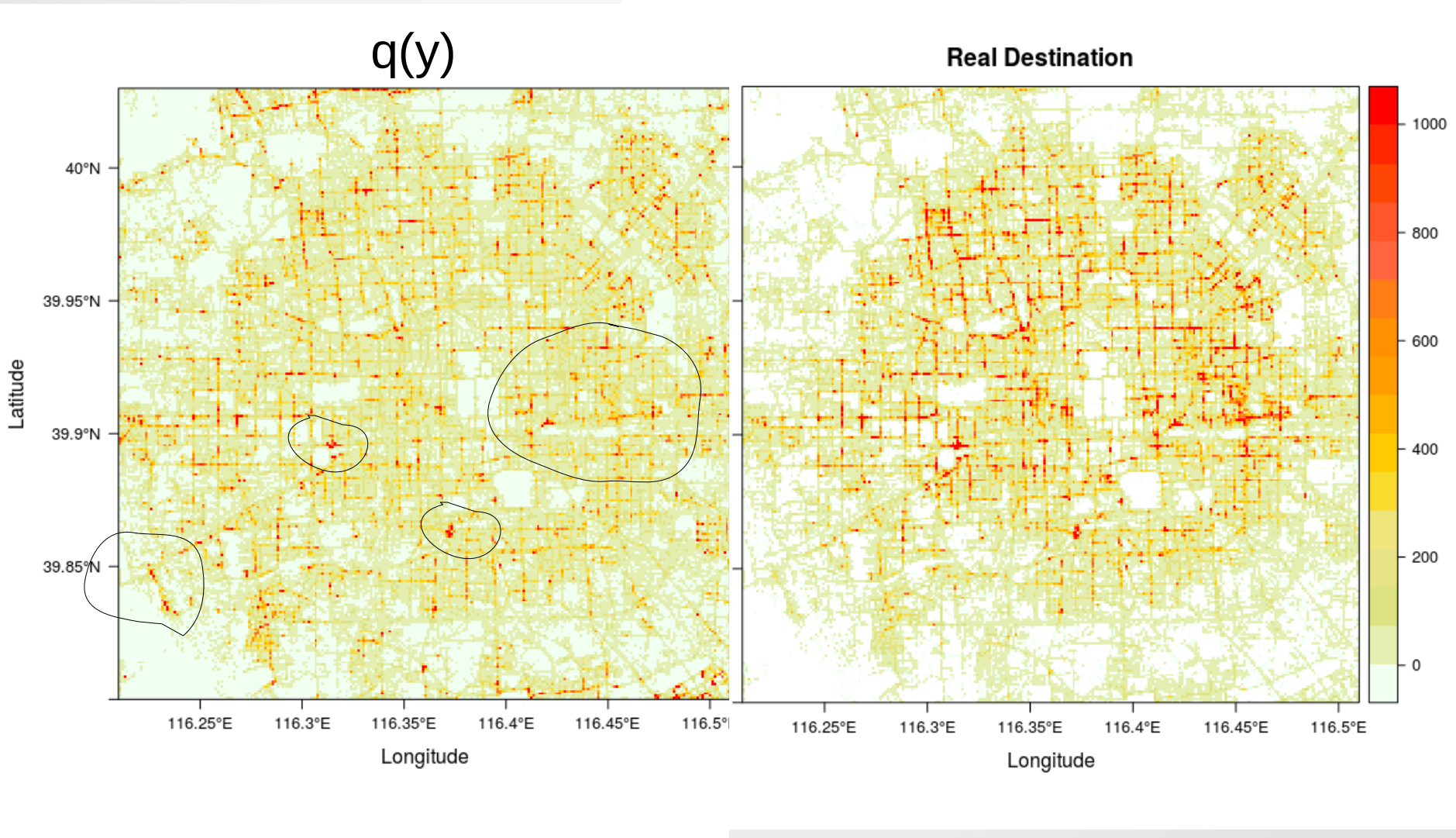
$$p_D(\mathbf{y}) = \int_{\mathbf{x}} p_O(\mathbf{x}) p_{\mathbf{x}}(r) \frac{q(\mathbf{y})}{\int_{\mathbf{y}' \in \Omega^r_{\mathbf{x}}} q(\mathbf{y}') d\mathbf{y}'} d\mathbf{x}$$

