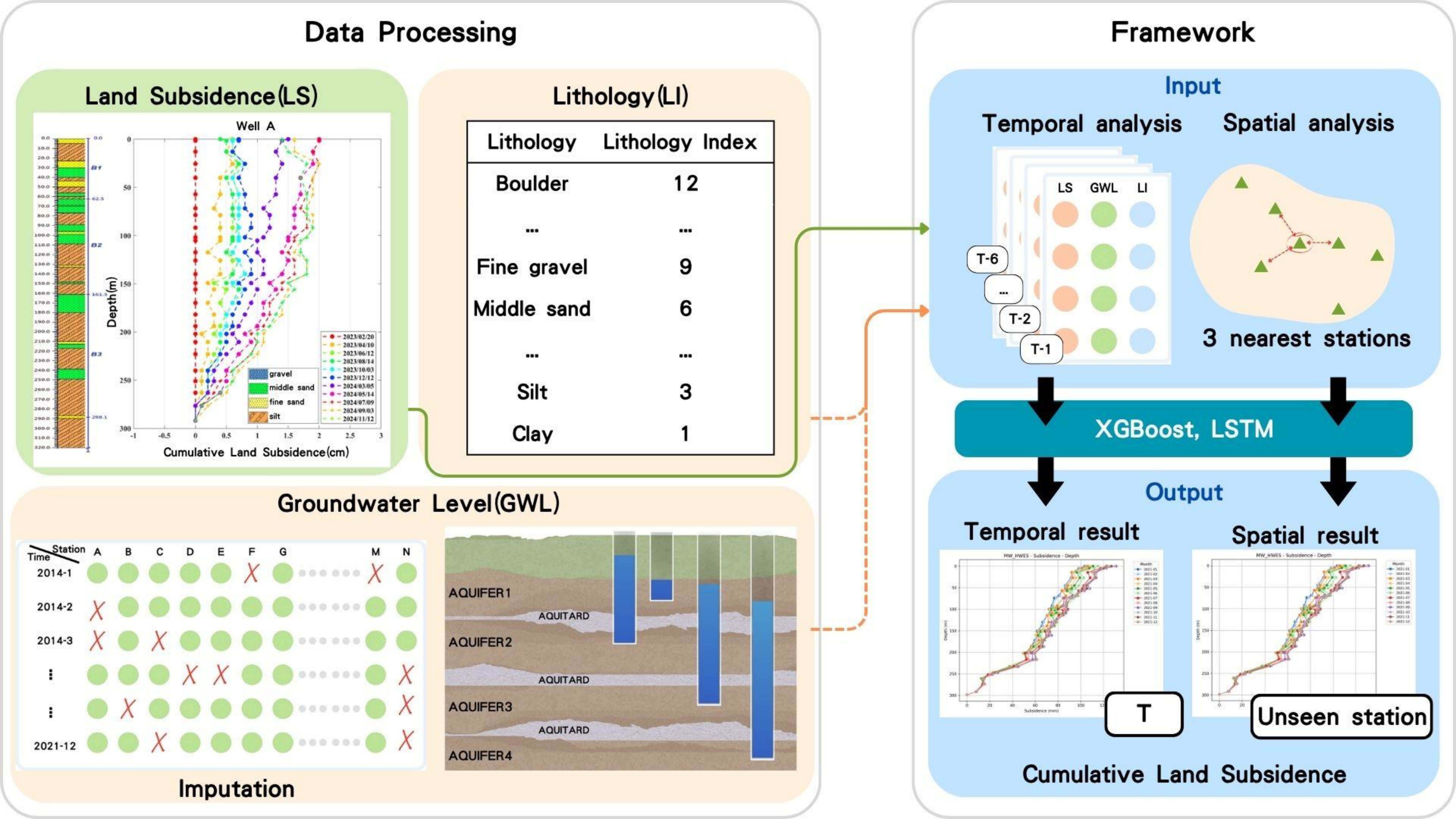
Predicting Cumulative Land Subsidence and Its Spatiotemporal Relationship Using Machine Learning

Presenter Yu-Yun Hsu

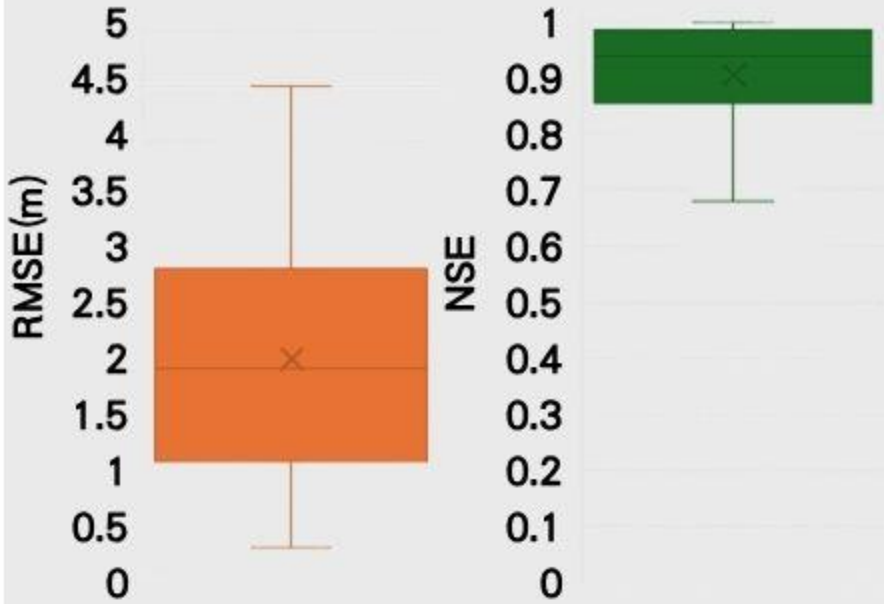
Advisor WeiCheng Lo
Jhe-Wei Lee
Chih-Tsung Huang



Data Processing
Land Subsidence(LS)
Well A
Depth(m)
Cumulative Land Subsidence(cm)
gravel
middle sand
fine sand
silt
Lithology(LI)
Lithology
Lithology Index
Boulder
12
Fine gravel
9
Middle sand
6
Silt
3
Clay
1
Groundwater Level(GWL)
Station
Time
2014-1
2014-2
2014-3
2021-12
AQUIFER1
AQUITARD
AQUIFER2
AQUIFER3
AQUIFER4
Imputation
Framework
Input
Temporal analysis
Spatial analysis
LS
GWL
LI
T-6
T-2
T-1
3 nearest stations
XGBoost, LSTM
Output
Temporal result
Spatial result
T
Unseen station
Cumulative Land Subsidence

