CPGIS Educational Webinars on Spatiotemporal Study of Urban Dynamics (1)

Co-sponsors:

International Association of Chinese Professionals in Geographical Information Sciences
NSF Spatiotemporal Innovation Center
Urban Institute, Tsinghua University
School of Geography, Jiangxi Normal University
China Data Institute & Future Data Lab
Journal of Computational Urban Science

"Tracking and modeling diseases in urban spaces"

Speaker: Dr. Xun Shi, Dartmouth College

Chair: Hui Lin, Jiangxi Normal University

9:00 PM-10:00 PM, Thursday, Feb 25, 2021 (US EDT)

Tracking and Modeling Diseases in Urban Spaces

Xun Shi[1] and Meifang Li[1, 2, 3]

1. Dartmouth College
2. Jiangxi Normal University
3. Nanchang Center for Disease Control and Prevention

Overview: GeoSpatial Analysis in Health Studies

On human health/diseases

- On chronic diseases (typical environmental-health studies)
  - To form etiological hypotheses
    - *Characterize spatial distribution of diseases (disease mapping)*
    - *Detecting spatial association between disease and environmental factors*
  - Model environmental exposures
  - Predict disease risk based on the disease-environment models
- On communicable diseases
  - Detect outbreaks
  - Model transmissions
  - Surveillance and prediction
  - Particularly on vector-borne diseases
    - *Detect spatial association between vector/host and environment*
    - *Model habitats of vectors and hosts*

On healthcare service

- Evaluate spatial access to healthcare services
- Evaluate demand to healthcare services
- Evaluate disparity in utilization of healthcare services
- Optimization and planning of healthcare resources

Overview: Urban Spaces

On human

- High population density
- High population mobility
  - Intra-space mobility
  - Inter-space mobility
- On population diversity
  - Demographic diversity
  - Socioeconomic diversity
  - Cultural diversity

On physical environment

- Pollutions (air, water, solid, …)
- Heat island effect
- Complicated land use
  - Habitats of vectors/hosts
  - The built-environment
  - Greenspace
  - Walkability
  - Access to healthcare services
  - Access to healthy food

Overview: Analysis

Spatiality

- Variability
- Autocorrelation
- Similarity/Repetitively
- Scale (aggregation)
- Network
  - Network about people
  - Network about location

Temporality

- Frequency (hourly, daily, seasonally, …)
- Term (temporarily, permanently)
- Scale (aggregation level)

Case studies ….

Land use change and cancer outcome in China

Basic idea

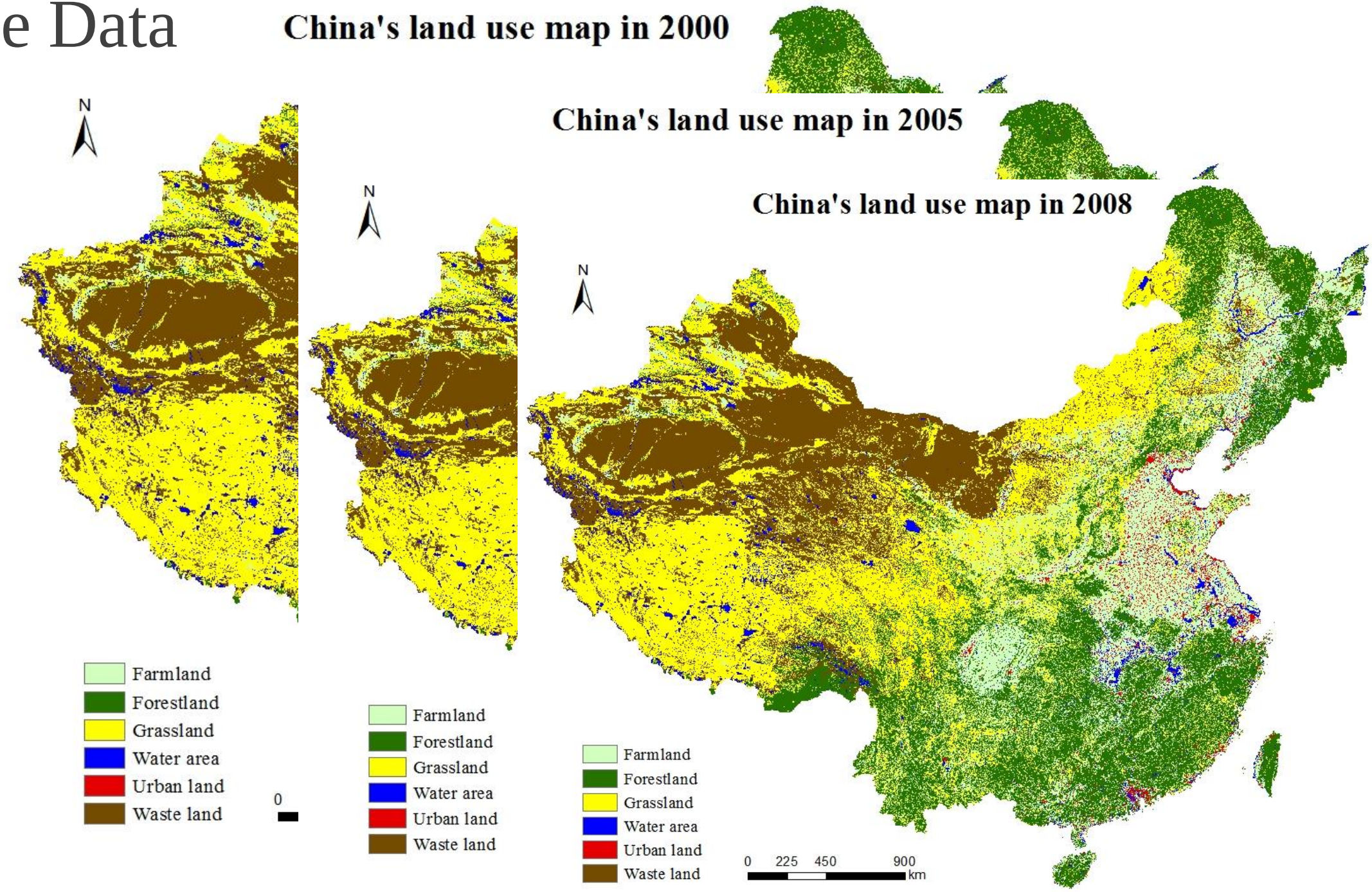
- Land use pattern affects/determines the emission, spread and variation of pollutions.
- Land use affects the built-environment.
- Land use affects access to healthcare services.

Cancer Data

Year	No. of registries	Population (10,000)	No. of provinces and municipalities	No. of cities	No. of counties
2003	35	5603	20		
2004	35/43	7191	20	16	19
2005	34/45	5492	19		
2006	34/49	5957	19	16	18
2007	38	5981	17		
2008	40/56	6614	18	20	21
2009	72/106	8547	24	31	41

Source: Chinese Cancer Registry Annual Reports

Land use Data



Analysis

- Calculate correlation between proportion of 6 land use types and 11 cancers' incidence and mortality.
- Calculate correlation between land use change and 11 cancers' incidence and mortality.
- Consider time lag of land use's effect on cancers.

Findings

1. General incidence/mortality of cancers only positively correlates with *urban area* and *farmland*, may be due to more human activities in these two land cover types.

2. Lung, colorectal, breast, and pancreas cancers, as well as malignant lymphoma, and leukemia positively correlate with the increase of urban area, and negatively correlate with increase of grassland and farmland, possibly caused by increased carcinogen emissions sourced from urban development.

3. Stomach, and esophagus cancers positively correlate with the increase of farmland, and negatively correlate with the increase of urban area, may be due to different dietary styles in rural and urban areas.

