Deep learning and process understanding for data-driven urban prediction

Feng Zhang, Hongye Zhou, Xuying Yang

Department of Geography, Zhejiang University

Institute for Geography and Spacial Information

What is Urban?

城

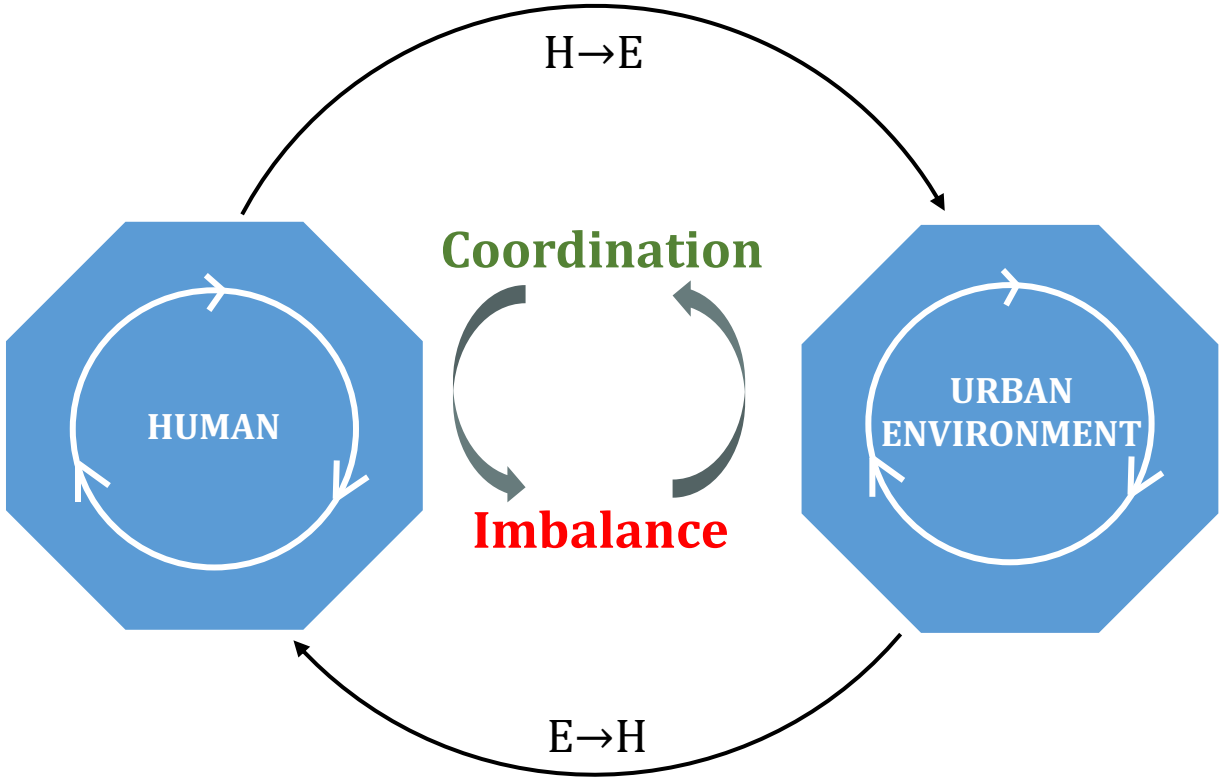
Physical environment

市

Human activities

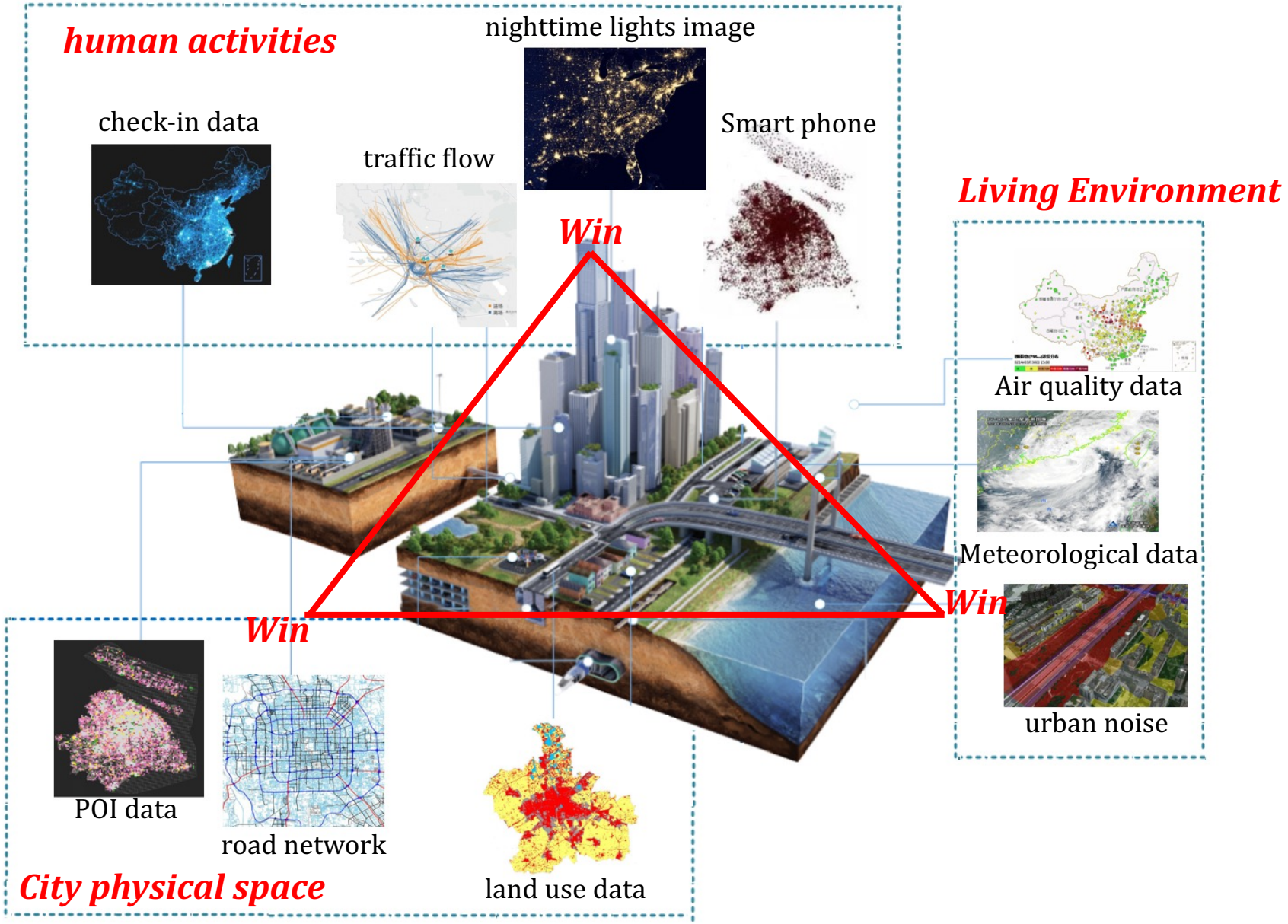
燕寻杭城|杭州古地图变迁
www.zcool.com.cn

Urban Dynamic system

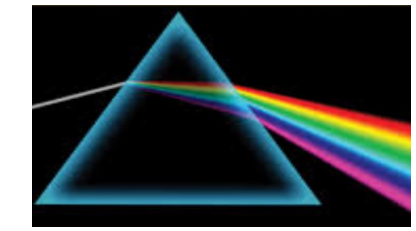


- Conditions for survival and development
- Environment control human action and activities
- Environment change human attitude and decision making

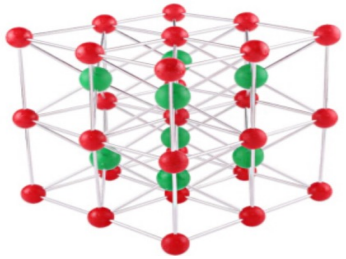
Urban Big Data

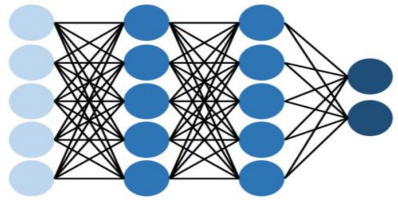


Data-driven: the fourth paradigm of scientific research



$$e^{i\pi} + 1 = 0$$
$$\mathbf{F} = m\mathbf{a}$$
$$E_0 = mc^2$$
$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r}, t) = \hat{H}\Psi(\mathbf{r}, t)$$





1st PARADIGM
Empirical Science

Experiments

2nd PARADIGM
Theoretical Science

Laws of classical mechanics, electrodynamics, etc.

3rd PARADIGM
Computational Science Simulations

Density-functional theory and beyond; molecular dynamics

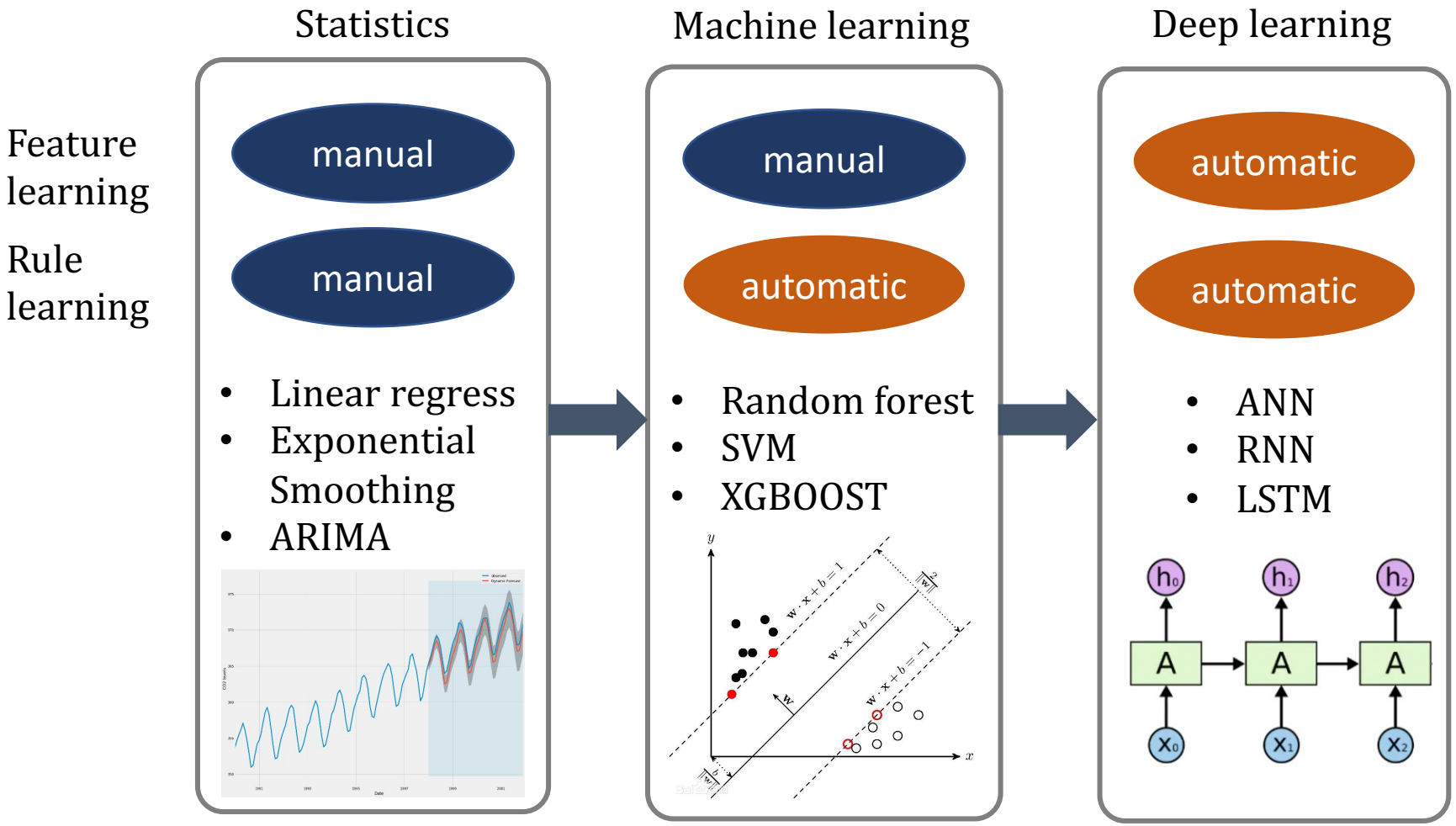
4th PARADIGM
Big-Data-Driven Science

Detection of patterns and anomalies in Big Data; artificial intelligence; etc.

1600 1950 2010

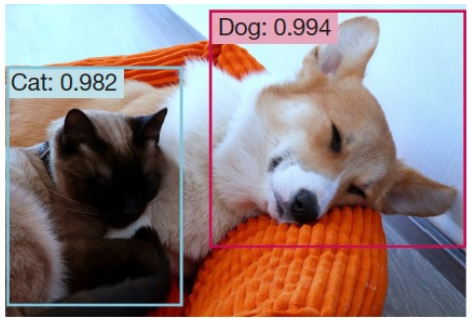
Tansley S, Tolle K M. The fourth paradigm: data-intensive scientific discovery[M]. Redmond, WA: Microsoft research, 2009.

Data-driven models for geographic research

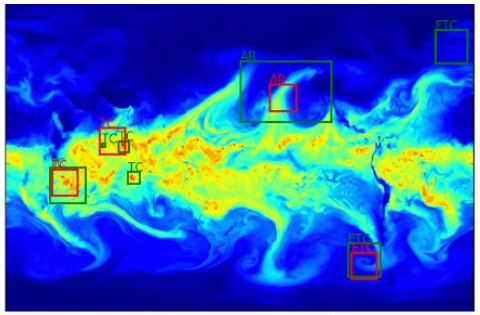


Deep learning in Geoscience

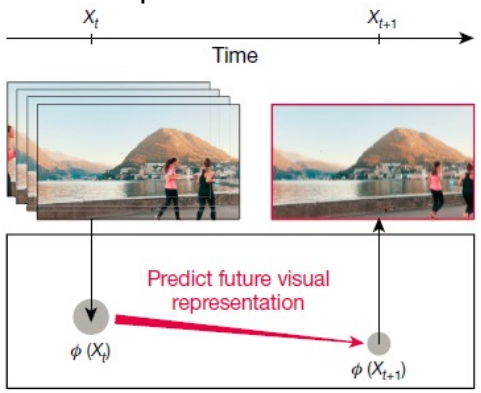
Object classification and localization



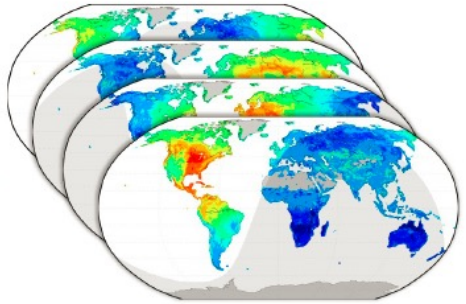
Pattern classification

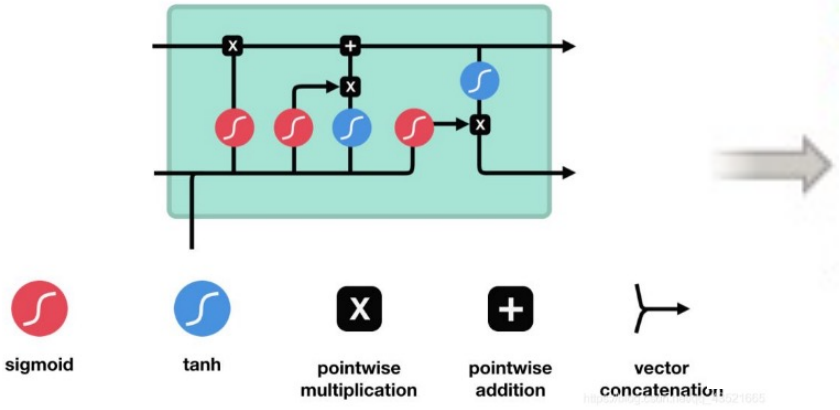


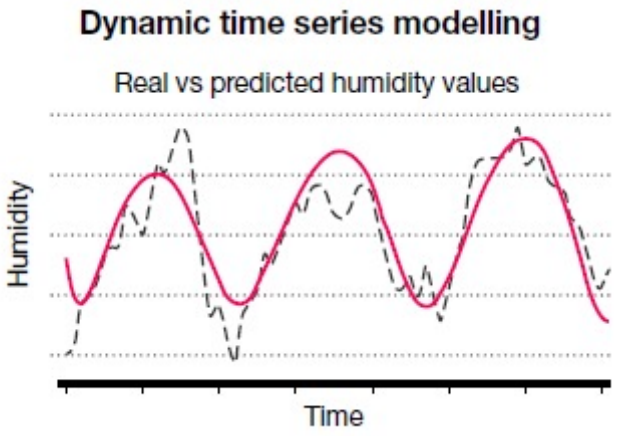
Video prediction

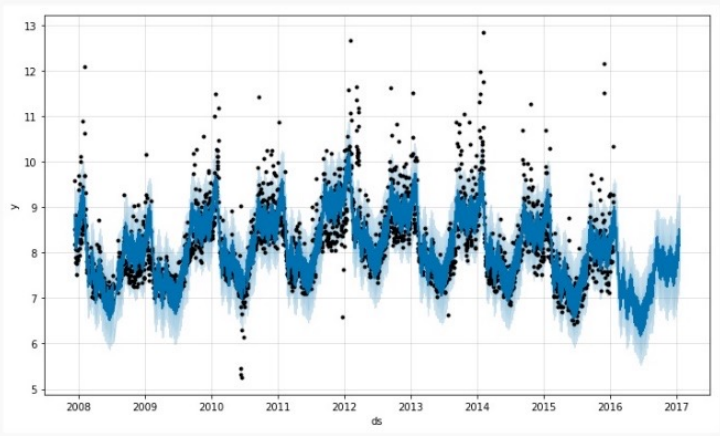


Short-term forecasting







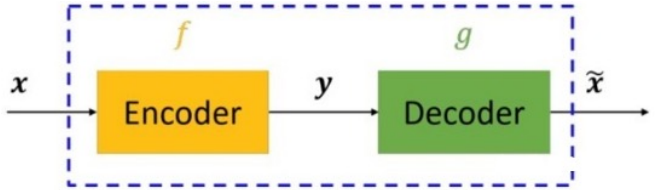


Reichstein, M., Camps-Valls, G., Stevens, B. *et al.* Deep learning and process understanding for data-driven Earth system science. *Nature* **566,** 195–204 (2019).

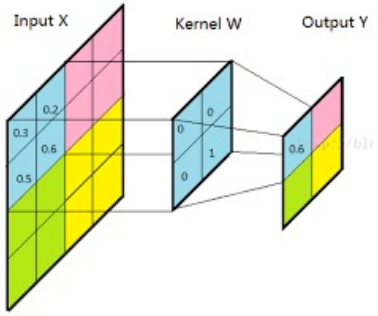
Urban prediction helps to

- improve urban planning
- easing traffic congestion
- controlling of environmental pollution
- reducing energy consumption
- …….