Key Findings

Groundwater Imputation Overall Performance



	RMSE(m)	NSE
Average	2.04	0.90
Max	4.50	0.97
Min	0.32	0.50

LSTM

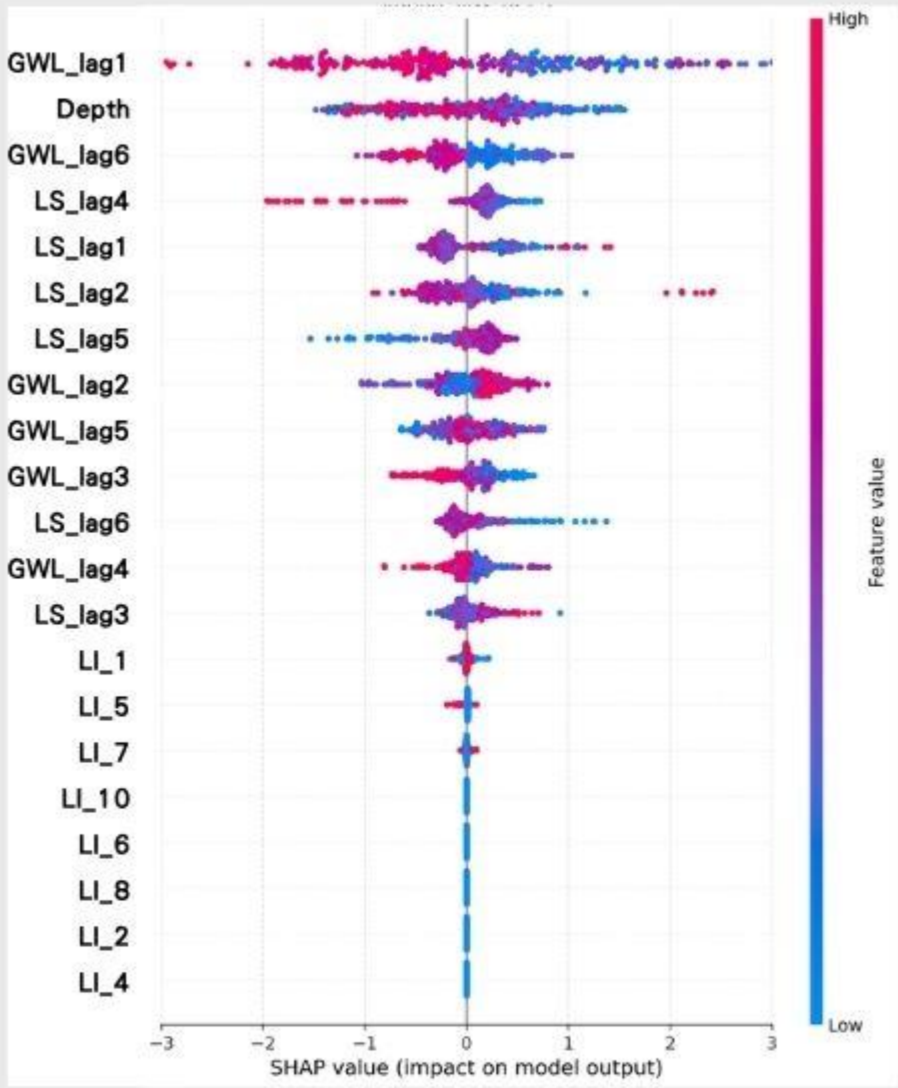
Rank	Training period/Inputs	RMSE(mm)	R^2
Best	7yr / LS, LI	8.41	0.94
Worst	1yr / LS, GWL, LI	17.92	0.81

XGBoost

Rank	Training period/Inputs	RMSE(mm)	R^2
Best	7yr / LS, GWL, LI	8.70	0.94
Worst	1yr / LS	10.88	0.91