4. There is an obvious time lag on the correlation between cancer data and land cover change.

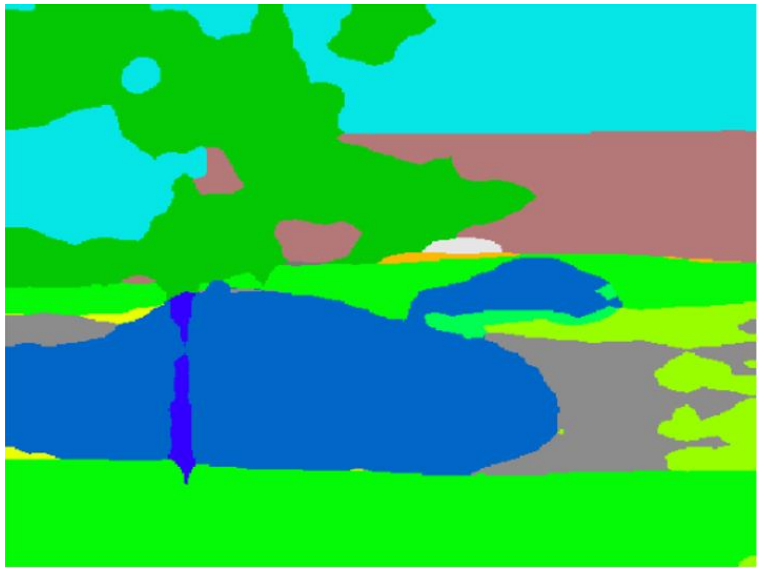
Comparison of StreetView images and satellite images in studies of greenspace

Basic idea

- Most studies of greenspace use measurement derived from satellite images, typically NDVI, to identify and quantify greenspaces, which takes a “bird’s view”.
- StreetView images is closer to a “human’s view”.
- Empirically to compare the two.

Google StreetView image segmentation

Input image



Output segmentation

Area ratios of different types

Provided by Dr. Fang Zhang

Ratio of tree:
0.283

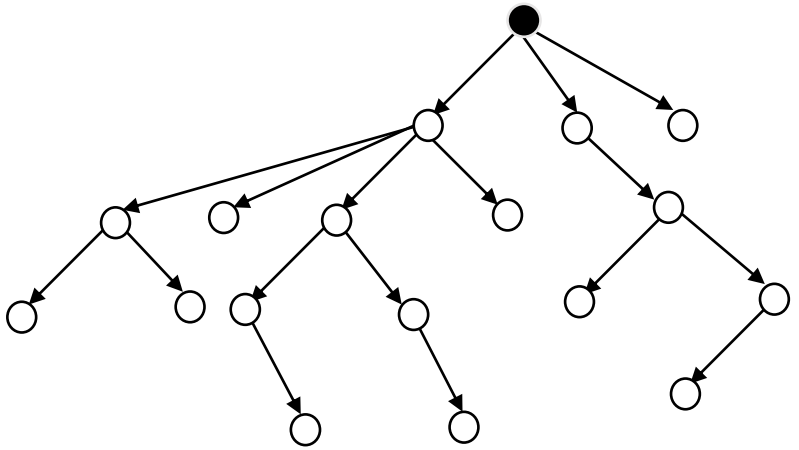
A panorama view at an address constructed from Google StreetView images

NDVI value:
0.153

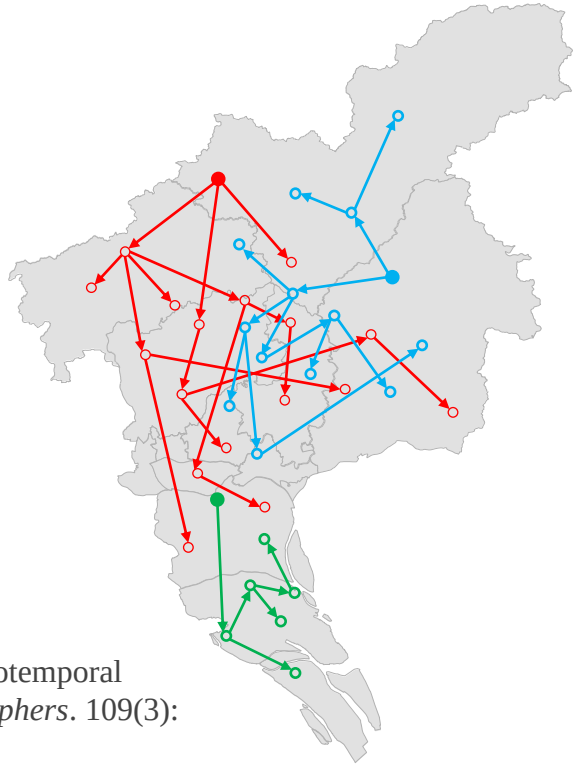
An NDVI image around the same address

Modeling individual-level transmission of dengue fever in Guangzhou

Building Individual-level Transmission Chains

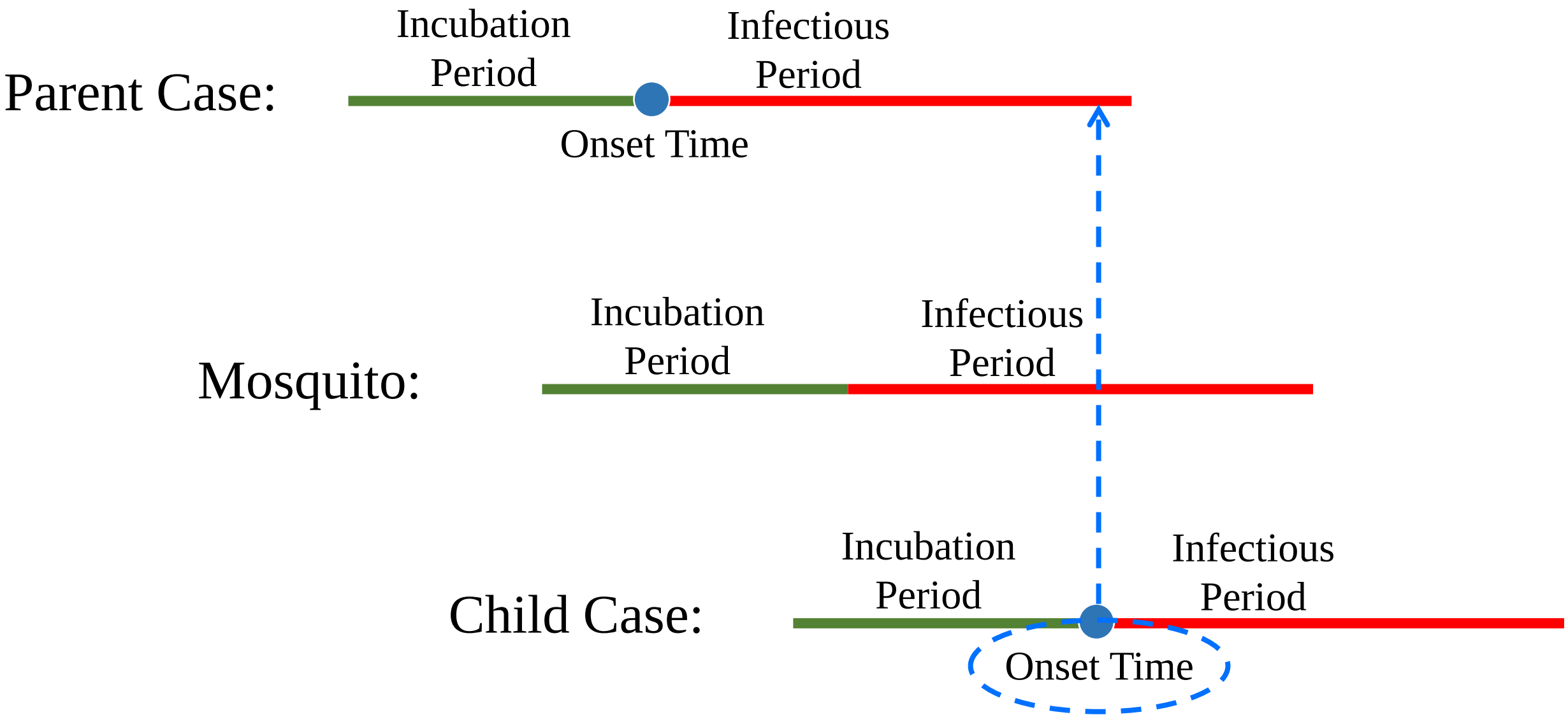


- Constructing the ***Epidemic Forest*** model of the epidemic.
- Deriving characteristics of the epidemic from the *Epi-Forest* model.



Li, M., Shi, X., Li, X., Ma, W., He, J., and Liu, T., 2019, Epidemic Forest: A Spatiotemporal Model of Communicable Diseases. *Annals of the American Association of Geographers*. 109(3): 812-836. DOI: 10.1080/24694452.2018.1511413.

