西安交通大學
XI'AN JIAOTONG UNIVERSITY

International Workshop on Deep Learning and Numerical Methods for PDEs

AI for Geophysics：Unlocking Insights into Subsurface and Climate

Zhiguo Wang

Xi'an Jiaotong University

June 22, 2024

Xi'an City

OUTLINE

1、Introduction

2、Our research

3、Conclusion and Future

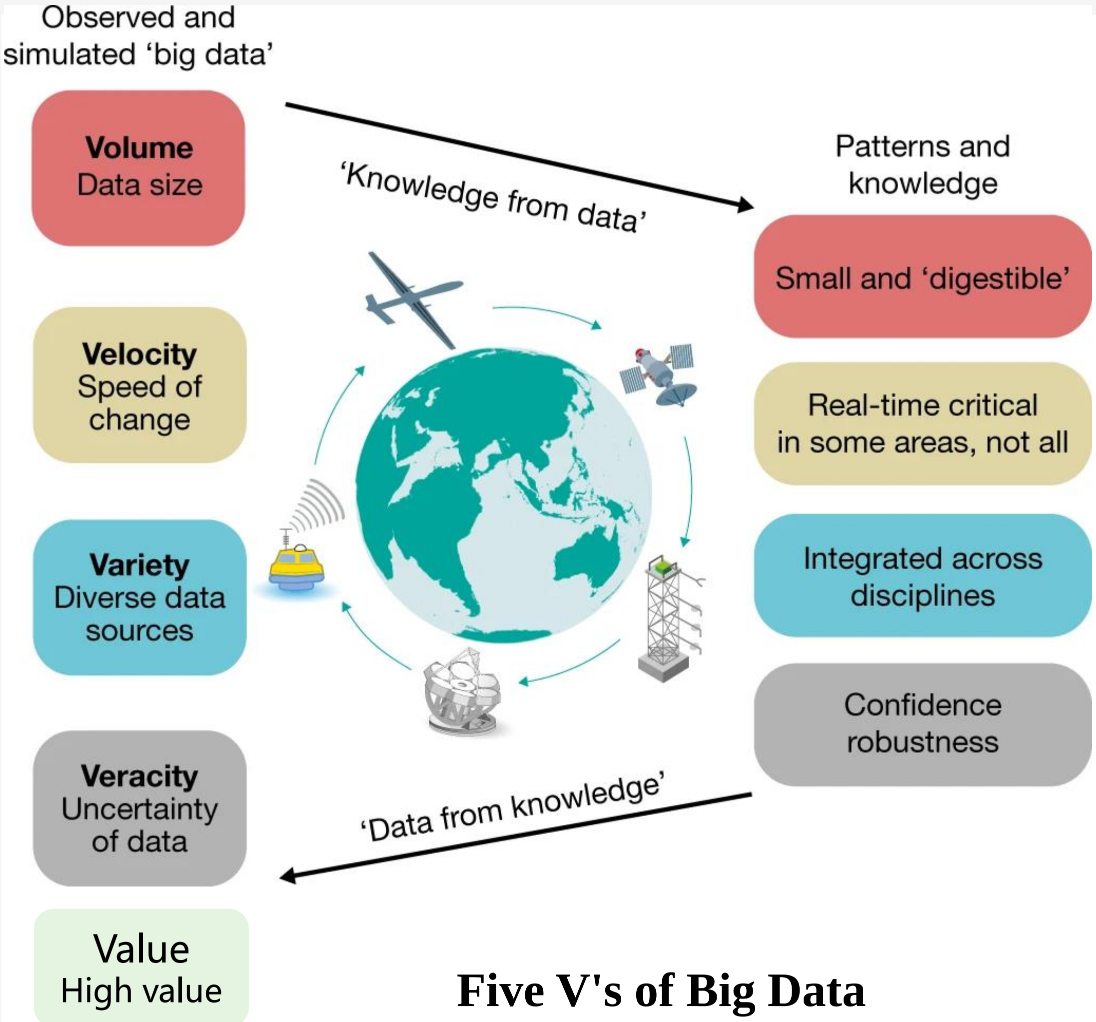
西安交通大學
XI'AN JIAOTONG UNIVERSITY

1、Introduction

[1]Tiffany C. Vance et al., **Big data in Earth science: Emerging practice and promise**. Science 383, eadh9607(2024)
[2]S. Mostafa Mousavi, Gregory C. Beroza, **Deep-learning seismology**. Science 377, eabm4470(2022)
[3]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)

Big data in the geoscience

- Data size now exceeds **100 petabytes**, and is growing quasi-exponentially
- Speed of change exceeds **5 petabytes a year**
- Data are taken **at frequencies of up to 10 Hz** or more

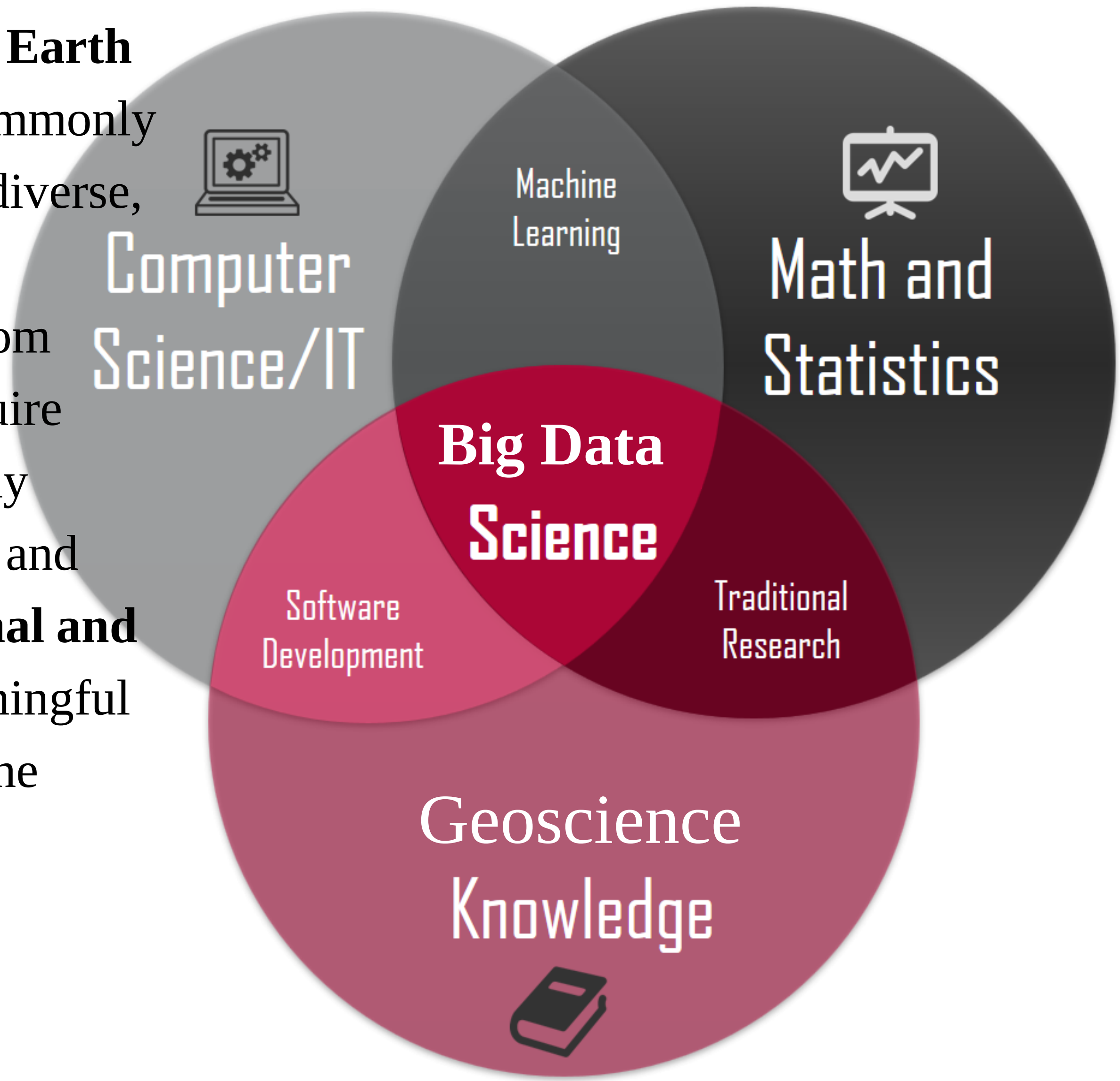


Five V's of Big Data

- **Data sources** are one- to four-dimensional, spatially integrated, from the organ level to the global level
- **Diverse observational systems**, from remote sensing to in situ observation
- **Uncertainty** of data can stem from observational errors or conceptual inconsistencies

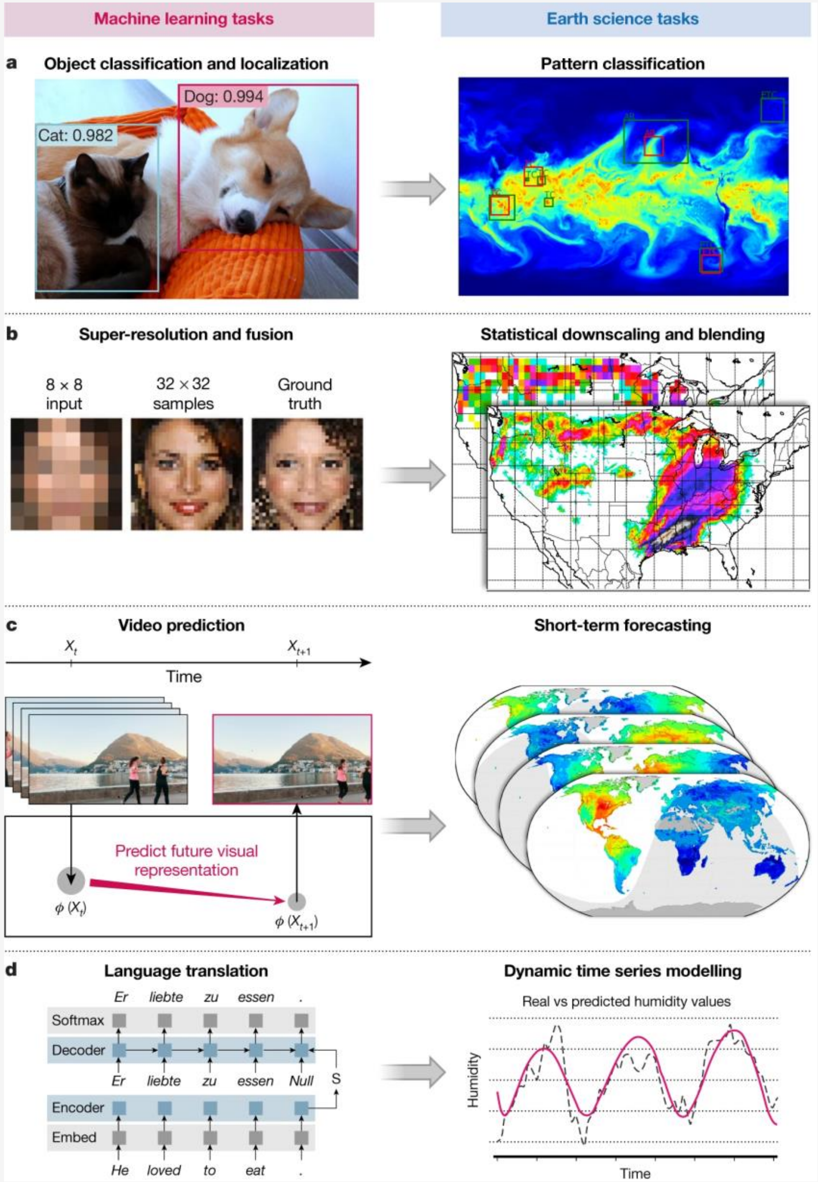
Big data in the geoscience

Tiffany et al. (2024) define **Big Earth Data** as massive (relative to commonly available datasets) amounts of diverse, complex, and continuously accumulating data generated from heterogeneous sources that require **advanced** (relative to commonly available computing resources) and **potentially novel computational and analytical tools** to extract meaningful insights and knowledge about the Earth system.



AI (Artificial Intelligence)->Machine Learning->Deep Learning->**Deep Neural Network**

AI in the geoscience



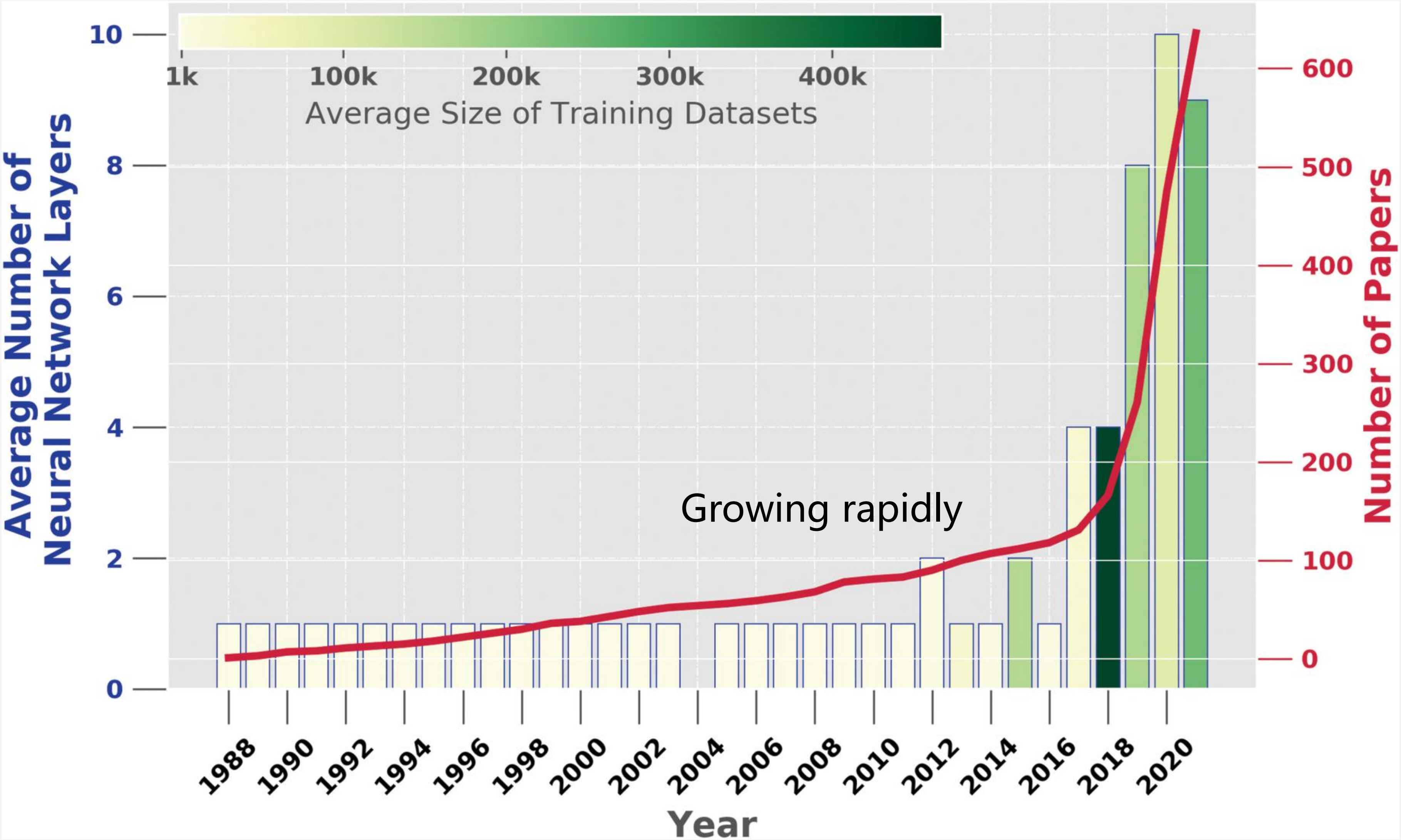
Object recognition in images links to **classification of extreme weather patterns** using a unified convolutional neural network on climate simulation data