LS: land subsidence, GWL: groundwater variation, LI: lithology

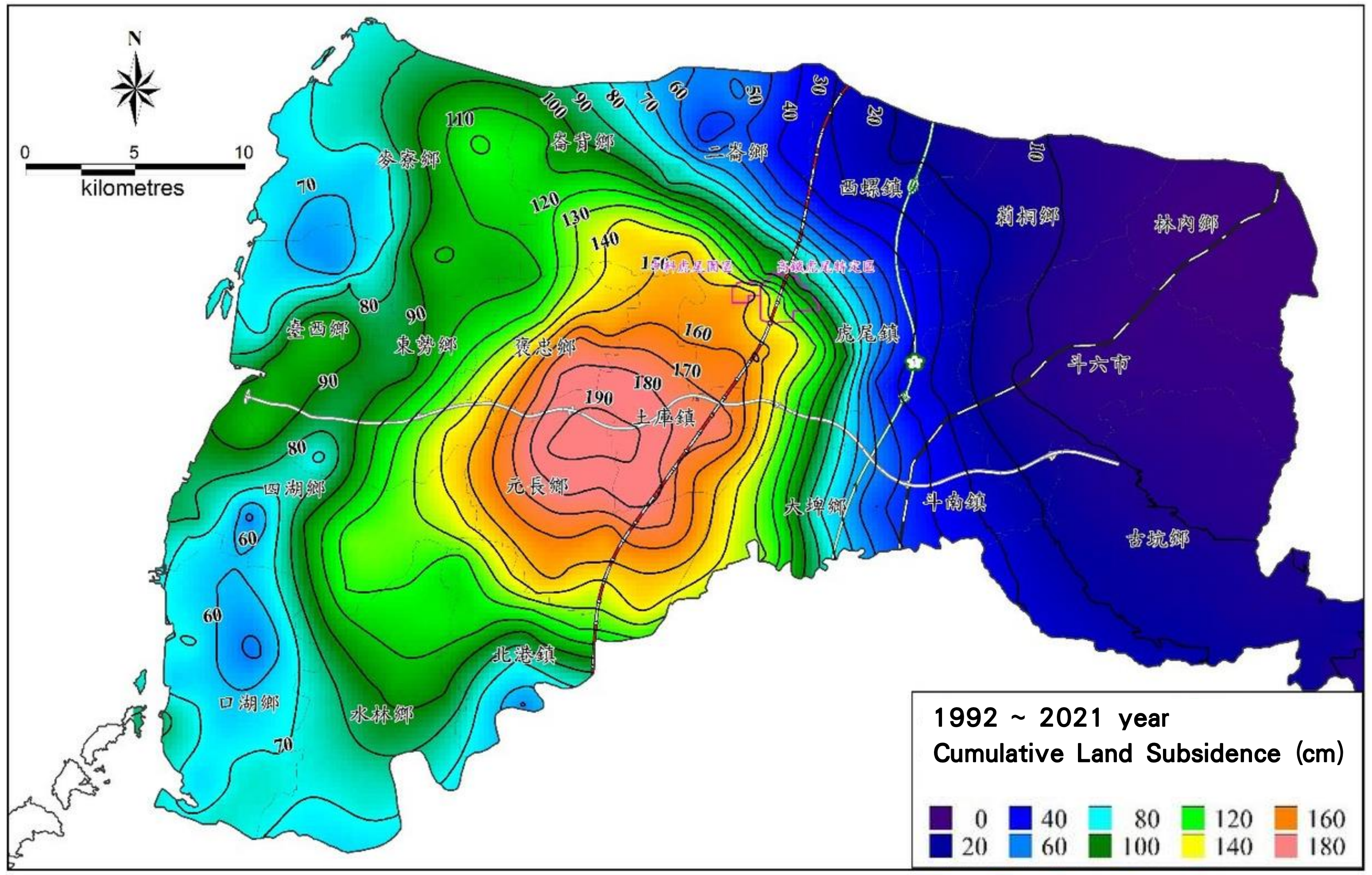
SHAP

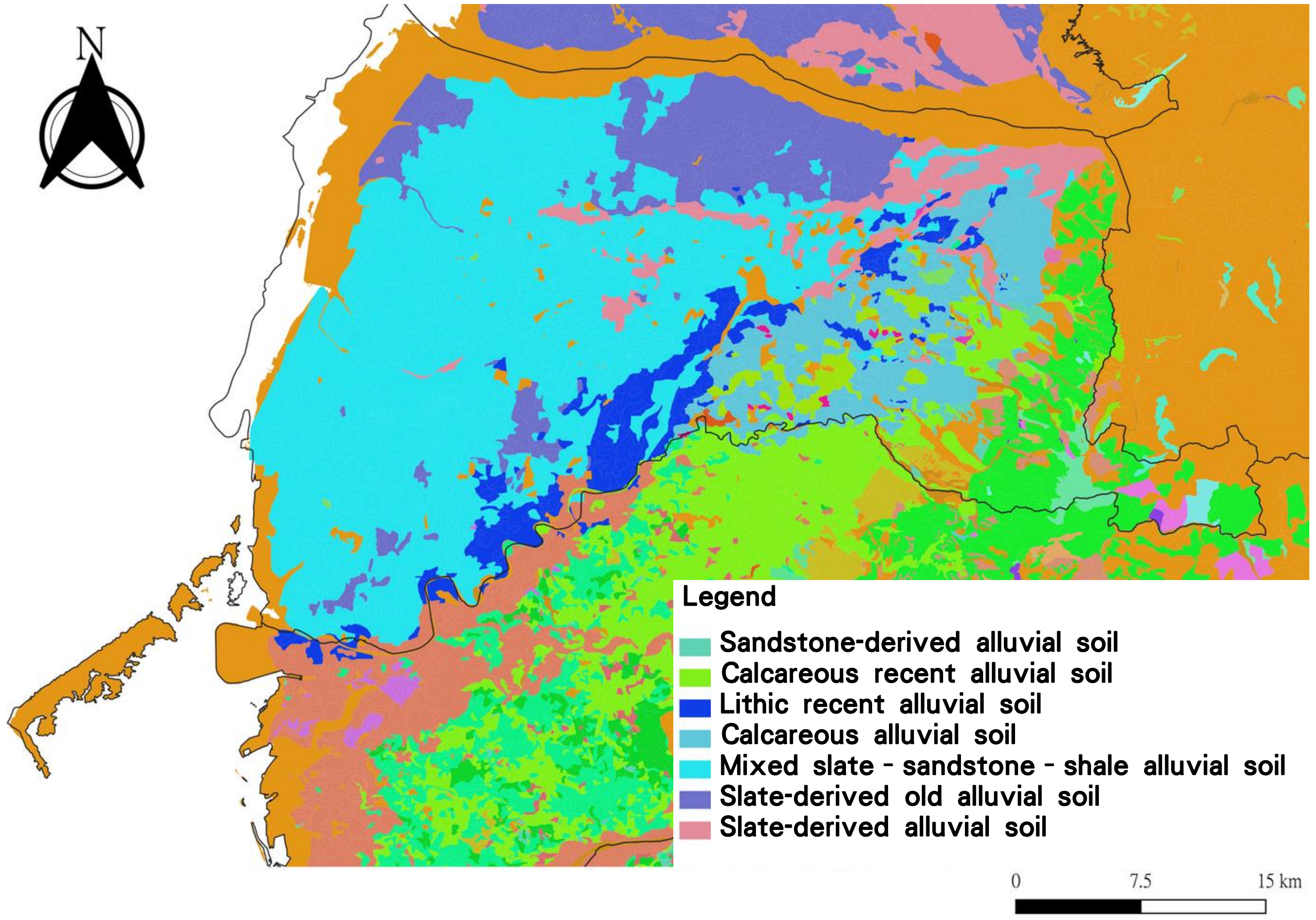


- Groundwater imputation achieved reliable overall performance and supported subsequent prediction tasks.
- LSTM was more sensitive to training period and input combination, while XGBoost showed more stable performance.
- SHAP analysis indicated that groundwater level change was among the most influential predictors.

Study Area—Yunlin County

- The maximum annual subsidence rate reached 7.8 cm/year.(2021 year)
- The area of significant subsidence was 502.7 km^2.(2021 year)
- Alluvial deposits dominate the study area, making the strata sensitive to groundwater-pressure changes.





Source：[Land Subsidence Monitoring Information Integration Platform]; [Agricultural Open Data Platform]

Monitoring Data and Model Variables

Land subsidence and groundwater monitoring data in Yunlin, Taiwan

Monitoring network

23 subsidence monitoring wells

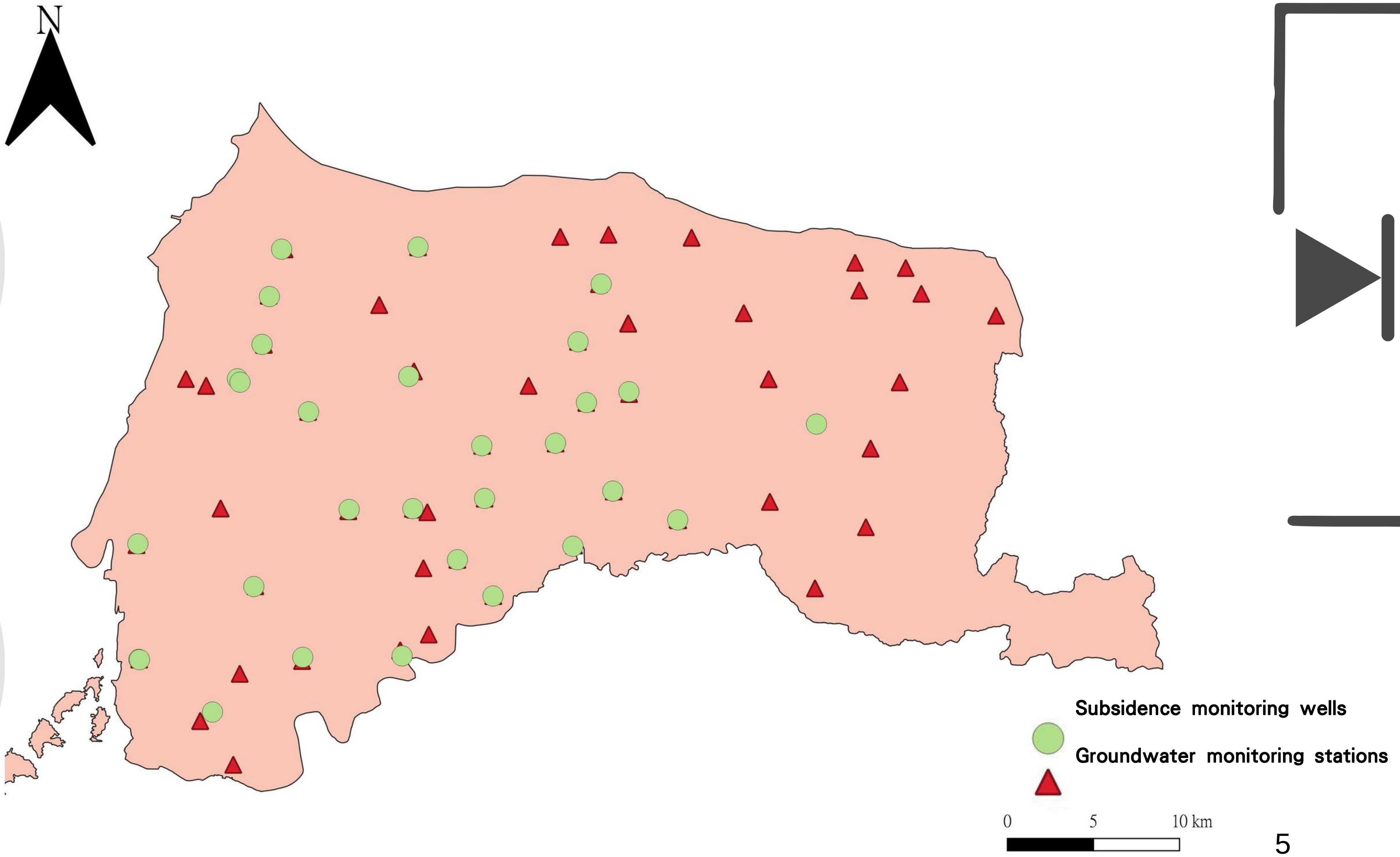
Matched groundwater monitoring stations

Model inputs

- Previous cumulative subsidence
- Monthly groundwater-level variation
- Lithology