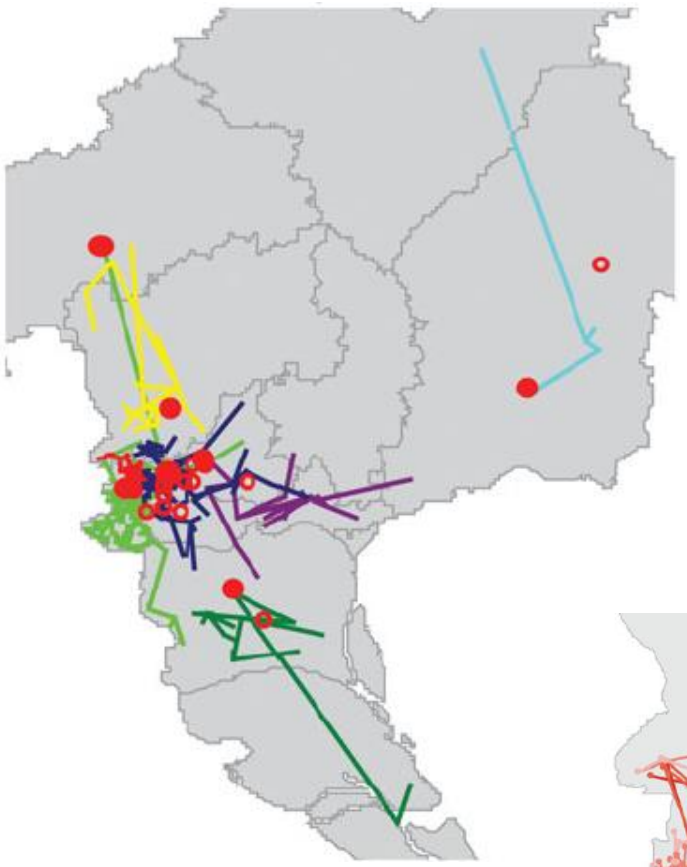
Identifying potential parent cases



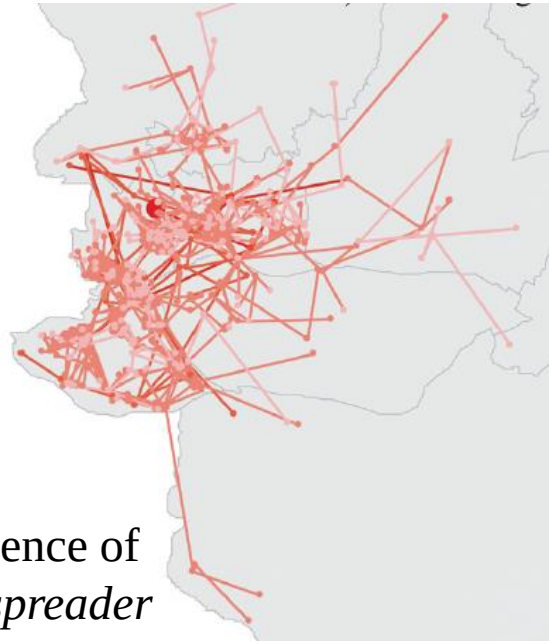
Determine the parent case for a child case

$$D_{s\text{-}t} = (1\text{-}W_t)D_s + W_tD_t$$

- $D_{s\text{-}t}$: the spatiotemporal distance between two cases
- D_s : the spatial distance between the two cases
- D_t : the time difference between the two cases
- W_t : the weight assigned to the time difference

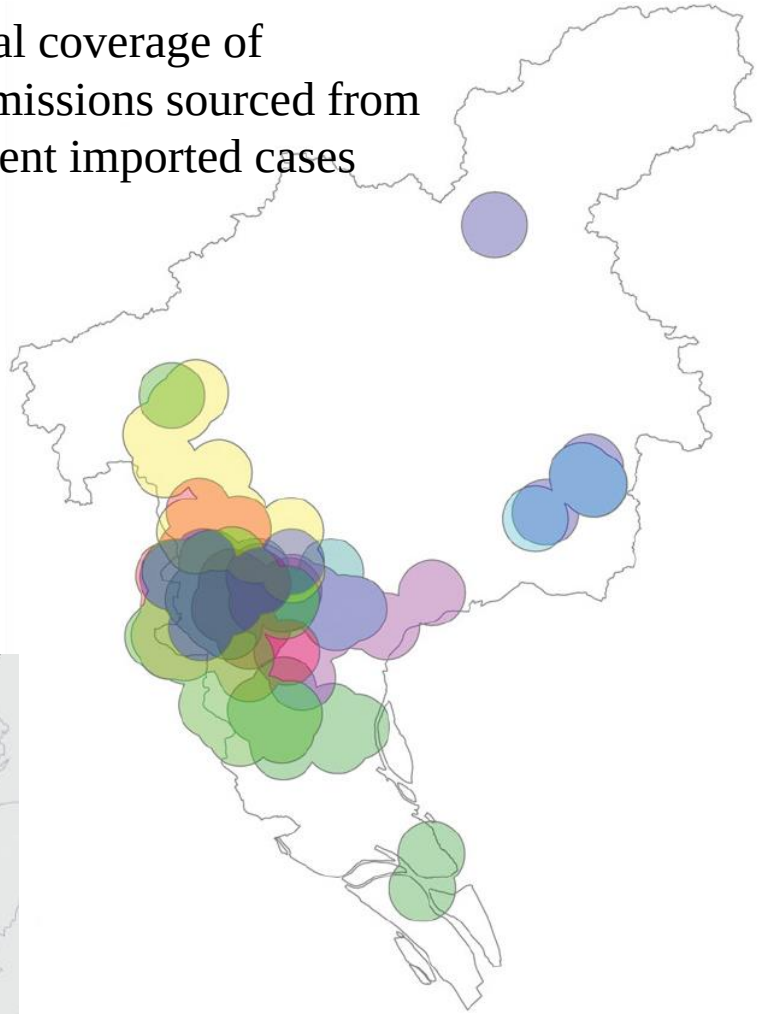


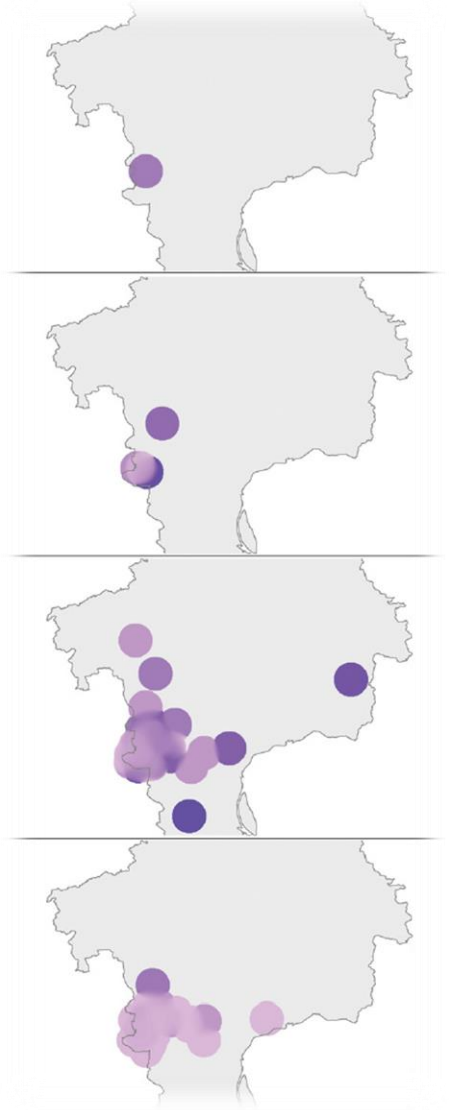
Infection consequences of different imported cases