Super-resolution applications relate to statistical downscaling of climate model output

Video prediction is similar to short-term forecasting of Earth system variables

Language translation links to modelling of dynamic time series

AI in the geoscience



For example： deep learning applications in seismology

AI in the geoscience

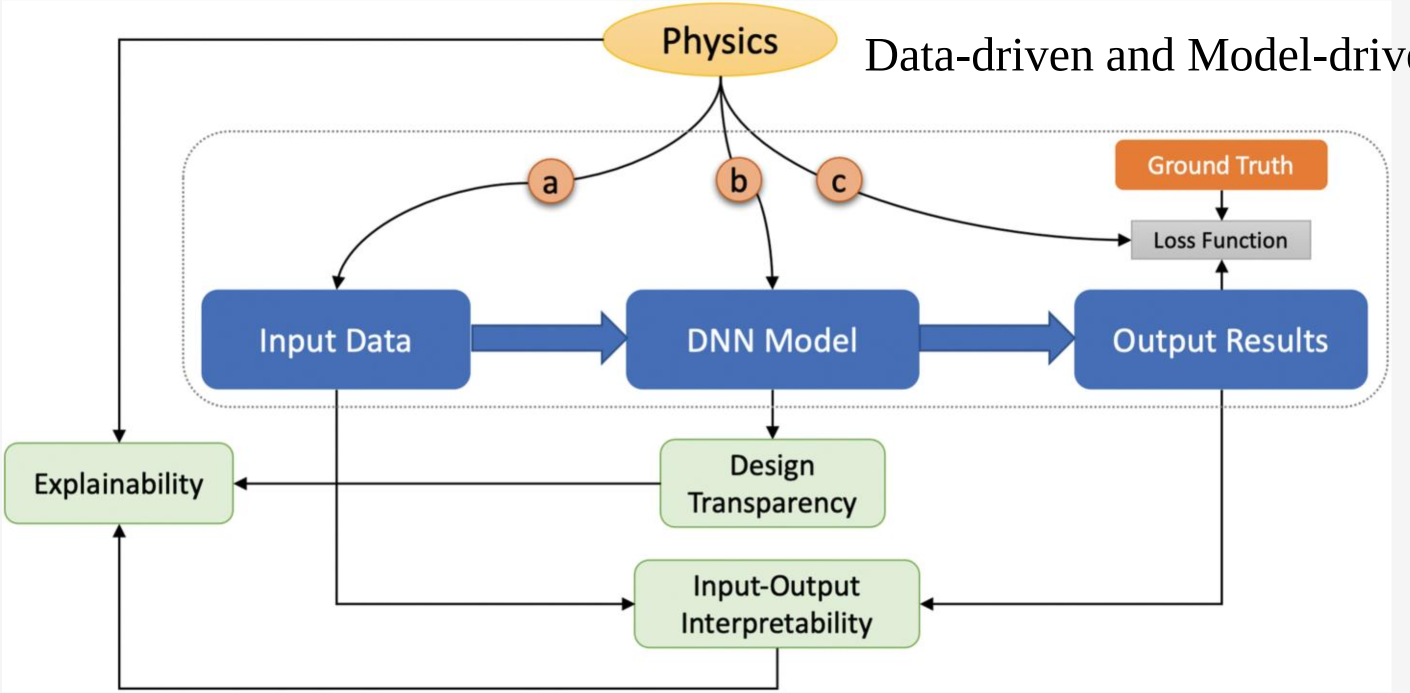
Deep Learning Challenges:

1) Interpretability; 2) Physical consistency; 3) Complex and uncertain data;

4) Limited labels; 5) Computational demand

Data-driven and Model-driven

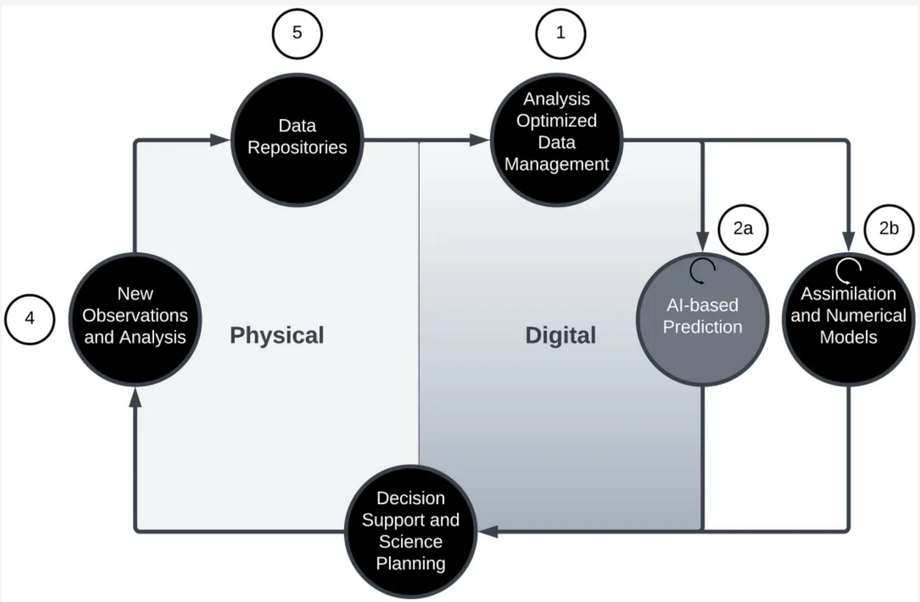
Integrate data and mathematical physics models, even if only partially understood.



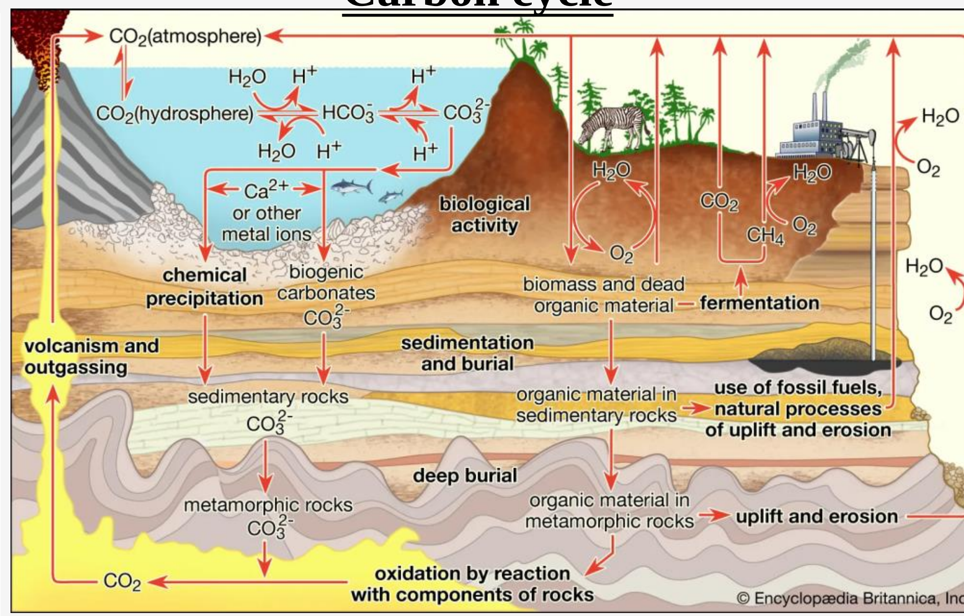
The domain knowledge, physics, can be incorporated into the training data through (a) data simulation, (b) the hypothesis space by designing specialized neural network architectures, and (c) the training procedure of a deep-learning model by enforcing consistency with governing equations.

AI in the geoscience

Digital twins of the Earth system are intended to establish highly accurate digital representations of the Earth



Carbon cycle



AI plays an essential role in DTs: relevant data, analysis, and numerical models as well as resource management.

AI formalizes the process to learn from the past to improve the accuracy of future predictions.

AI-based models require ongoing training and continuous validation.

Physics-based models are essential to forecast the environment’s reactions.

2、Our research

- **Full Wave Inversion**
- **Magnitude Estimation in Earthquake**
- **Temperature Prediction in Climate**

- **Full Wave Inversion**
- **Magnitude Estimation in Earthquake**
- **Temperature Prediction in Climate**

Full Wave Inversion

Full waveform inversion (FWI) is a data matching procedure which attempts to use the whole wavefield to obtain quantitative information about the subsurface.

Tarantola (1984); Bunks et al.(1995); Hu et al. (2009); Burstedde et al. (2009); Symes (2009); Virieux et al.(2009); Tran et al. (2013); Lin et al. (2015), ……

Model-driven

Nonlinear Inverse Problem

- Low computational efficiency
- Depend on the initial model
- Ill-posed

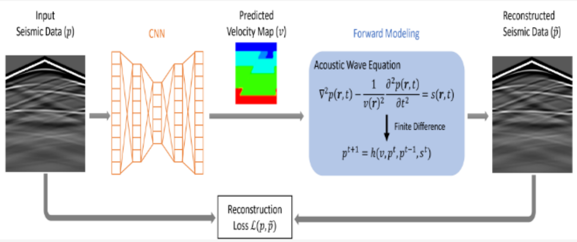
- Efficient
- Not dependent on initial model

Data-driven

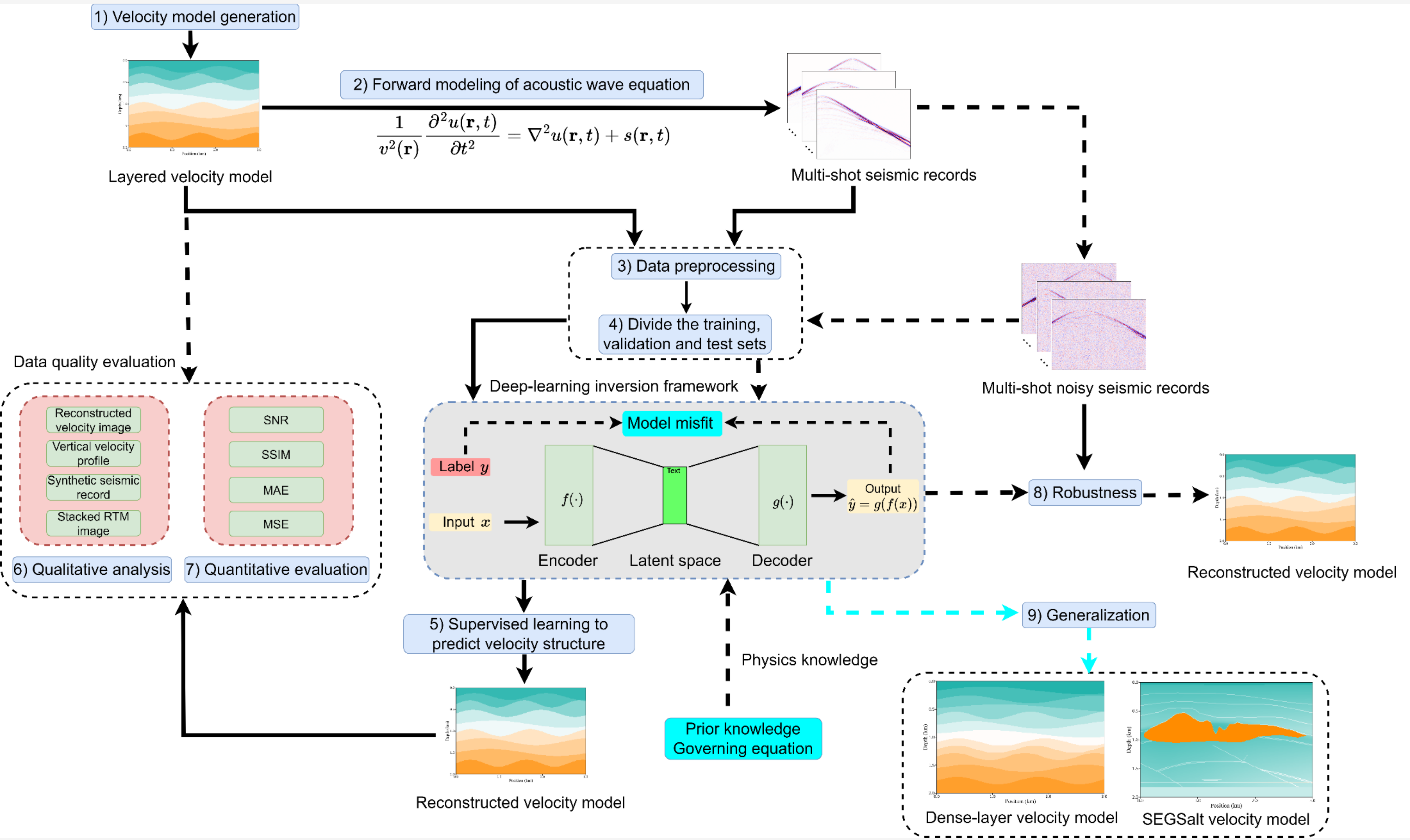
CNN: Wang et al. (2018); Yang et al. (2019), Wu et al. (2020)
CNN+FC: Li et al. (2020), Wang et al. (2020), Liu et al. (2021)
GAN: Zhang et al. (2020)
Transformer: Wang et al. (2023)
……