Prediction target

Cumulative land subsidence



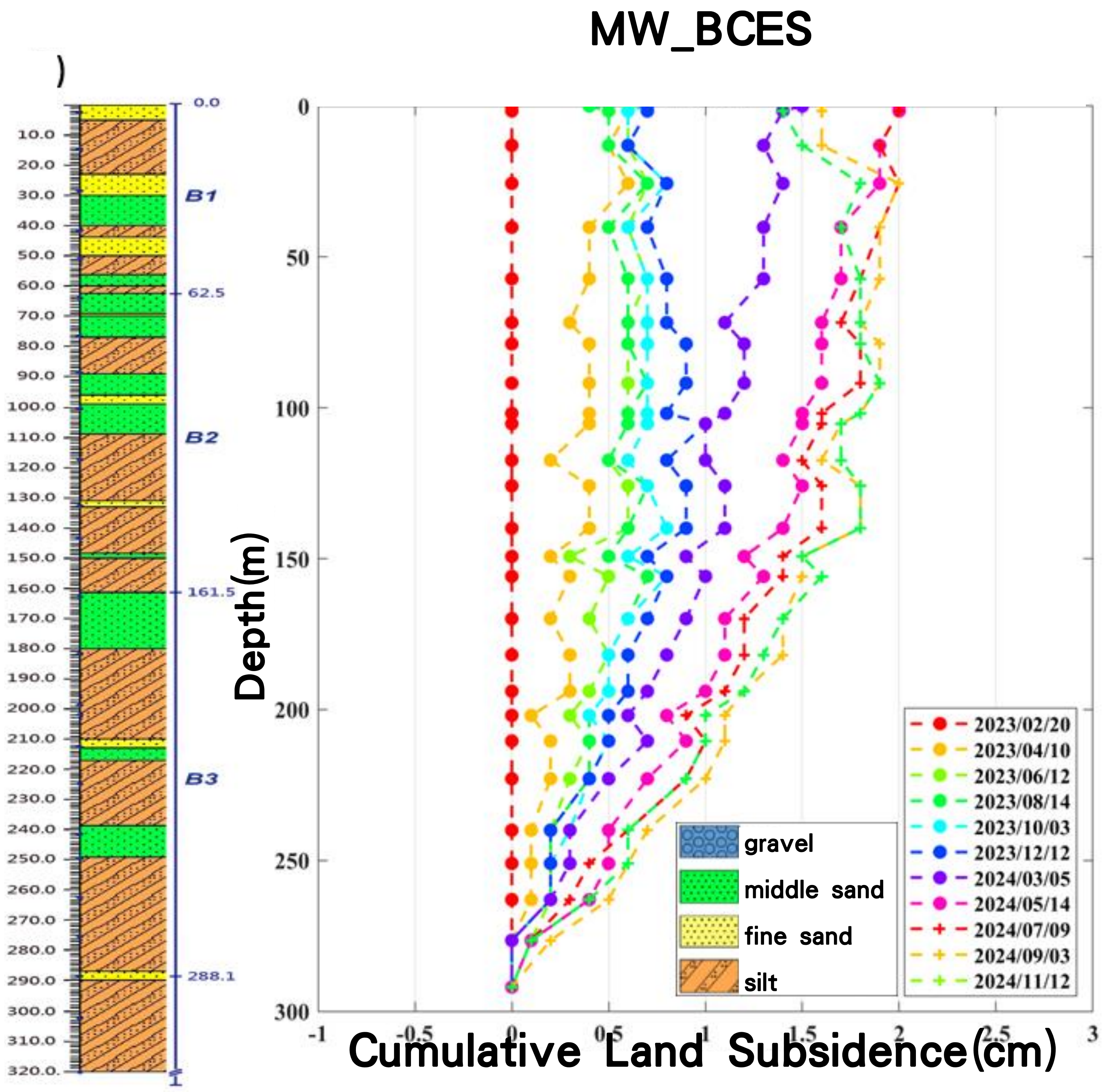
Cumulative Land Subsidence Processing

1.Match ring IDs and depths over time

2.Calculate depth differences between rings

3.Derive layer compaction from adjacent ring pairs

4.Accumulate layer compaction relative to the well bottom



Source：Land Subsidence Monitoring Information Integration Platform

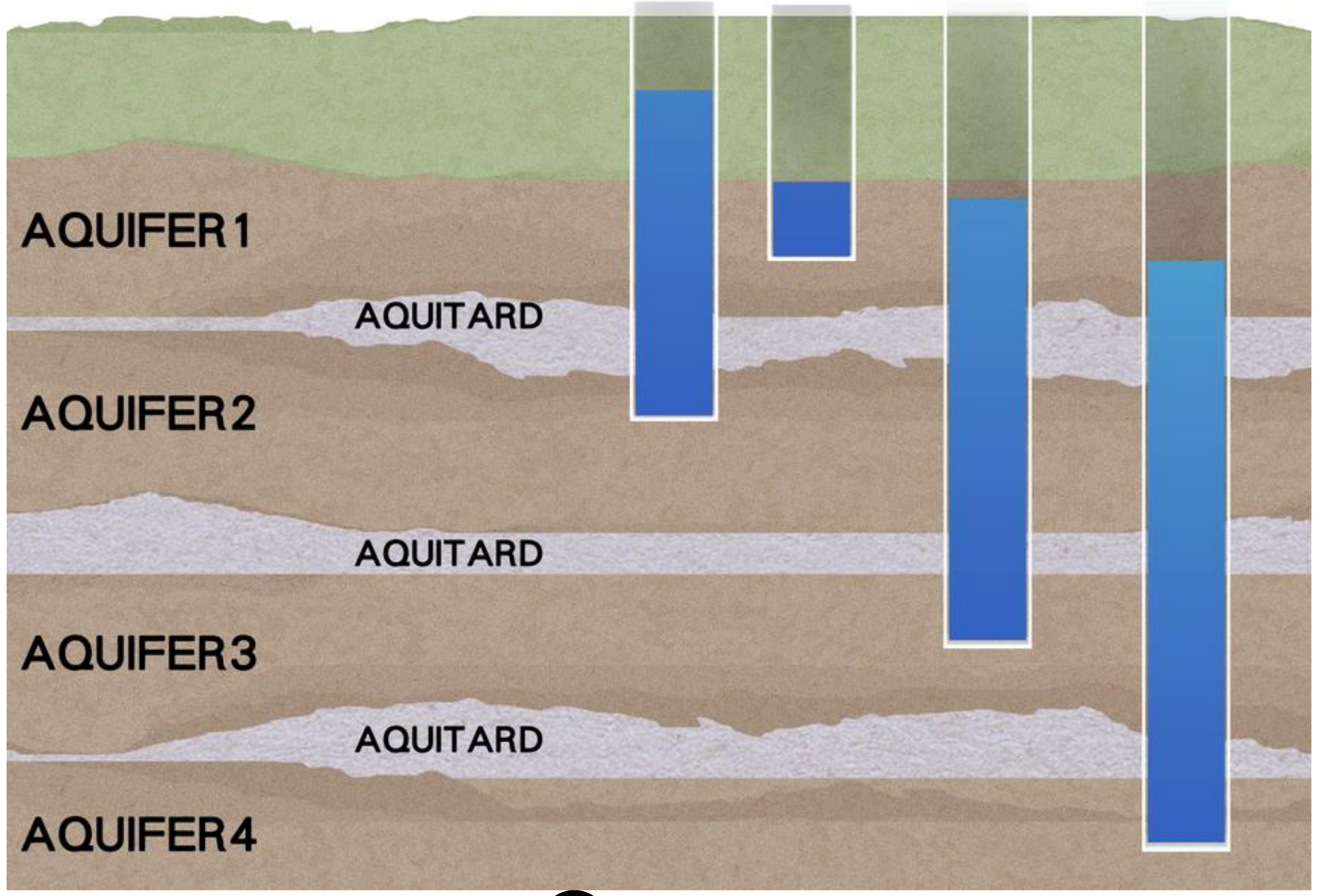
Groundwater-Level Variation Processing

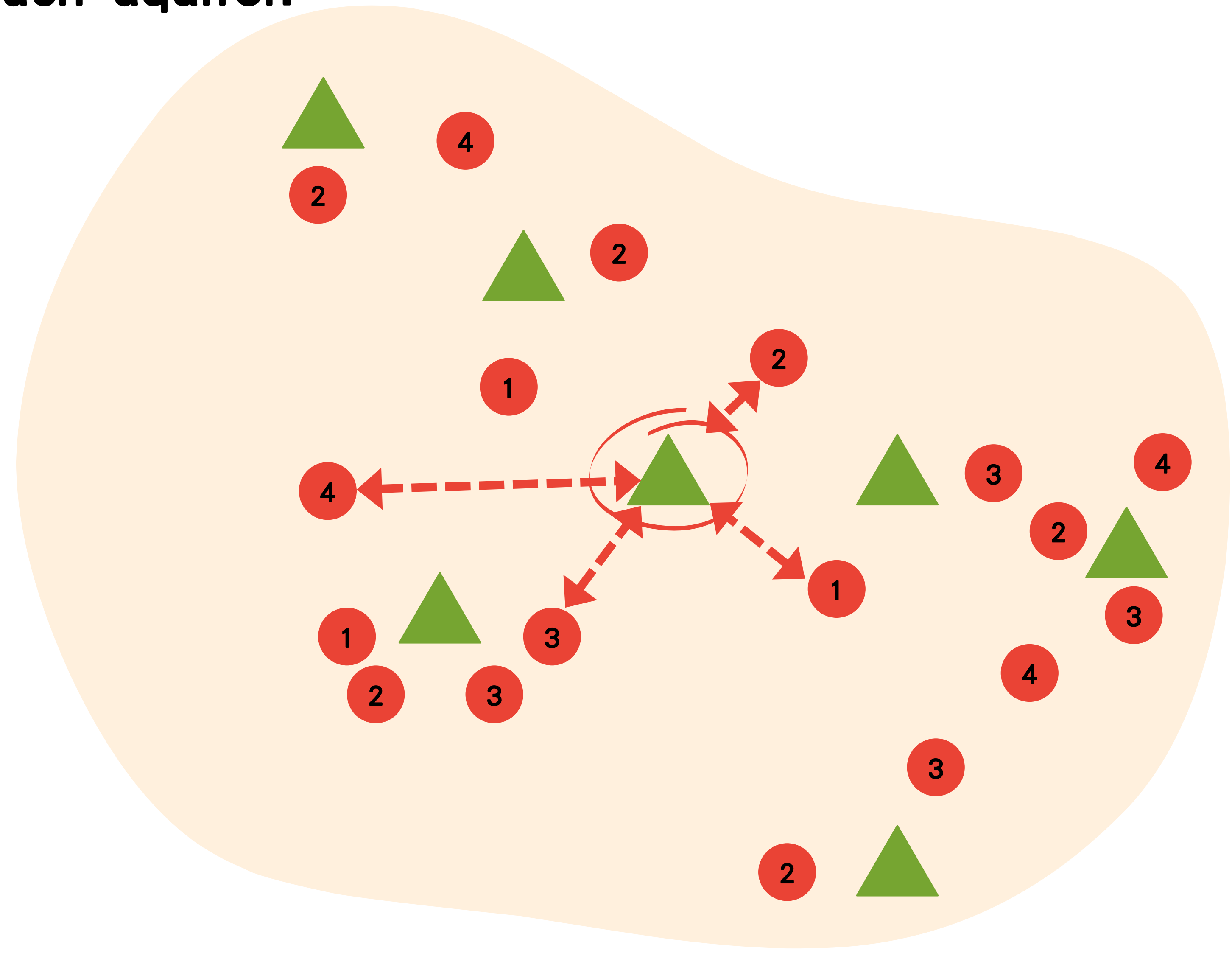
1.Impute missing values

2.Calculate monthly groundwater-level variation

3.Match geological boundaries (B1/B2/B3) of subsidence monitoring wells with aquifer depths

4.Select groundwater monitoring stations within 6 km for each aquifer.





Lithology Processing

1. Extract the upper and lower depths and lithology of each borehole layer
2. Obtain the hydraulic conductivity range for each lithology type