Consequence of a *super spreader*

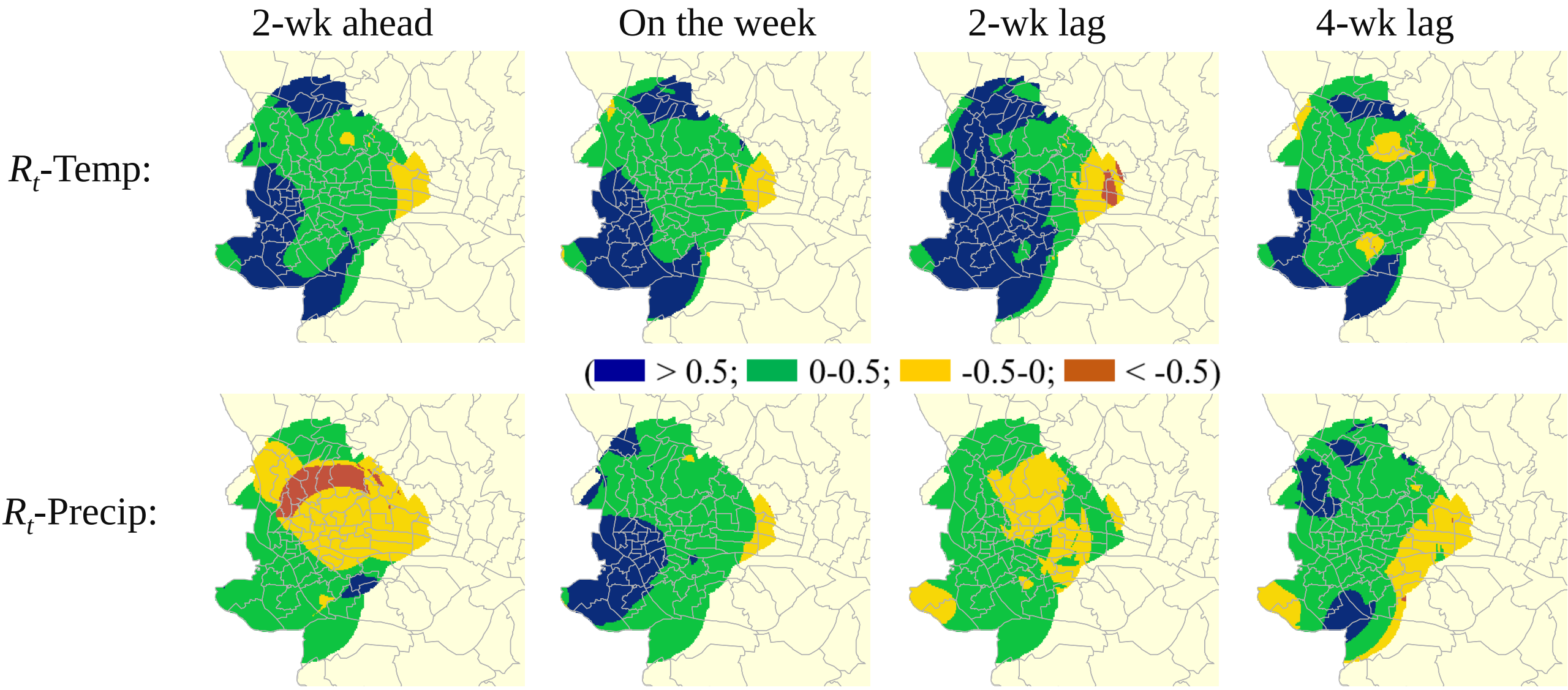
Spatial coverage of transmissions sourced from different imported cases





Local reproduction number（Rt）varying with time

Correlation with climate (pixel-wise analysis)



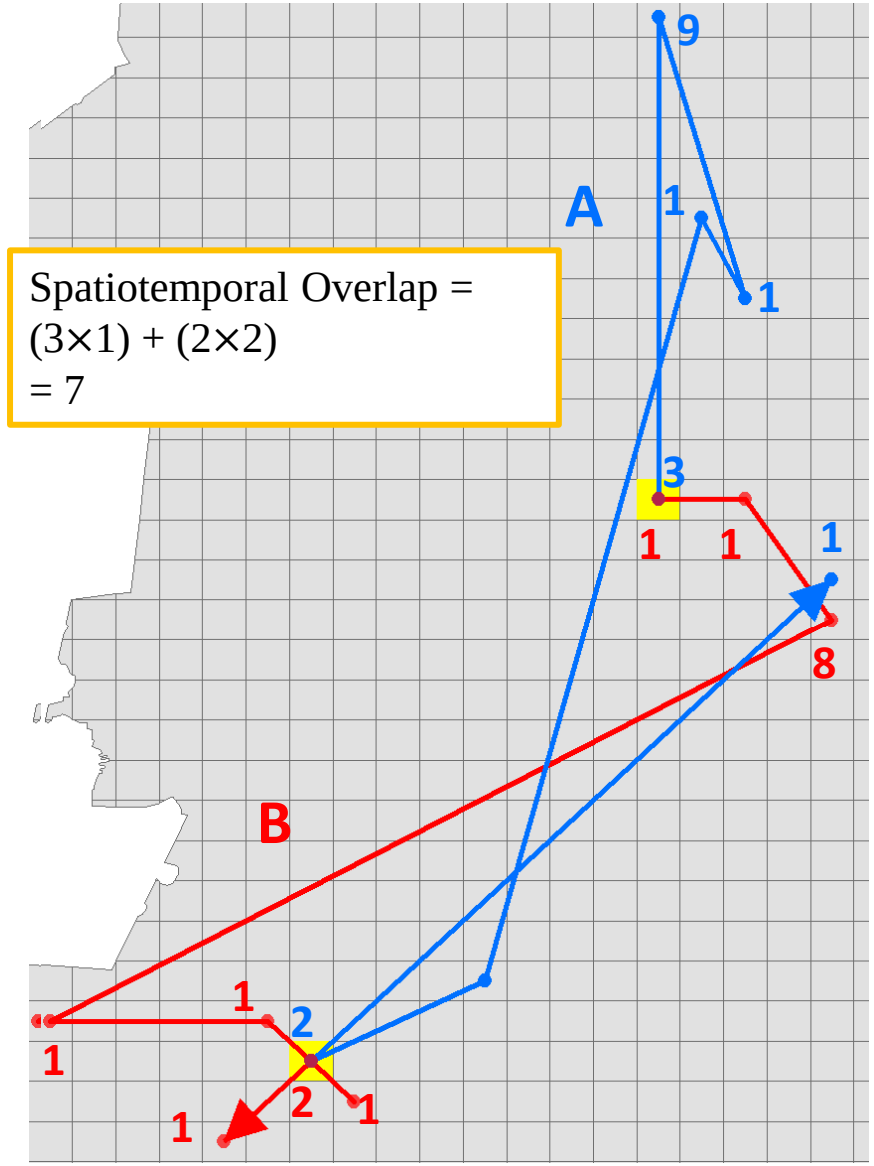
Tracking individual-level transmission of COVID-19 in Nanchang

Basic idea

- A human-human communicable disease, so directly use the overlap of moving trajectories of individuals to estimate the transmission risk, considering the incubation and infectious period.
- Construct individual moving trajectories based on big data.
- Track “close contacts”.
- Build the *Epidemic Forest* model.

Evaluating Individual-level Contacts

- Consider both spatial and temporal overlaps and their buffers.
- Take into account information from epidemiological investigations.