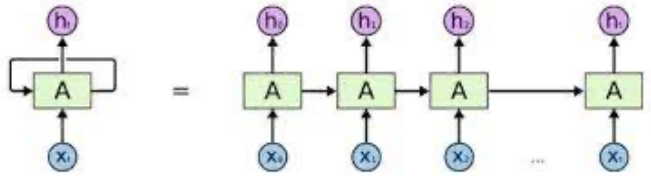
Deep learning models for prediction



Auto Encoder (AE)

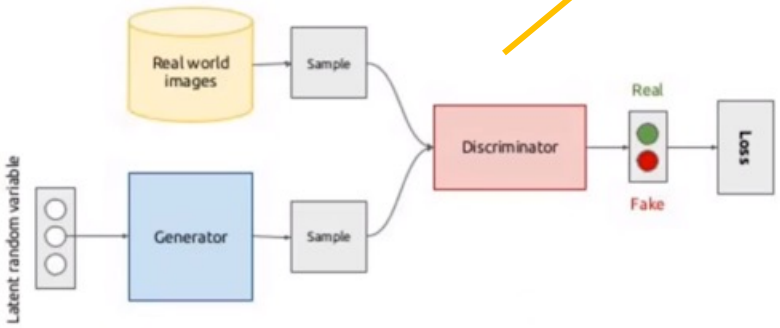


Convolutional Neural Network (CNN)

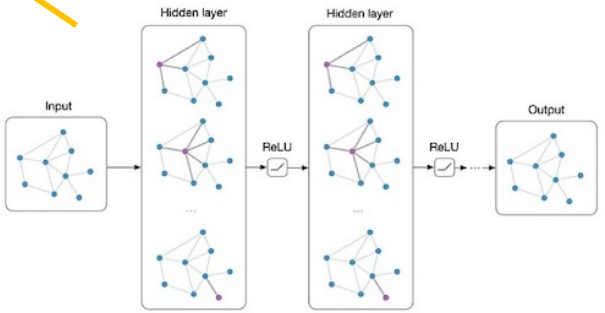


Recurrent Neural Network (RNN)

Variations & Combinations



Generative Adversarial Networks (GAN)



Graph Neural Networks (GNN)

Challenges

1. **Interpretability**

   Black-box deep learning models suffer from poor interpretability since they are not explicitly designed for representing physical relationships and providing mechanistic insights.

2. **Physical consistency**

   Deep learning models can fit observations very well, but predictions may be physically inconsistent or implausible.

3. **Accuracy dilution**

   Strongly nonlinear features tend to cause underfitting or overfitting in deep learning modeling, which makes it difficult to find the global optimal solution and accumulate significant errors.

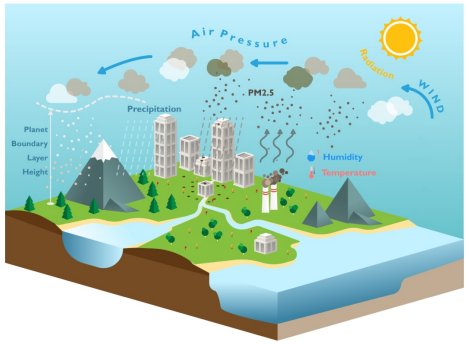
4. **Sample demand**

   The data available for several scientific problems are far smaller than what is needed to effectively train advanced data-driven models.

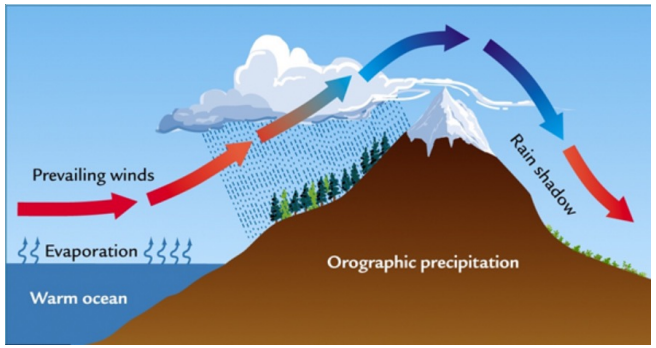
Spatio-temporal processes in the city

Theory-driven

Process understanding

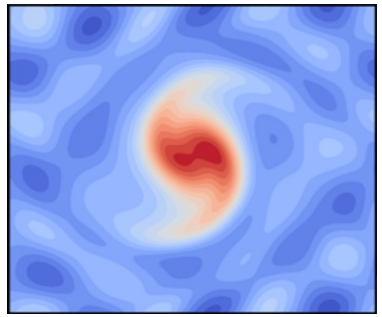


Pollution dispersion



Precipitation processes

traffic wave theory



Fluid dynamics

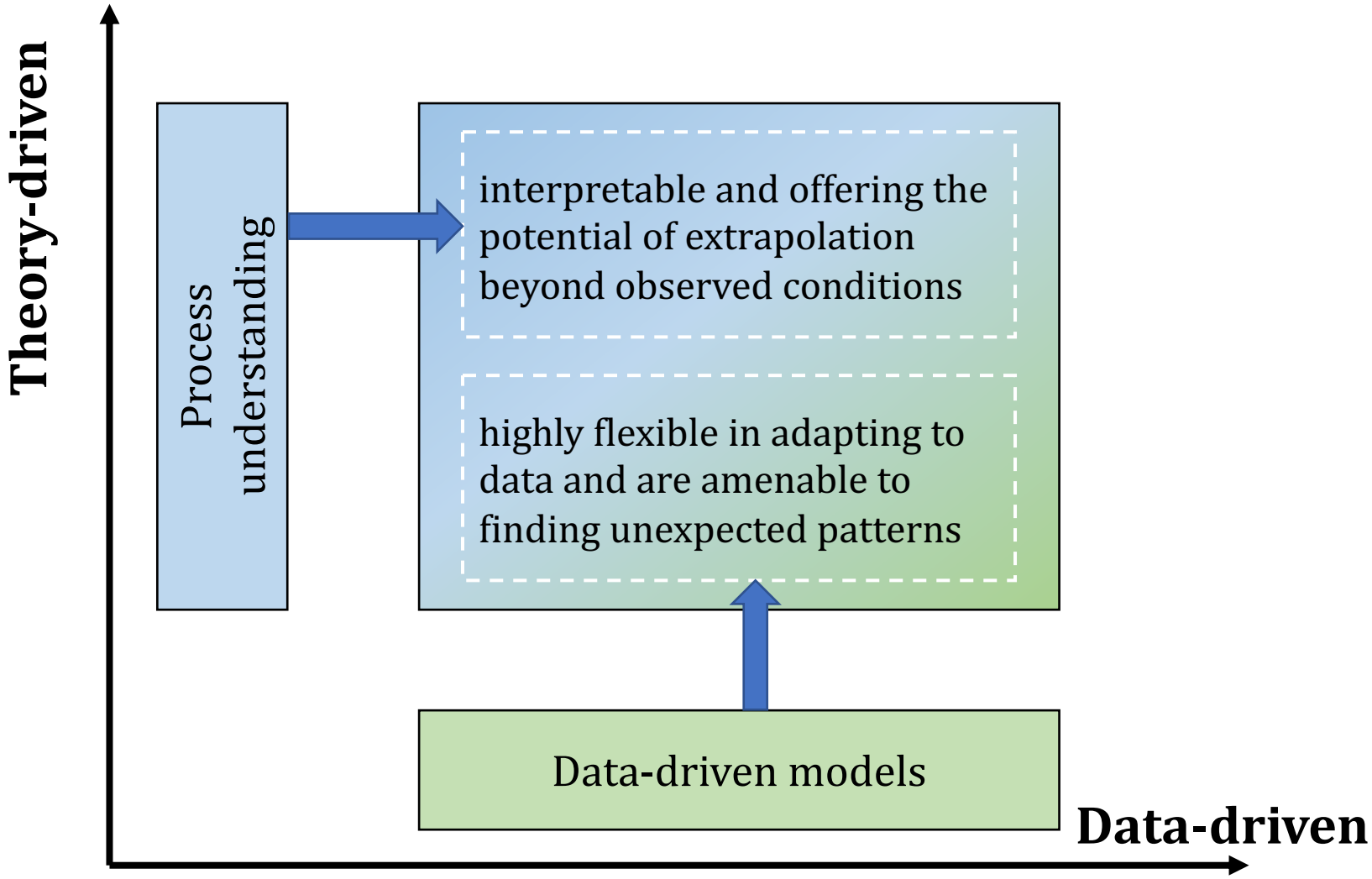
fine-grained parametric data ✓

enormous computational resources ✓

detect complex and unknown high-dimensional relationships ✗

Data-driven

Spatio-temporal processes in the city



The two research paradigms can complement each other, and hybrid models can integrate the strengths of both approaches

How to integrate theory-driven and data-driven models?

1. Optimizing theory-driven models with data-driven models

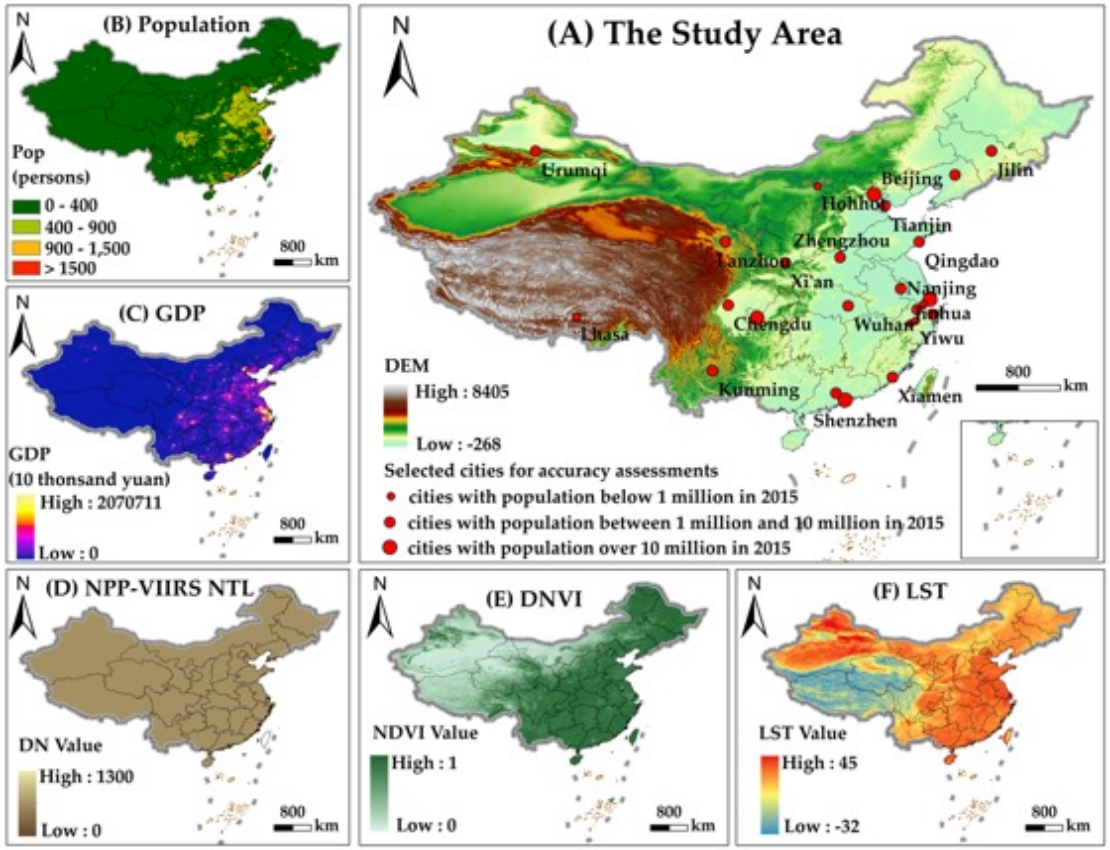
2. Theory-guided data driven models

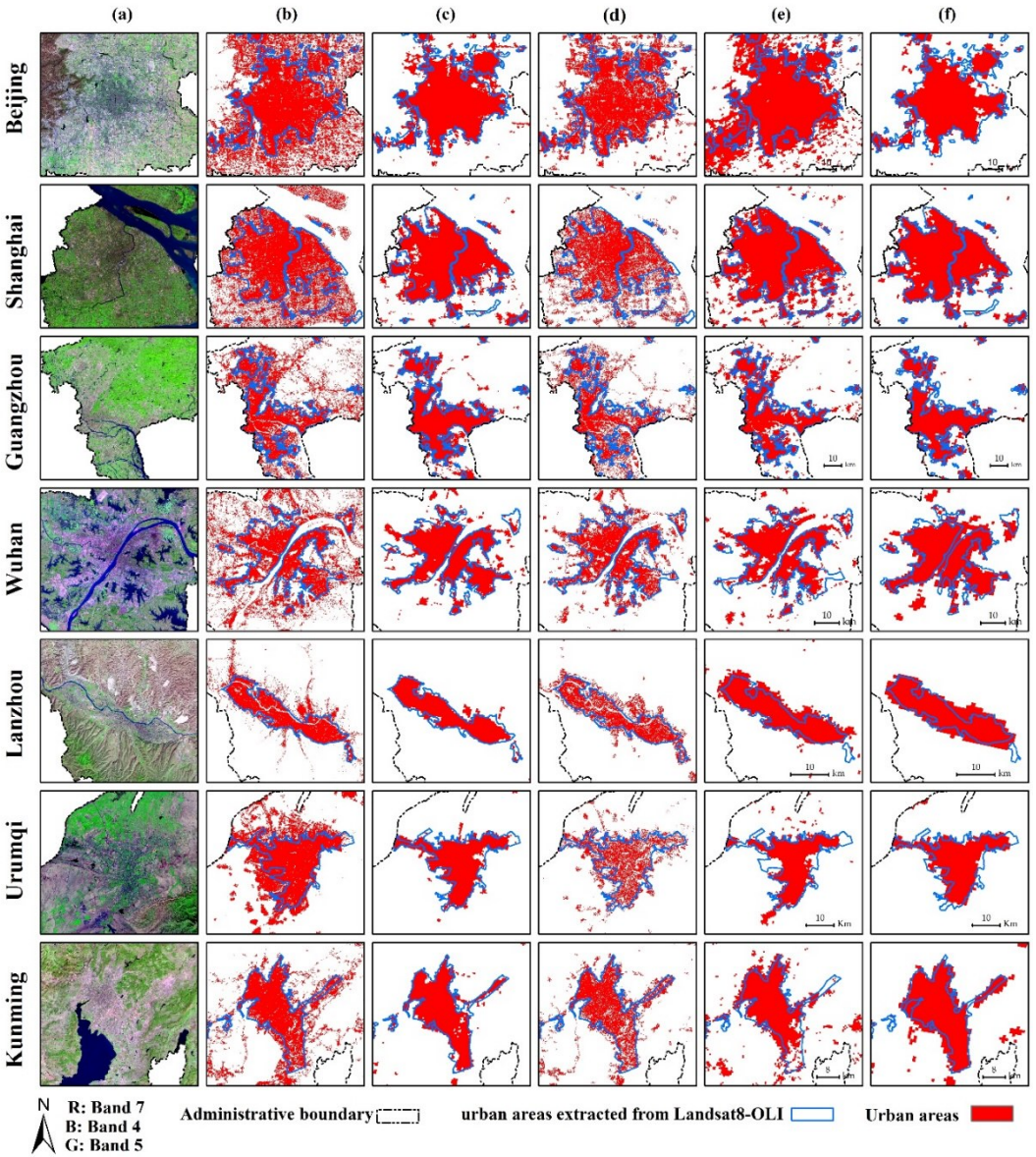
3. Data-driven prediction that uncovers dynamic patterns

How to integrate theory-driven and data-driven models?

1. **Optimizing theory-driven models with data-driven models**
2. Theory-guided data driven models
3. Data-driven prediction that uncovers dynamic patterns

Urban Extent recognition

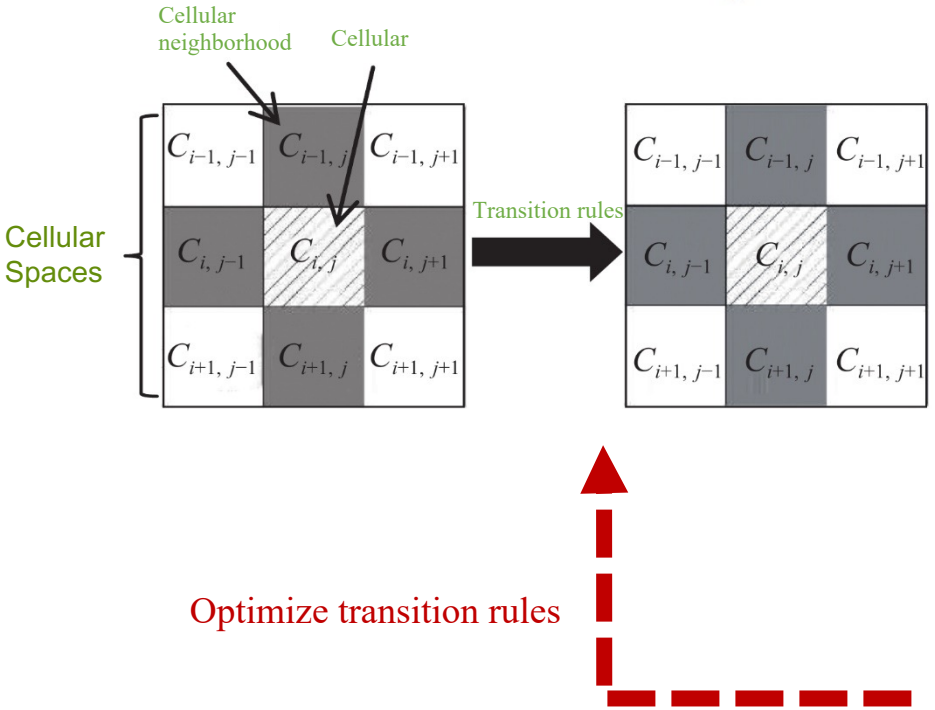


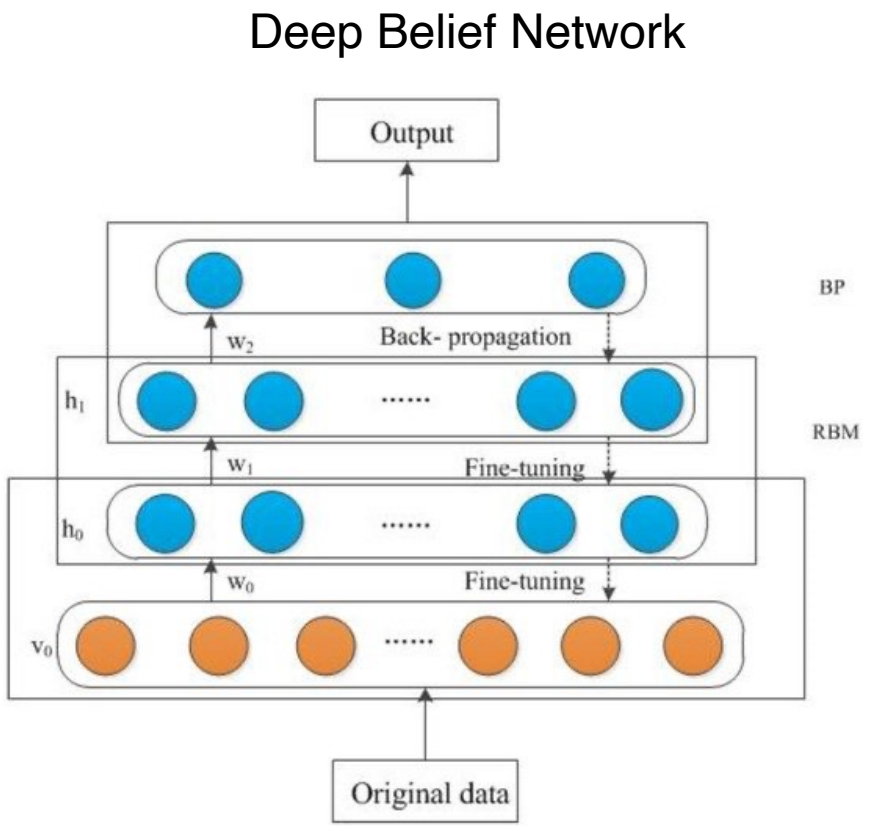


Chen X , Zhang F , Du Z , et al. An Unsupervised Urban Extent Extraction Method from NPP–VIIRS Nighttime Light Data[J]. Remote Sensing, 2020