3. Rank lithology types based on hydraulic conductivity ranges
4. Simplify lithology classifications

Depth	Lithology	Simplified
0	Very fine sandy mud	mud
0.67	Medium sandy coarse gravel	Coarse gravel
1	Fine gravelly coarse sand	Coarse sand
...	...	...
400	Bottom	Bottom

Lithology	Lithology Index
Boulder	12
Coarse gravel	11
Middle gravel	10
Fine gravel	9
Very coarse sand	8
Coarse sand	7
Middle sand	6
Fine sand	5
Very fine sand	4
Silt	3
Mud	2
Clay	1

Cue Wasserstein GAN with Gradient Penalty, CWGAIN-GP

Rubin(1976)：

Missing Completely At Random, MCAR

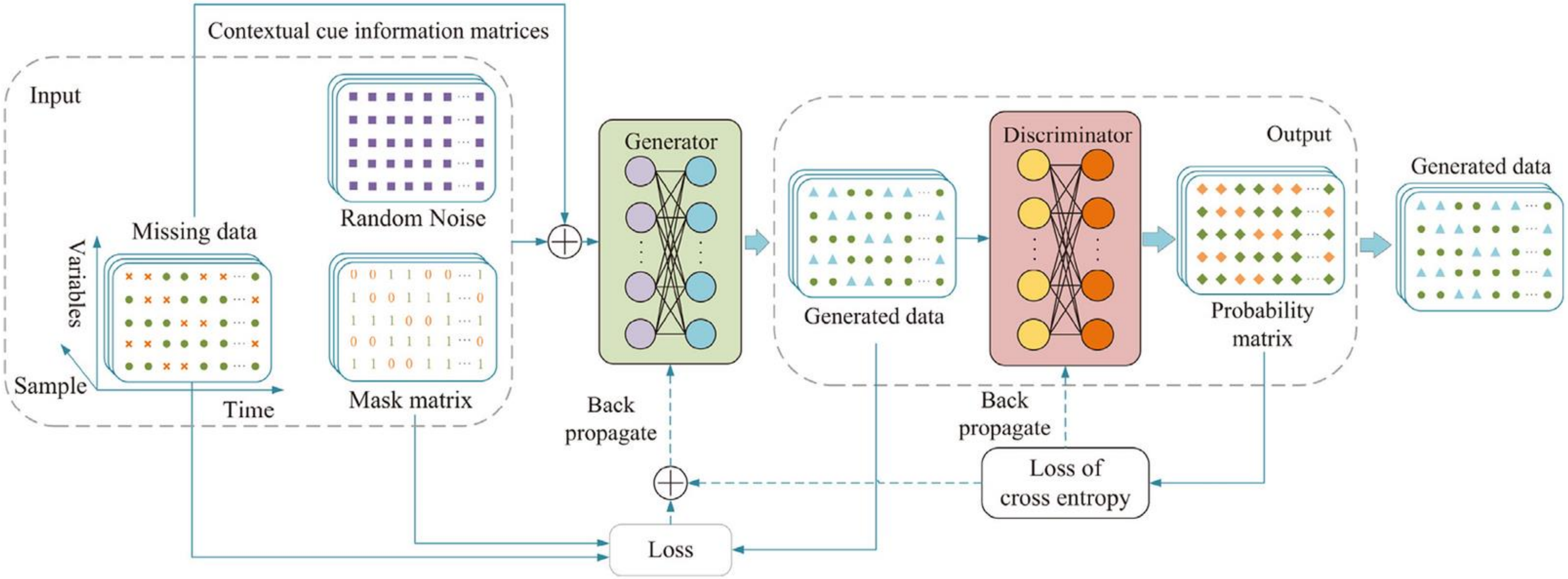
Missing At Random, MAR

Missing Not At Random, MNAR

- Groundwater records often contain contiguous missing periods
- Standard GAIN assumes MCAR
- Therefore, a method considering temporal context is needed

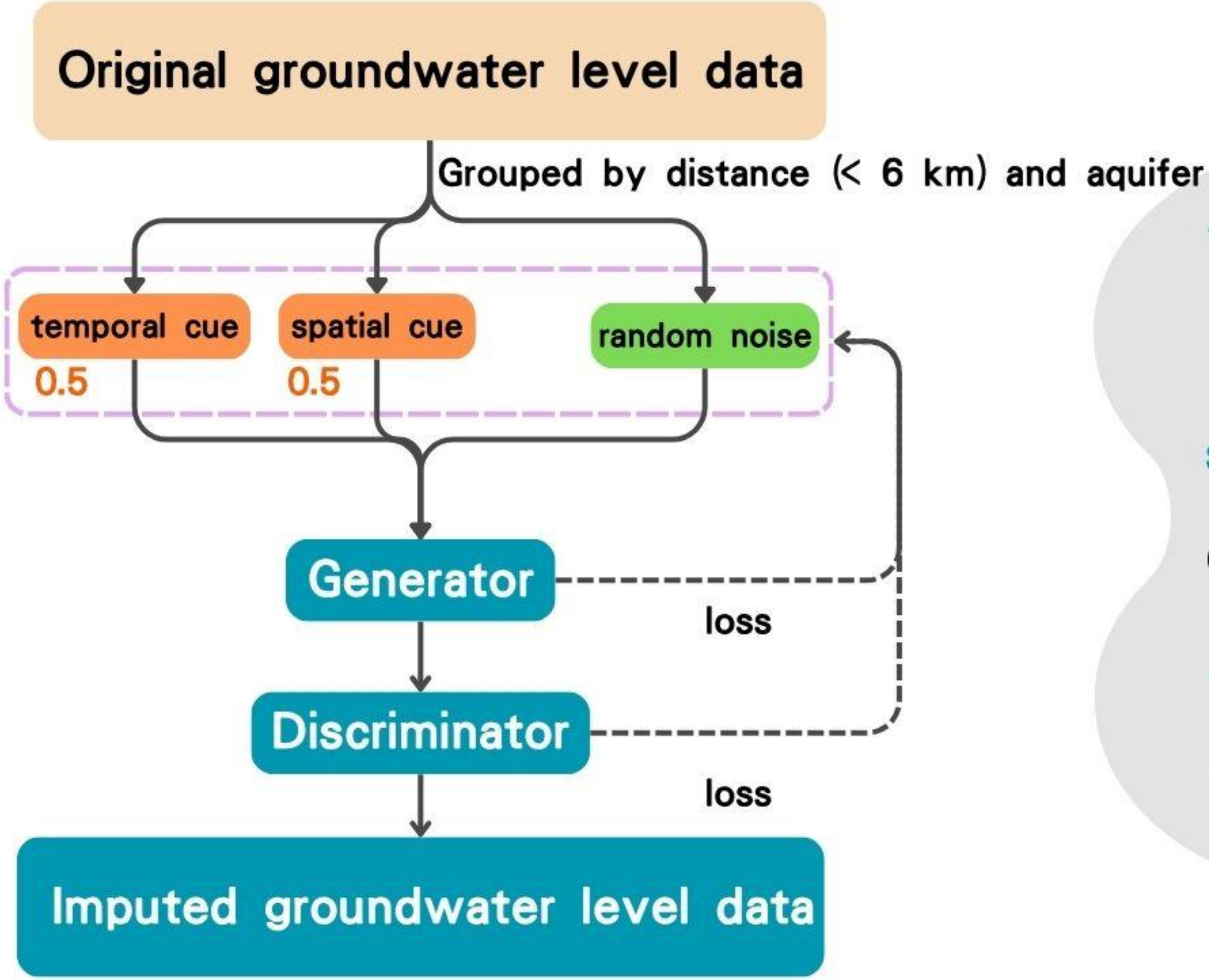
Cue Wasserstein GAN with Gradient Penalty, CWGAIN-GP