PINN: Sun et al. (2021); Jin et al. (2021); Zhu et al. (2022)
Data augmentation: Rojas-Gómez et al. (2022), Feng et al. (2022)
Regularization: Zhang et al. (2020), He et al. (2021), Feng et al. (2022), Zhu et al. (2022)
……

Physical- and data- driven



Full Wave Inversion

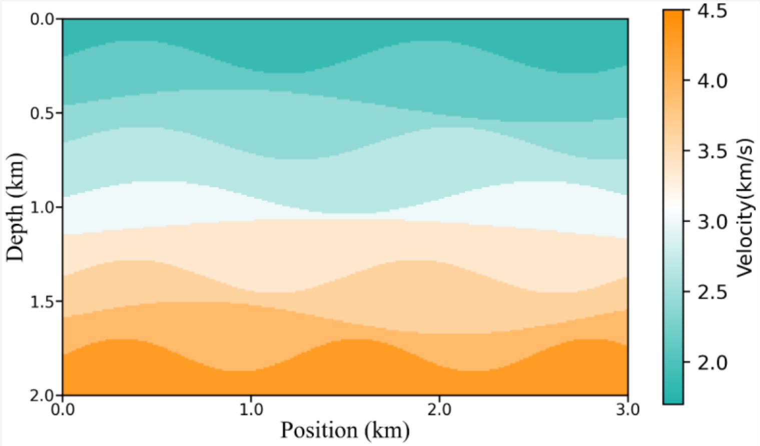


Workflow of our work

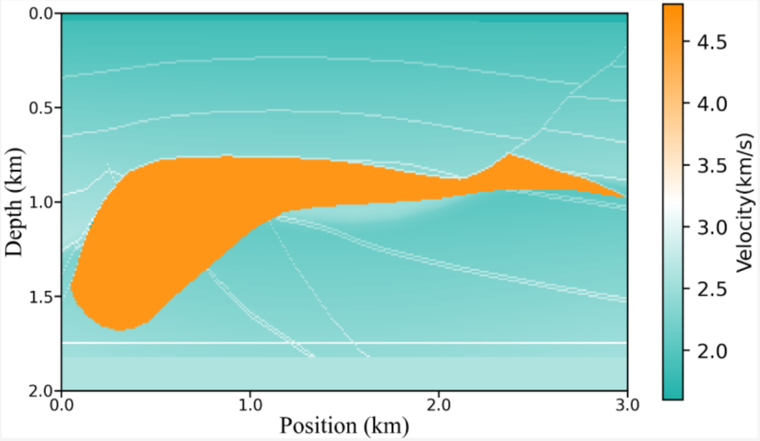
Full Wave Inversion

Wave equation $$\frac{1}{v^2(r)}\frac{\partial^2 u(r,t)}{\partial t^2} = \nabla^2 u(r,t) + s(r,t)$$

Velocity model



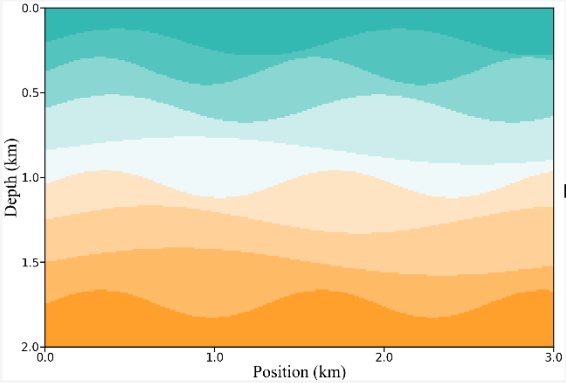
Layered model



SEG Salt model

$$F = \sum_{j=1}^{N} k_j \sin\left(f_j \times \pi \times ix/nx\right)$$

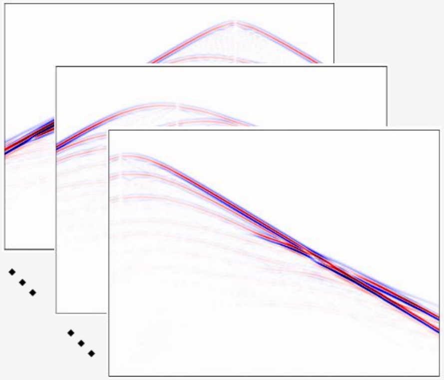
Full Wave Inversion



2D velocity model

Forward modeling

$$\frac{1}{v^2(r)}\frac{\partial^2 u(r,t)}{\partial t^2} = \nabla^2 u(r,t) + s(r,t)$$



Seismic records

Time-domain staggered-grid finite-difference method

- First-order partial derivatives in space

$$\frac{\partial \mathrm{u}}{\partial x} \approx \frac{1}{\Delta x}\sum_{k=1}^{\mathrm{N}} c_k\left(u\left(t, x + \frac{(2k-1)\Delta x}{2}\right) - u\left(t, x - \frac{(2k-1)\Delta x}{2}\right)\right)$$

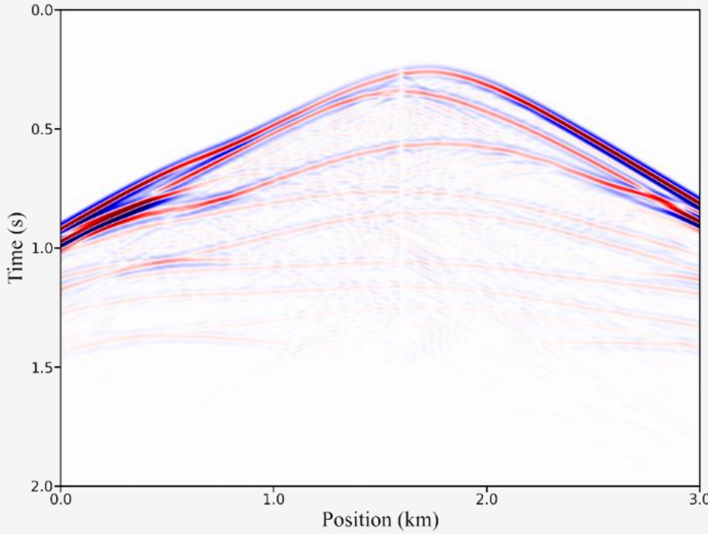
- Second-order partial derivative in time

$$\frac{\partial^2 u}{\partial t^2} \approx \frac{u(t+\Delta t) - 2u(t) + u(t-\Delta t)}{\Delta t^2}$$

Full Wave Inversion

Data preprocessing

$D \in \mathbb{R}^{T \times W}$

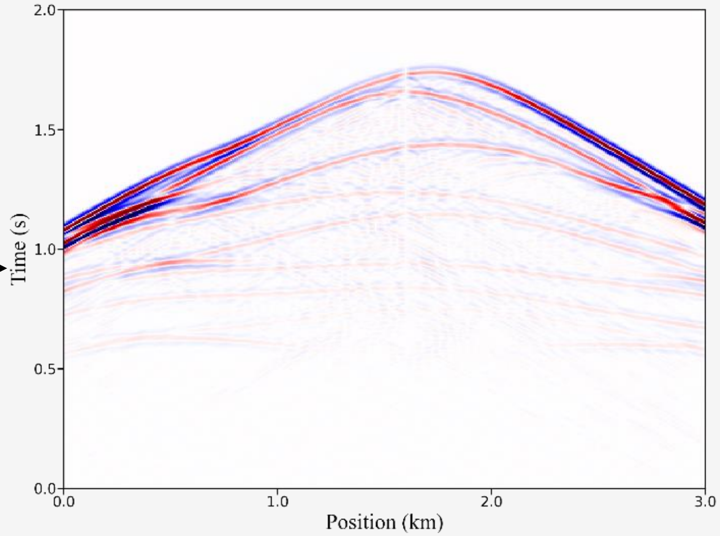


Original seismic record

Downsampling

uniform sampling along the time

$D \in \mathbb{R}^{\frac{T}{r} \times W}$



Sampling seismic records

Reduce the computational load of deep networks in training

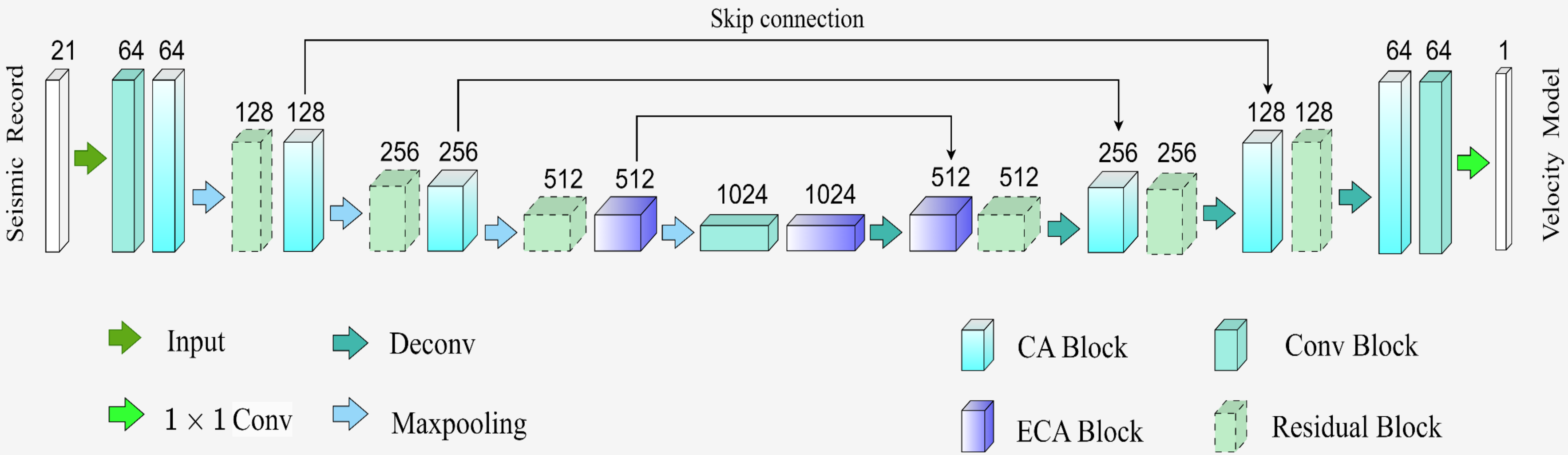
Full Wave Inversion

Method

Supervised learning, nonlinear inversion operator

$$M \approx \Gamma(D)$$

$$D \in \mathbb{R}^{C\times\frac{T}{r}\times W}, M \in \mathbb{R}^{H\times W}$$



Residual network embedded with dual attention mechanisms (DA-ResNet)

- Efficient channel attention
- Coordinate attention

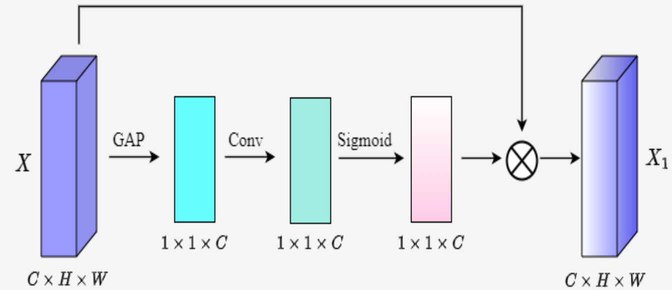
Full Wave Inversion

Method

➢ ECA (Efficient Channel Attention): capture channel information from seismic recordings

$$z_c = \frac{1}{H \times W} \sum_i \sum_j x_c(i,j)$$

$$\omega = \sigma\left(C_{1D}(z)\right) \quad X_1 = X \cdot \omega$$



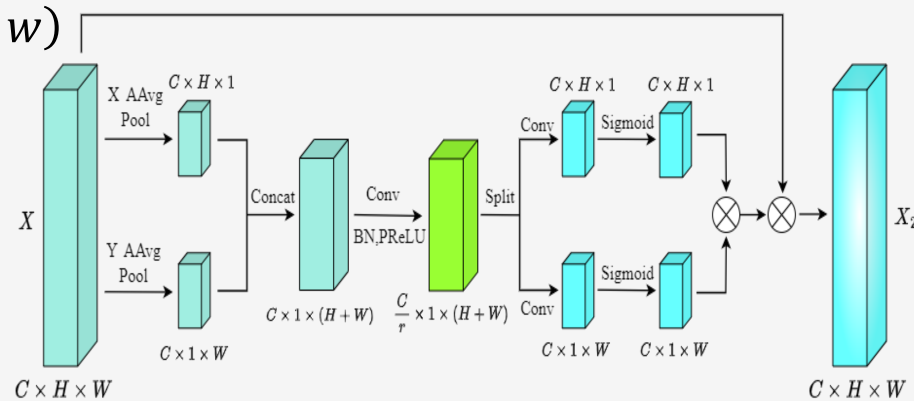
➢ CA (Coordinate Attention): encode the position information of waveform data

$$z_c^h(h) = \frac{1}{W} \sum_j x_c(h,j) \qquad z_c^w(w) = \frac{1}{H} \sum_j x_c(i,w)$$

$$F = PReLU\left(C_1\left([z^h, z^w]\right)\right)$$

$$g^h = \sigma\left(C_h(F^h)\right) \qquad g^w = \sigma\left(C_w(F^w)\right)$$

$$(x_2)_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j)$$



Full Wave Inversion

- Cyclic learning rat

$$\eta(\tau) = \frac{\eta_{max}}{2}\left(1 + \cos(\frac{\pi \ mod \ (\tau - 1, \lceil T/M \rceil)}{\lceil T/M \rceil})\right)$$

$$\hat{\eta}(\tau) = \begin{cases} \eta_0 & 0 < \tau \leq \quad T_0 \\ \eta(\tau) & T_0 < \tau \leq T \end{cases}$$