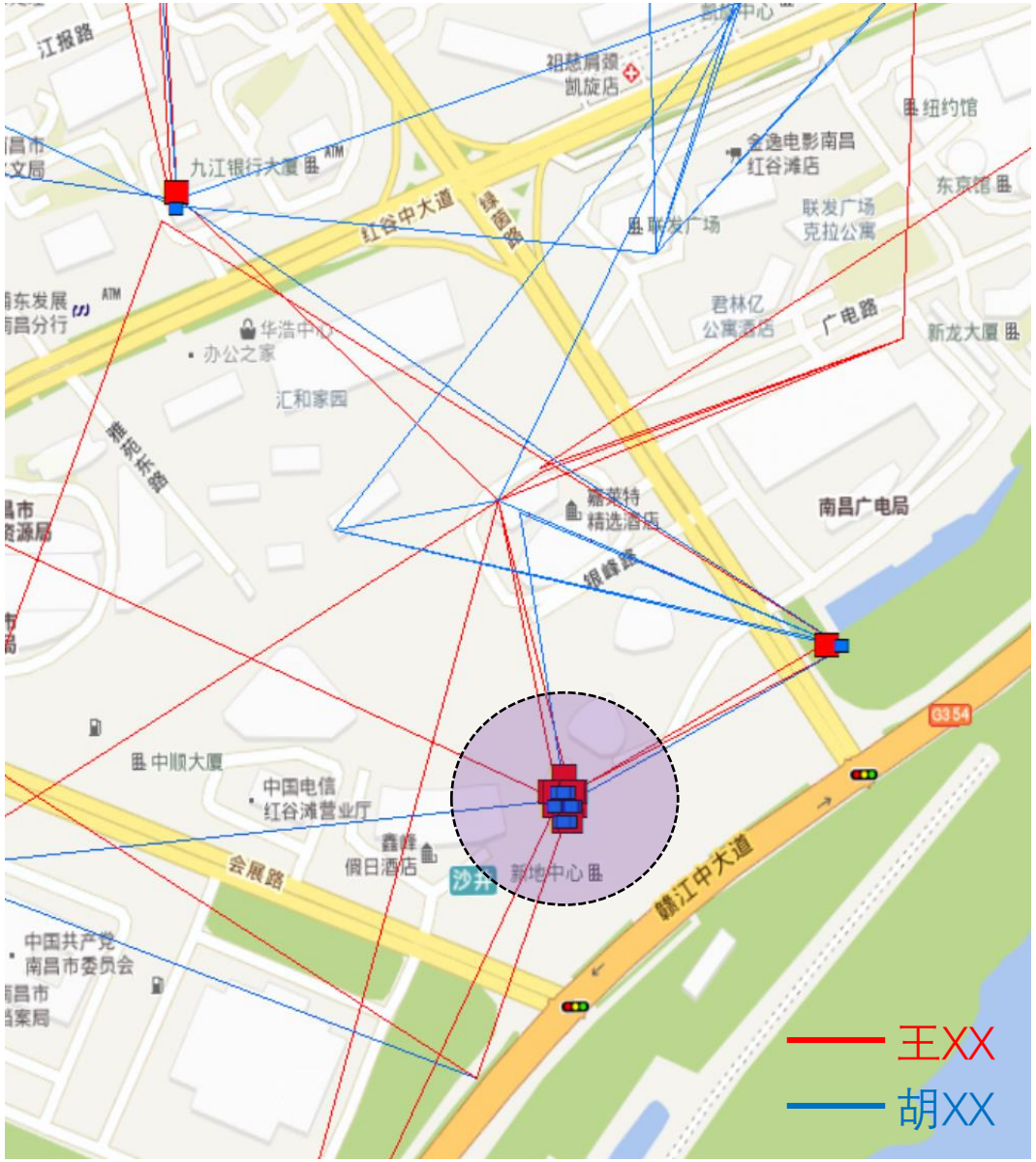


Meifang Li, Yu Liu, Xia Li, Xun Shi, Agent-based Modeling of Communicable Disease Transmissions with Real Human Mobility and Disease Data, paper presented in AAG 2019 Annual Meeting for the SAM Student Paper Competition.

Results of identifying close contacts

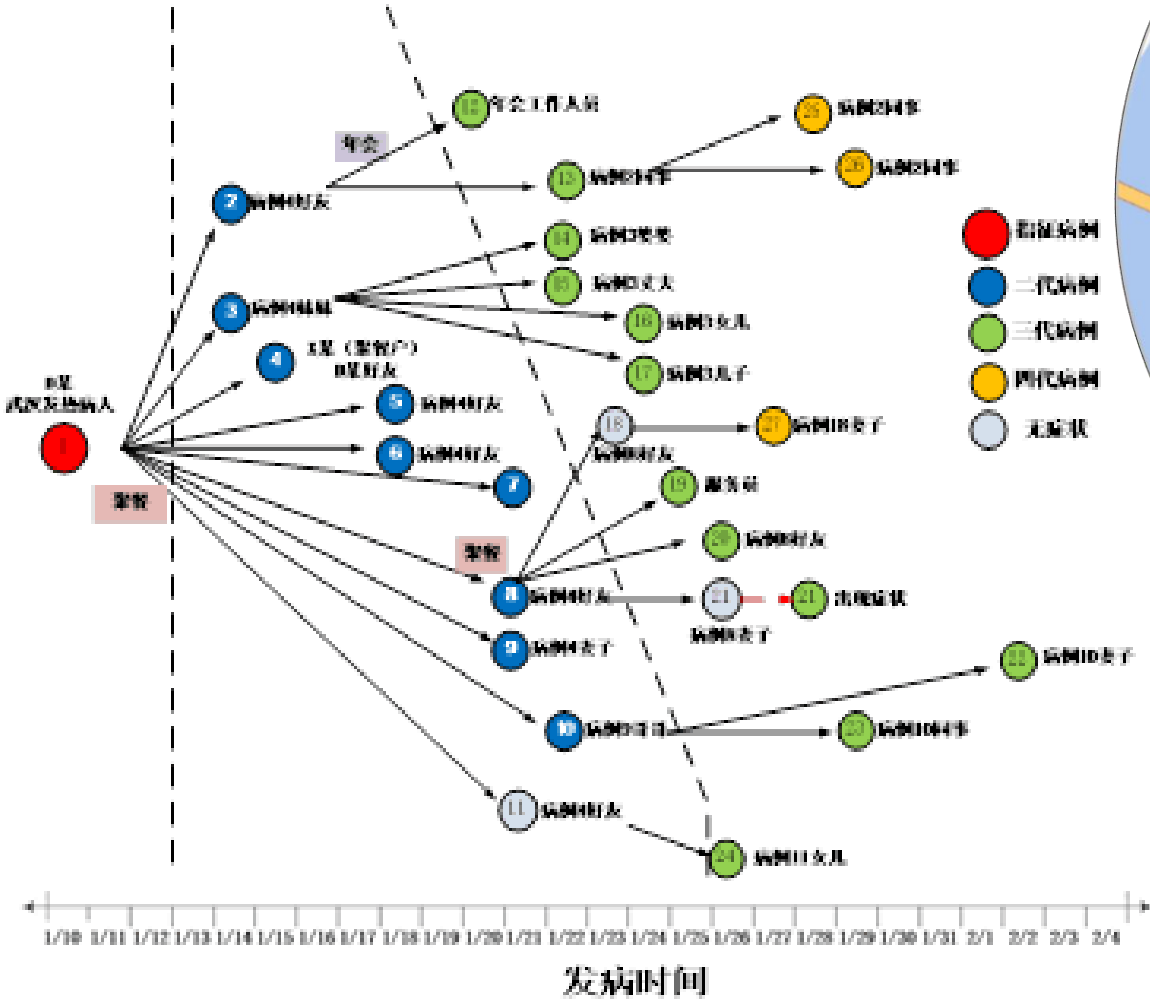
Method	Number of Close-contacts identified	Number of infected in the close-contacts	Percentage of infected in close-contacts
Only by trajectory analysis	1,718	13	0.8%
Only by epi-investigation	878	8	0.9%
By both	1,127	92	8.2%
Total	**3,723**	**113**	**3.0%**