CWGAIN-GP architecture diagram



Reference：Wang et al.(2024)

Cue Wasserstein GAN with Gradient Penalty, CWGAIN-GP



temporal cue

Observations from nearby months

spatial cue

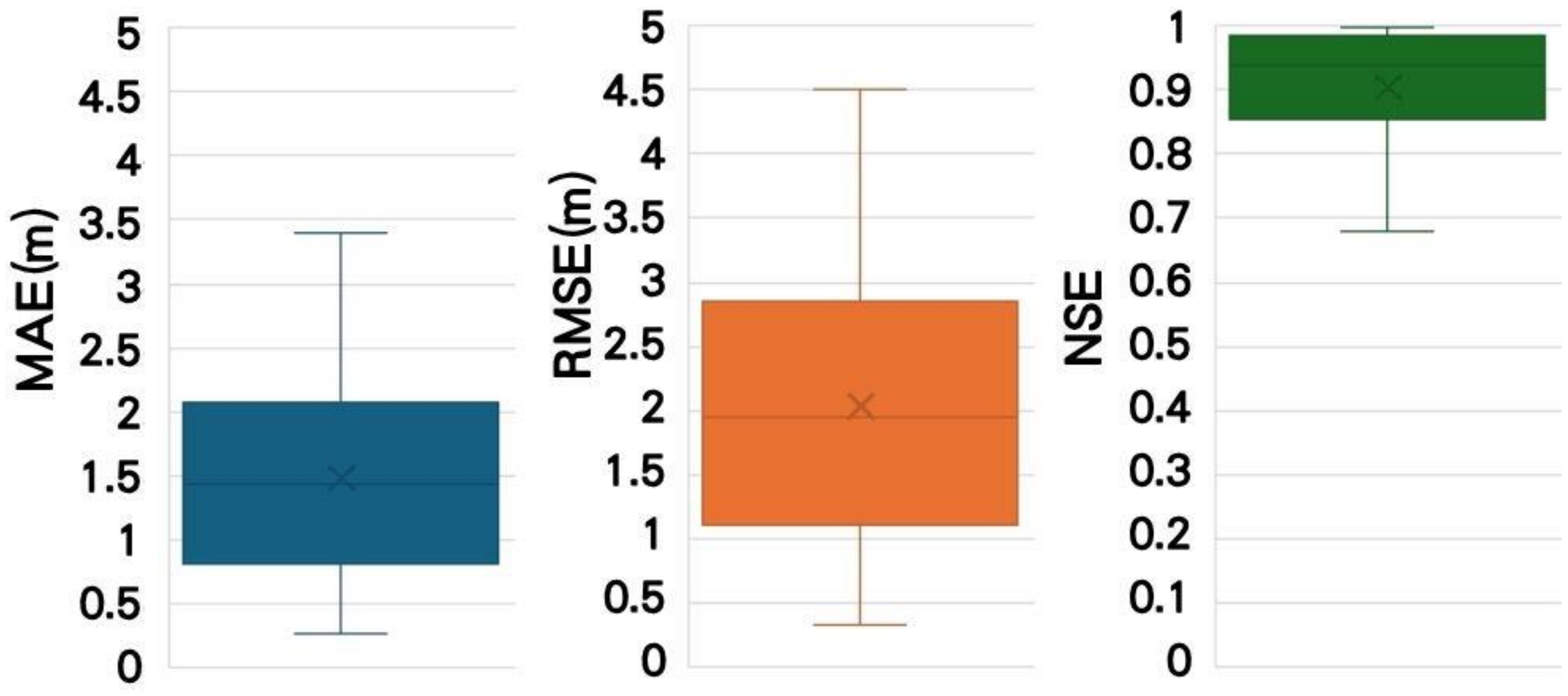
Observations from nearby stations

cue weights

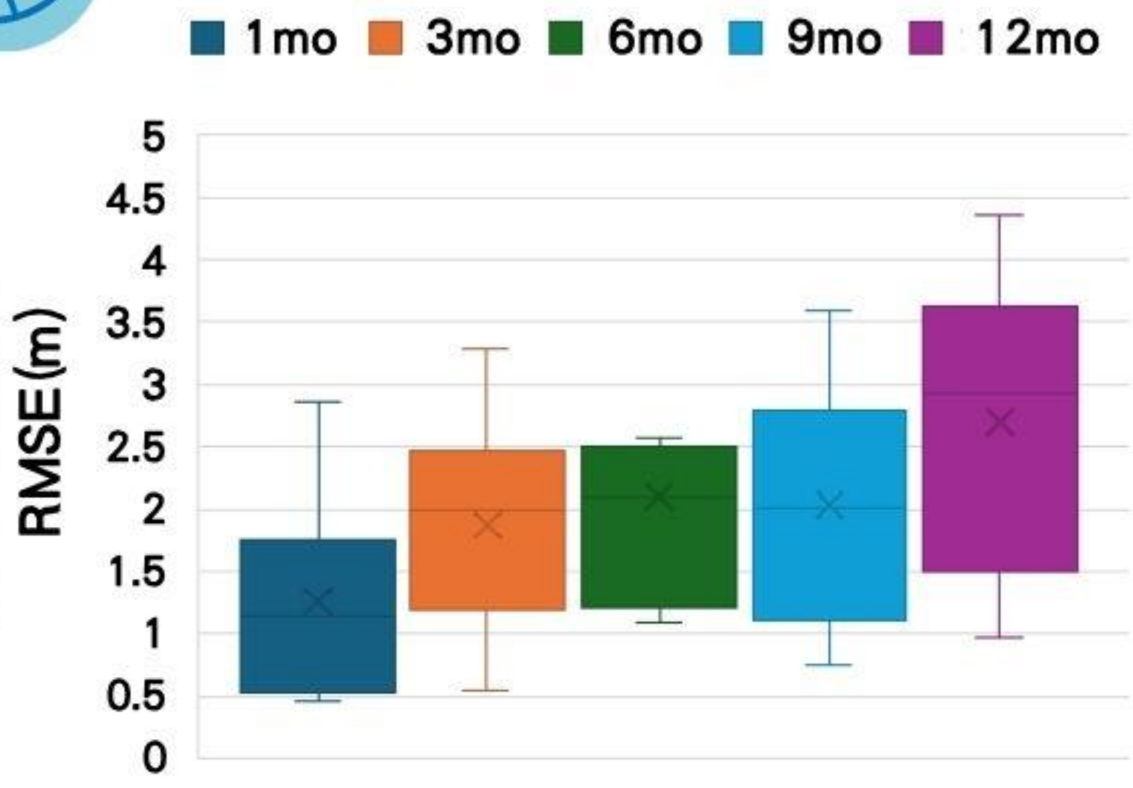
0.5 (temporal) +0.5 (spatial)