- Loss function

$$L(m, \widehat{m}) = \frac{1}{K \times H \times W} \sum_i (\lambda \|m_i - \widehat{m}_i\|_1 + (1 - \lambda)\|m_i - \widehat{m}_i\|_2^2)$$

K: the mini-batch size;

H and W:height and width of the velocity model

L_1 norm:
sparse reflection

CLR-FWI

- Activation function

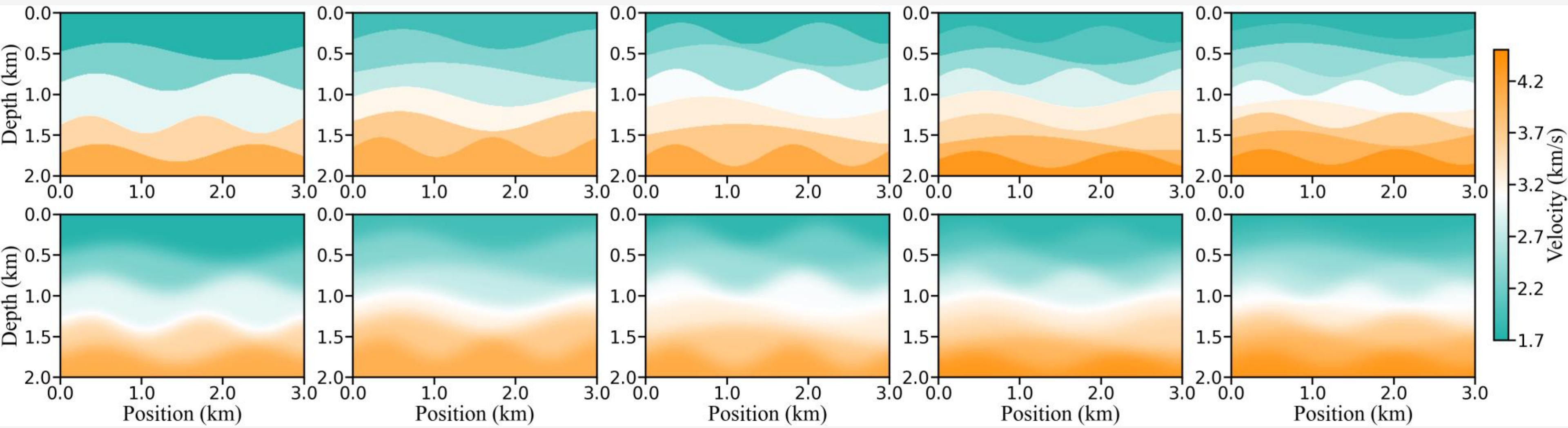
$$PReLU(x) = \max(0, x) + \beta_i min(0, x)$$

β_i is the learnable parameter of the ith channel.

Full Wave Inversion

Experiment data (velocity model and seismic data**)**

- **Velocity Model**: Randomly synthesize 6000 velocity models with varying complexity. each with a 201×301 grid (10 m × 10 m spacing). 5 to 9 layers, with interval velocities from 1700 to 4500 m/s.

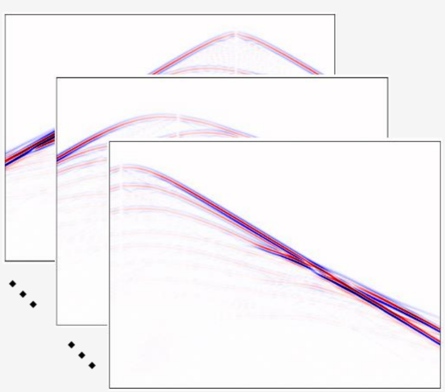


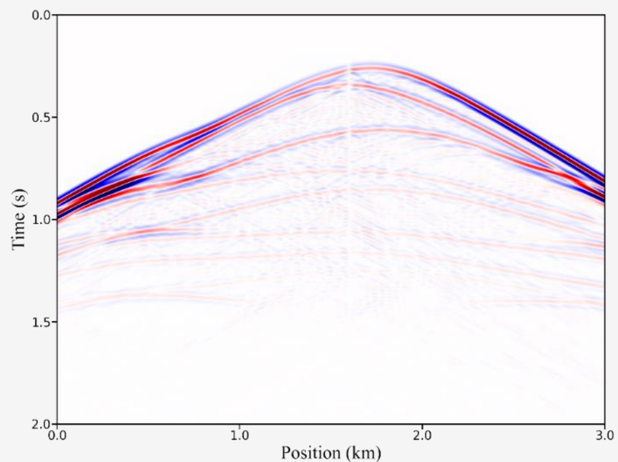
Training set: 5000 samples**, validation set:** 500 samples**, test set**: 500 samples

Full Wave Inversion

Experiment data (velocity model and seismic data**)**

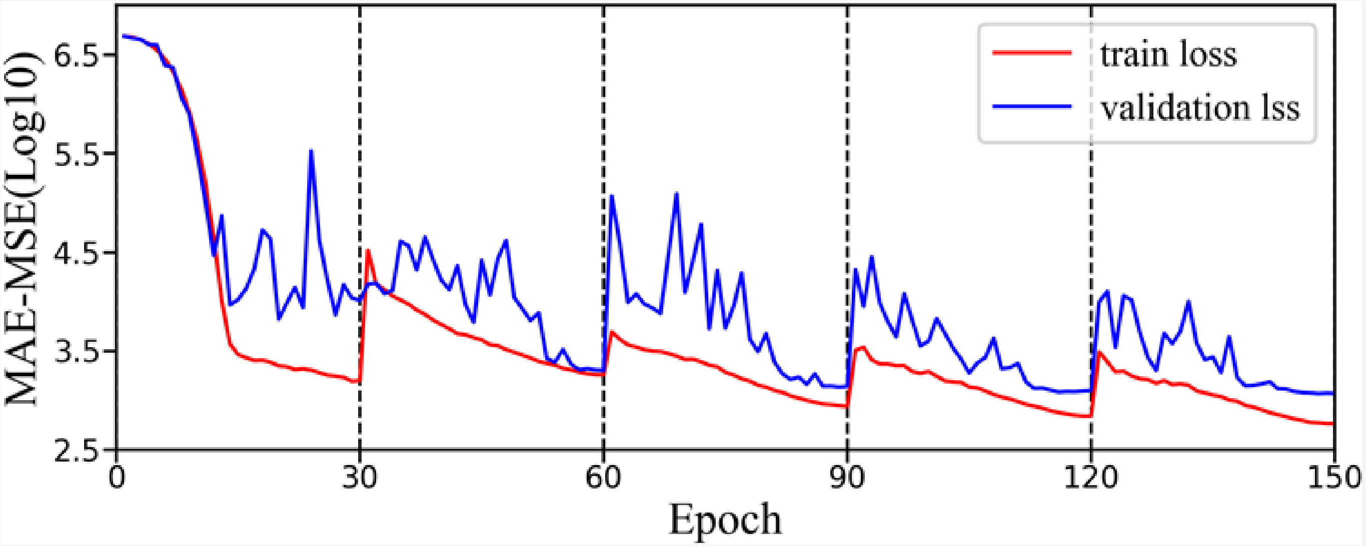
- **Seismic data**: For each model, 30 seismic sources with a 100 m interval. 301 receivers with a space interval of 10 m. 20 Hz Ricker wavelet as the source wavelet. The simulated seismograms have a 2 s record length sampled at 1 ms (2000 samples).





Full Wave Inversion

Experiment results



CLR-FWI loss curve

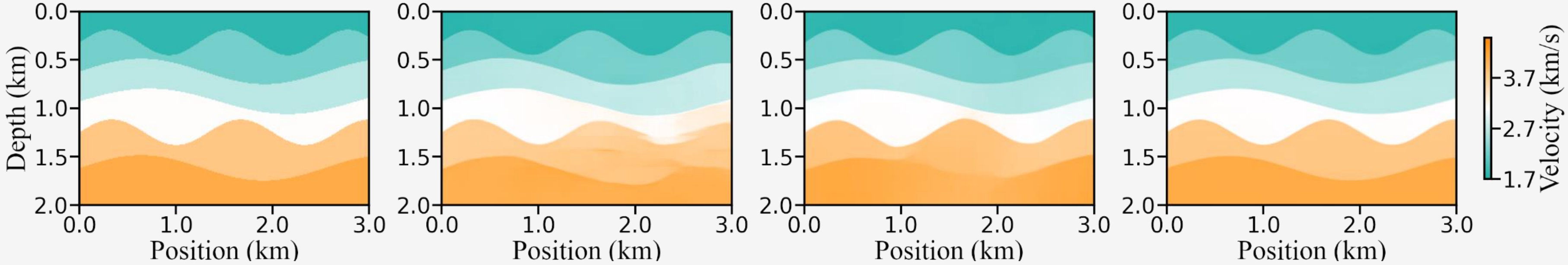
Evaluation metrics: MAE (Mean absolute error), MSE (mean square error)), SNR(signal-to-noise ratio), and SSIM (structural similarity index)

$$MAE = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \left| p_{ij} - g_{ij} \right| \quad MSE = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \left| p_{ij} - g_{ij} \right|^2$$

$$SNR = 20 \log_{10} \frac{\|g\|_2}{\|g - p\|_2} \quad SSIM(p, g) = \frac{(2\mu_g \mu_p + C_1)(2\sigma_{gp} + C_2)}{(\mu_g^2 + \mu_p^2 + C_1)(\sigma_g^2 + \sigma_p^2 + C_2)}$$

Full Wave Inversion

Experiment results



Ground Truth

SNR: 31.50
SSIM: 0.7441

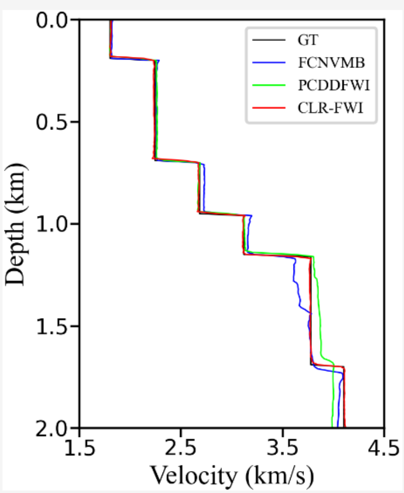
SNR: 33.13
SSIM: 0.8061

SNR: 39.50
SSIM: 0.9230

FCNVMB[1]

PCDDFWI[2]

CLR-FWI



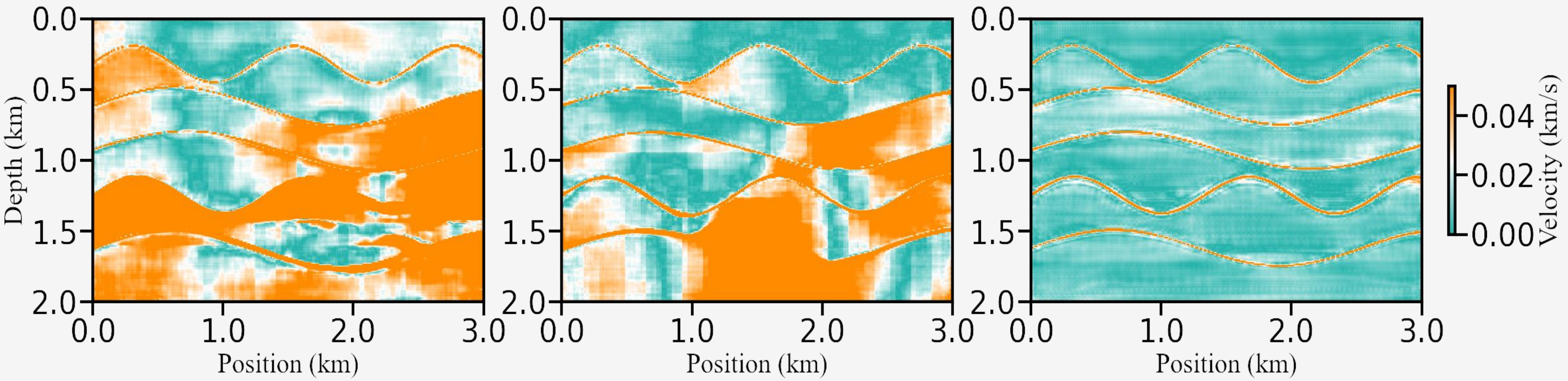
Methods	MAE	MSE	SNR	SSIM
FCNVMB[1]	51.45	8276.76	31.23	0.7292
PCDDFWI[2]	44.40	6556.67	32.27	0.7774
CLR-FWI	19.21	2083.54	37.15	0.8861

[1] F. Yang and J. Ma. Deep-learning inversion: A next-generation seismic velocity model building method. Geophysics, 84, R583–R599, 2019
[2] R. Rojas-Gomez. Physics-consistent data-driven waveform inversion with adaptive data augmentation. IEEE Geoscience and Remote Sensing Letters, 19, 2022, Art no. 8001305