Urban Growth Simulating (DBN-CA)

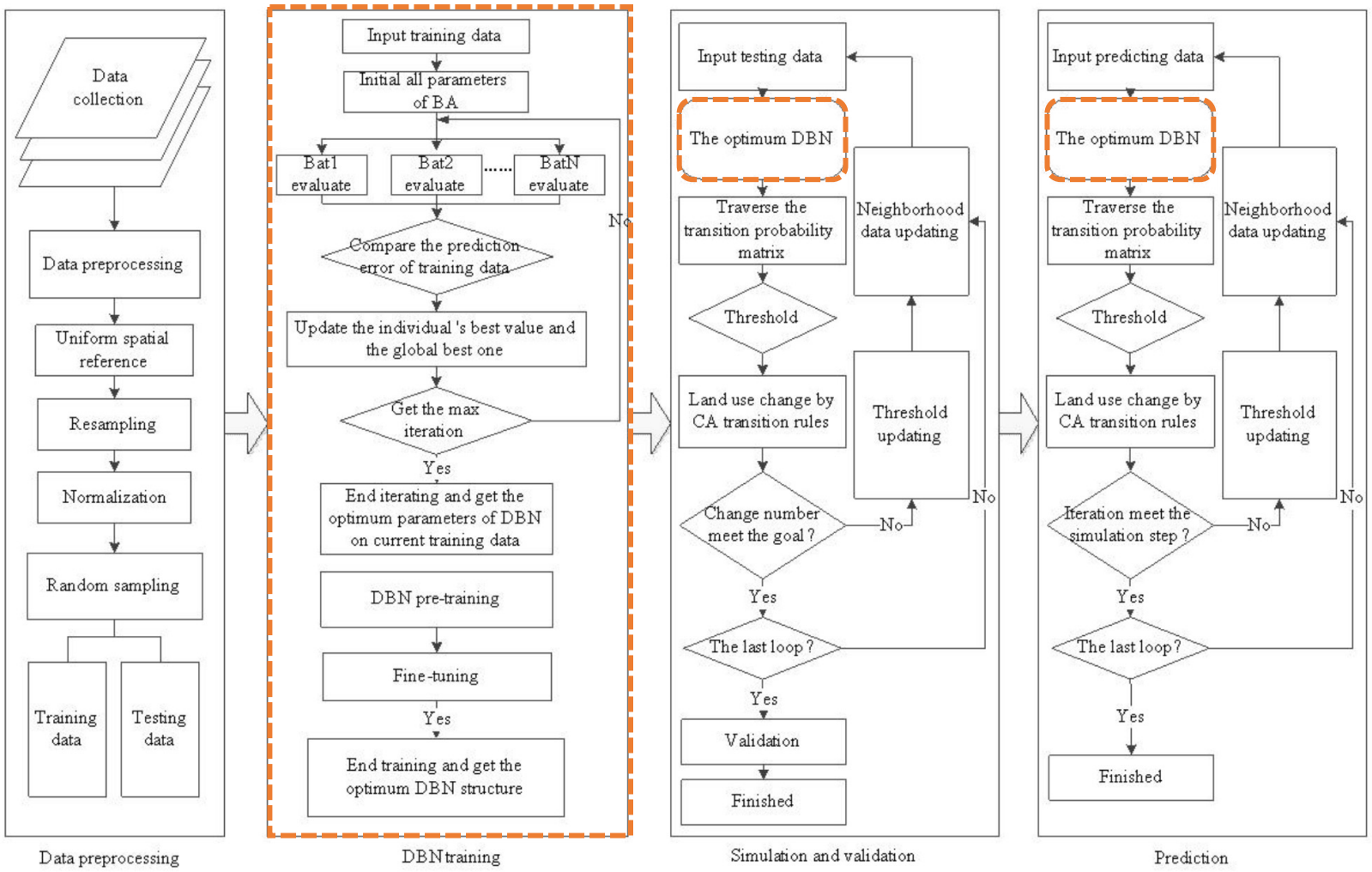
$$State_{ij}^{t+1} = f\left(Neighbour_{ij}^{t}, Global_{ij}^{t}, Constraint_{ij}^{t}\right) = \begin{cases} urban, & p_{ij}^{t} > pthreshold \\ non-urban, & p_{ij}^{t} \le pthr \end{cases}$$





Zhou, Y.; Zhang, F.; Du, Z.; Ye, X.; Liu, R. Integrating Cellular Automata with the Deep Belief Network for Simulating Urban Growth. Sustainability 2017,

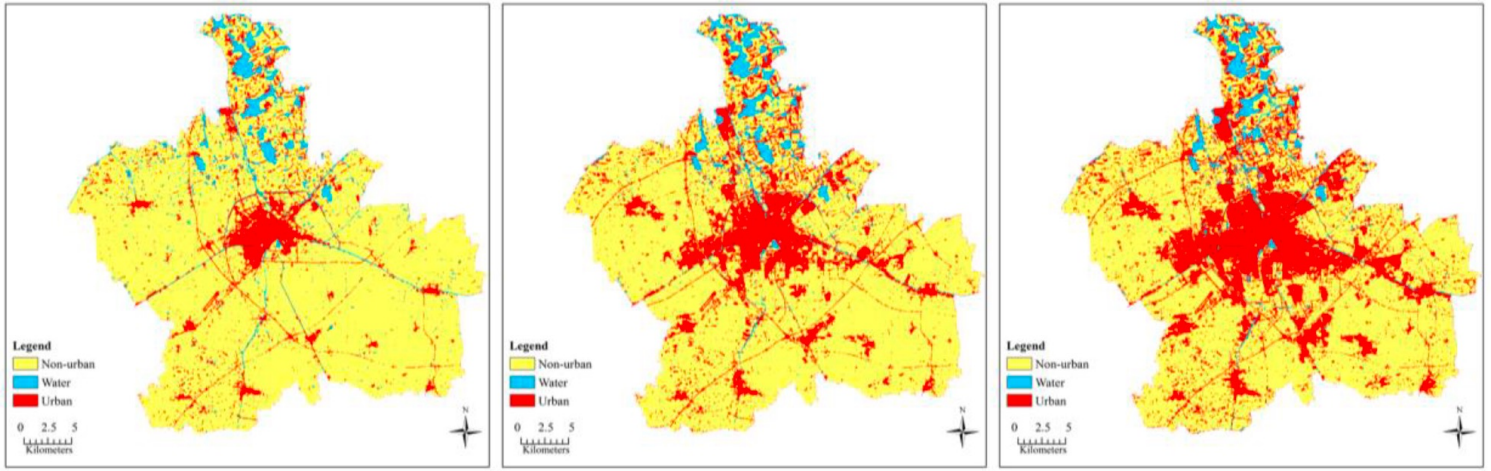
Urban Growth Simulating (DBN-CA)

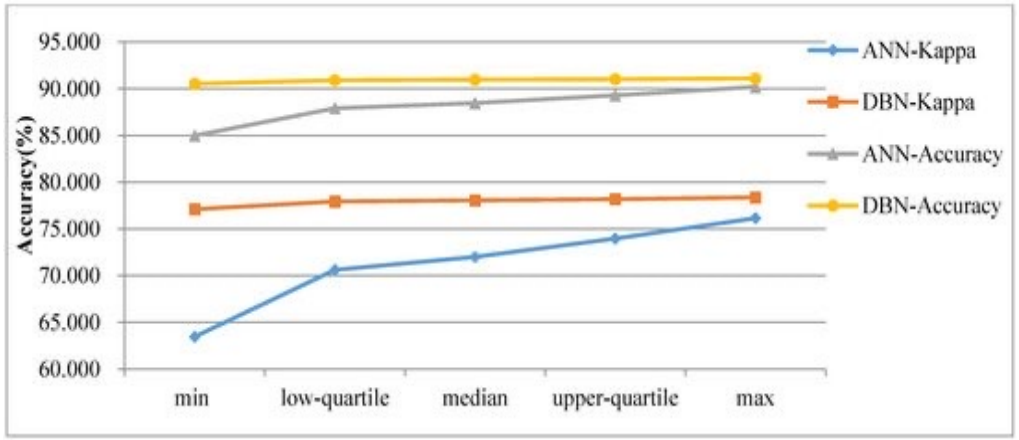


Zhou, Y.; Zhang, F.; Du, Z.; Ye, X.; Liu, R. Integrating Cellular Automata with the Deep Belief Network for Simulating Urban Growth. Sustainability 2017,

Urban Growth Simulating (DBN-CA)

Land of Jiaxing City in 2000, 2008 and 2015

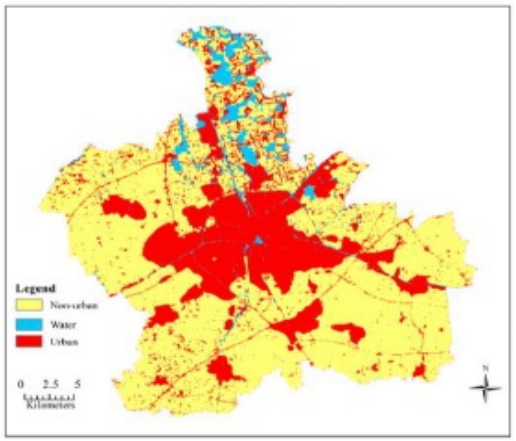




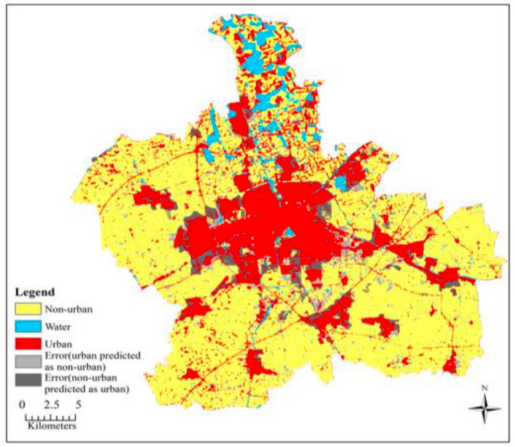
The Accuracy and the Kappa coefficient of DBN-CA and ANN-CA

Zhou, Y.; Zhang, F.; Du, Z.; Ye, X.; Liu, R. Integrating Cellular Automata with the Deep Belief Network for Simulating Urban Growth. Sustainability 2017,

Urban Growth Simulating (DBN-CA)

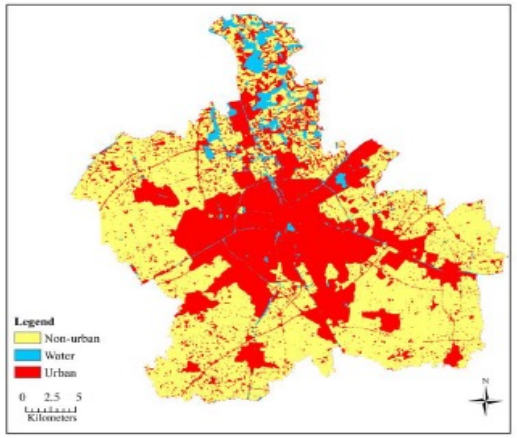


DBN-CA in 2015 (simulation)

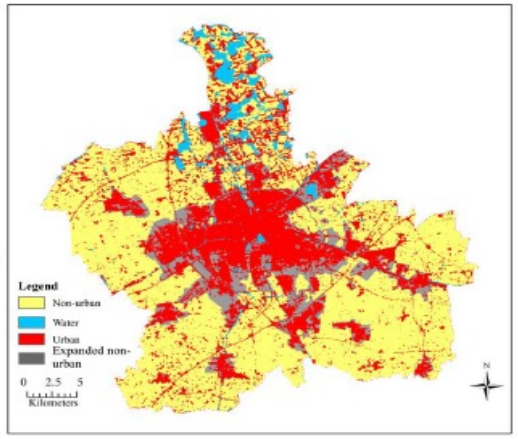


Error

The simulation and the error of DBN-CA in 2015



DBN-CA in 2024 (prediction)



Expansion area in 2024

The prediction and expansion area of DBN-CA in 2024

Zhou, Y.; Zhang, F.; Du, Z.; Ye, X.; Liu, R. Integrating Cellular Automata with the Deep Belief Network for Simulating Urban Growth. Sustainability 2017,

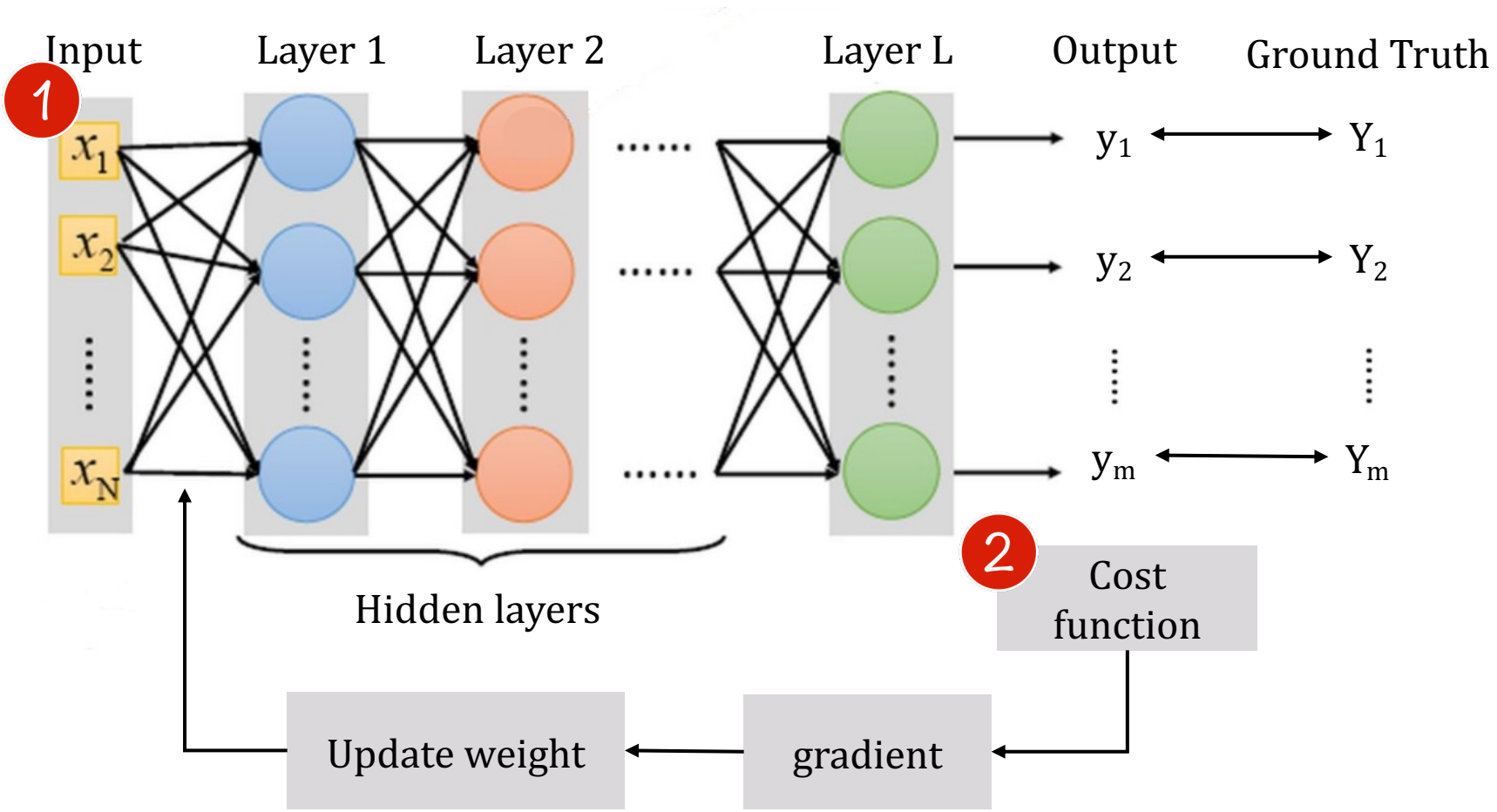
How to integrate theory-driven and data-driven models?