Algorithm:

1. Generate **x**;
2. Generate r for this **x**;
3. Generate **y** for this **x** and r;

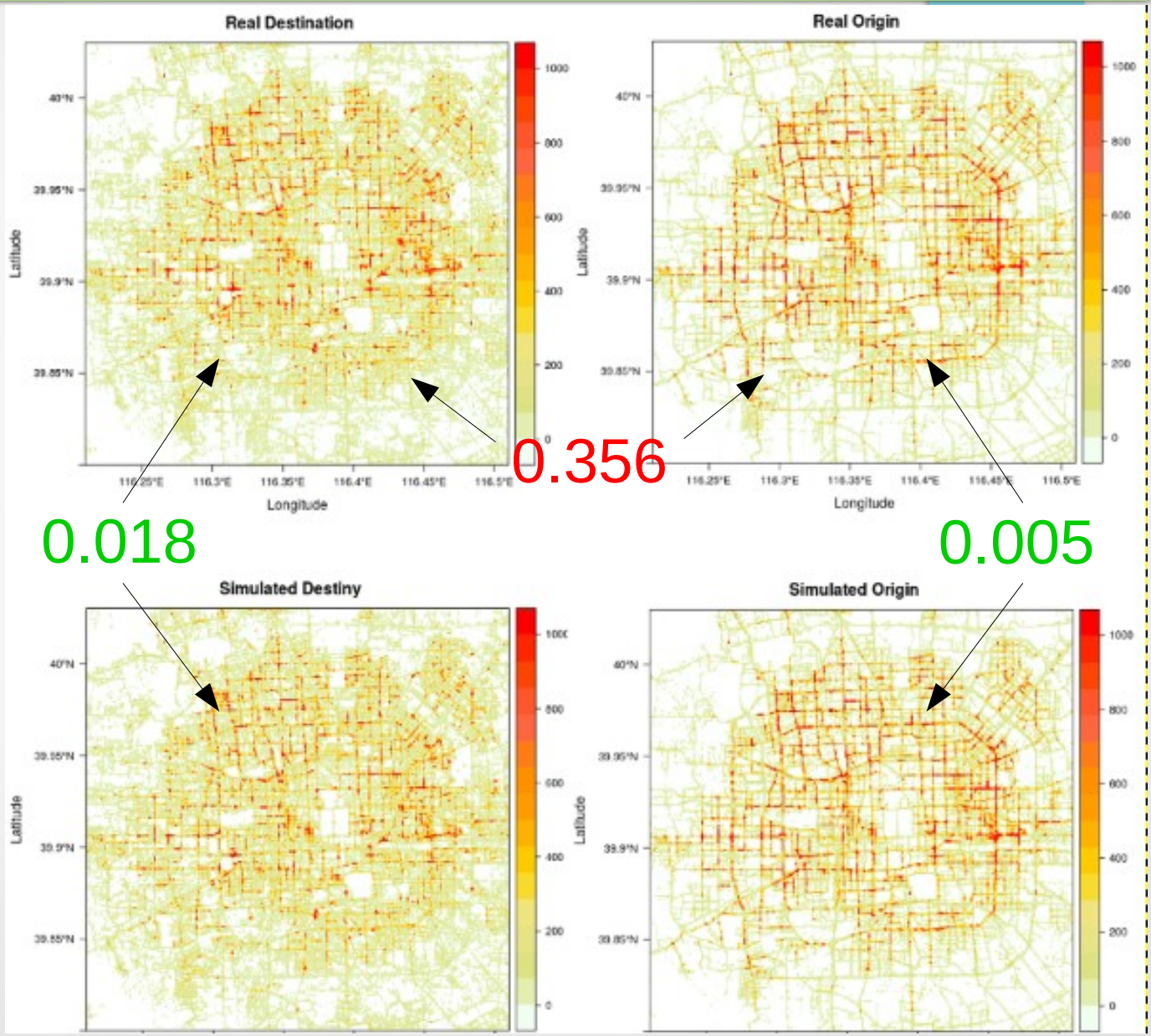
Approximation: $$q(\mathbf{y}) \sim \frac{p_D(\mathbf{y})}{\int_{\mathbf{x}} p_O(\mathbf{x}) p_{\mathbf{x}}(r) \frac{1}{r} d\mathbf{x}}$$