Full Wave Inversion

Experiment results

Absolute model misfits between ground truth models and reconstructed results



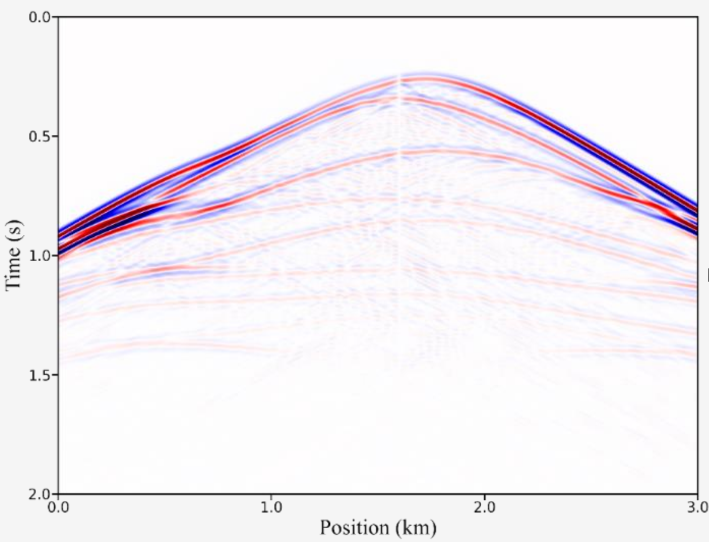
FCNVMB PCDDFWI CLR-FWI

Full Wave Inversion

Experiment results

Noisy data：

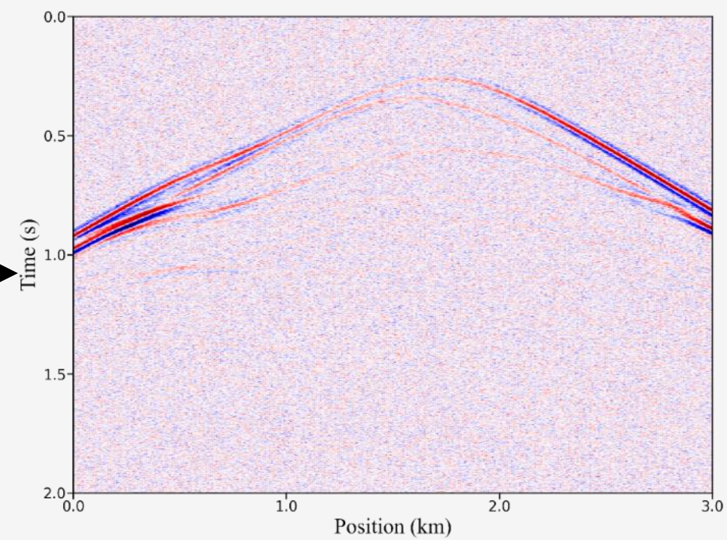
$$D_{noise} = \left(\frac{D_{clean}}{\max(abs(D_{clean}))} + G_{noise}(\mu, \sigma)\right) \times \max(abs(D_{clean}))$$



Original data

Gaussian noise

mean $\mu = 0$
standard deviation
$\sigma = 0.05$



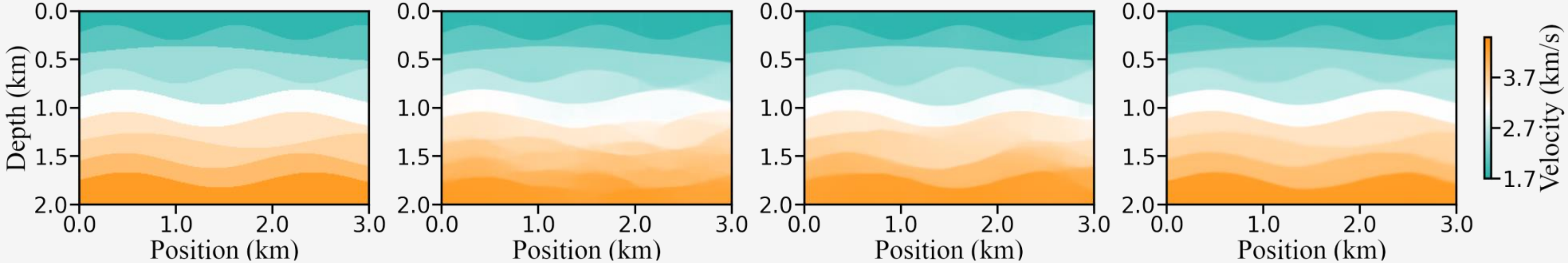
Noisy data

Full Wave Inversion

Experiment results

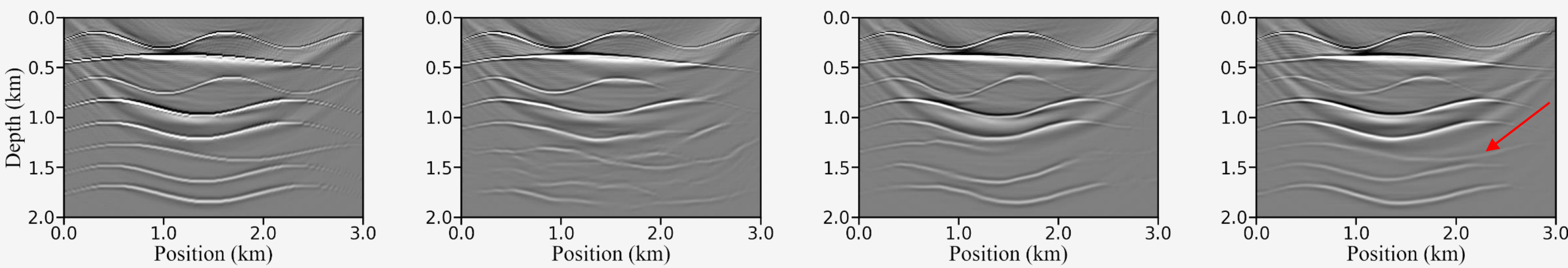
Robustness

- Reconstructed velocity



Ground Truth FCNVMB PCDDFWI CLR-FWI

- Reverse time migration (RTM)



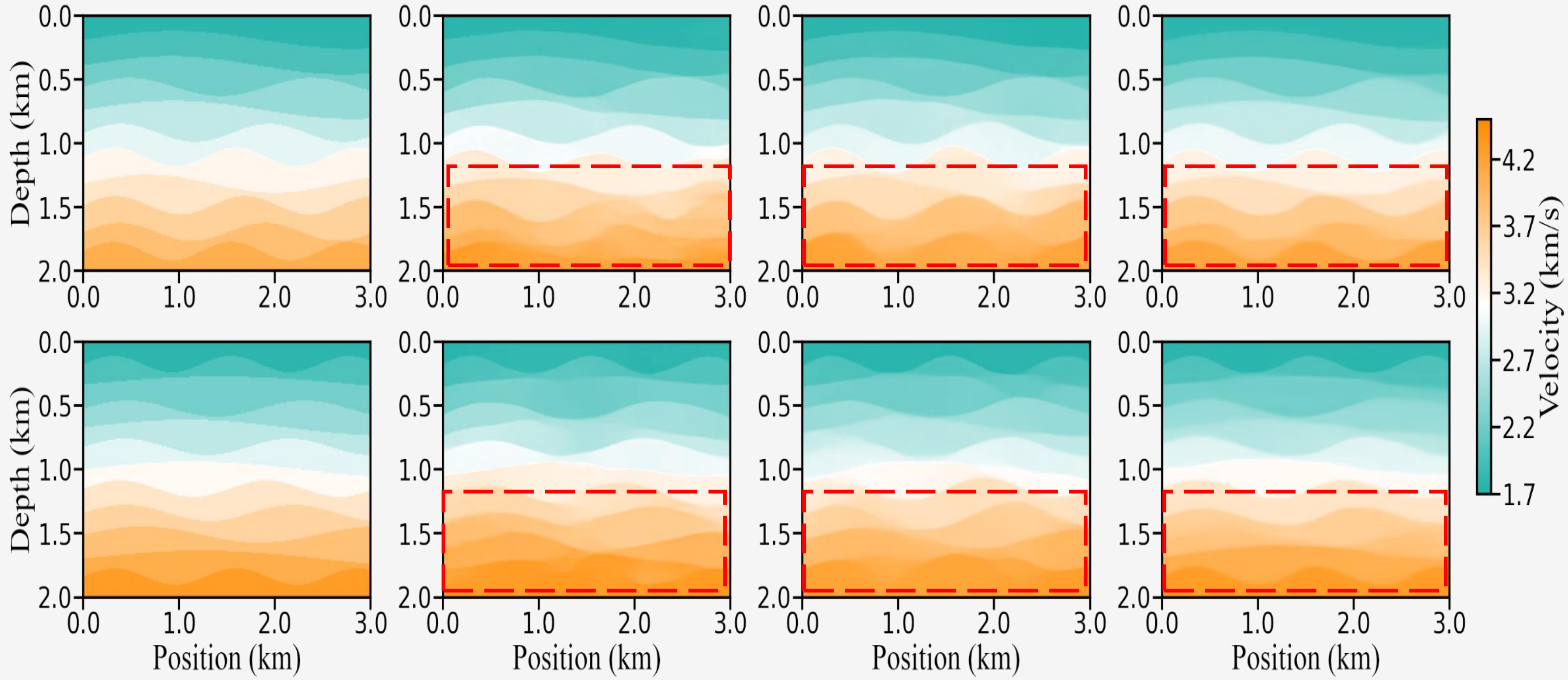
Ground Truth FCNVMB PCDDFWI CLR-FWI

Full Wave Inversion

Experiment results

Generalization

Dense-layer models: 440 velocity models, 10 to 12 layers



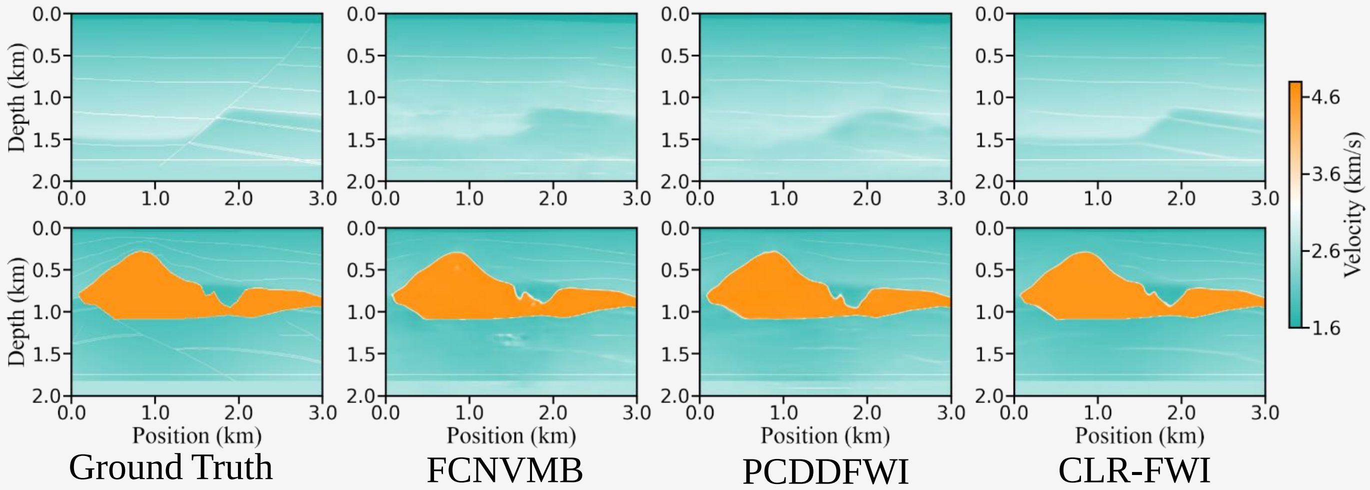
Ground Truth FCNVMB PCDDFWI CLR-FWI

Full Wave Inversion

Experiment results

Generalization

SEG salt models: velocity from 1500 to 4482 m/s, 440 training samples, 110 test samples

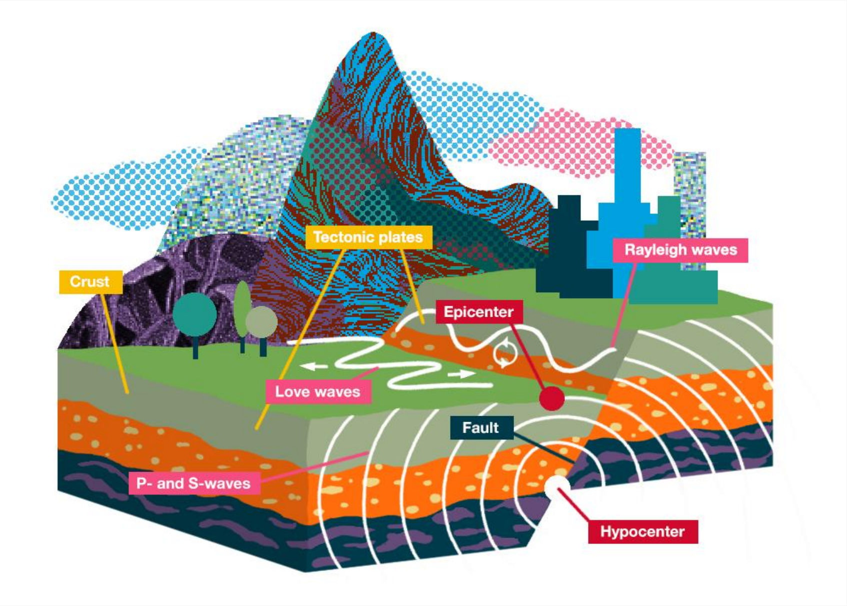


Methods	MAE	MSE	SNR	SSIM
FCNVMB	36.77	16632.83	27.28	0.7258
PCDDFWI	40.96	16185.02	27.38	0.7202
CLR-FWI	26.36	11571.68	28.82	0.8206

- Full Wave Inversion
- **Magnitude Estimation in Earthquake**
- Temperature Prediction in Climate

Earthquake magnitude estimation

Magnitude: the relative amount of energy released by an earthquake event



An earthquake is the shaking of the Earth's surface resulting from a sudden release of energy in the lithosphere that creates seismic waves.