1. Optimizing theory-driven models with data-driven models

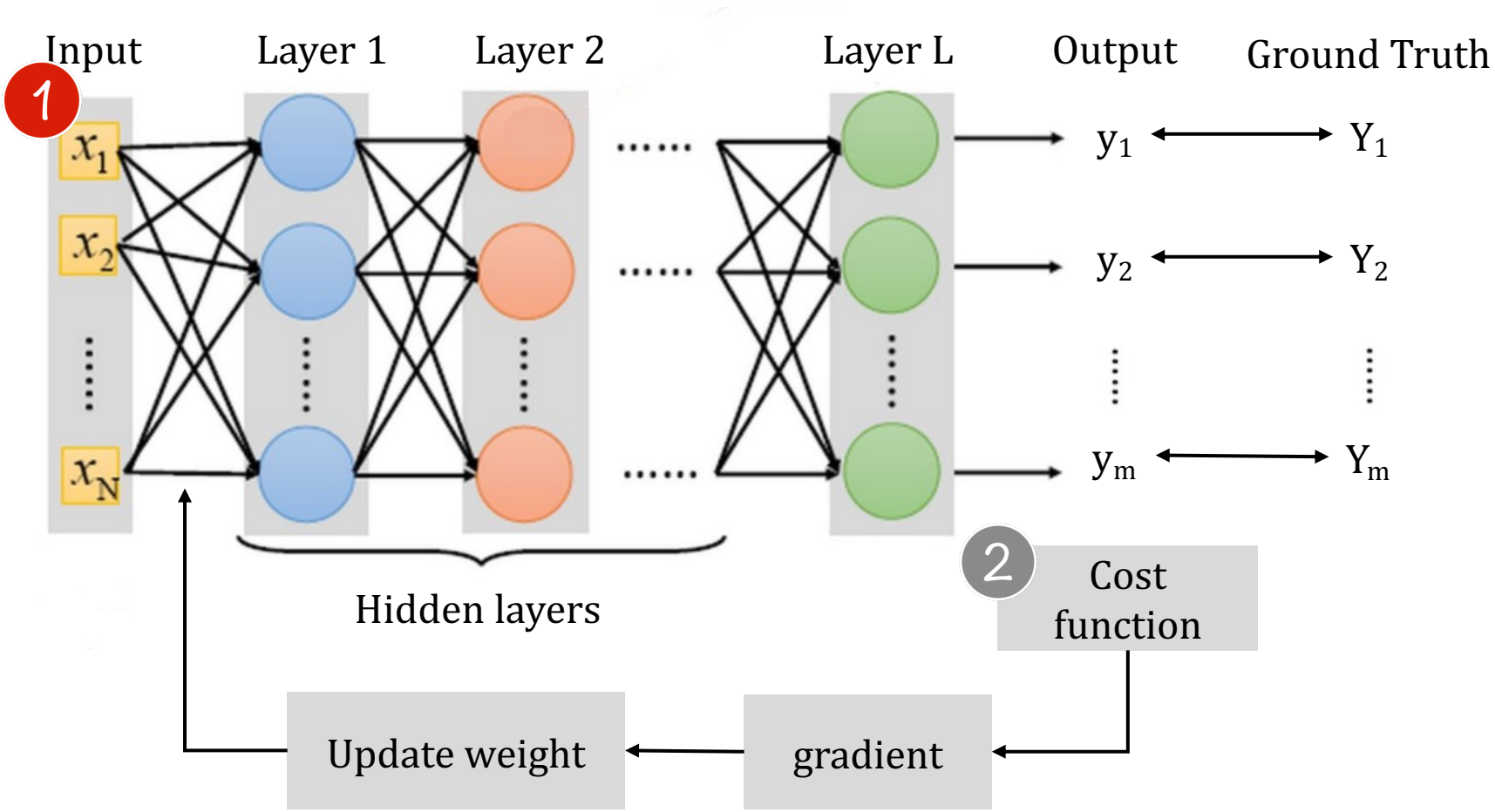
2. **Theory-guided data driven models**

3. Data-driven prediction that uncovers dynamic patterns

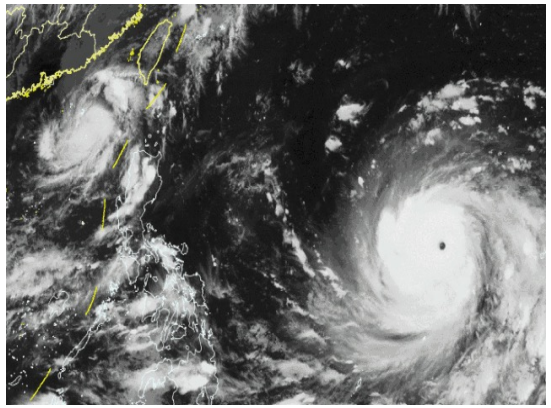
Theory-guided data driven models

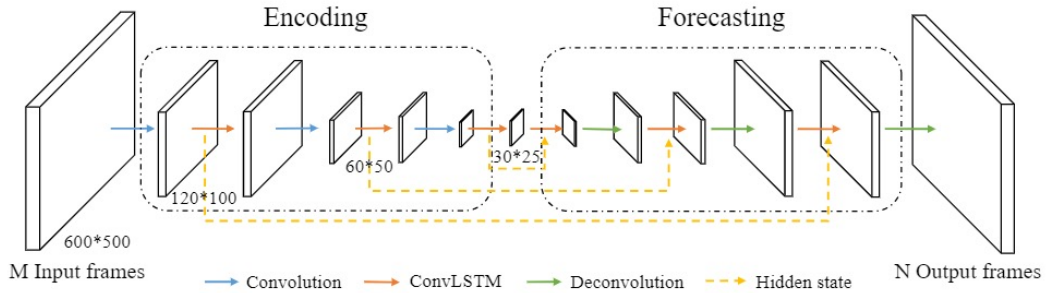


Theory-guided data driven models

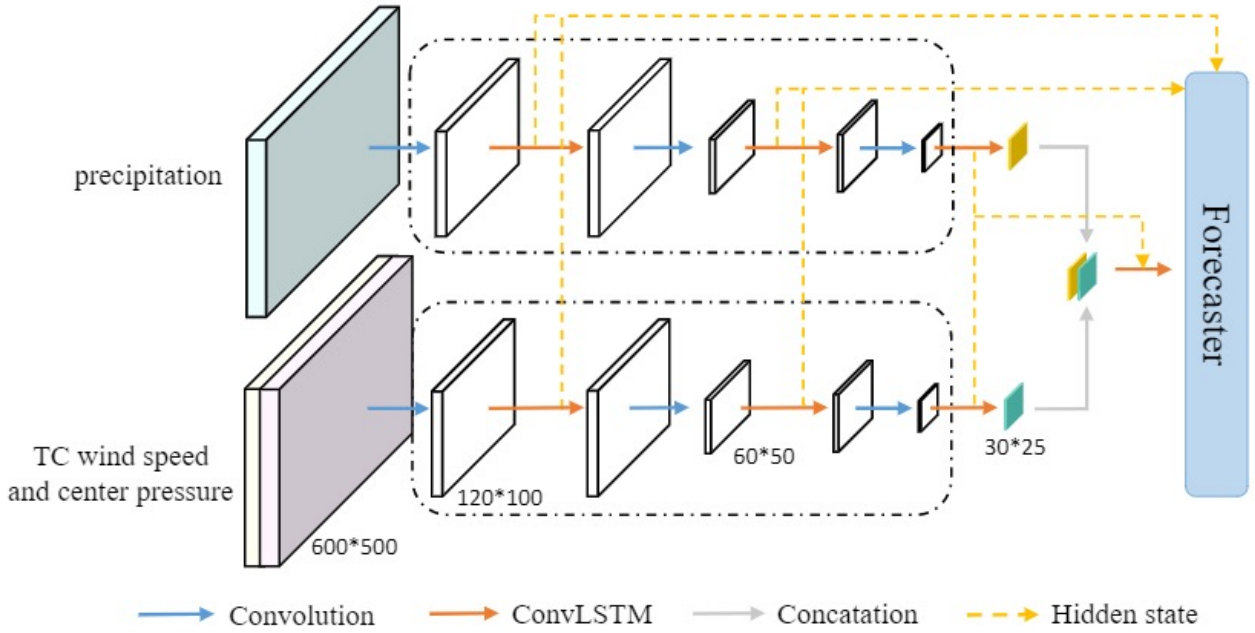


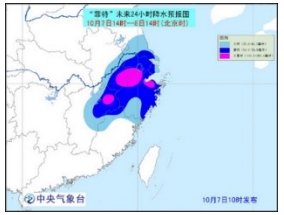
Typhoon Precipitation Nowcasting





The fusion of typhoon data

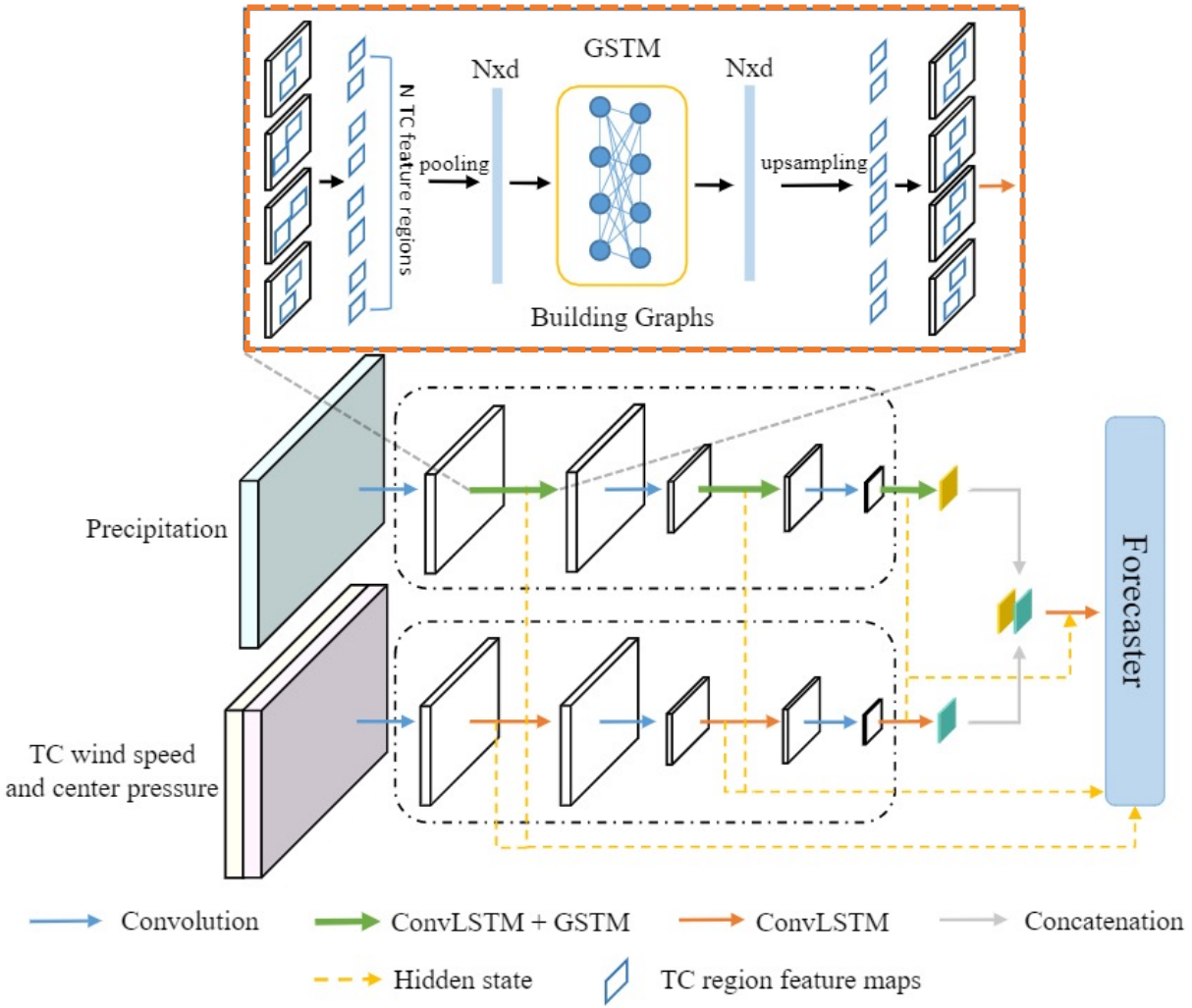




Yang X, Zhang F., et al. A Dynamic Graph-guided Network for Tropical Cyclone Precipitation Nowcasting [J]. Applied soft computing (under review)

Typhoon Precipitation Nowcasting

Graph-guided Spatio-Temporal Module （GSTM）



Extract the regional series and construct the spatio-temporal inference graph

$$NF = \text{AvgPool}(RF)$$

$$Adj = Softmax((NF)(NF^T))$$

Propagate the information on the graph

$$NF' = (Adj)(NF)W + NF$$

$$RF' = Unsampling(NF')$$

$$FRF = \lambda RF' + RF$$

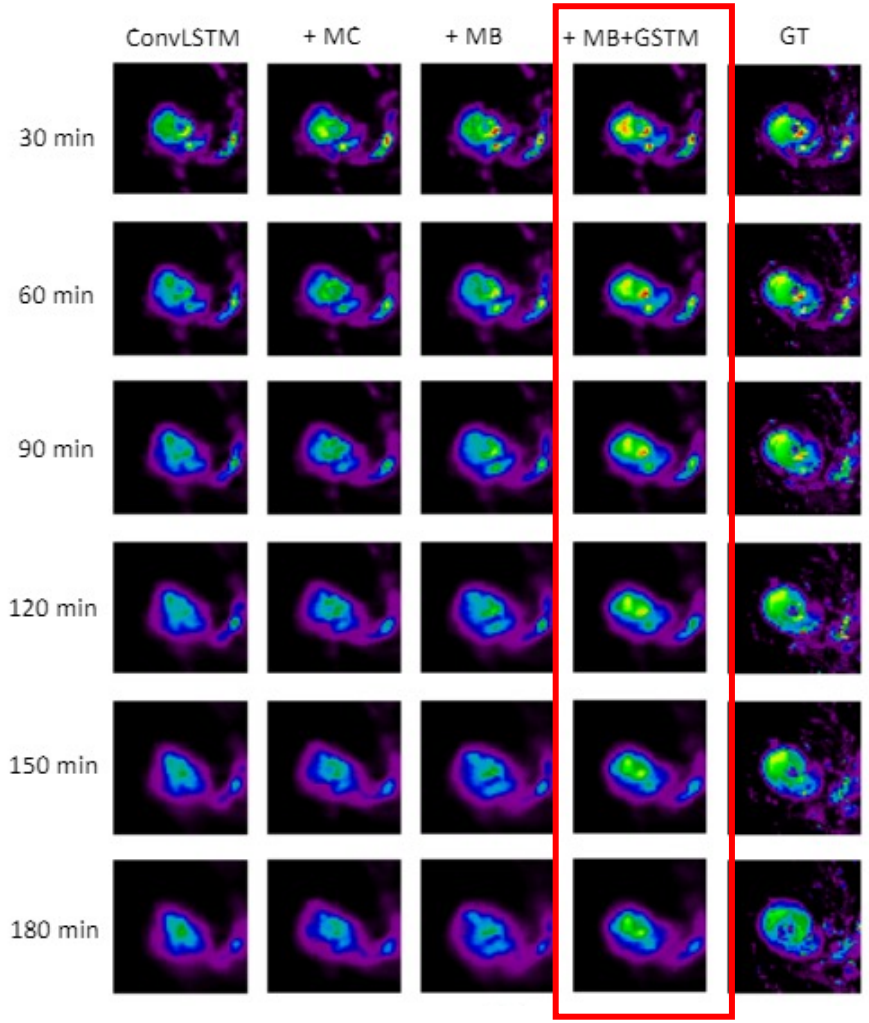
Yang X, Zhang F., et al. A Dynamic Graph-guided Network for Tropical Cyclone Precipitation Nowcasting [J]. Applied soft computing (under review)

Typhoon Precipitation Nowcasting

Algorithms	CSI ↑					
	$r \geq 0.5$	$r \geq 2$	$r \geq 5$	$r \geq 10$	$r \geq 20$	$r \geq 30$
2D CNN	0.4456	0.374	0.291	0.2194	0.138	0.082
TrajGRU	0.4650	0.3913	0.2971	0.2274	0.1620	0.1133
ConvLSTM	0.4643	0.3913	0.3003	0.2307	0.1596	0.1067
ConvLSTM+MC	0.469	0.4011	0.3092	0.2352	0.1697	0.1149
ConvLSTM+MB	**0.4728**	0.4027	0.3191	0.2553	0.1859	0.1236
ConvLSTM+MB+GSTM	0.4716	**0.4027**	**0.3246**	**0.2632**	**0.2015**	**0.149**

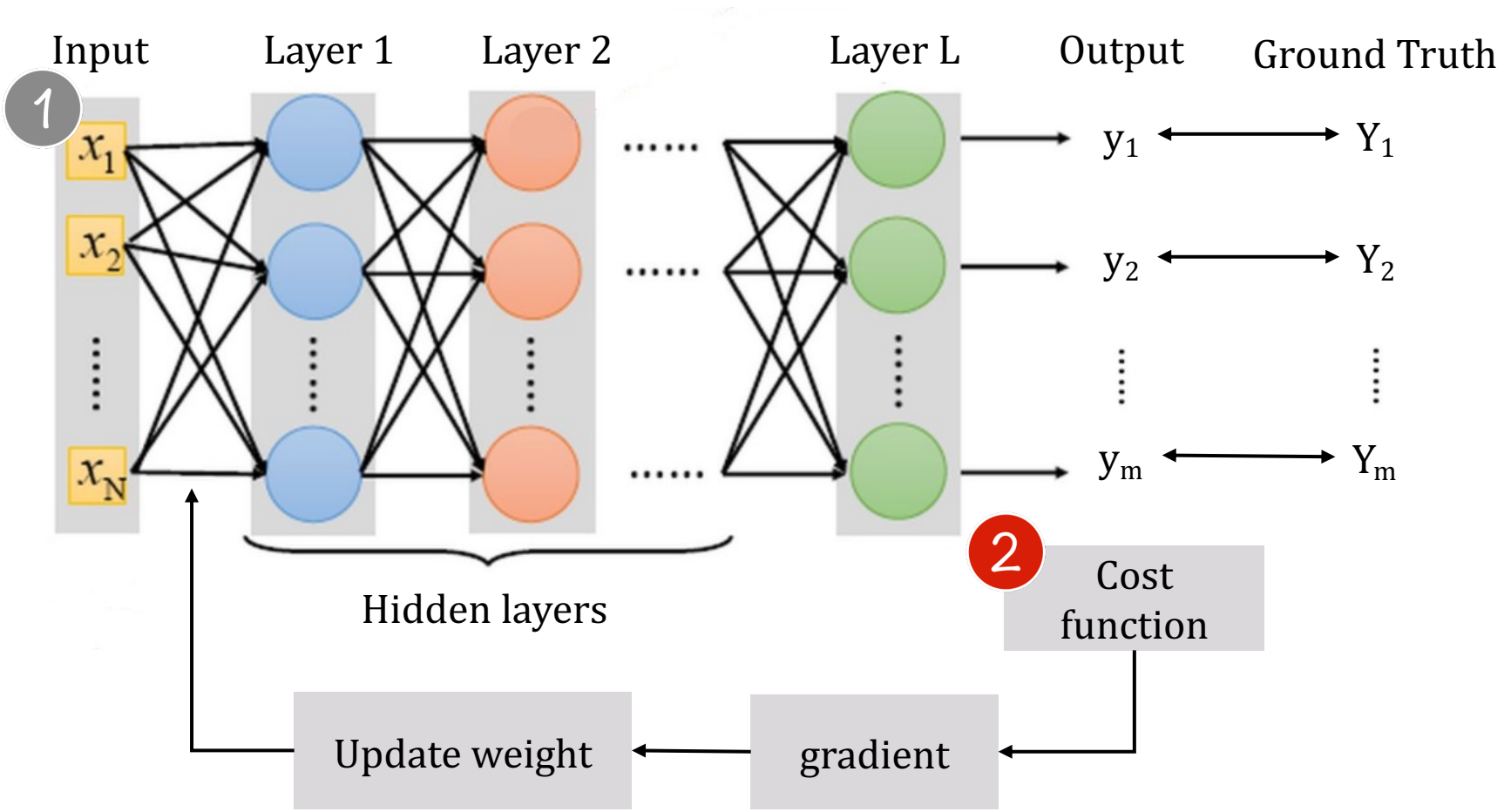
Algorithms	HSS ↑					
	$r \geq 0.5$	$r \geq 2$	$r \geq 5$	$r \geq 10$	$r \geq 20$	$r \geq 30$
2D CNN	0.58	0.5273	0.4402	0.3506	0.2328	0.1432
TrajGRU	0.6009	0.5452	0.4458	0.3597	0.2668	0.1918
ConvLSTM	0.5998	0.5452	0.4501	0.3651	0.265	0.1835
ConvLSTM+MC	0.6036	0.5555	0.4616	0.3719	0.2814	0.1985
ConvLSTM+MB	**0.6072**	0.557	0.4727	0.3982	0.305	0.2114
ConvLSTM+MB+GSTM	0.6055	**0.5581**	**0.4793**	**0.4085**	**0.328**	**0.2523**

Algorithms	F1 score ↑					
	$r \geq 0.5$	$r \geq 2$	$r \geq 5$	$r \geq 10$	$r \geq 20$	$r \geq 30$
2D CNN	0.6147	0.541	0.4454	0.3524	0.2332	0.1433
TrajGRU	0.6323	0.5579	0.4506	0.3614	0.2673	0.1920
ConvLSTM	0.6315	0.5581	0.455	0.3668	0.2655	0.1837
ConvLSTM+MC	0.6361	0.5687	0.4665	0.3736	0.2819	0.1987
ConvLSTM+MB	**0.6394**	0.5697	0.4776	0.4	0.3055	0.2116
ConvLSTM+MB+GSTM	0.6374	**0.5699**	**0.4844**	**0.4106**	**0.3286**	**0.2525**