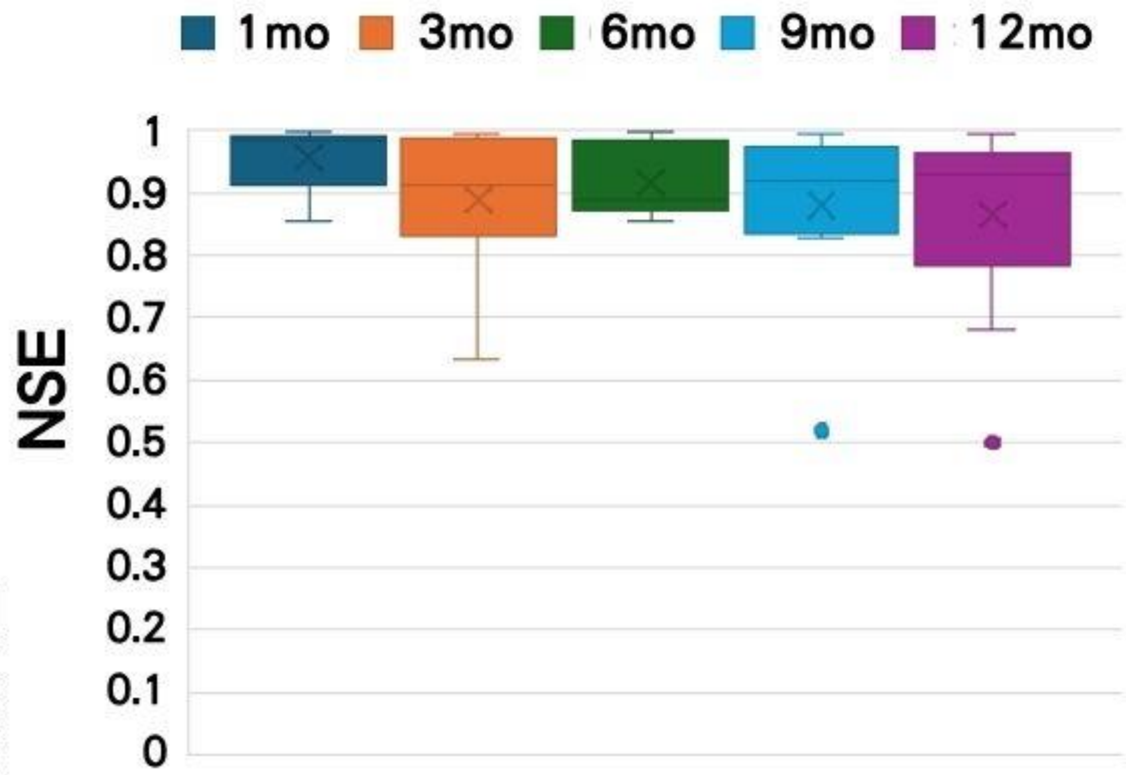
Groundwater Imputation: Overall Performance

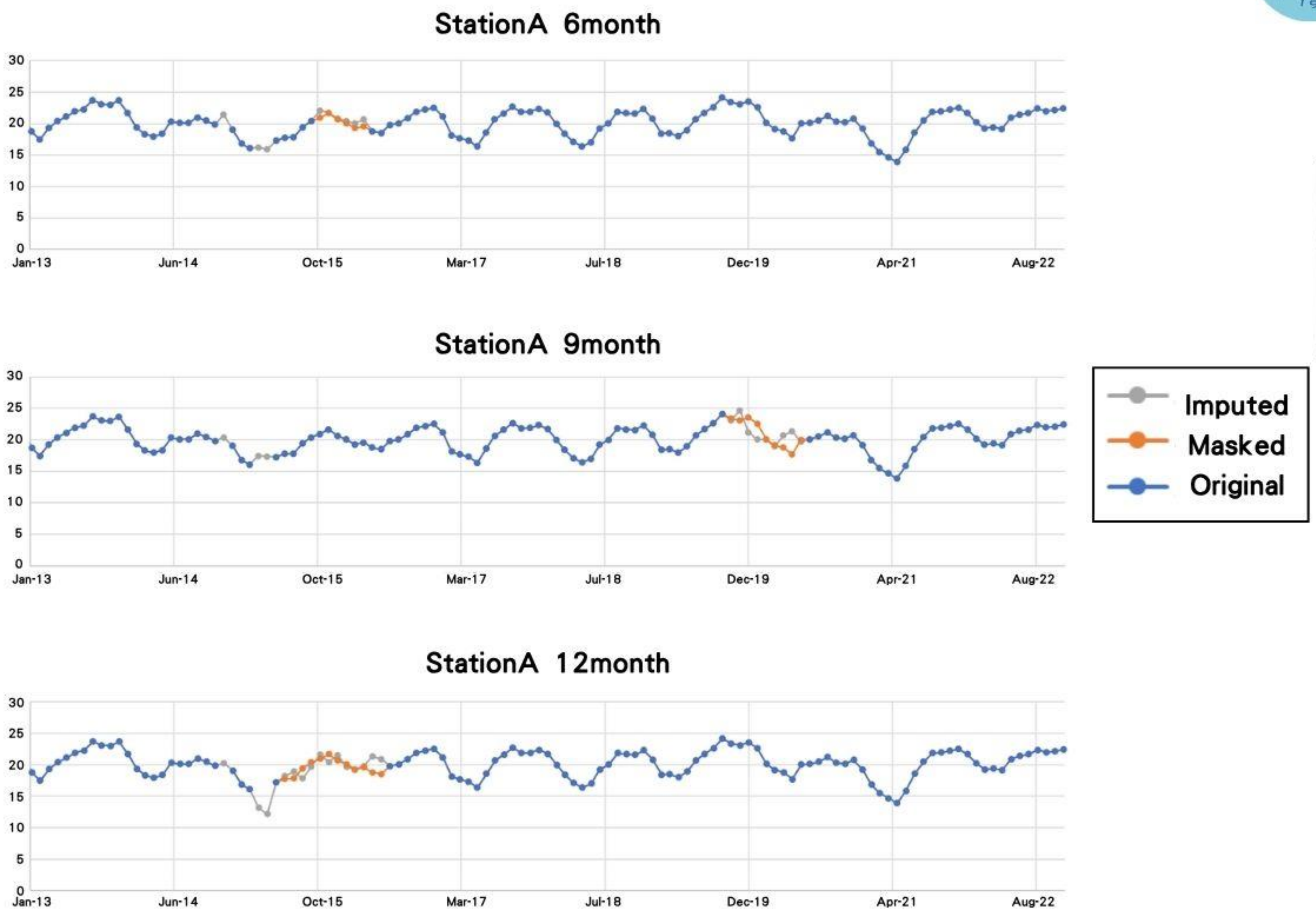
	MAE(m)	RMSE(m)	NSE
Average	1.48	2.04	0.90
Max	3.39	4.50	0.99
Min	0.27	0.32	0.50



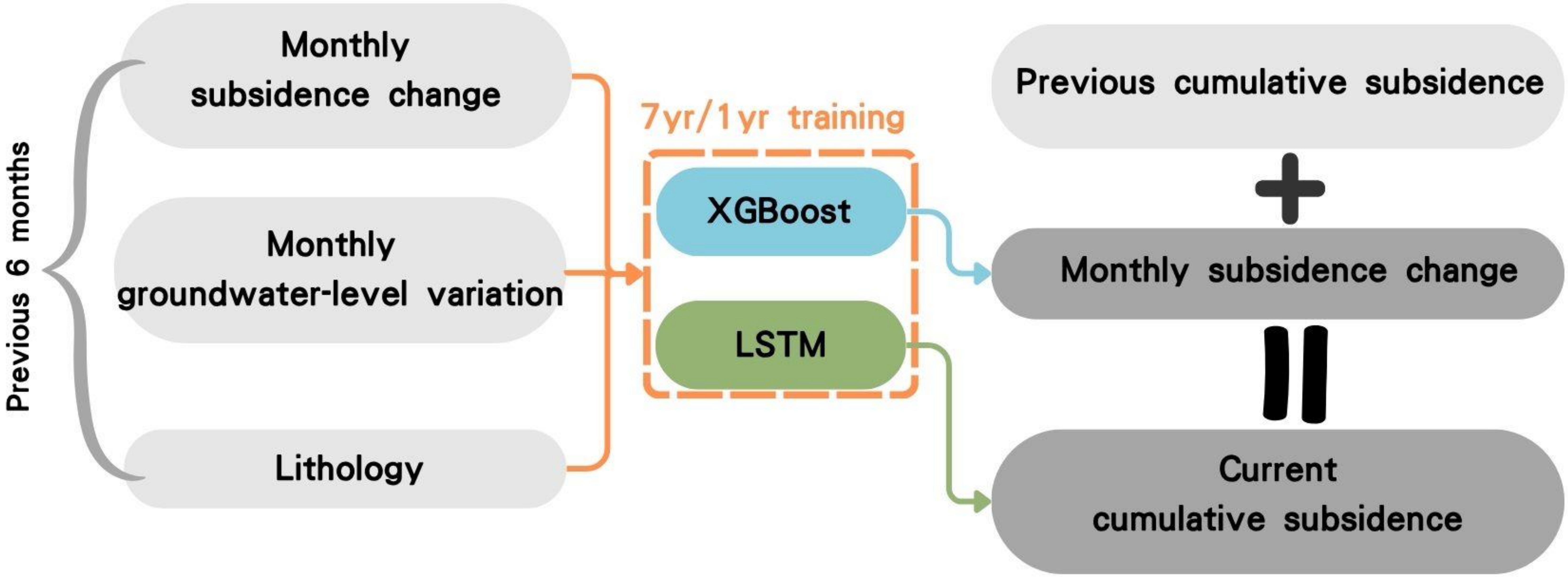
Groundwater Imputation: Contiguous missingness performance