Earthquake magnitude estimation

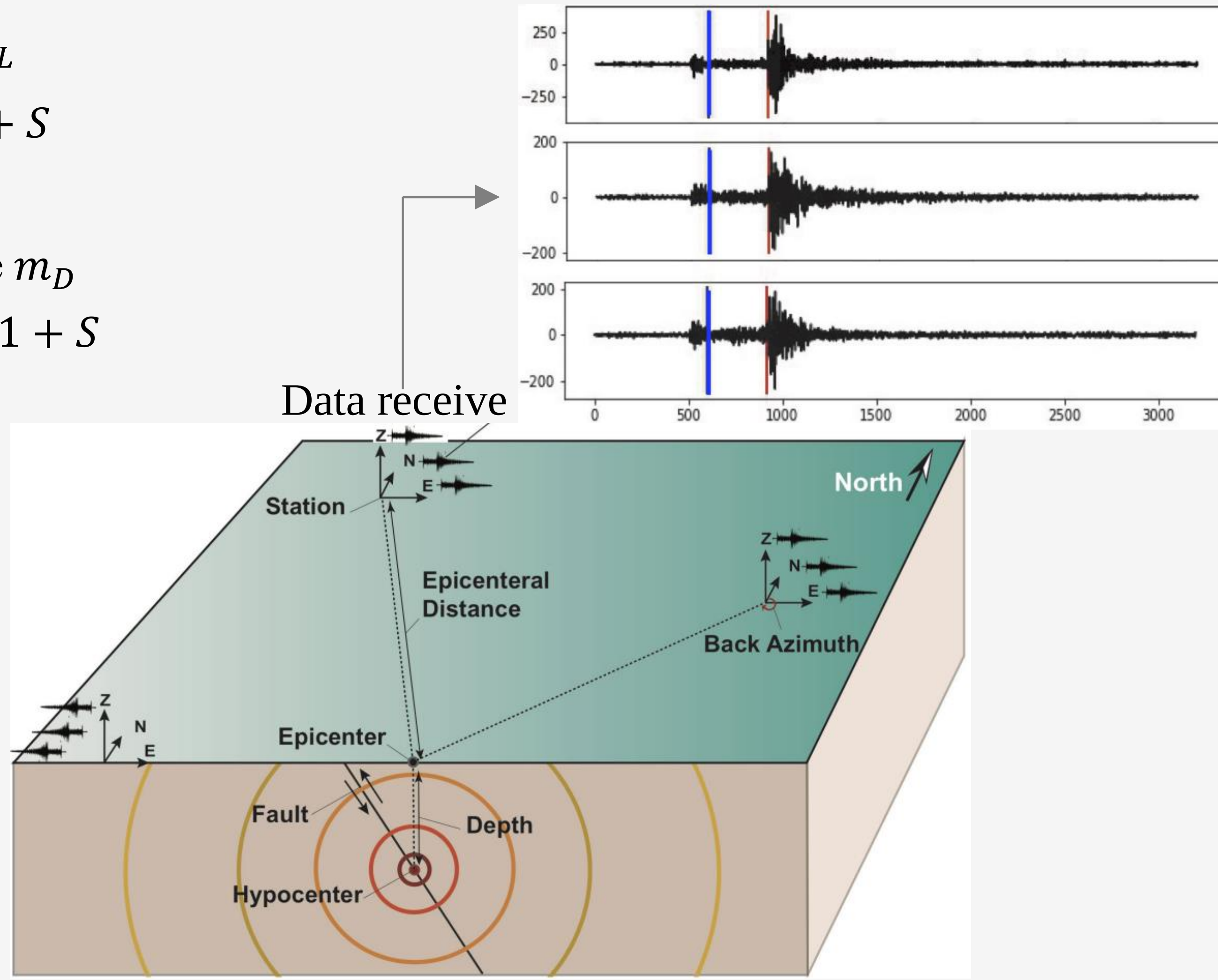
Magnitude: the relative amount of energy released by an earthquake event

- Local magnitude m_L

  $\log(A) - \log(A_0) + S$

- Duration magnitude m_D

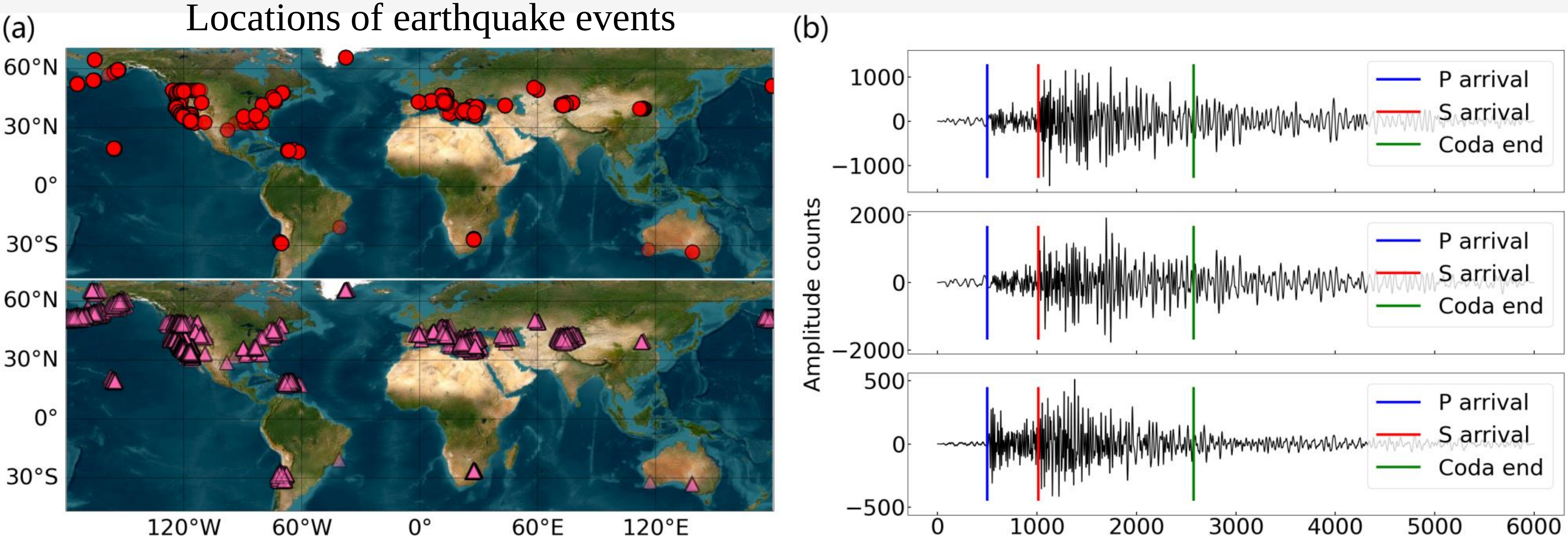
  $2.49 \times \log(T) - 2.31 + S$



Earthquake magnitude estimation

STanford EArthquake Dataset (STEAD)

Locations of earthquake events



Receiver stations

Examples of earthquake signal: P wave arrival time, S wave arrival time and the end of signals dominant energy.

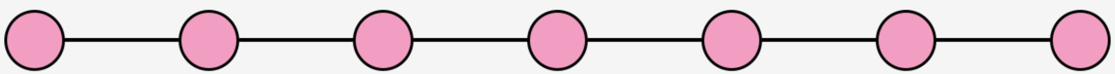
Training set 75%, test set 25%

Mousavi SM, et al. STanford EArthquake Dataset (STEAD): A global data set of seismic signals for AI. IEEE Access, 2019, 7

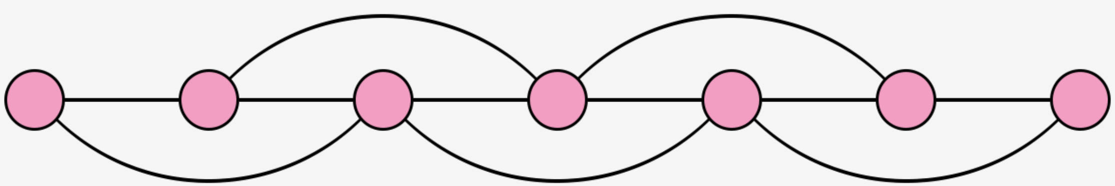
Earthquake magnitude estimation

Symmetric Graph-K, SG-K

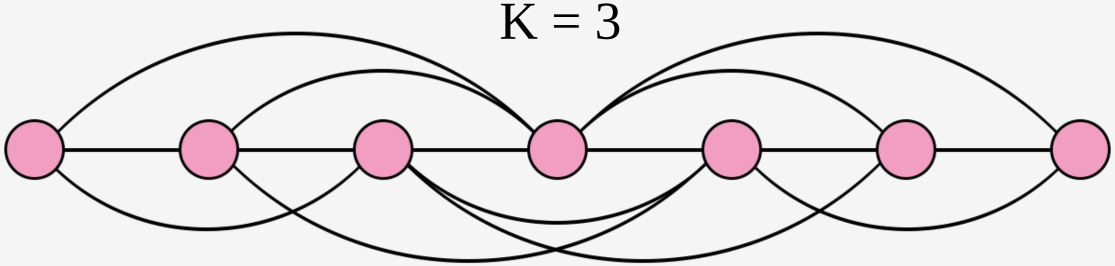
K = 1



K = 2



K = 3



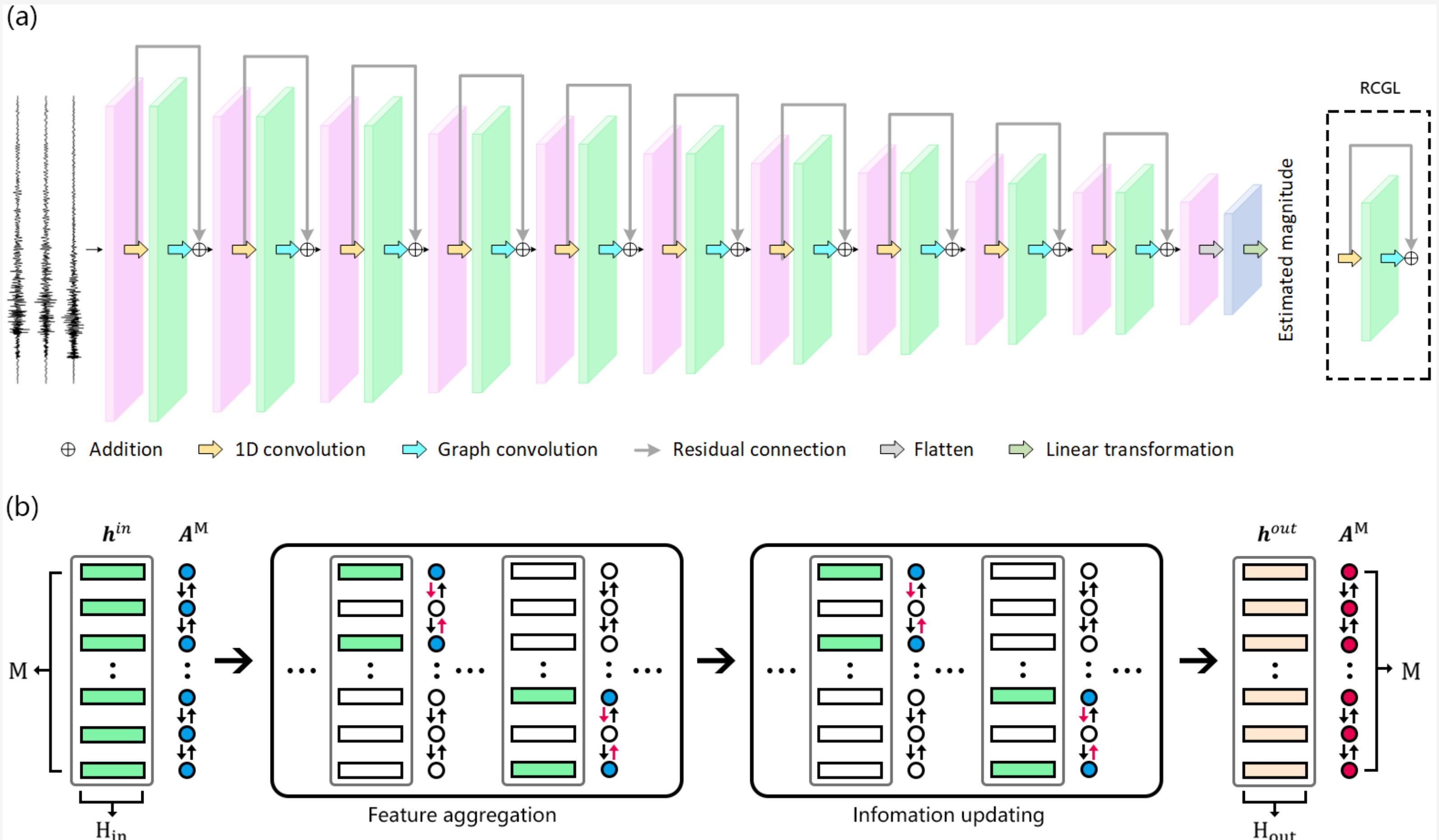
For SG-1, adjacency matrix $\boldsymbol{A}^{\mathrm{M}} \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$:

$$\boldsymbol{A}^{\mathrm{M}} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