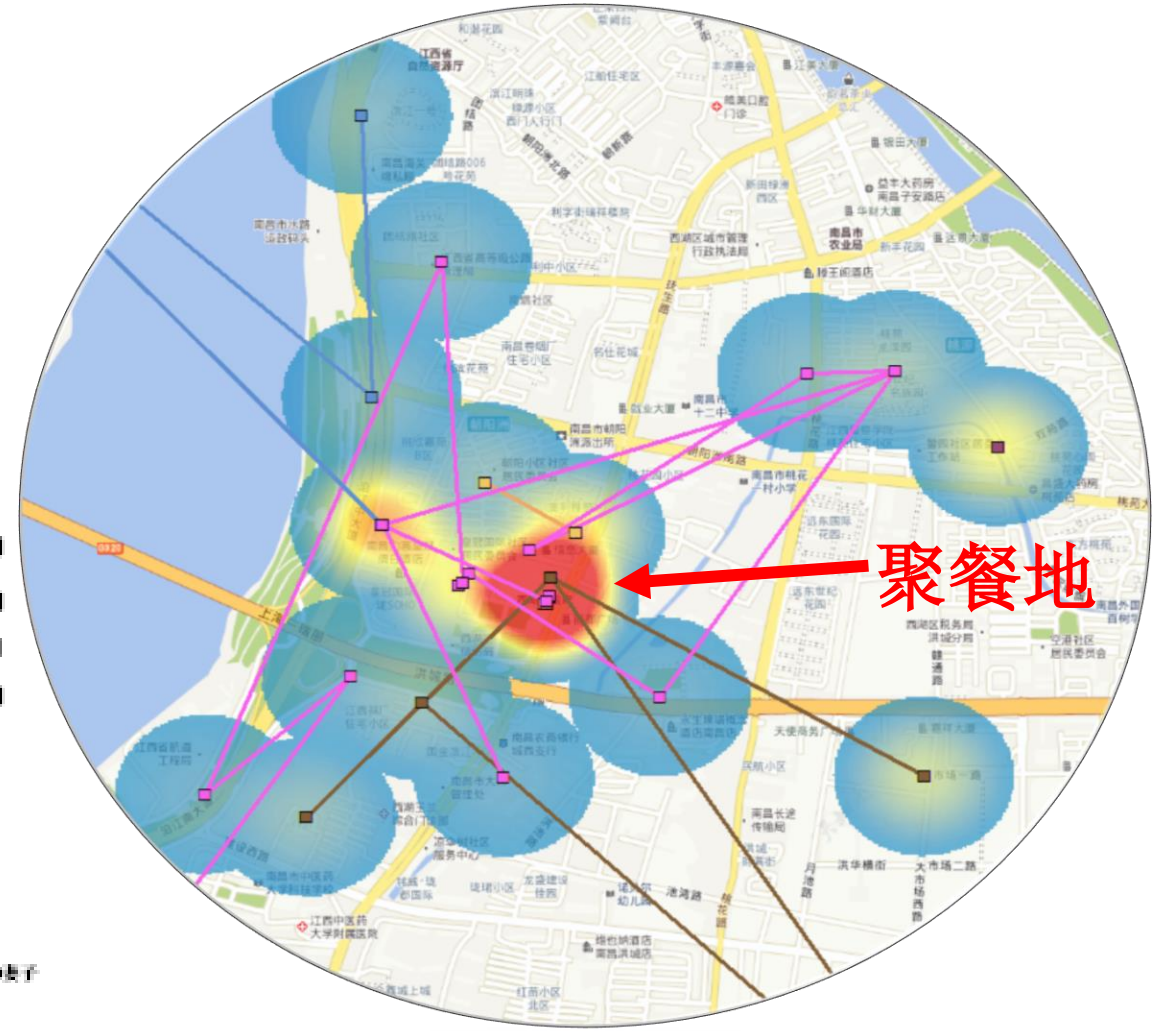
A real example of tracing the infection source of cases



Date	Infected	Infector	Overlap location	Overlap measurement
1月21日	胡XX	王XX	新地中心	470
1月21日	胡XX	王XX	新地中心	7068
1月21日	胡XX	王XX	新地中心	696
1月21日	胡XX	王XX	新地中心	656
1月21日	胡XX	王XX	新地中心	15842
1月21日	胡XX	王XX	新地中心	608
1月21日	胡XX	王XX	新地中心	607
1月21日	胡XX	王XX	新地中心	615
1月21日	胡XX	王XX	新地中心	617
1月21日	胡XX	王XX	新地中心	137
1月21日	胡XX	万XX		4
1月22日	胡XX	王XX	新地中心	10688
1月22日	胡XX	王XX	新地中心	12067
1月22日	胡XX	王XX	新地中心	400
1月22日	胡XX	王XX	新地中心	198
1月22日	胡XX	王XX	新地中心	33
1月23日	胡XX	马XX		74
1月23日	胡XX	王XX	新地中心	295
1月23日	胡XX	王XX	新地中心	5
1月23日	胡XX	王XX	新地中心	8694
1月23日	胡XX	王XX	新地中心	4

A real example of identifying a super spreader





Results of tracing infection source

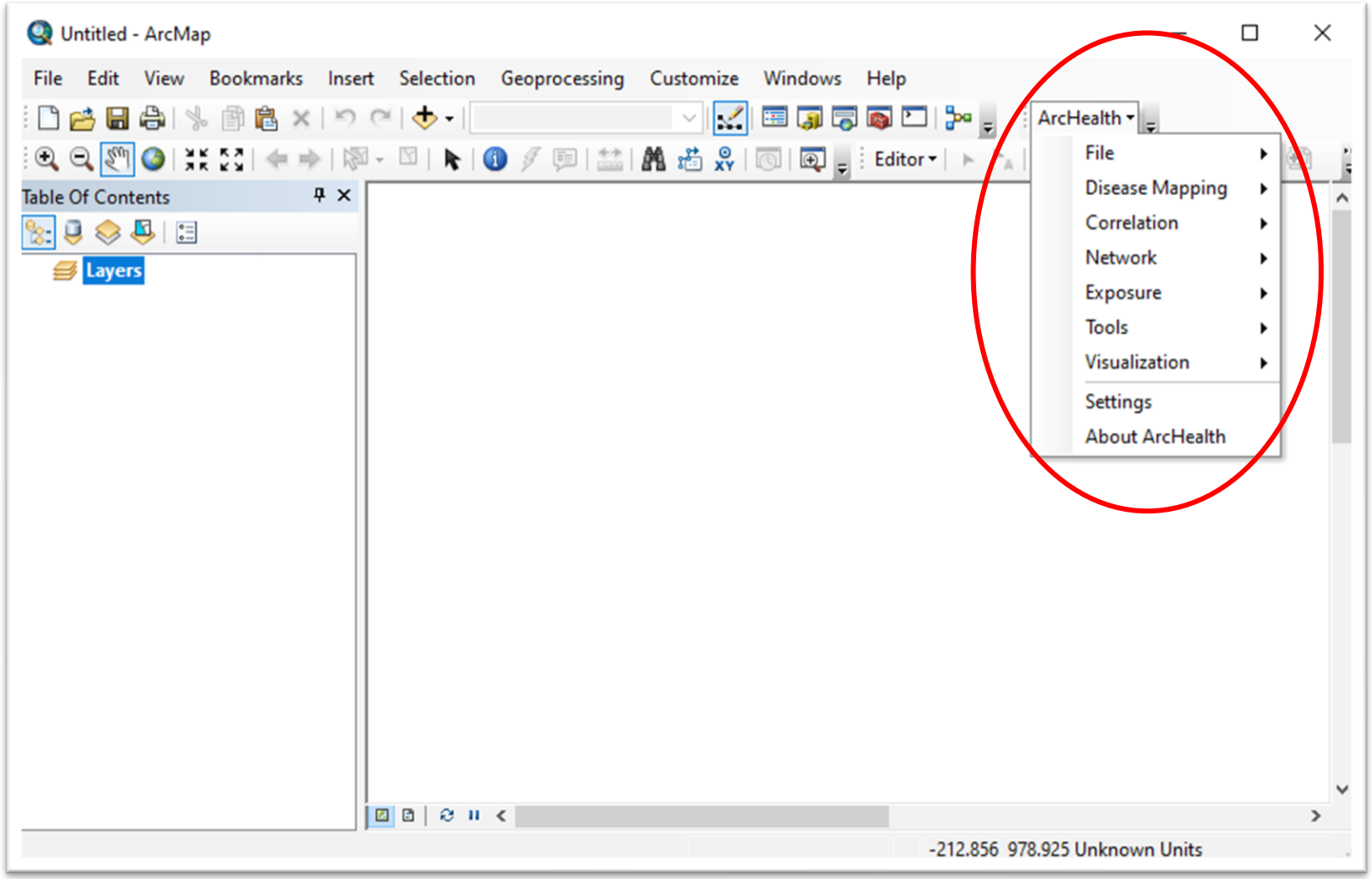
- Up to now, a total of 248 cases have been identified in Nanchang, of which 107 are imported, 141 are local.
- The conventional epidemiological investigation successfully found infection sources for 97 local cases, whereas the trajectory analysis found sources for all 141 local cases.
- An Epi-Forest model has been built for the cases. Epidemiological features have been derived from the model, and can be used for further scenario simulations.