Attractivity

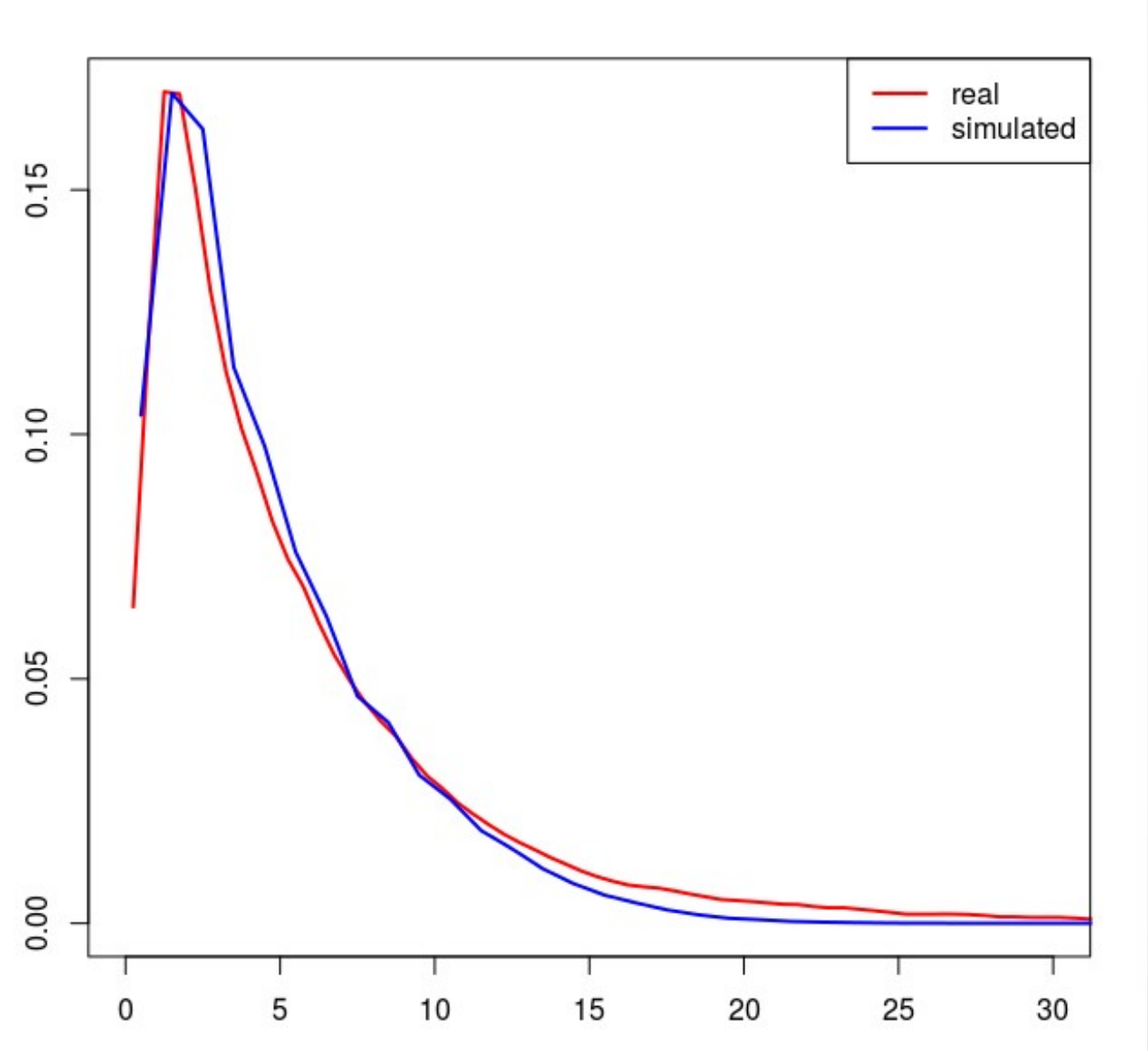


Numerical results

KL distance



Numerical results



Total OD-distance distribution

Summary

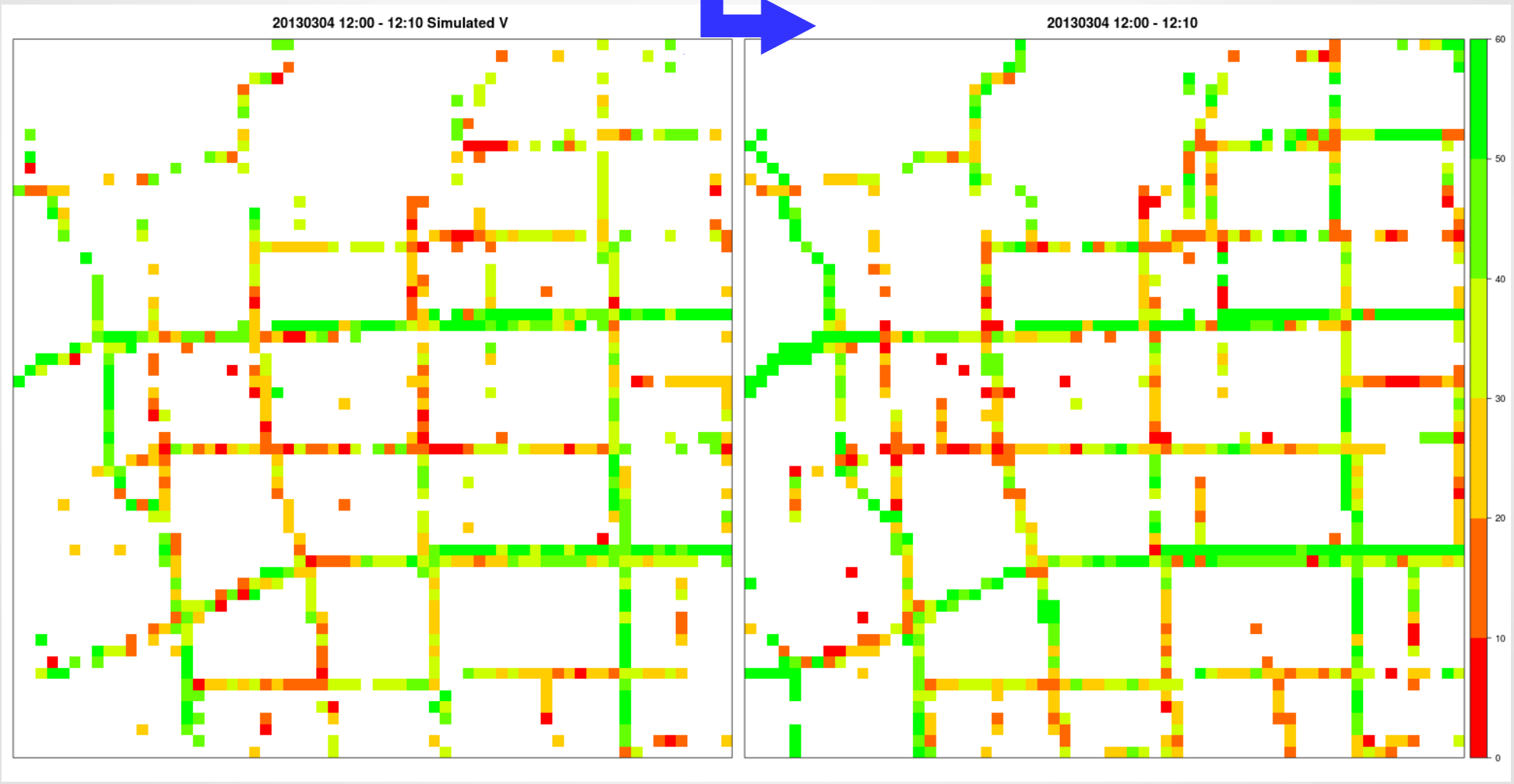
- A model-based OD sampling approach.
- Accounts for spatial heterogeneous.

Acknowledgements

- Peking University: Weinan E (Princeton), Tiejun Li, Xiaowei Wang, Xiaolu Guo, Huizhuo Yuan
- BTIC: Hao Liu
- CAS: Haijun Yu
- PrinceTechs: Haoshu Tian
- BIU: Yong Zhang

Thank You!

Results: 2D small



Simulation **Real data**

Road space rationing


HTTP://DARKKENJIE.DEVIANTART.COM/