N is the number of graph nodes

Earthquake magnitude estimation

Architecture of EQGraphNet

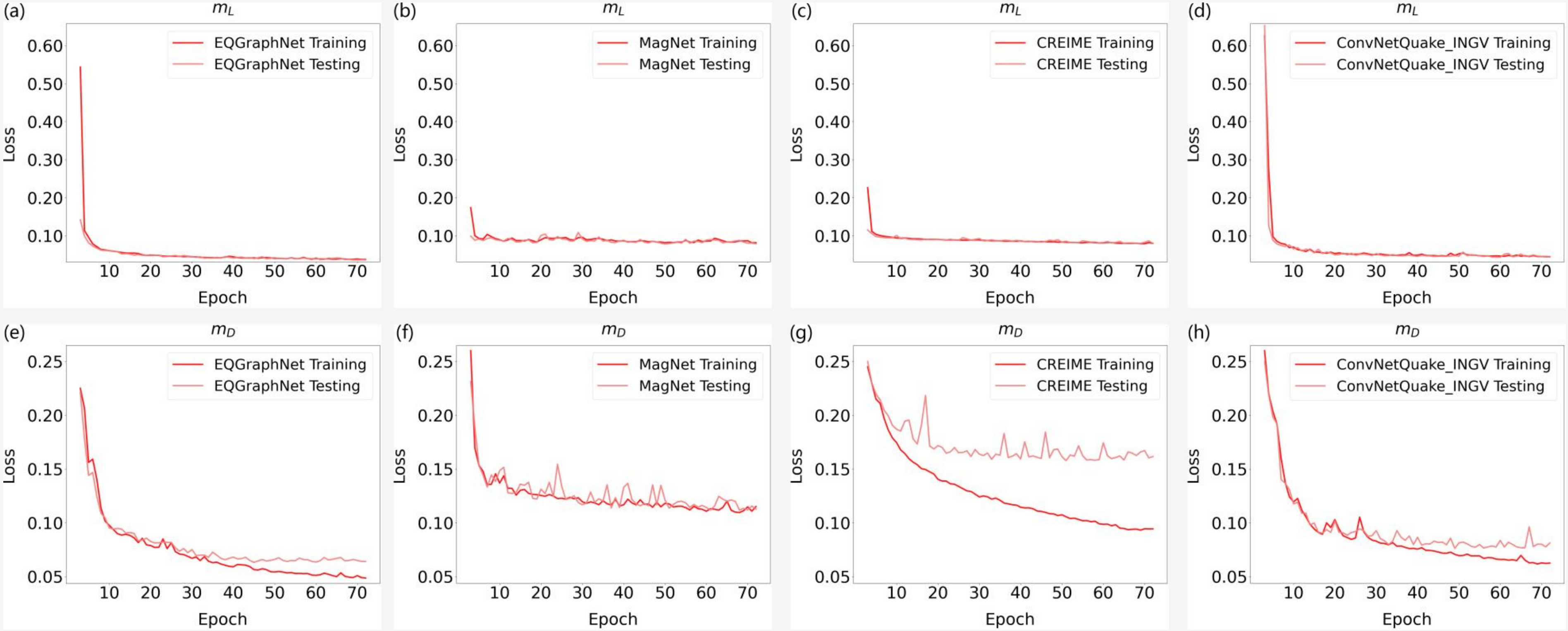


Graph convolutional network (GCN):
feature aggregation and information updating

$$Y = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}X\Theta\right)$$

Earthquake magnitude estimation

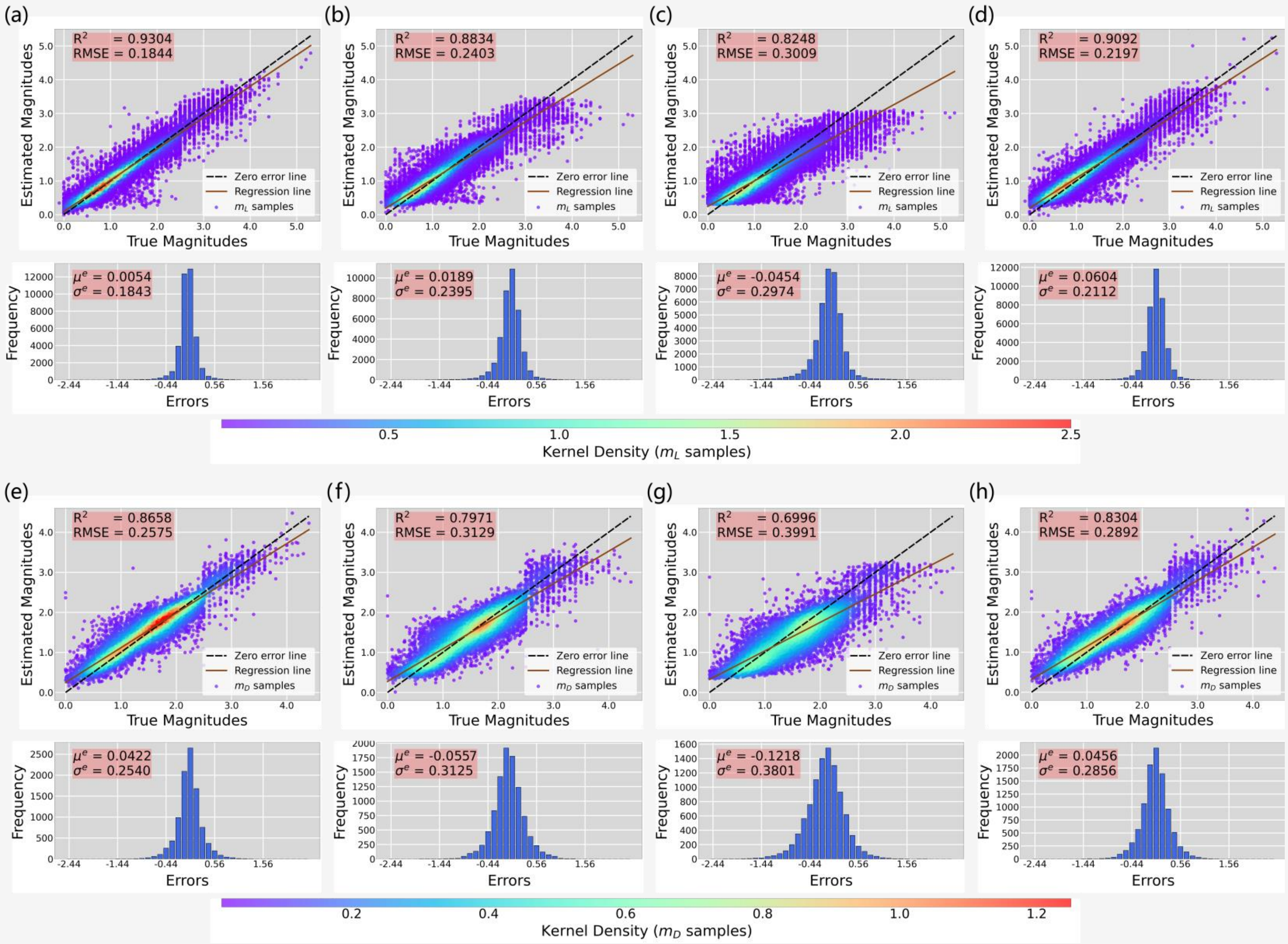
EQGraphNet MagNet CREIME CNQ



Loss curves of training and testing during training procedures for m_L and m_D samples

Earthquake magnitude estimation

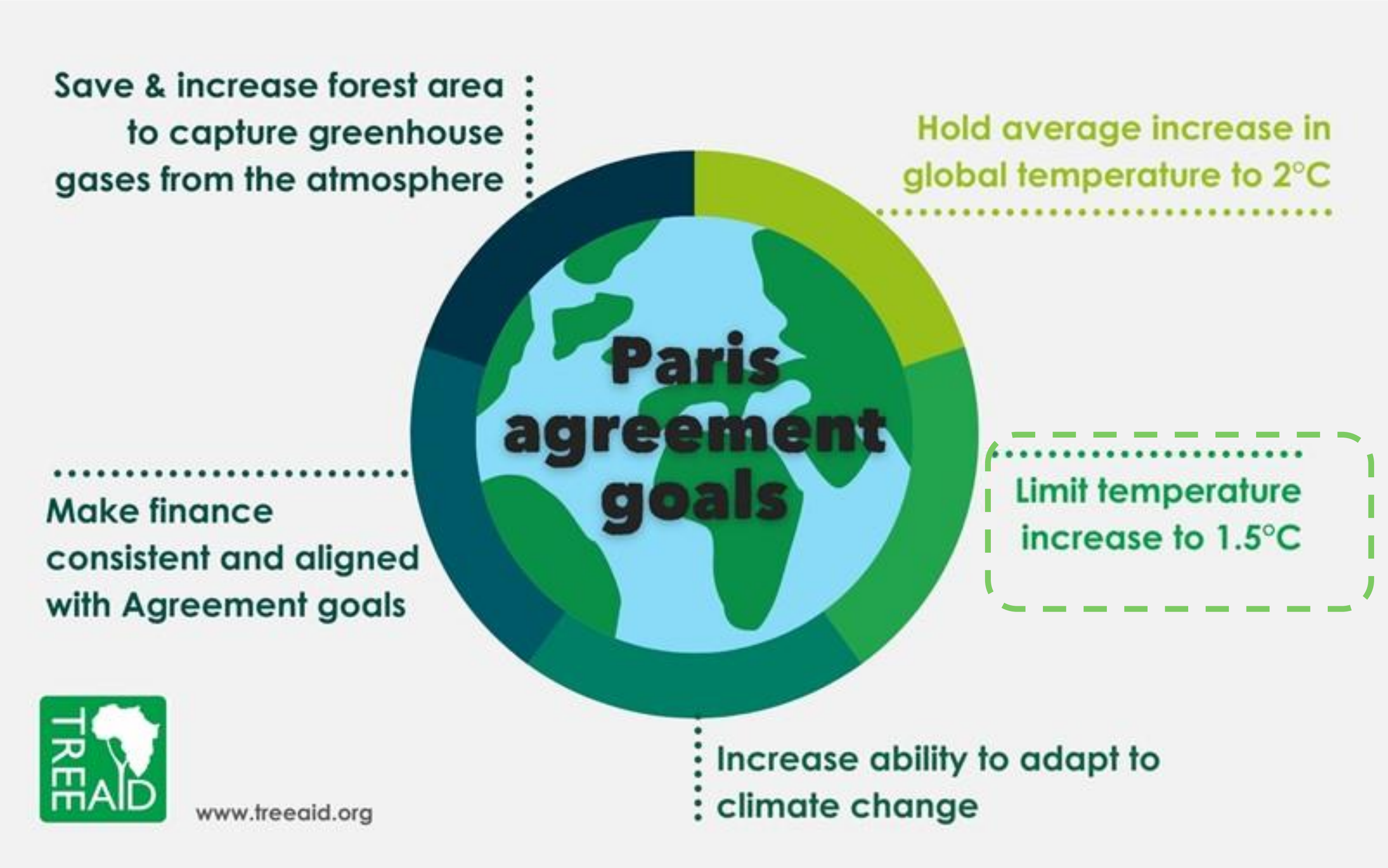
EQGraphNet MagNet CREIME CNQ



- **Full Wave Inversion**
- **Magnitude estimation in earthquake**
- **Temperature prediction in climate**

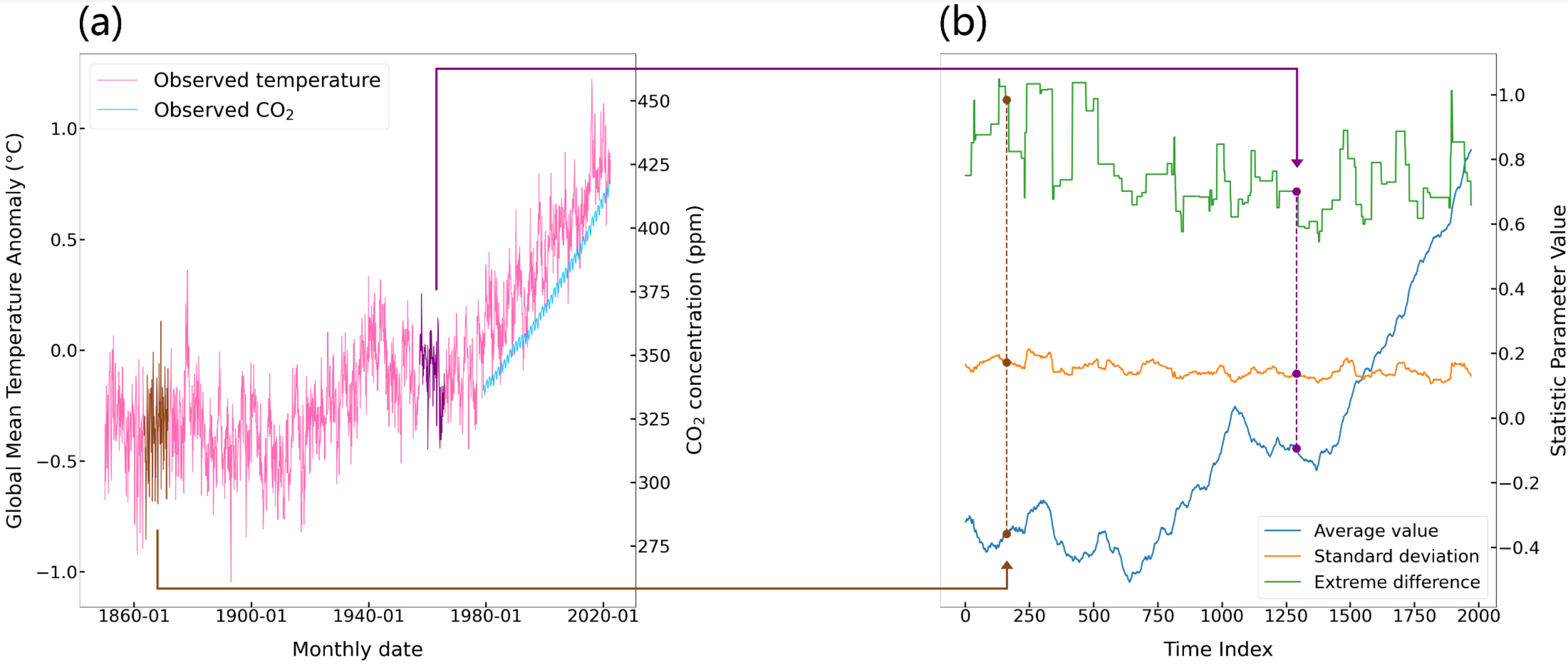
Temperature prediction in Climate

Global average temperature (GAT) represents the mean temperature recorded worldwide, typically expressed as an “anomaly” value denoting the deviation from the average over a designated period, such as the pre-industrial climate era (1850–1900).