Implementation with *ArcHealth*

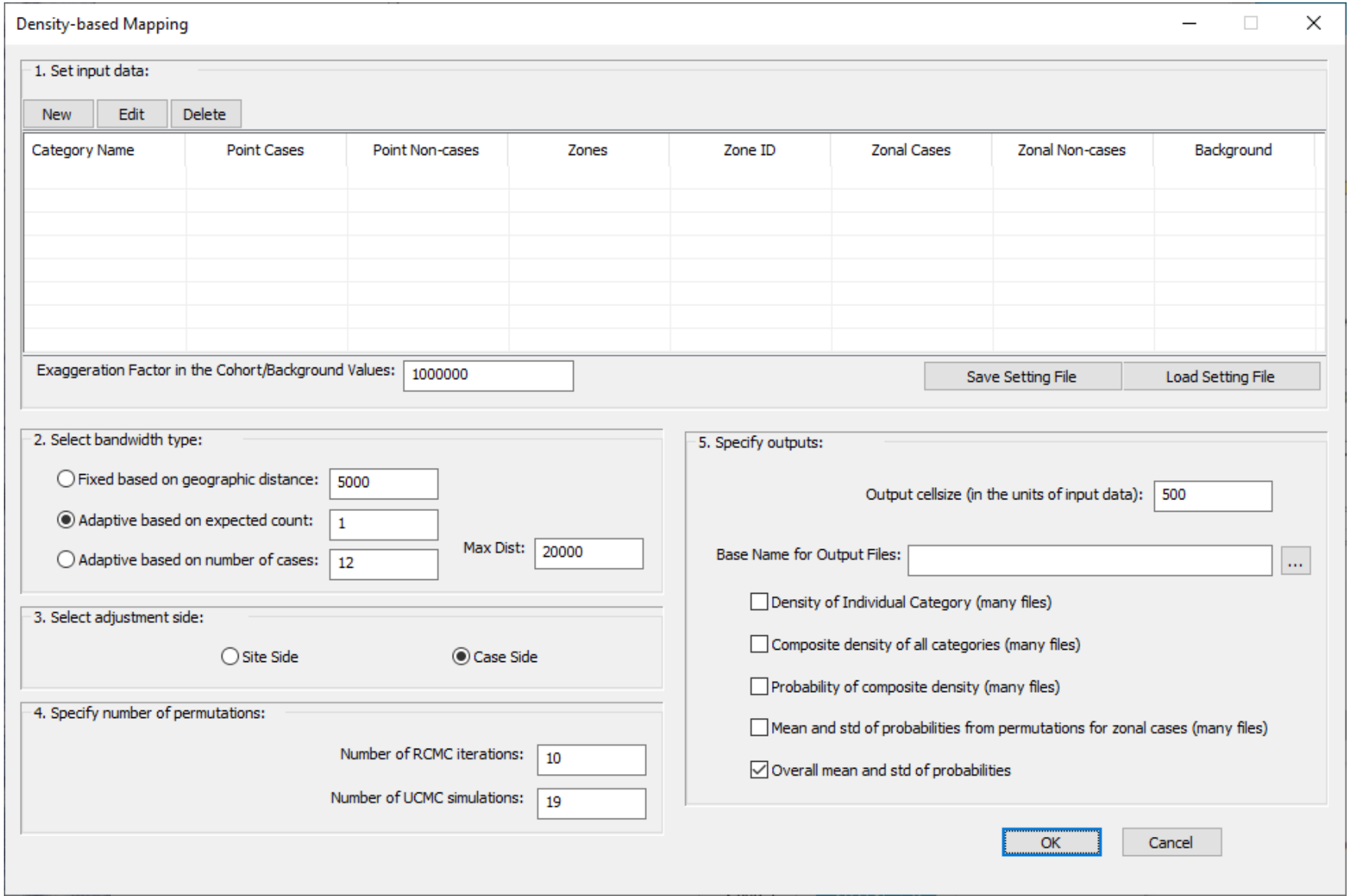
ArcHealth (www.Dartmouth.edu/~xunshi/ArcHealth)

- ArcHealth is a tool package for health-related spatiotemporal analyses.

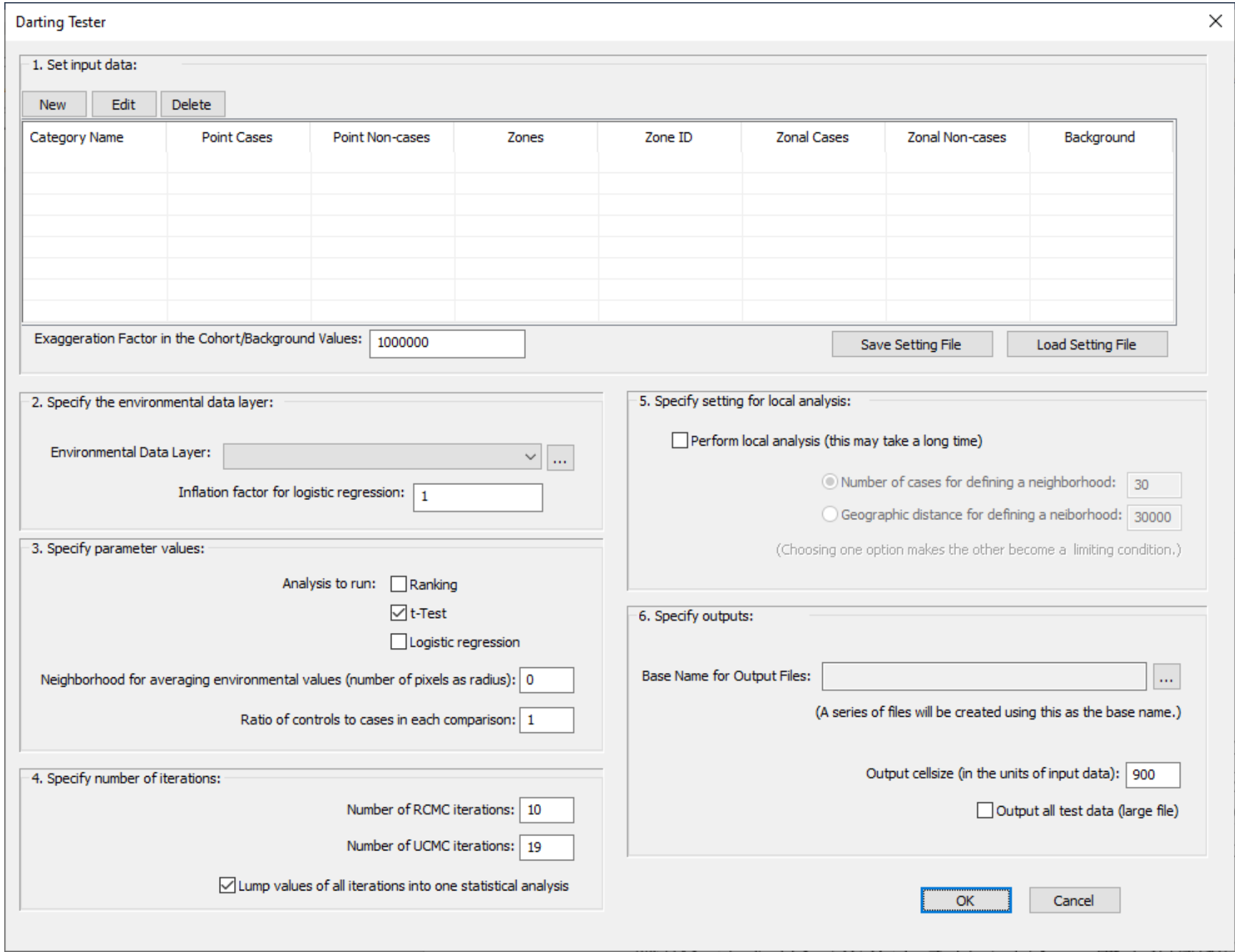
- ArcHealth runs as an Extension of ArcMap; will be a toolbox in ArcGIS Pro, and will be migrated to HPC.