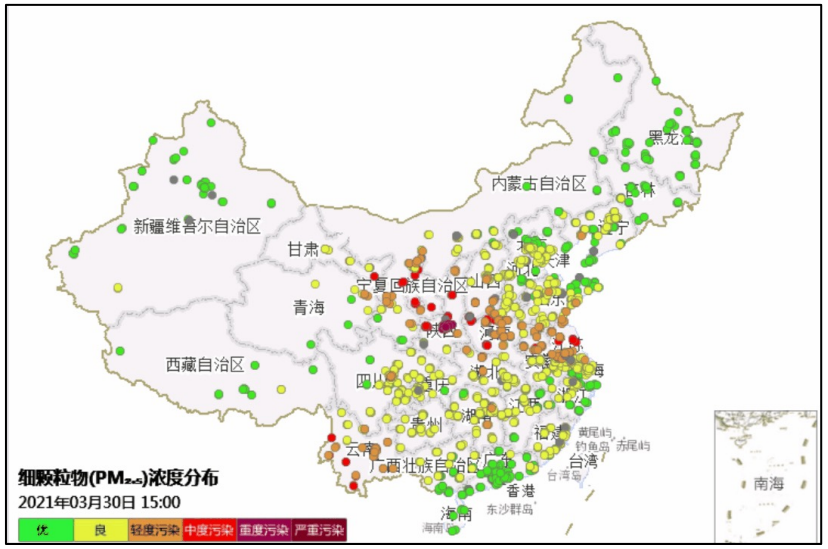


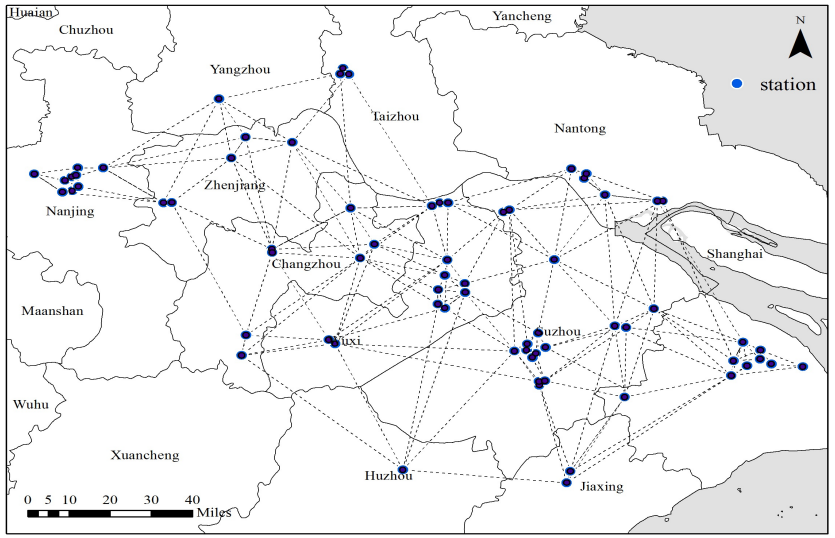
Yang X, Zhang F., et al. A Dynamic Graph-guided Network for Tropical Cyclone Precipitation Nowcasting [J]. Applied soft computing (under review)

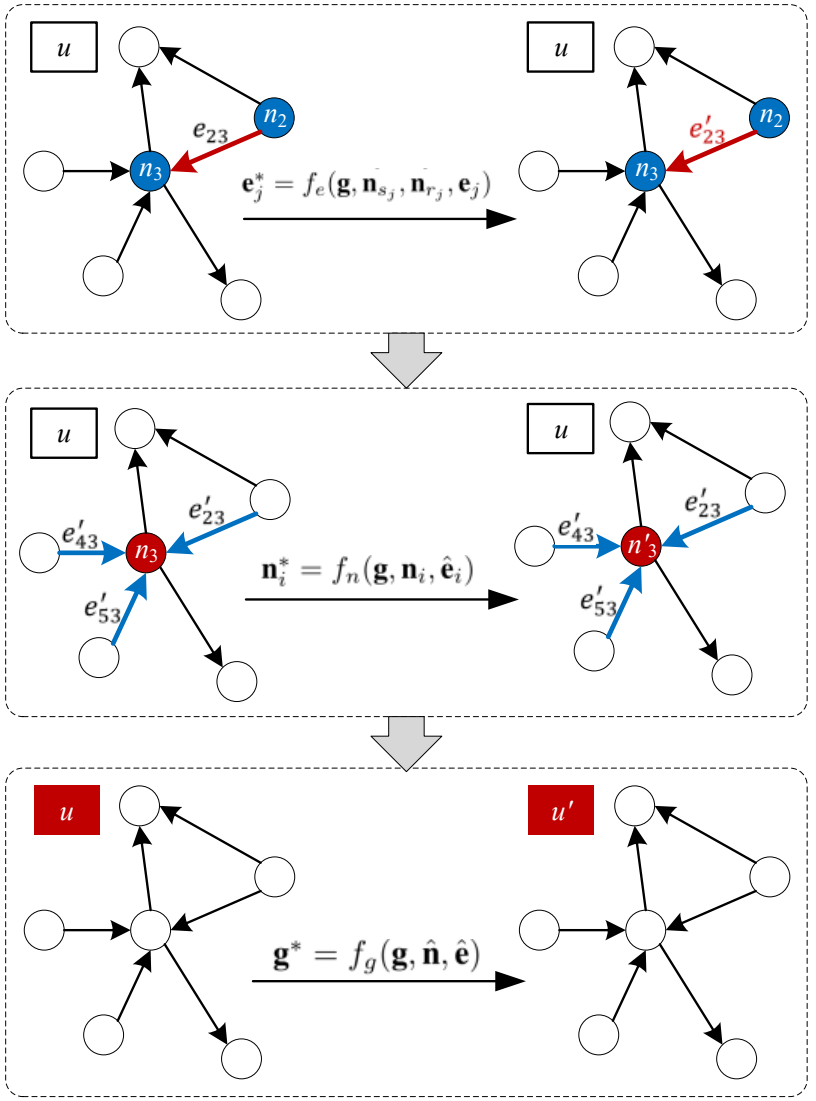
Theory-guided data driven models



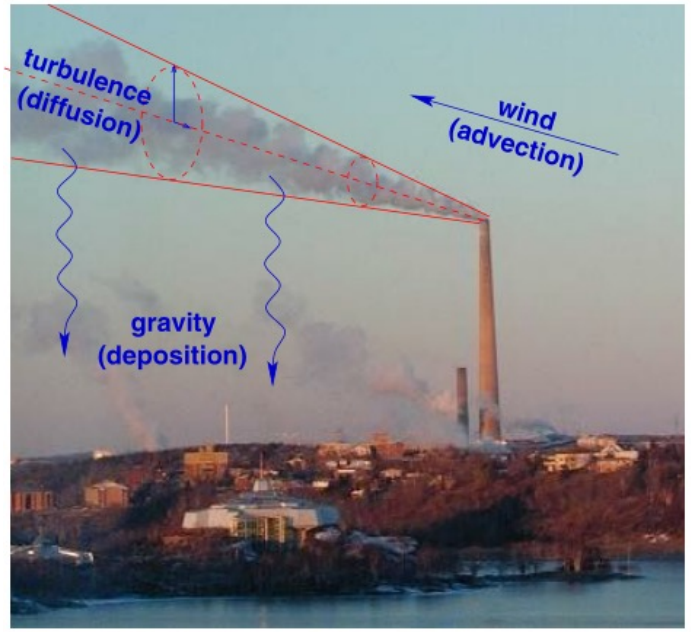
PM2.5 forecasting

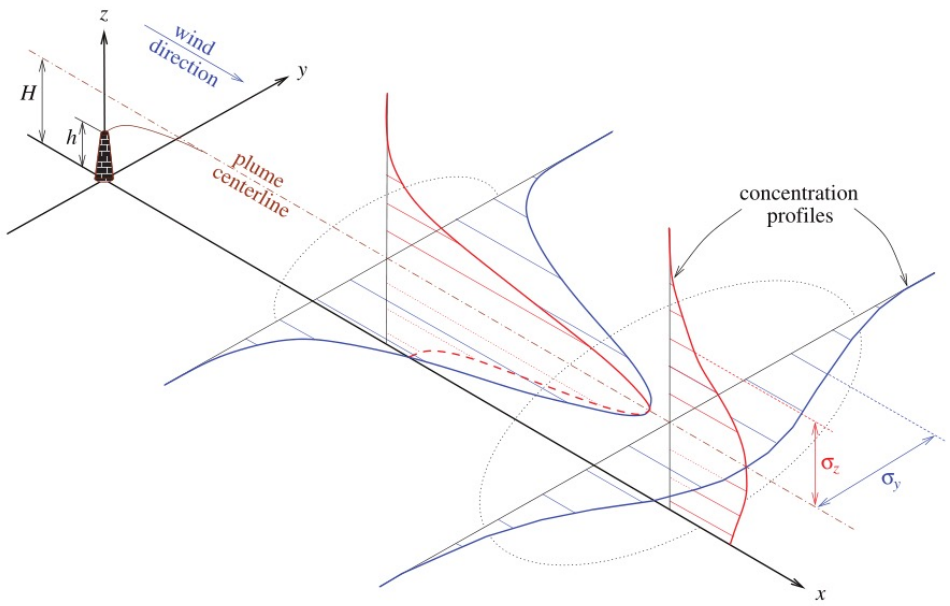






The atmospheric dispersion model

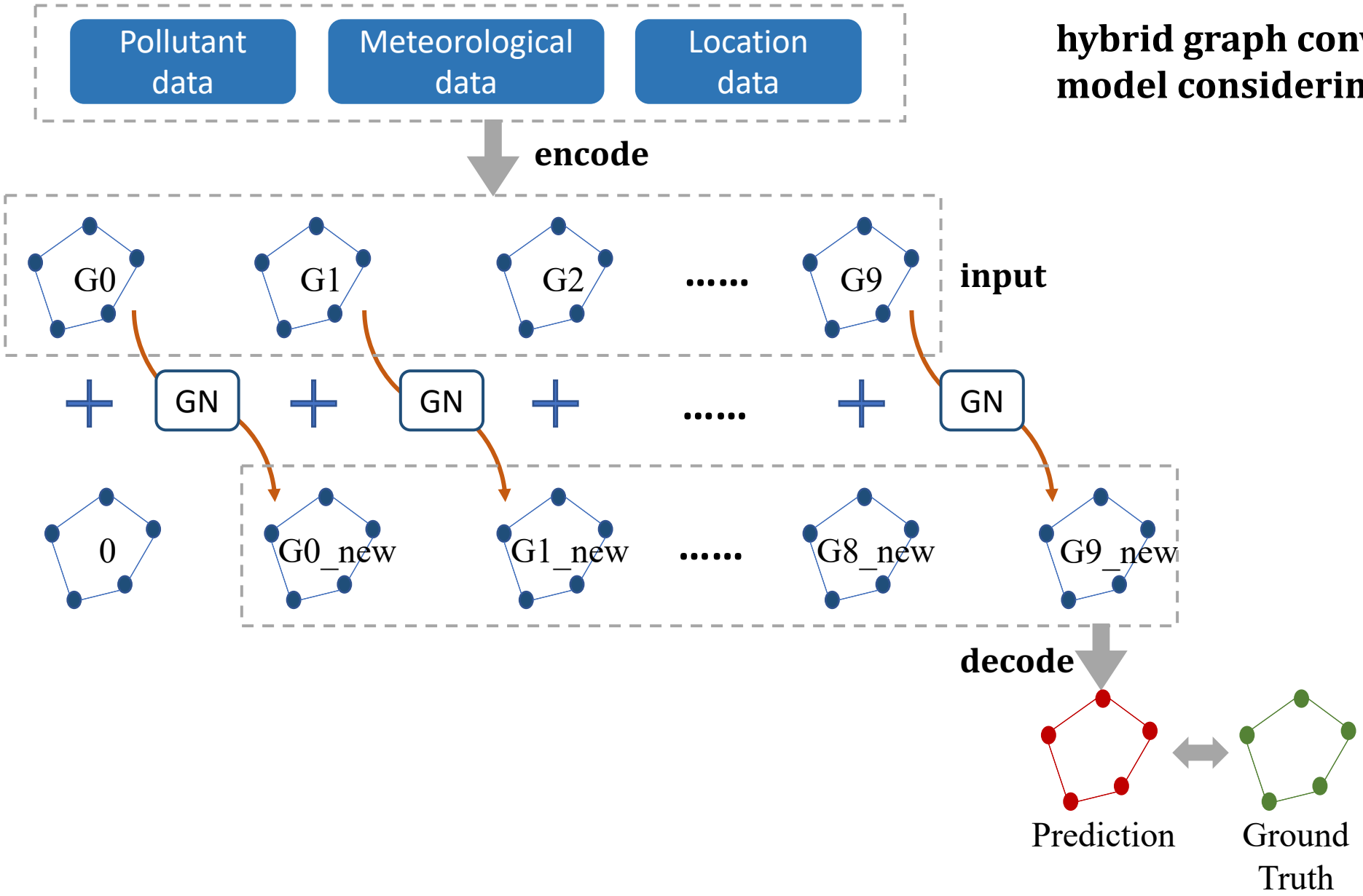




$$\frac{\partial c}{\partial t}+u_x\frac{\partial c}{\partial x}+u_y\frac{\partial c}{\partial y}+u_z\frac{\partial c}{\partial z}=\frac{\partial}{\partial x}\left(D_x\frac{\partial c}{\partial x}\right)+\frac{\partial}{\partial y}\left(D_y\frac{\partial c}{\partial y}\right)+\frac{\partial}{\partial z}\left(D_z\frac{\partial c}{\partial z}\right)+S(x,y,z,c,t)$$

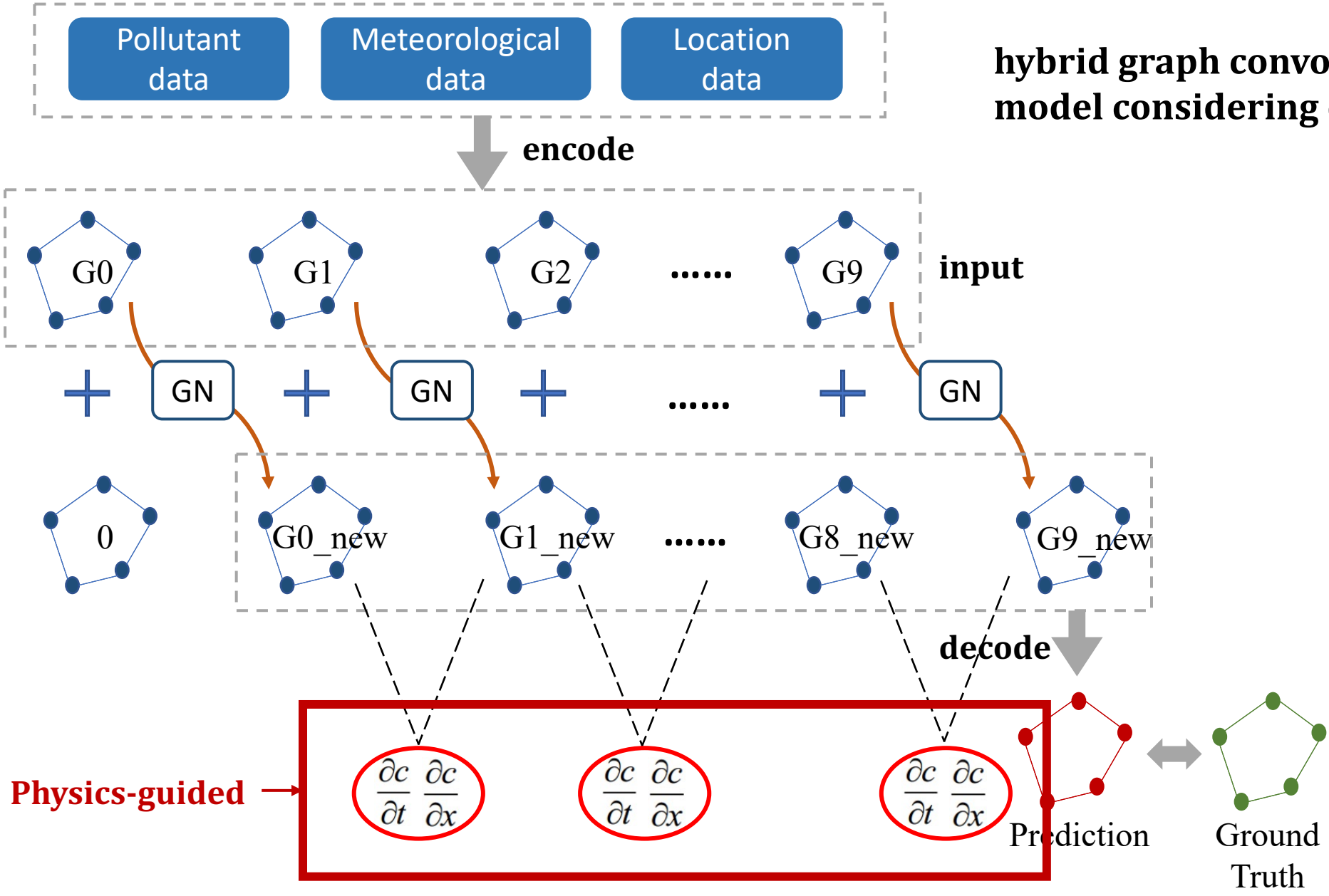
advection effect **diffusion effect** **Pollution source**

Zhou H , Zhang F , et al. Forecasting PM2.5 using hybrid graph convolution–based model considering dynamic wind–field to offer the benefit of spatial interpretability[J]. Environmental Pollution, 2021



hybrid graph convolution-based model considering dynamic wind-field

Zhou H , Zhang F , et al. Forecasting PM2.5 using hybrid graph convolution–based model considering dynamic wind–field to offer the benefit of spatial interpretability[J]. Environmental Pollution, 2021



hybrid graph convolution-based model considering dynamic wind-field

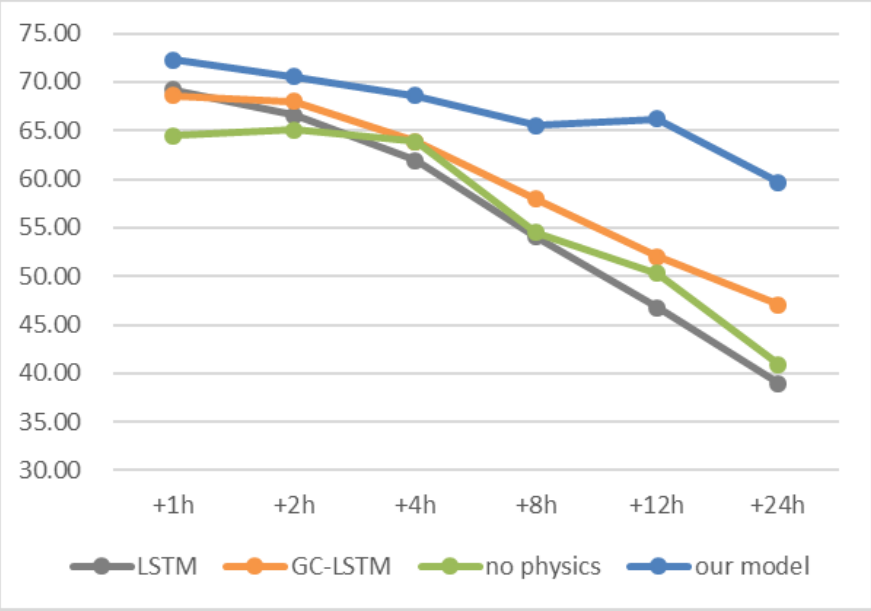
Zhou H , Zhang F , et al. Forecasting PM2.5 using hybrid graph convolution–based model considering dynamic wind–field to offer the benefit of spatial interpretability[J]. Environmental Pollution, 2021