Temperature prediction in Climate

Monthly temperature anomaly as well as CO_2 concentration from 1979 to 2024



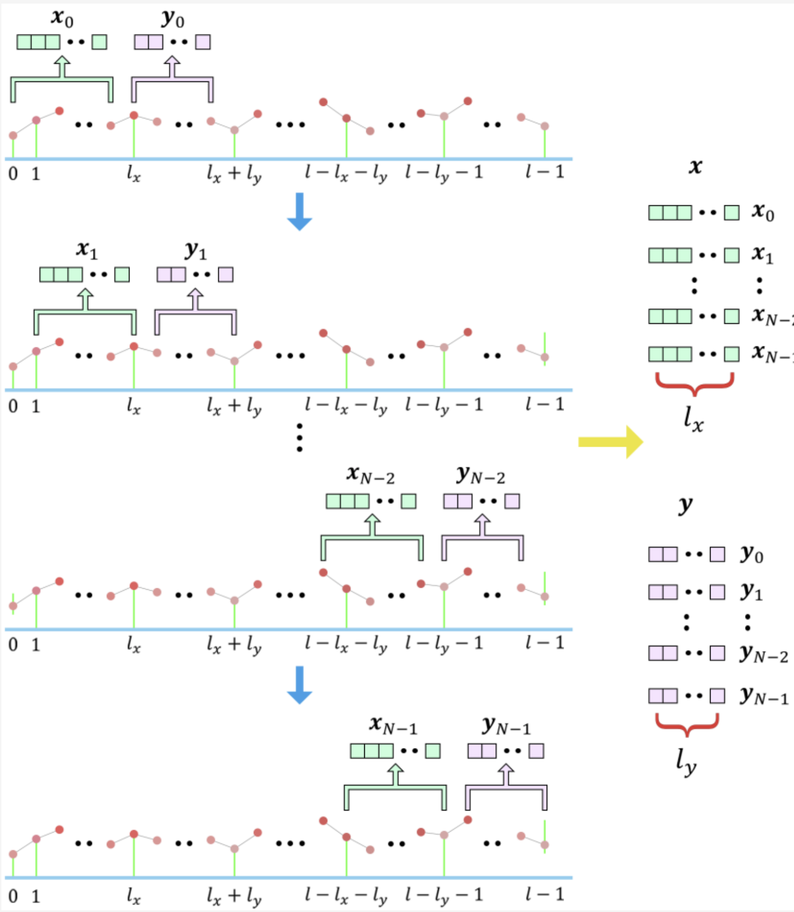
GAT

statistic parameters

Non-stationary

Temperature prediction in Climate

Time symmetric graph

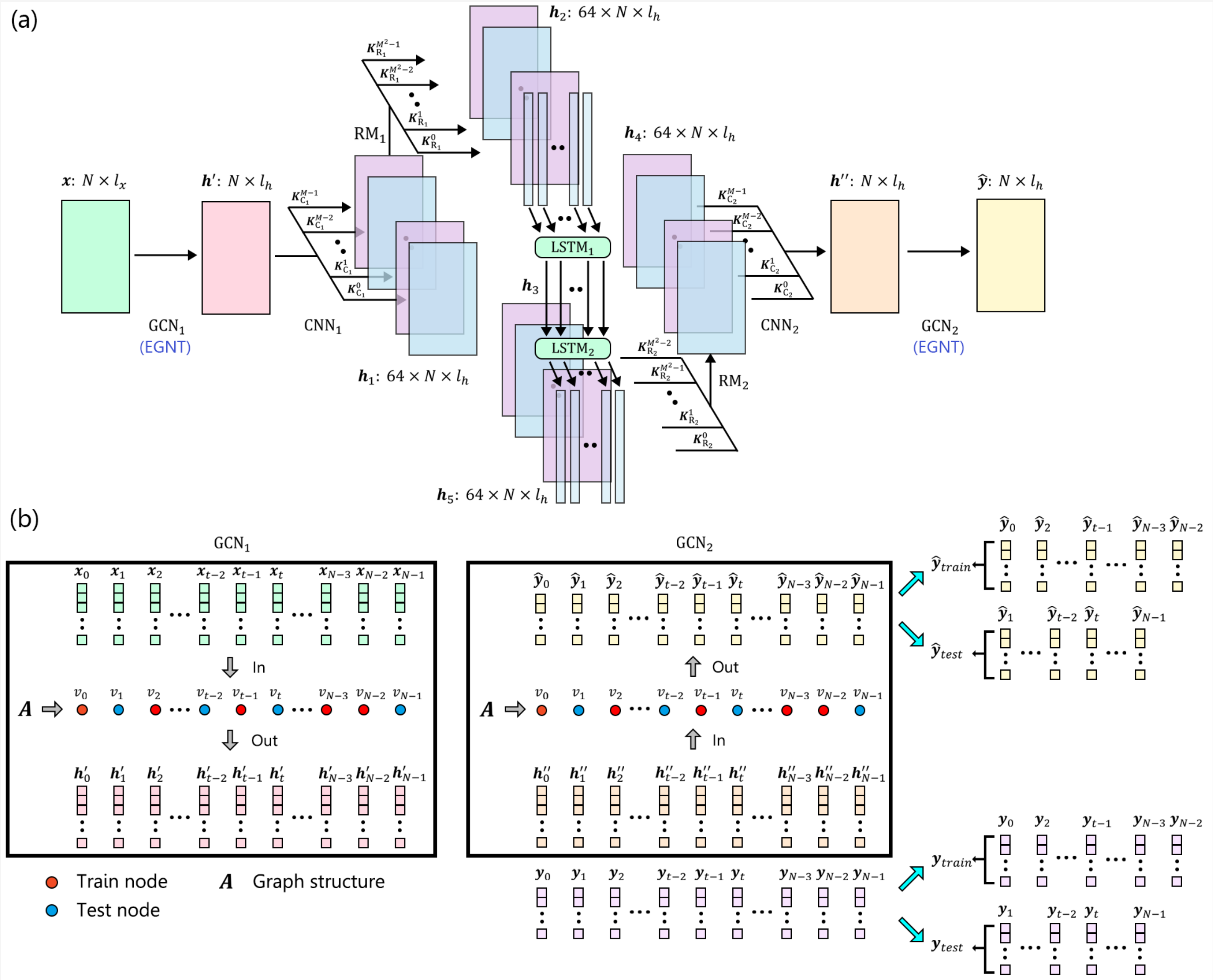


Adjacency matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

Sample construction process:
Non-stationary -> Stationary

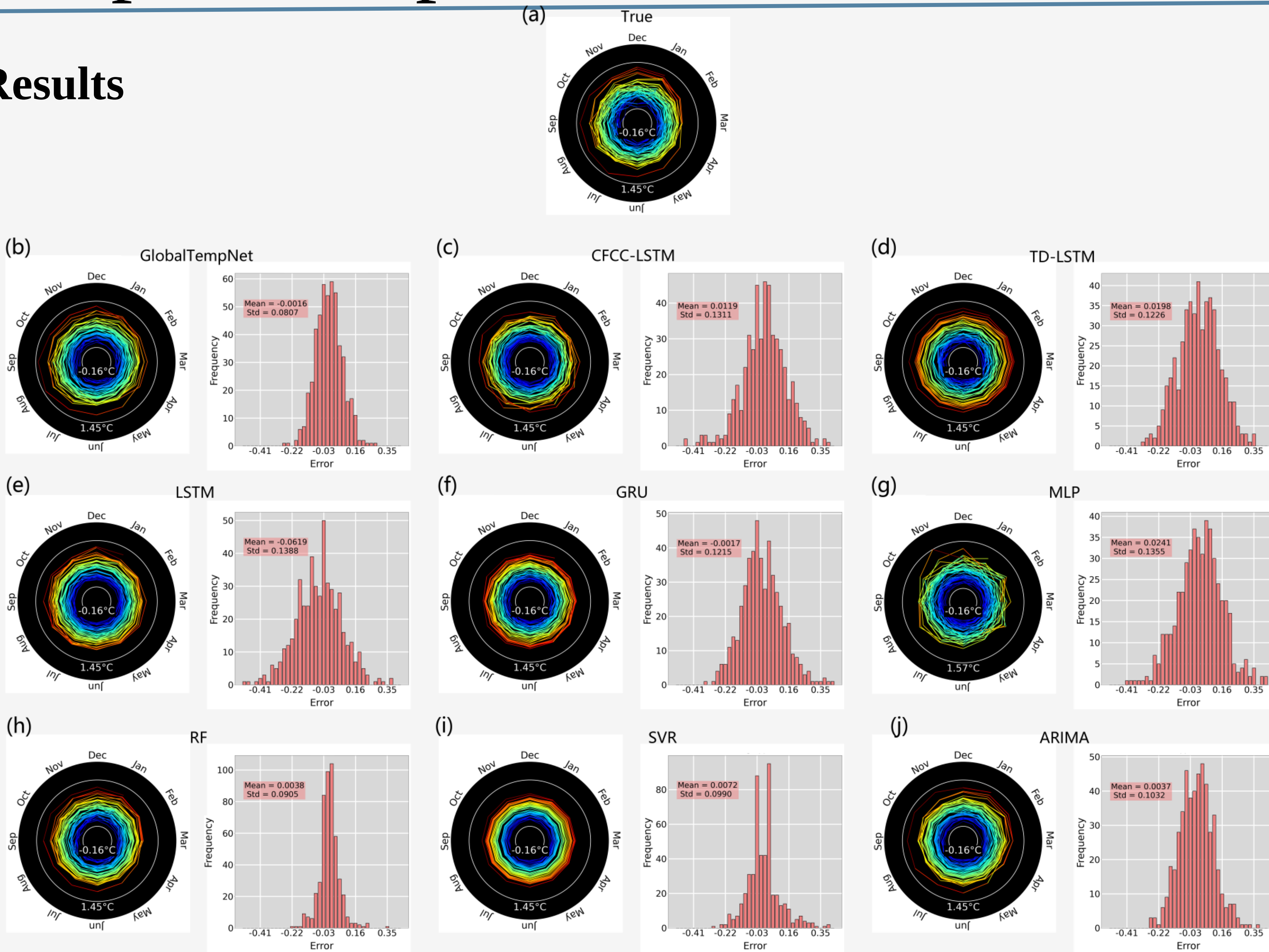
Temperature prediction in Climate



Our proposed solution: GlobalTempNet and ENGT (entire graph node training)

Temperature prediction in Climate

Results



Temperature prediction in Climate

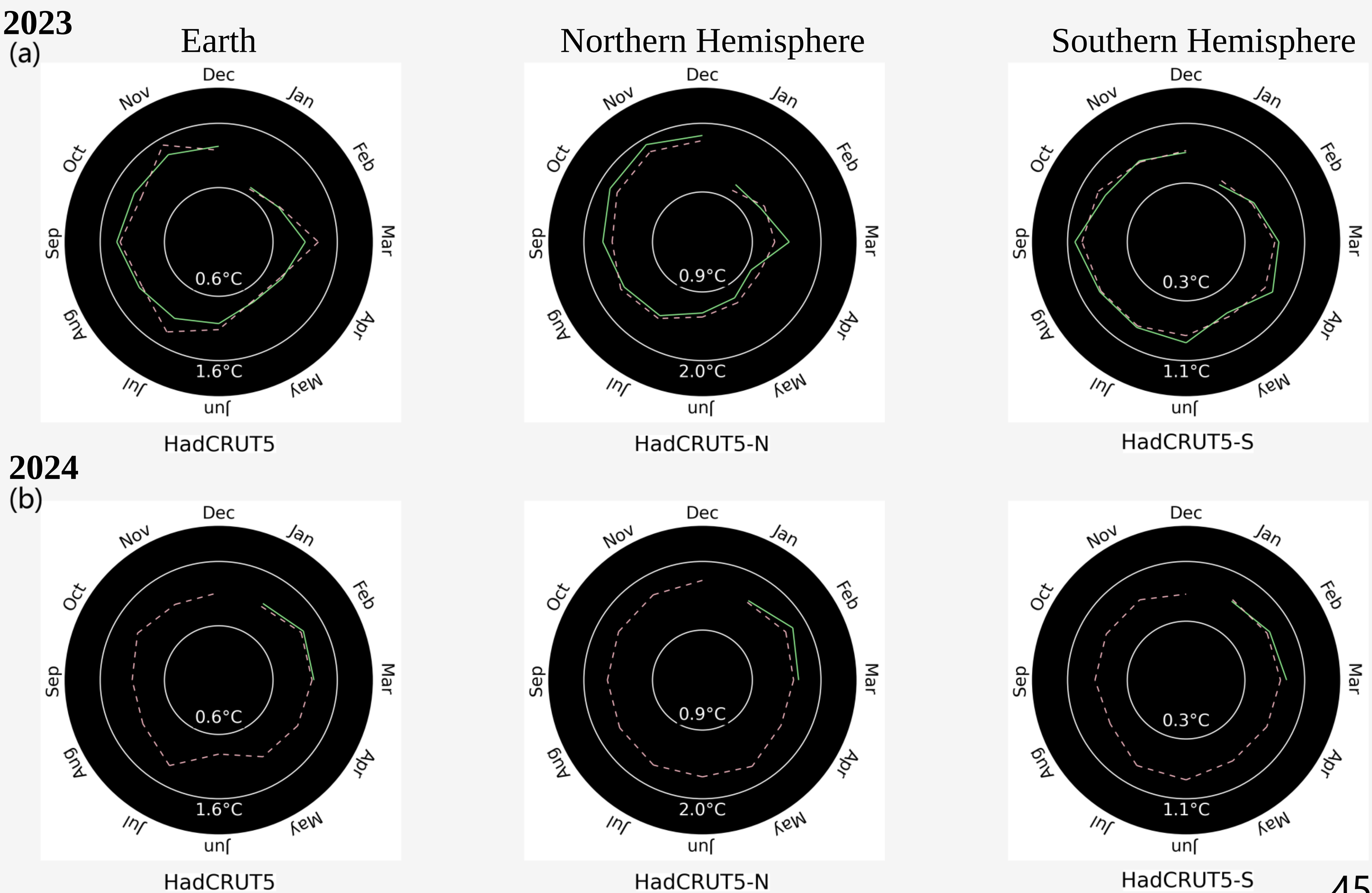
Generalization

Our model represents a valuable applicability in diverse weather measurement systems

RMSE on different temperature datasets

Data	GlobalTempNet	CFCC-LSTM	TD-LSTM	LSTM	GRU	MLP	RF	SVR	ARIMA
ERSSTv4 [55]	**0.0476**	0.0714	0.0518	0.0673	0.0850	0.0501	0.0667	0.0554	0.0728
ERSSTv3b [56]	**0.0327**	0.0524	0.0369	0.0488	0.0790	0.0354	0.0565	0.0419	0.0595
ERSSTv5 [57]	**0.0429**	0.0627	0.0506	0.0977	0.0875	0.0476	0.0633	0.0517	0.0692
HadSST3 [58]	**0.0581**	0.0853	0.0776	0.0863	0.0900	0.0830	0.0754	0.0750	0.0747
HadSST3-N [58]	**0.0877**	0.1071	0.1031	0.1131	0.1050	0.1042	0.0997	0.0999	0.1027
HadSST3-S [58]	**0.0911**	0.1031	0.1076	0.1151	0.1103	0.1051	0.1007	0.1025	0.1066
NCEP/NCAR [59]	**0.2892**	0.3301	0.2809	0.3444	0.3878	0.2910	0.3146	0.3446	0.3486
NCEP/DOE [60]	**0.2842**	0.3293	0.2915	0.4830	0.3786	0.3391	0.3359	0.3137	0.3254
NCEP/CIRES [61]	**0.1918**	0.2211	0.2103	0.2276	0.2682	0.2078	0.2066	0.2147	0.2834
ERA5 [62]	**0.2458**	0.2905	0.2671	0.3058	0.3573	0.2728	0.2658	0.2512	0.2727

Temperature prediction in Climate



3、Conclusion and Future

Conclusion and Future

- **Intelligent analytics of big earth data in a non-Euclidean framework**

- **AI for Science paradigm in Geoscience and Mathematics**

西安交通大學
XI'AN JIAOTONG UNIVERSITY

Thanks!

E-mail: emailwzg@mail.xjtu.edu.cn