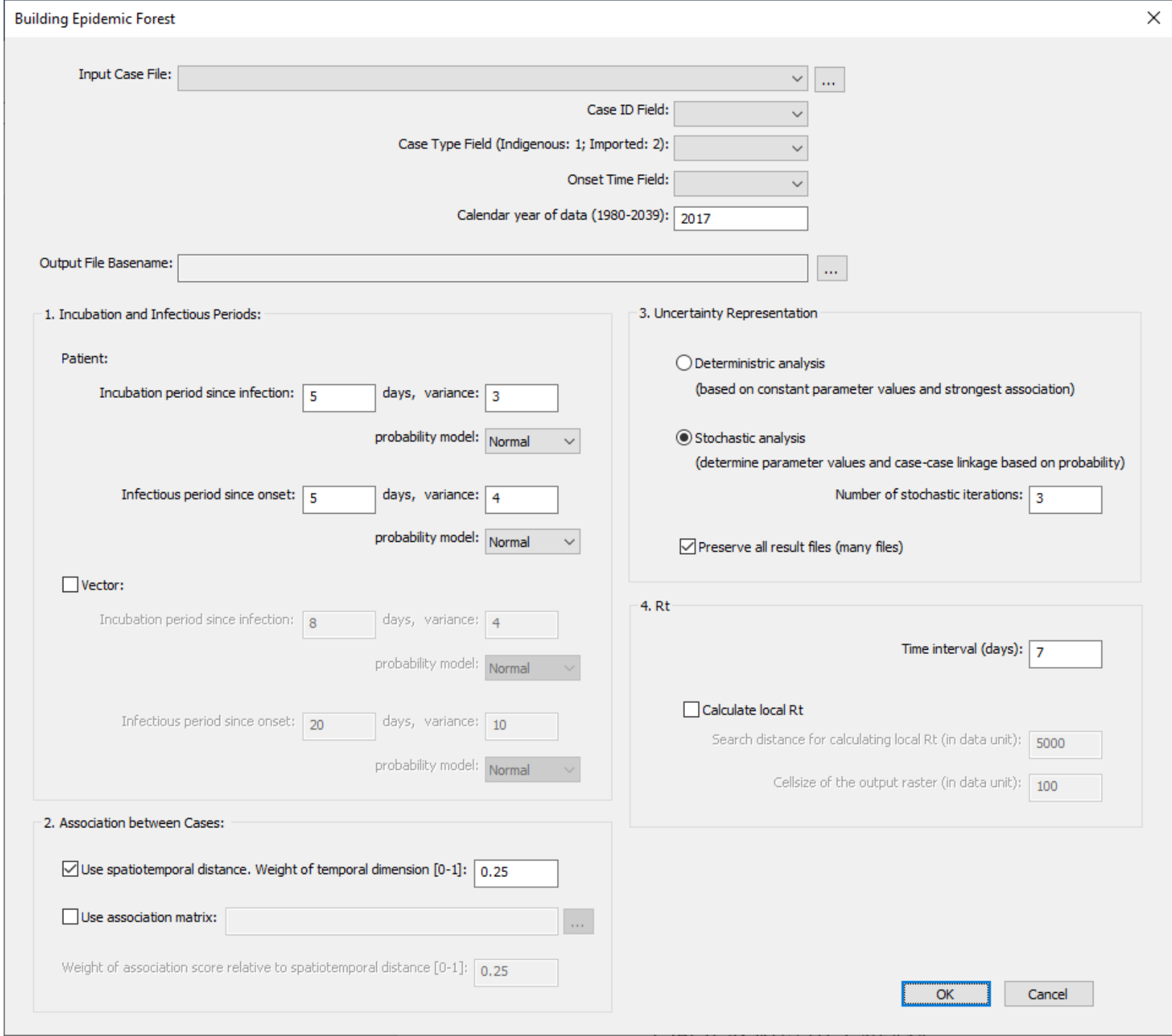
ArcHealth's tool for density-based disease mapping (cluster detection)



ArcHealth's tool for disease-environment association detection



ArcHealth tool for *Epidemic Forest* modeling

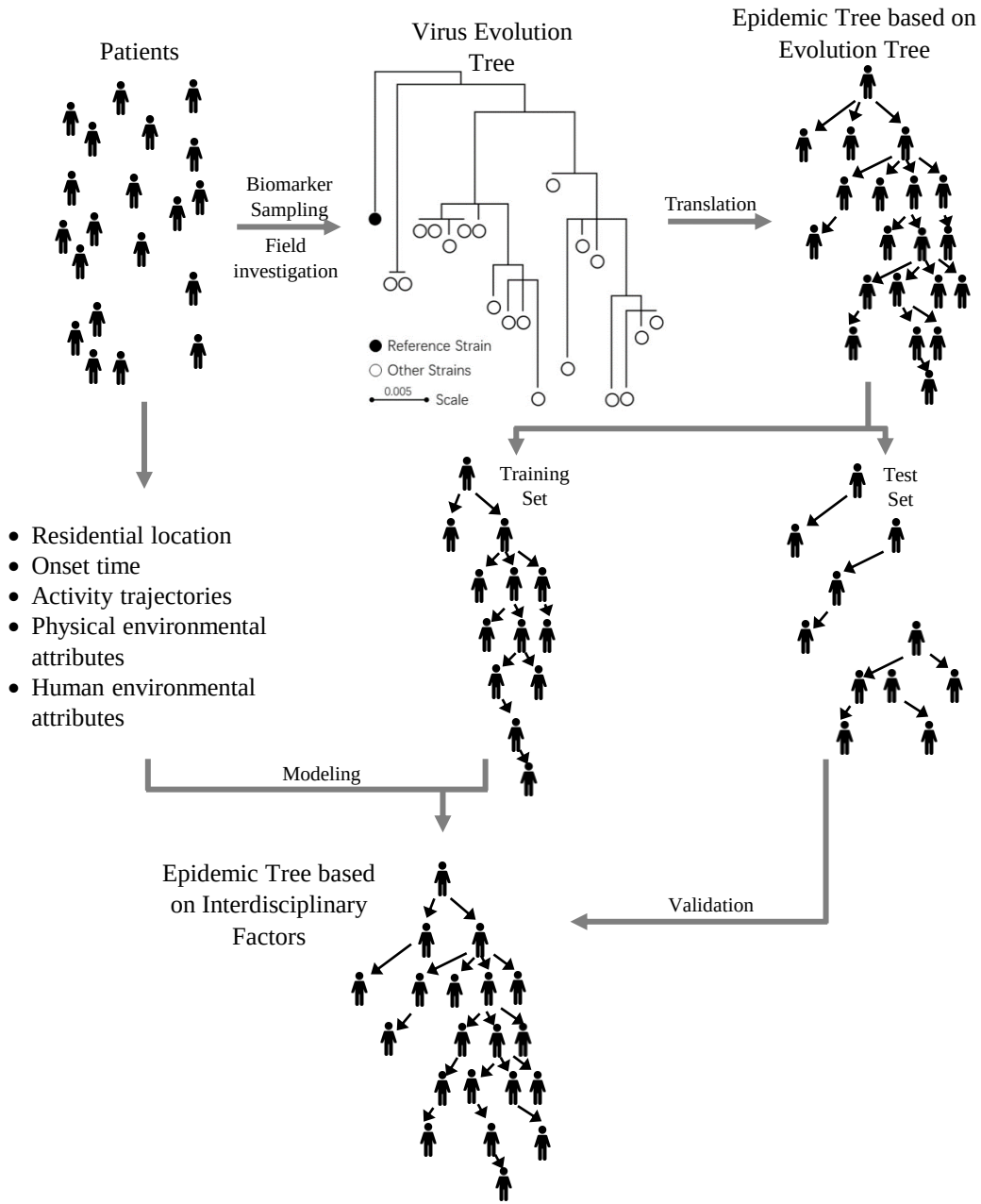


Concluding remarks

- Urban spaces have the most intensive human health problems and also have the most detailed data to study those problems.
- Land use is a representation and a result of human-environment interaction, and therefore is highly related to human health, but it is far from fully studied from the health perspective.
- Spatialization and individualization is a future of epidemic modeling, which leads to and facilitates the bottom-up approach. The bottom-up approach is particularly suitable for the within-city epidemic modeling.
- The restriction to access to individual-level data, as well as the quality issues of the big data, are bottlenecks in the bottom-up approach.

Integration of *small*, *medium*, and *big* data for spatial epidemiology

- Small data: data with small sample size, e.g., genetic and biomarker data.
- Medium data: data with medium sample/population size, e.g., spatial, temporal, and demographic data of individual patients.
- Big data: disaggregated data of a large sample or entire population, e.g., moving trajectory data of population in a city.



Thank you!

Questions and comments?