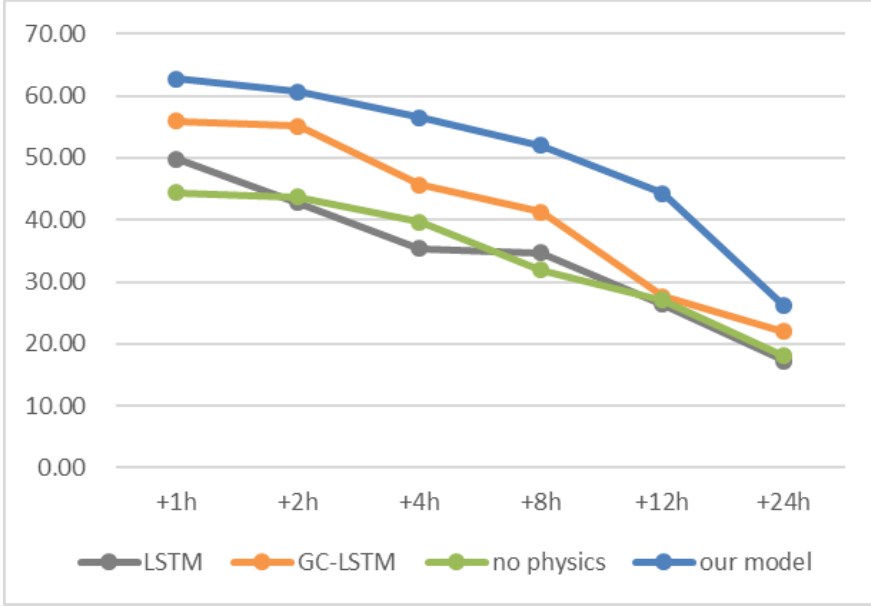
Result

Model		1	2	4	8	12	24
LSTM	rmse	20.6393	21.0103	22.3271	23.8861	25.1641	29.3749
	mae	15.2236	15.4459	16.2739	17.3312	18.1747	21.4094
	IA	0.8906	0.8850	0.8631	0.7988	0.7505	0.6026
	r	0.8266	0.8138	0.7776	0.7079	0.6506	0.5050
GC-LSTM	rmse	19.2963	19.8383	21.3861	23.4004	24.7533	28.5459
	mae	13.3595	13.7157	14.9543	17.1641	17.9145	21.4850
	IA	0.8806	0.8774	0.8591	0.8281	0.7852	0.7059
	r	0.8235	0.8168	0.7780	0.7284	0.6924	0.5583
GN-no-physics	rmse	19.2379	19.7214	21.0835	23.6588	25.3558	28.5466
	mae	13.0323	13.2661	14.4151	16.7451	18.4607	20.9568
	IA	0.8914	0.8853	0.8614	0.8033	0.7686	0.6560
	r	0.8325	0.8243	0.7950	0.7348	0.6864	0.5709
Our model	rmse	**18.1880**	**18.3571**	**19.6208**	**21.7959**	**23.3570**	**27.3578**
	mae	**12.1313**	**12.4879**	**13.5322**	**15.4026**	**16.7409**	**20.2831**
	IA	**0.9140**	**0.9087**	**0.8935**	**0.8670**	**0.8411**	**0.7863**
	r	**0.8532**	**0.8451**	**0.8198**	**0.7719**	**0.7324**	**0.6319**

0-35	Excellent	No effect on health
35-75	Good	Weak effects on the health of a very small number of susceptible people
75-115	Light pollution	Irritation symptoms in healthy people
115-150	Moderate pollution	Healthy people's heart and respiratory system are affected
150-250	Severe pollution	Significant increase in cardiopulmonary symptoms
250-500	extremelySevere pollution	Visible symptoms, disease onset

threshold=75



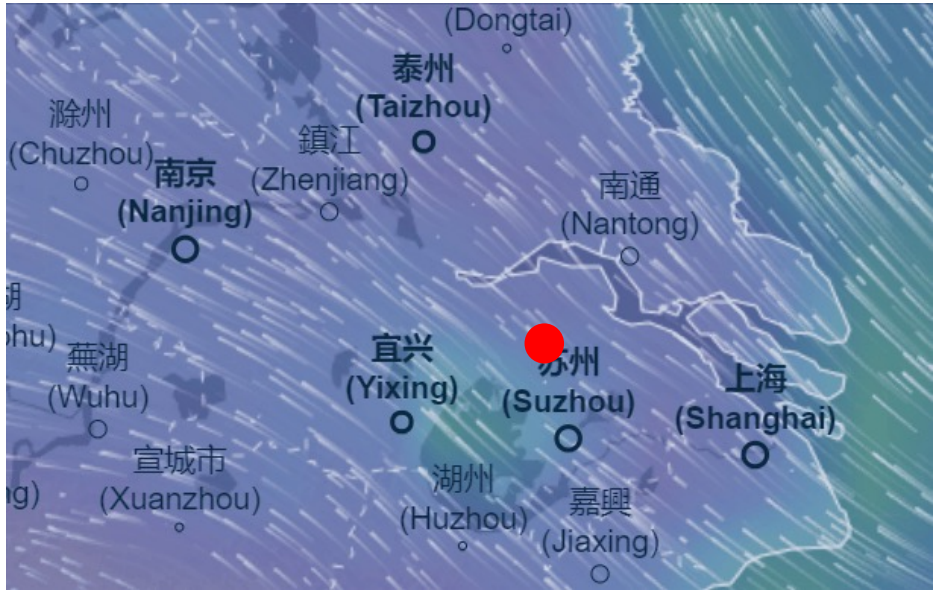


threshold=115

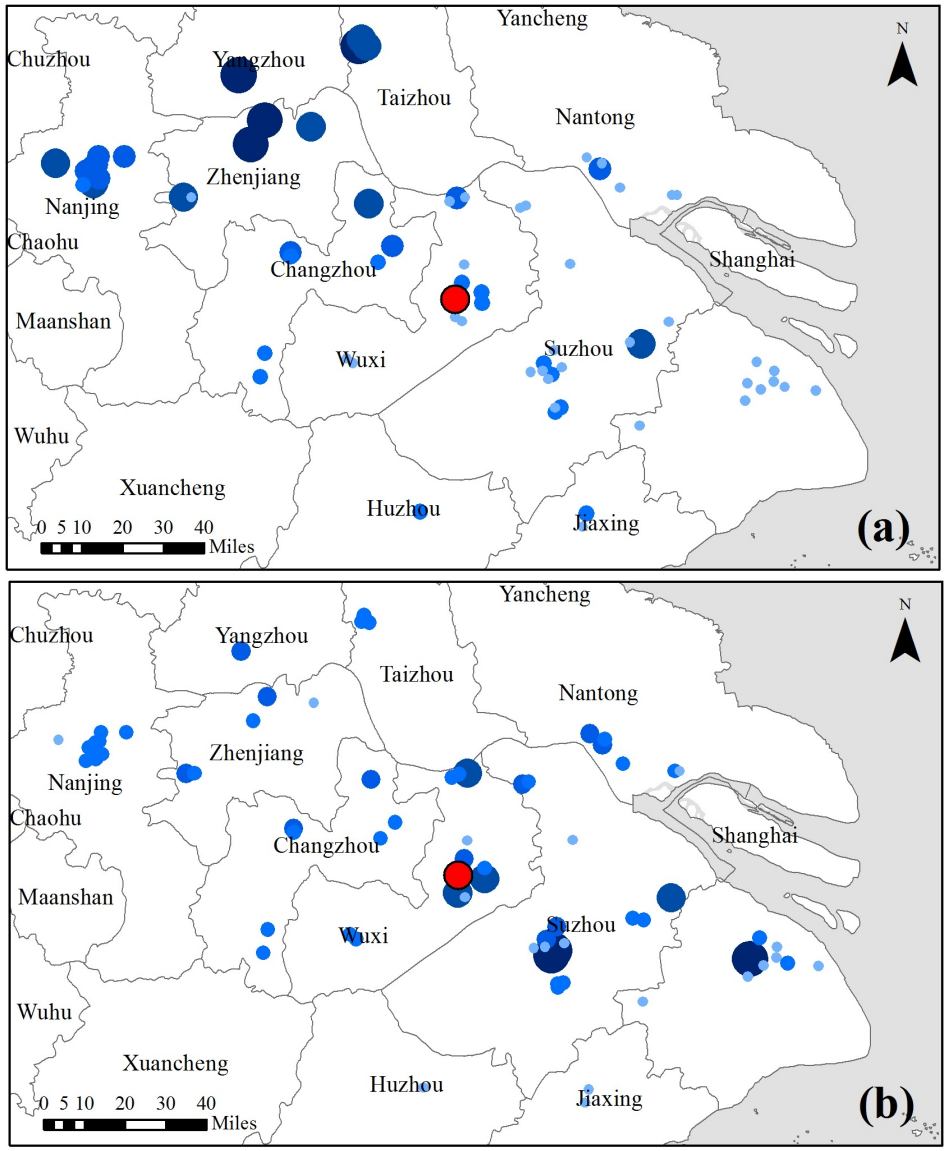
Zhou H , Zhang F , et al. Forecasting PM2.5 using hybrid graph convolution–based model considering dynamic wind–field to offer the benefit of spatial interpretability[J]. Environmental Pollution, 2021

Explain the model

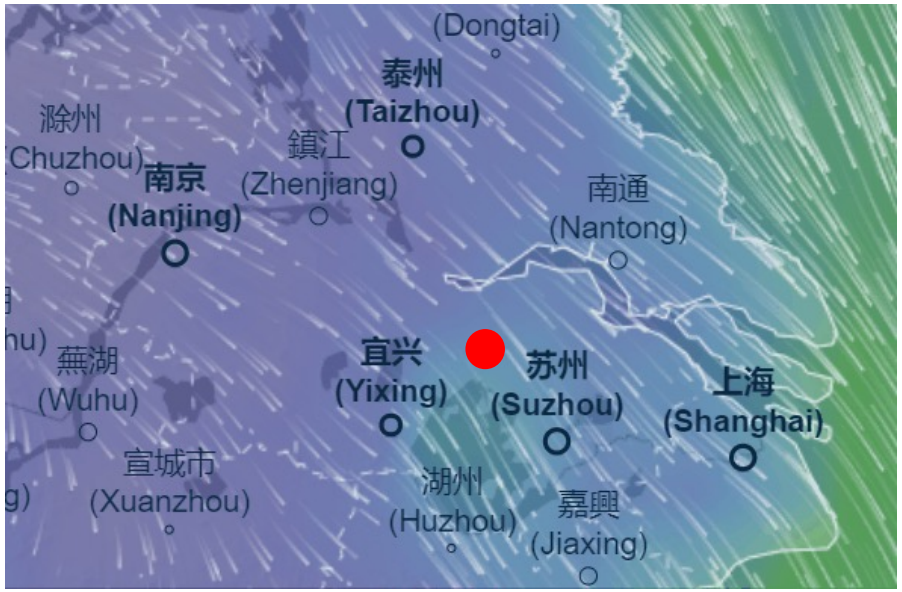
GNNEXPLAINER (Ying, Zhitao et al)



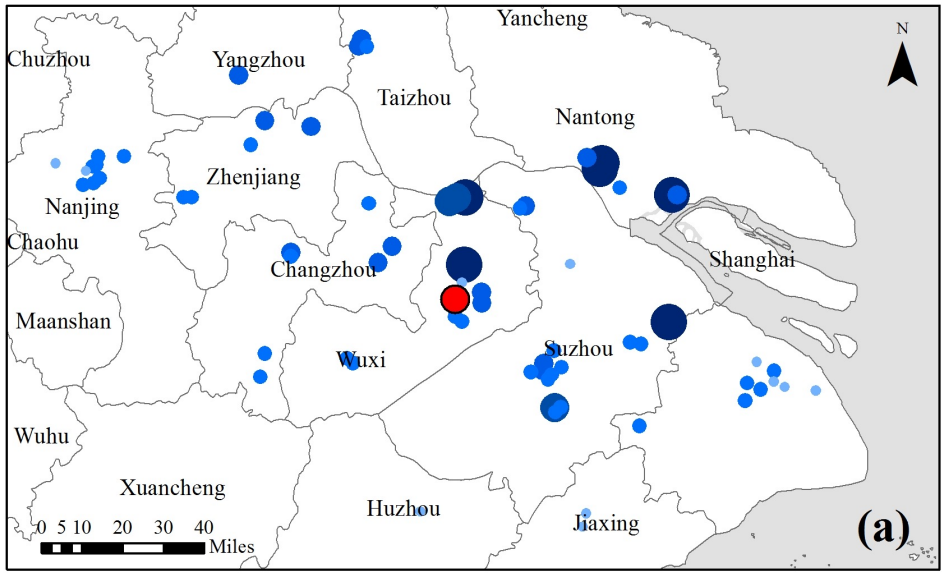
2016.08.01

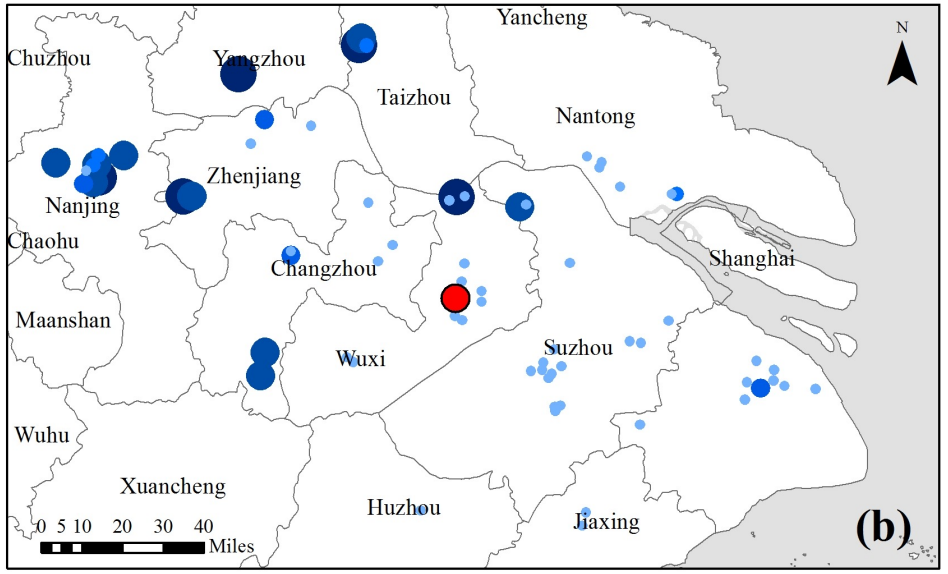


Explain the model

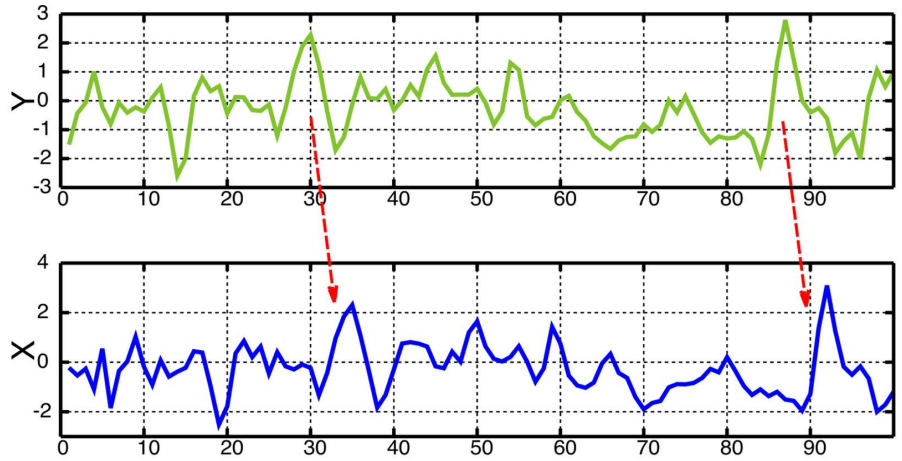


2016.11.23



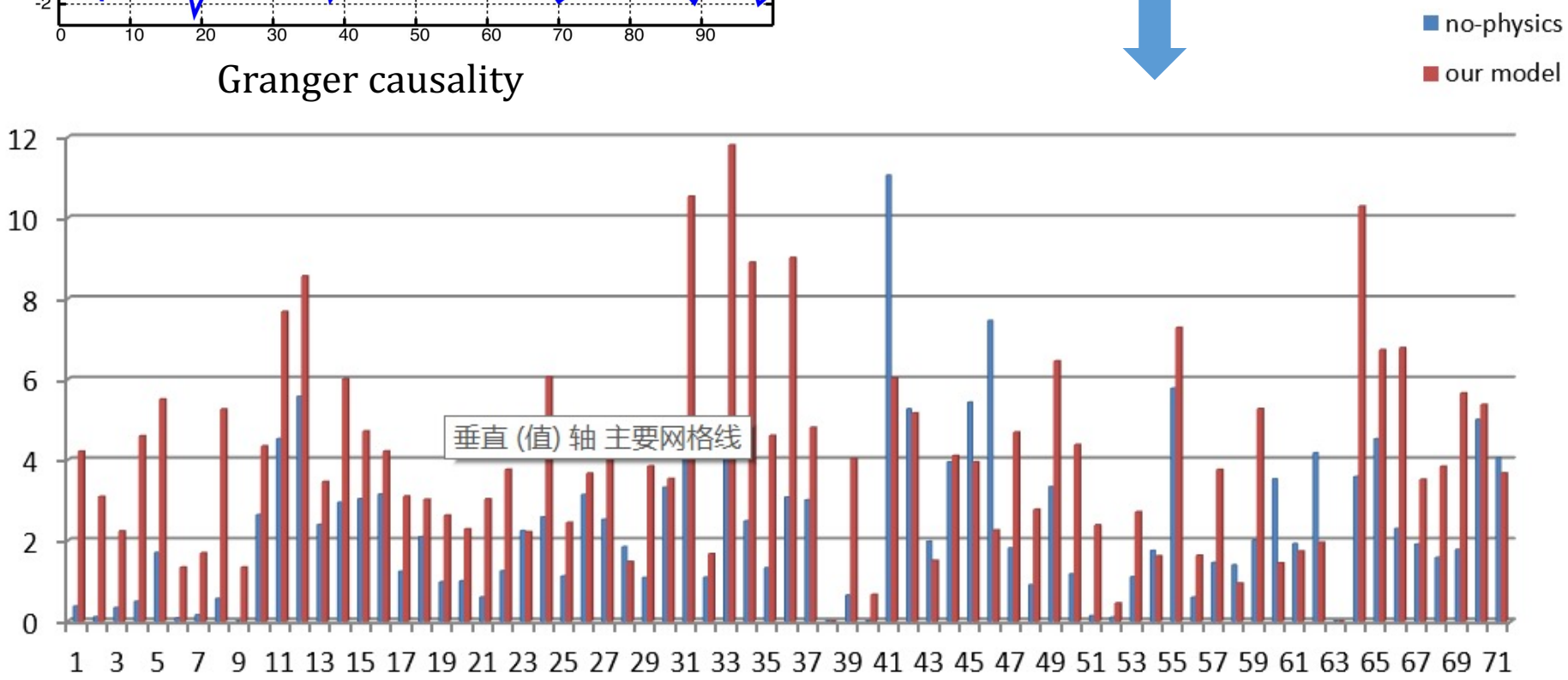


Explain the model



Granger causality

Granger causality
×
Explain Weight



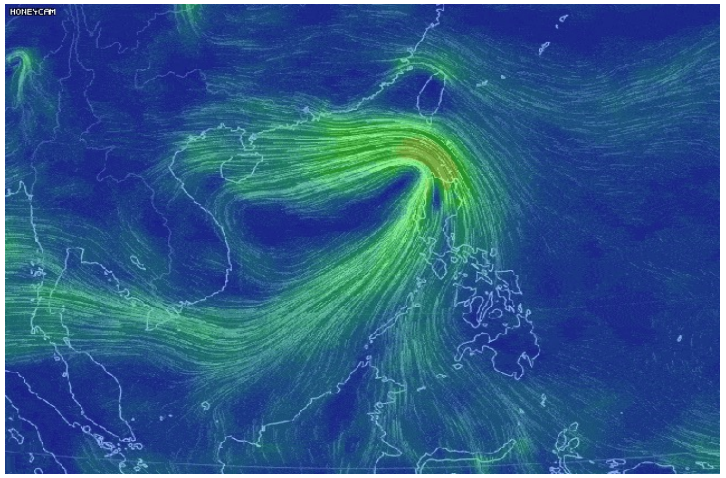
How to integrate theory-driven and data-driven models?

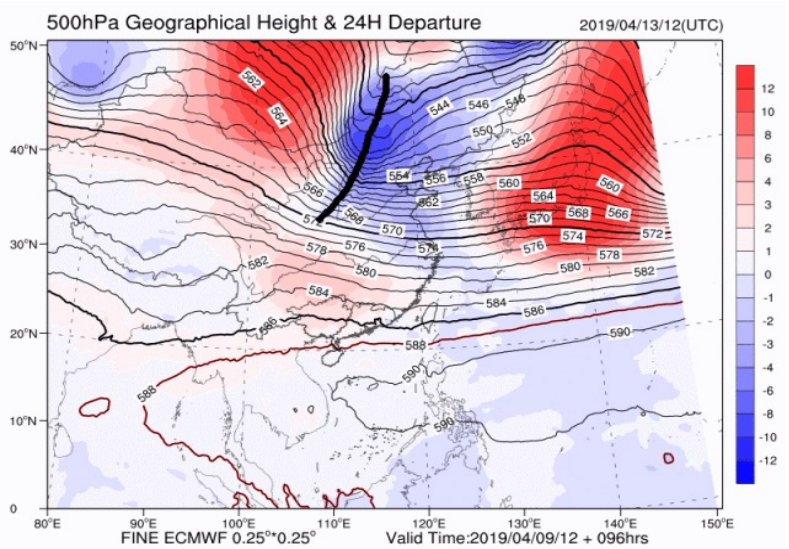
1. Optimizing theory-driven models with data-driven models

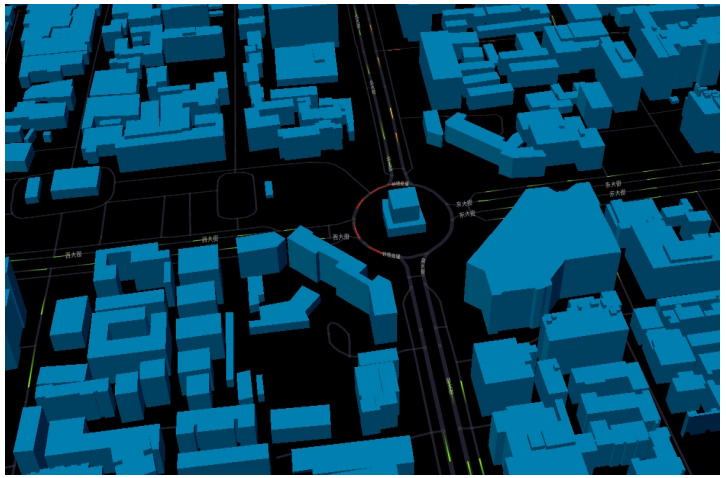
2. Theory-guided data driven models

3. **Data-driven prediction that uncovers dynamic patterns**

Dynamic System



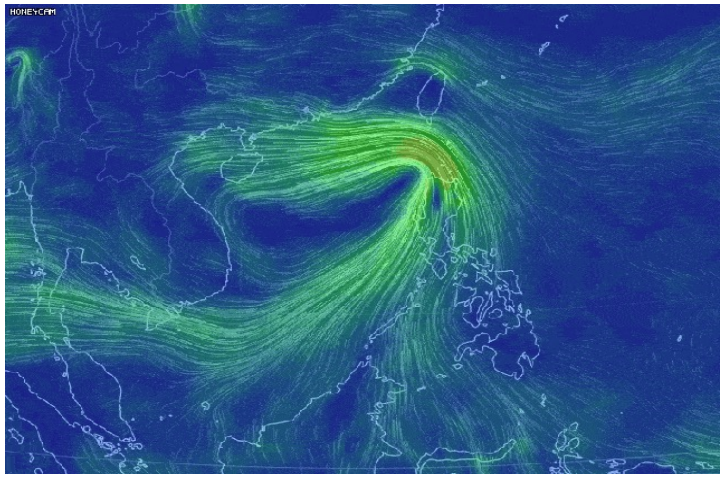


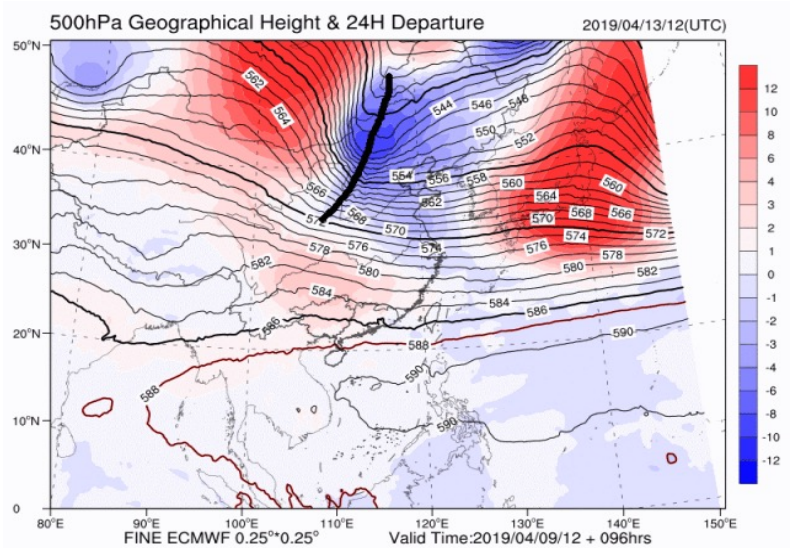


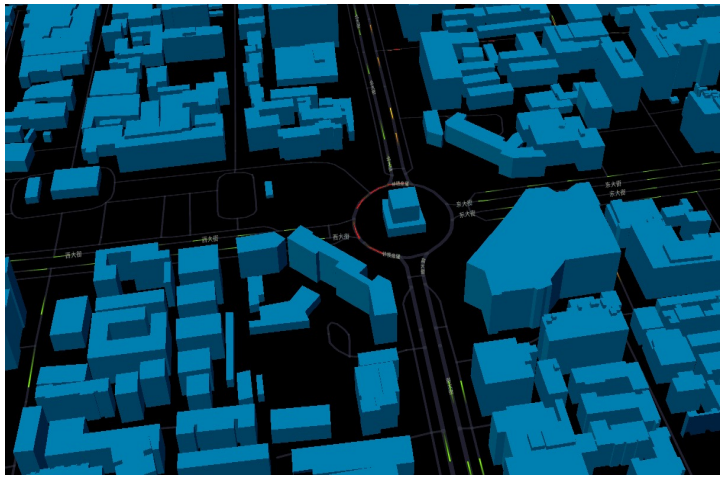
Dynamic System

Gaussian diffusion model

$$c(x,y,z)=\frac{Q}{4\pi u(D_yD_z)^{1/2}}\exp[-\frac{u}{4x}(\frac{y^2}{D_y}+\frac{z^2}{D_z})]$$



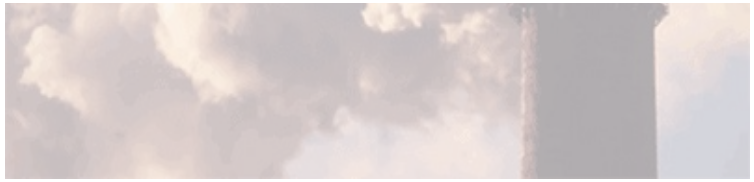




Dynamic System

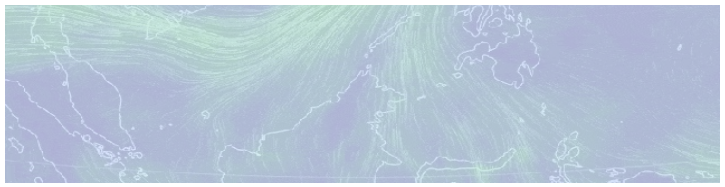
Gaussian diffusion model

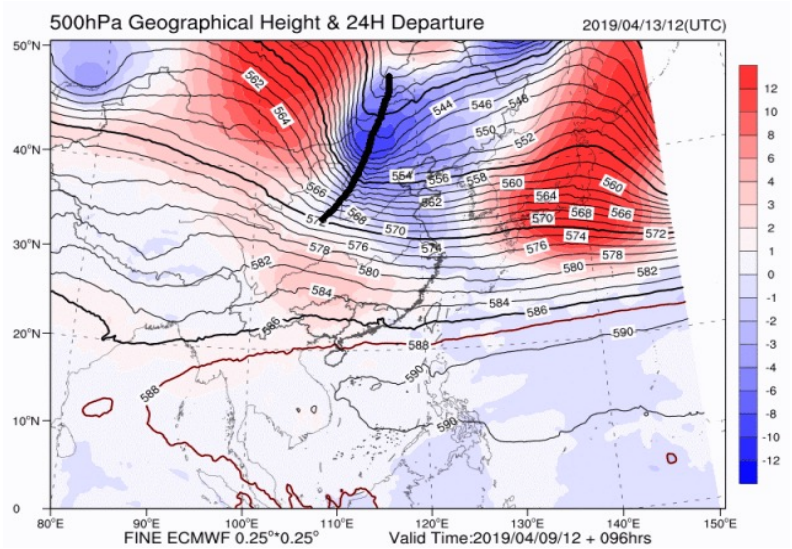
$$c(x,y,z)=\frac{Q}{4\pi u(D_yD_z)^{1/2}}\exp[-\frac{u}{4x}(\frac{y^2}{D_y}+\frac{z^2}{D_z})]$$

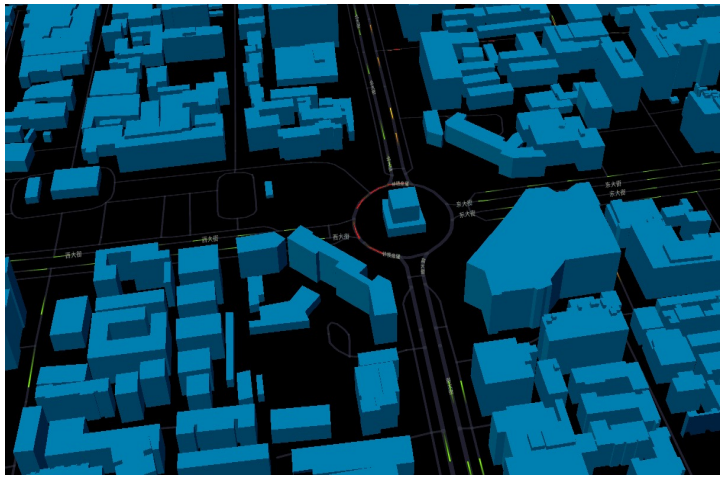


Navier-Stokes equation

$$\partial_t u+(u\cdot\nabla)u=-\nabla p+v\Delta u$$



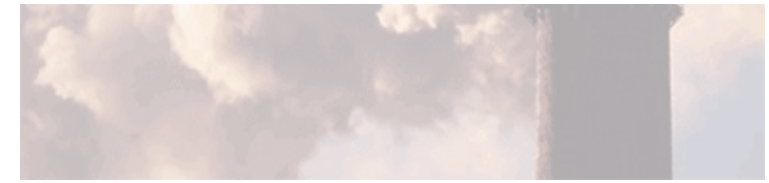




Dynamic System

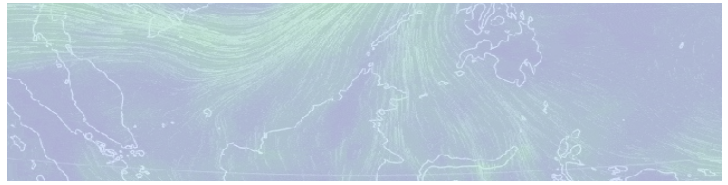
Gaussian diffusion model

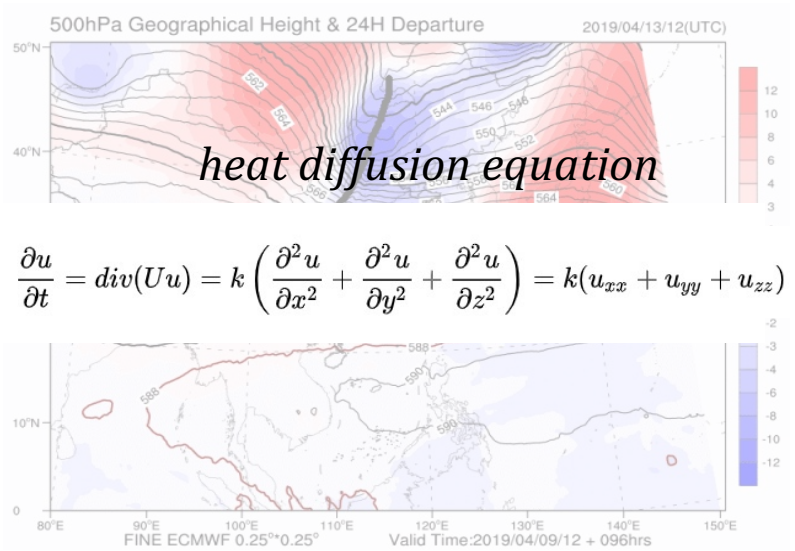
$$c(x,y,z)=\frac{Q}{4\pi u(D_yD_z)^{1/2}}\exp[-\frac{u}{4x}(\frac{y^2}{D_y}+\frac{z^2}{D_z})]$$



Navier-Stokes equation

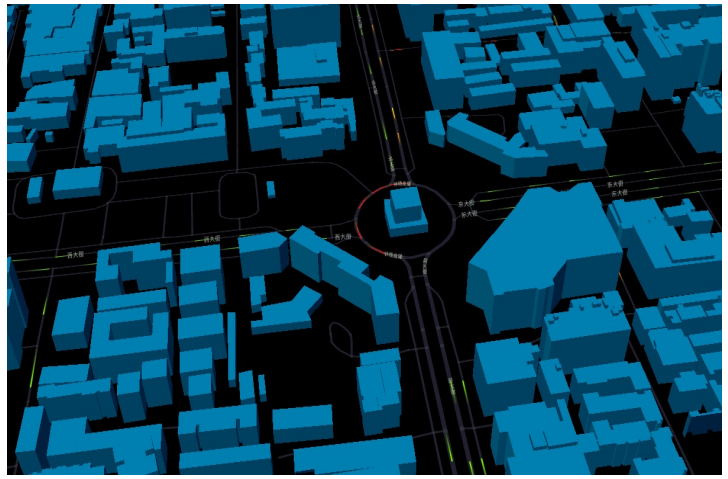
$$\partial_t u+(u\cdot\nabla)u=-\nabla p+v\Delta u$$





heat diffusion equation

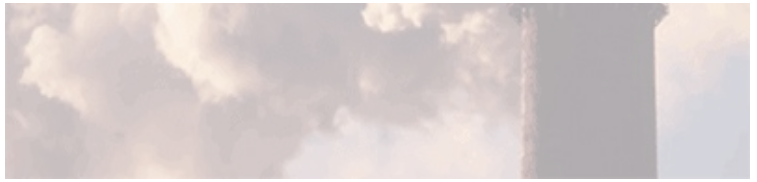
$$\frac{\partial u}{\partial t}=div(Uu)=k\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right)=k(u_{xx}+u_{yy}+u_{zz})$$



Dynamic System

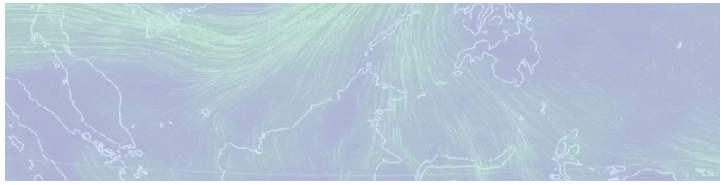
Gaussian diffusion model

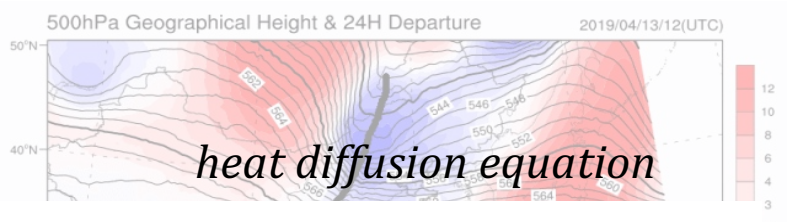
$$c(x,y,z)=\frac{Q}{4\pi u(D_yD_z)^{1/2}}\exp[-\frac{u}{4x}(\frac{y^2}{D_y}+\frac{z^2}{D_z})]$$



Navier-Stokes equation

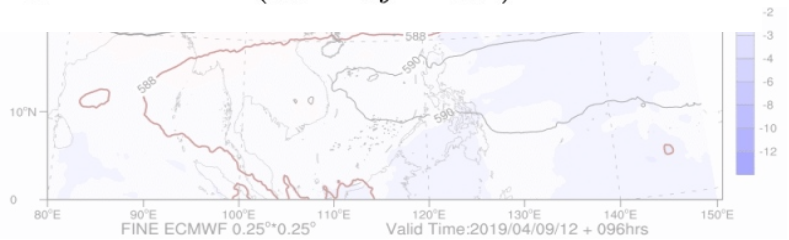
$$\partial_t u+(u\cdot\nabla)u=-\nabla p+v\Delta u$$





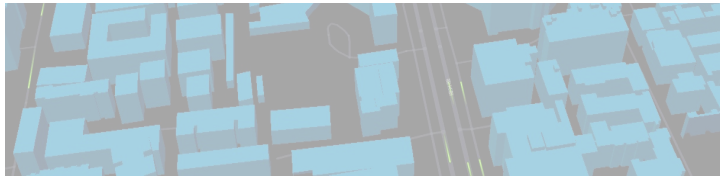
heat diffusion equation

$$\frac{\partial u}{\partial t}=div(Uu)=k\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right)=k(u_{xx}+u_{yy}+u_{zz})$$



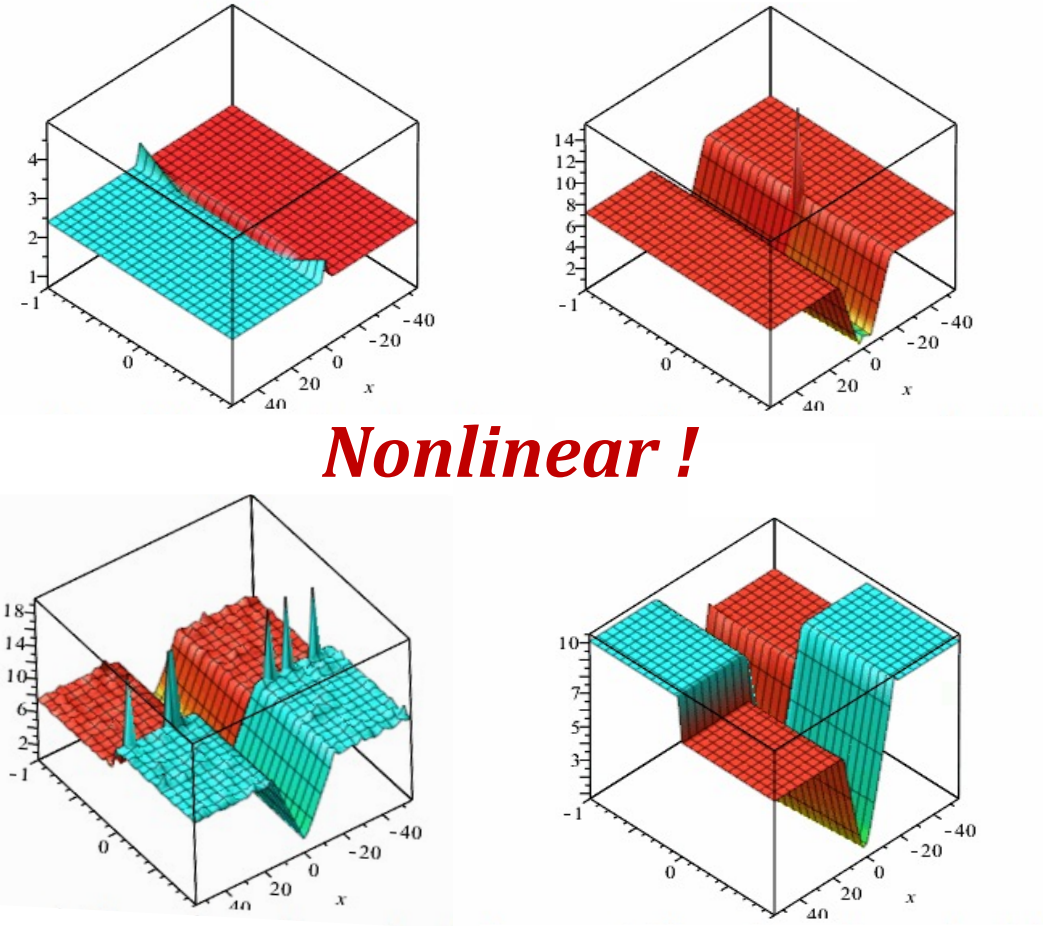
Galloping model

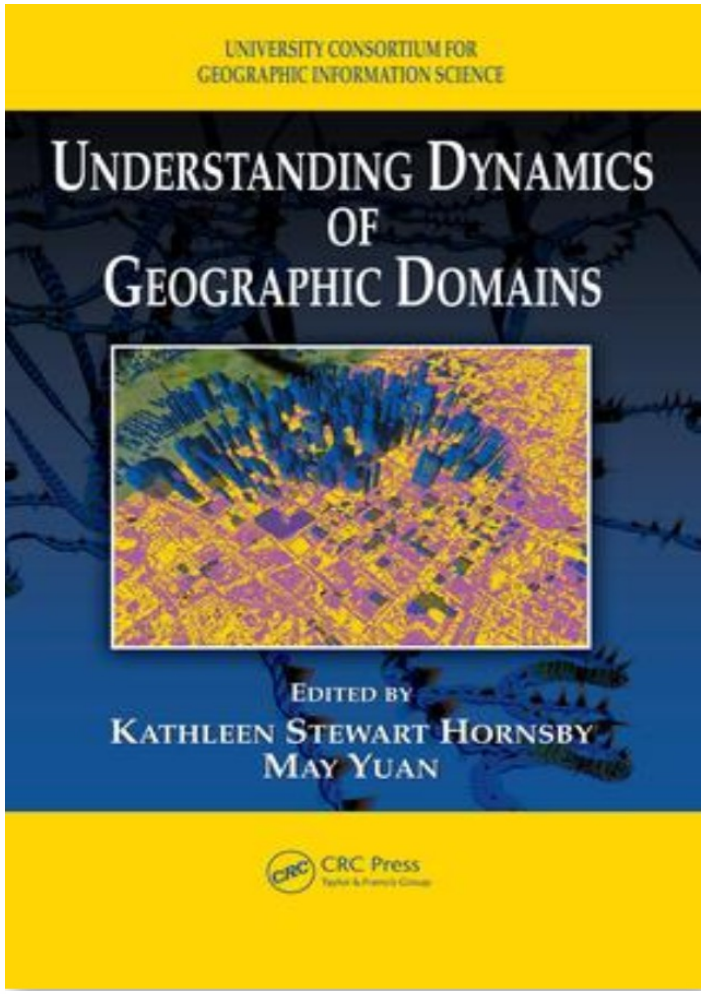
$$\ddot{x}_{n+1}(t+T)=\alpha[\dot{x}_n(t)-\dot{x}_{n+1}(t)]$$



Dynamic System

Different modes in spatio-temporal process





Data-driven forecasting for dynamical system

nature communications

Explore Content ▾ Journal Information ▾ Publish With Us ▾

nature › nature communications › articles › article

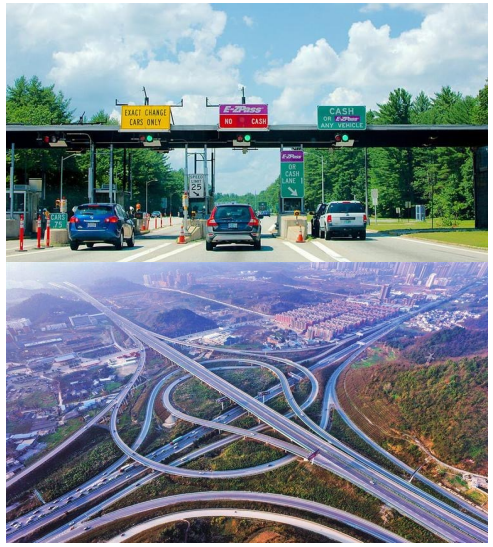
Article | Open Access | Published: 29 April 2020

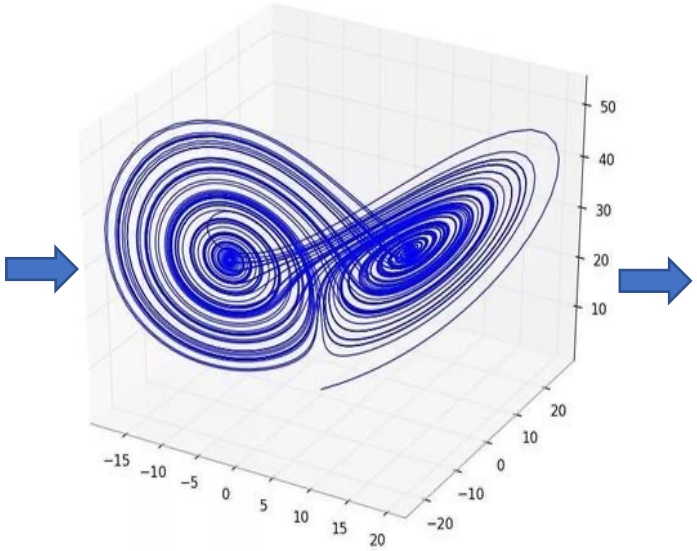
Data-driven analysis and forecasting of highway traffic dynamics

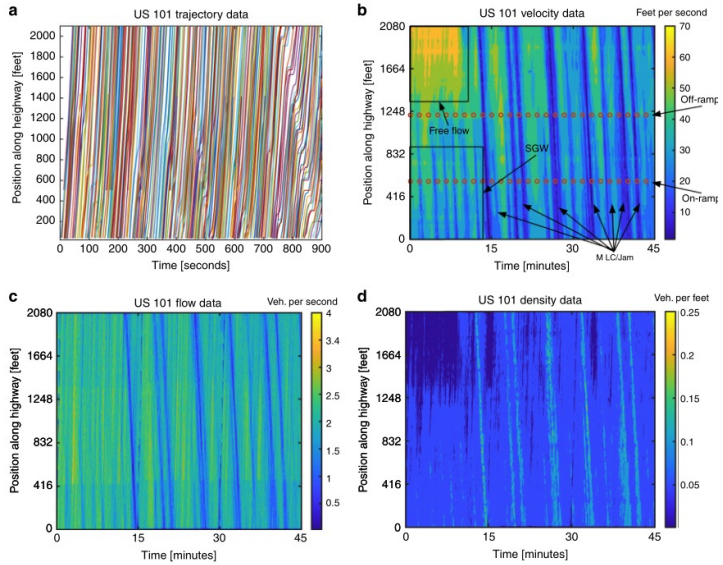
A. M. Avila & I. Mezić

Nature Communications **11**, Article number: 2090 (2020) | Cite this article

4149 Accesses | **4** Citations | Metrics







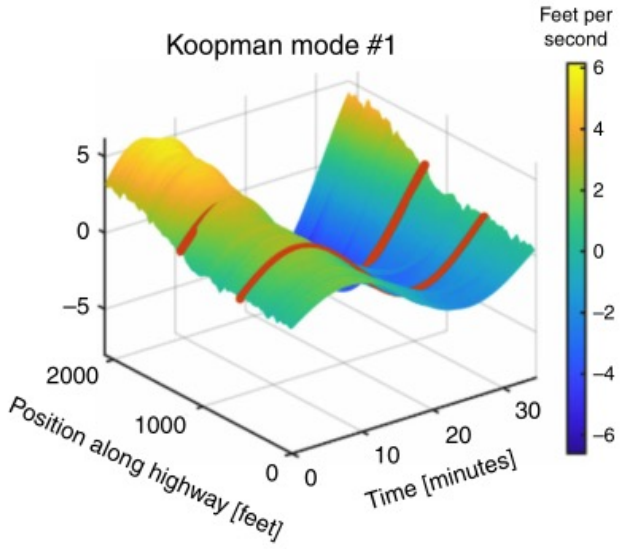
model-free

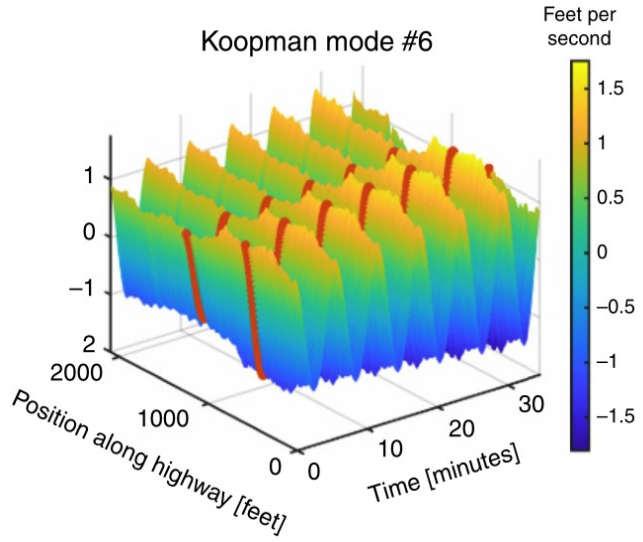
data-driven

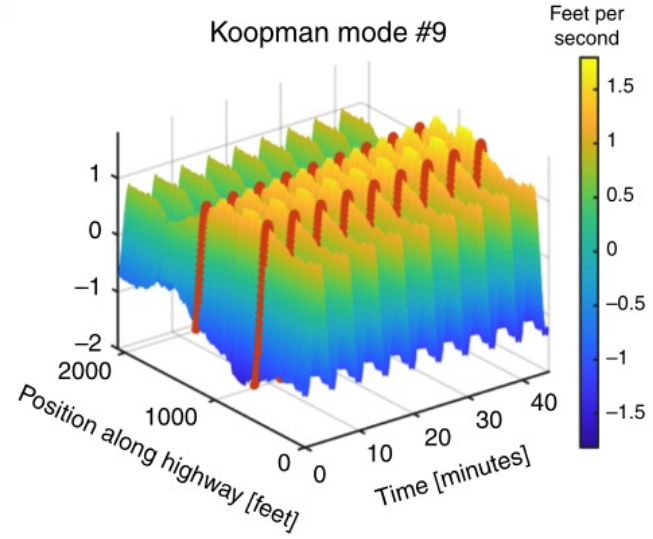
Koopman pattern decomposition

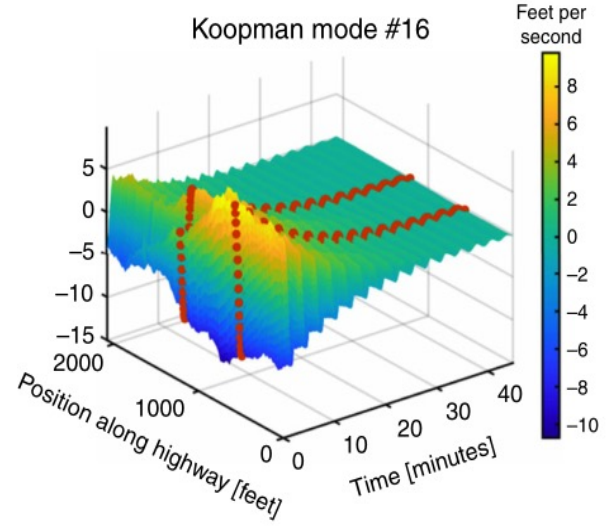
Data-driven forecasting for dynamical system

The Koopman modes uncover complex spatiotemporal wave structures that are hidden within traffic data.



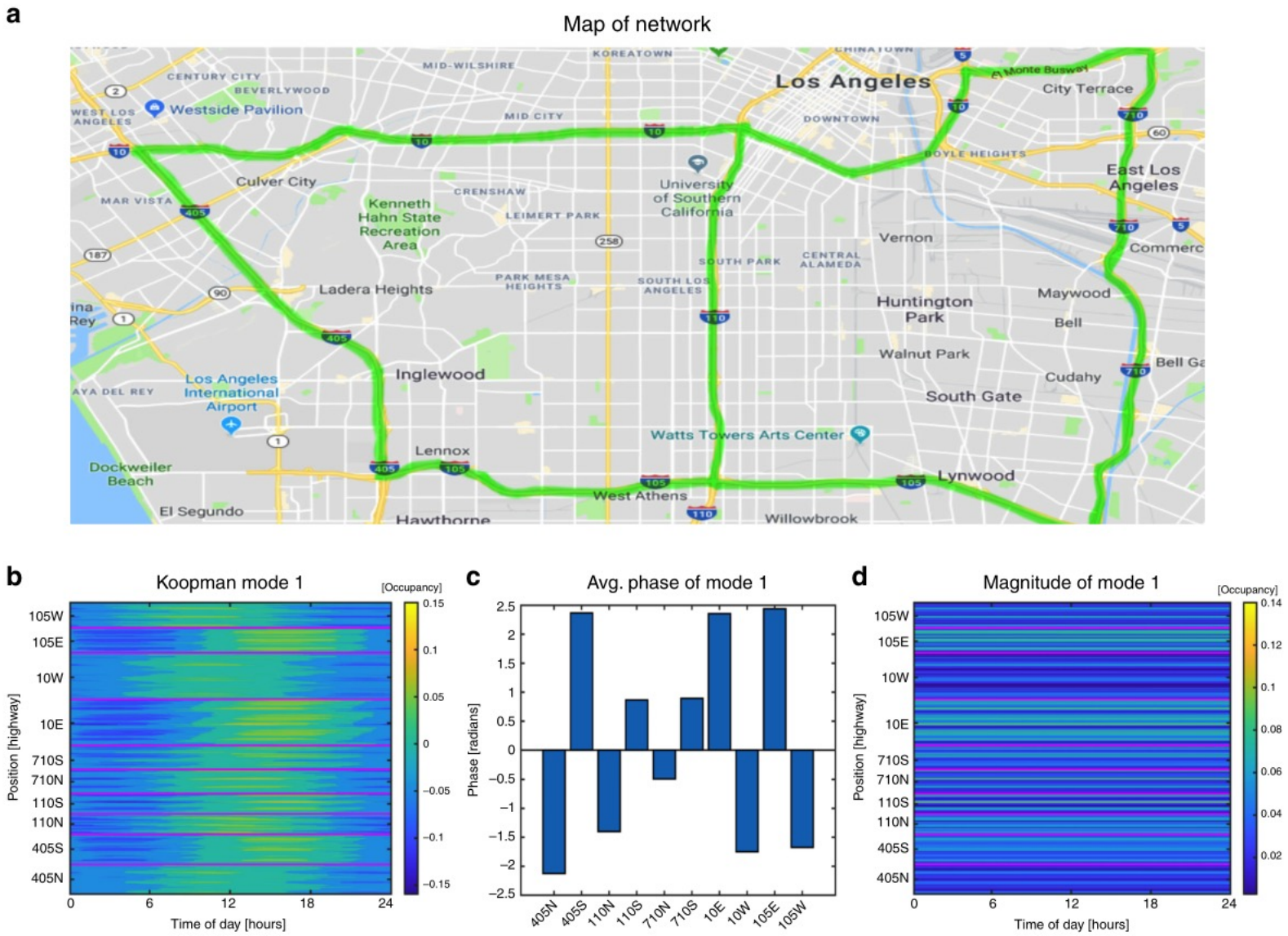






Data-driven forecasting for dynamical system

The 24-hour Koopman mode revealing the order of congestion within the network.



Conclusion

- Data-driven models will not replace theory-driven models, but strongly complement and enrich it.
- Data-driven models will benefit from plausible physically based relationships derived from the natural sciences.
- In the absence of theoretical guidance, data-driven results can guide the direction of future theoretical research

Computational Urban Science

Editors-in-Chief: Hui Lin, Xinyue Ye

Special Issue:
Urban Safety and Resilient Society

Guest editors:
Feng Zhang, Zhejiang University (zfcarnation@zju.edu.cn)
Shuming Bao, China Data Institute (sbao@umich.edu)
Fan Zhang, MIT (zhangfan@mit.edu)

Submission:
Deadline: June 30, 2021
Online System: https://www.editorialmanager.com/cusc/

Springer

In view of these challenges and opportunities, and responding to one of the UN Sustainable Development Goals (SDGs) — Make cities and human settlements inclusive, safe, resilient, and sustainable, this special edition invites original papers on topics related to the theories, methods, solutions, and empirical studies of urban safety in public health, disaster, public security, crime, etc.

Submitted manuscripts could cover but not limited to the following themes:

- Dataset and computational tools for the topics below
- Data-driven approaches and empirical works in disease, epidemic, and pandemic research
- Risk and vulnerability assessment of urban systems with urban big data
- Prevention and early warning systems for urban disaster
- Urban disaster reduction and emergency management
- Road traffic safety and transit-oriented development
- Human behavior and safety perception analysis
- Quantitative methods for built-environment measurement
- Security assessment for critical public infrastructure
- Urban crime mapping and analysis
- Urban resilience and sustainable development practices

浙江大学
ZHEJIANG UNIVERSITY

地球科学学院
人文学院
艺术与考古学院
建工学院
计算机科学与技术学院

十一届空间综合人文学与社会科学论坛

地点 | 浙江·杭州

时间 | 2021年 7月7日—9日