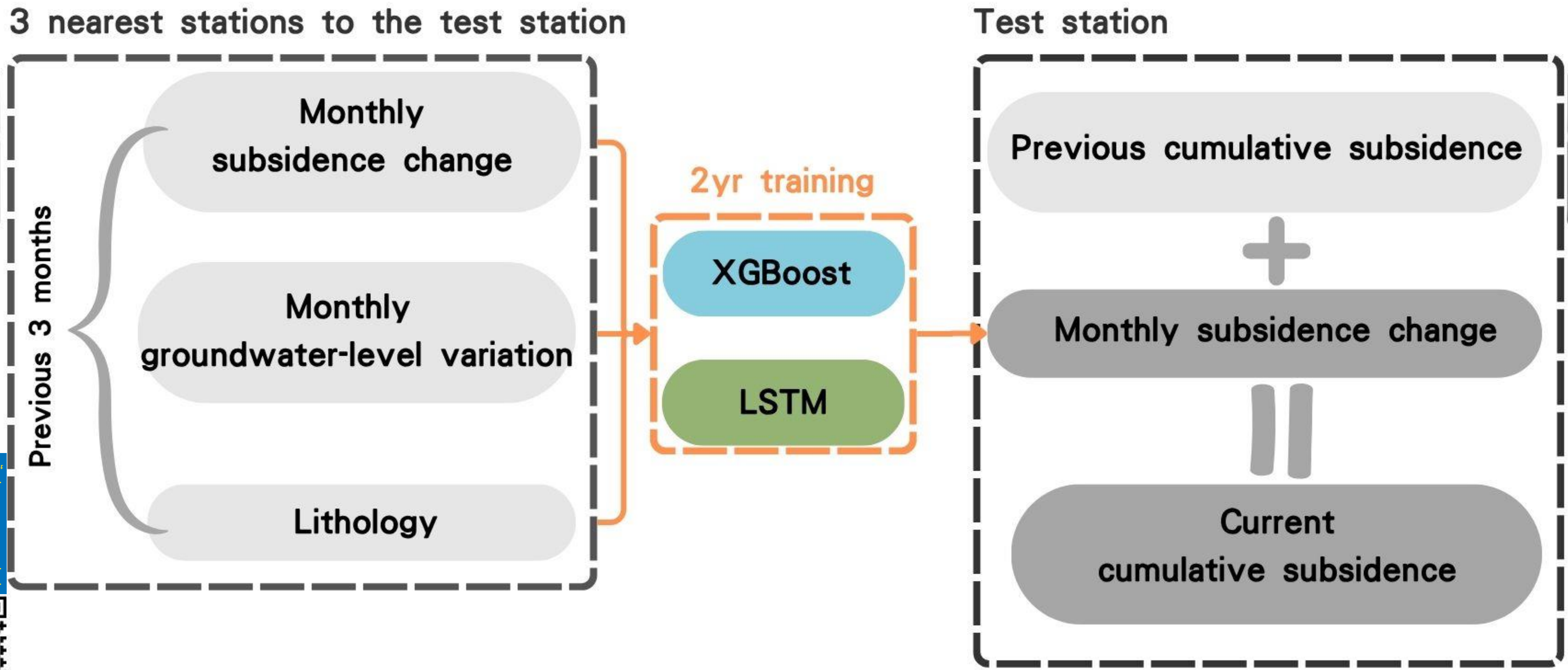




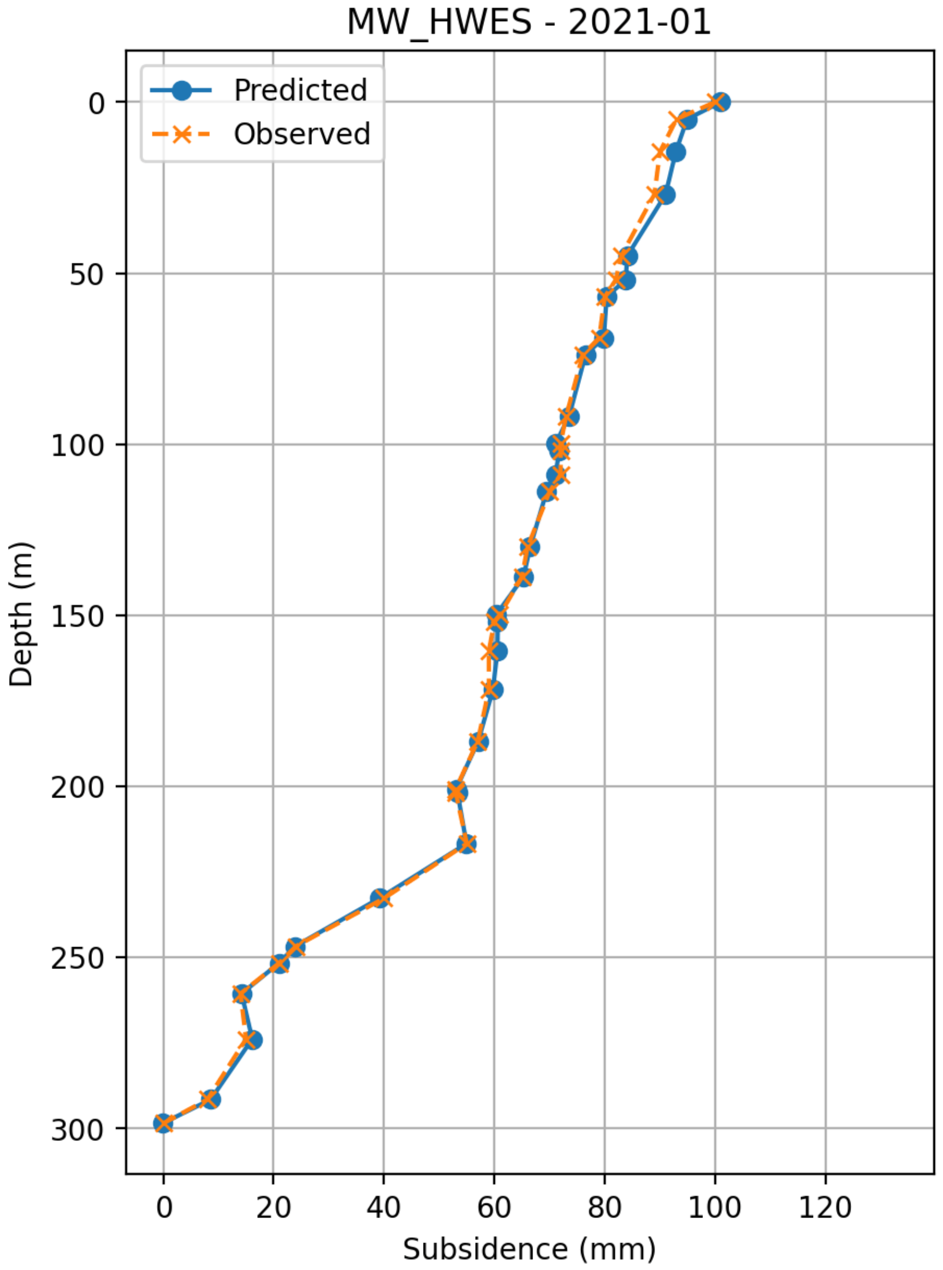
Temporal prediction

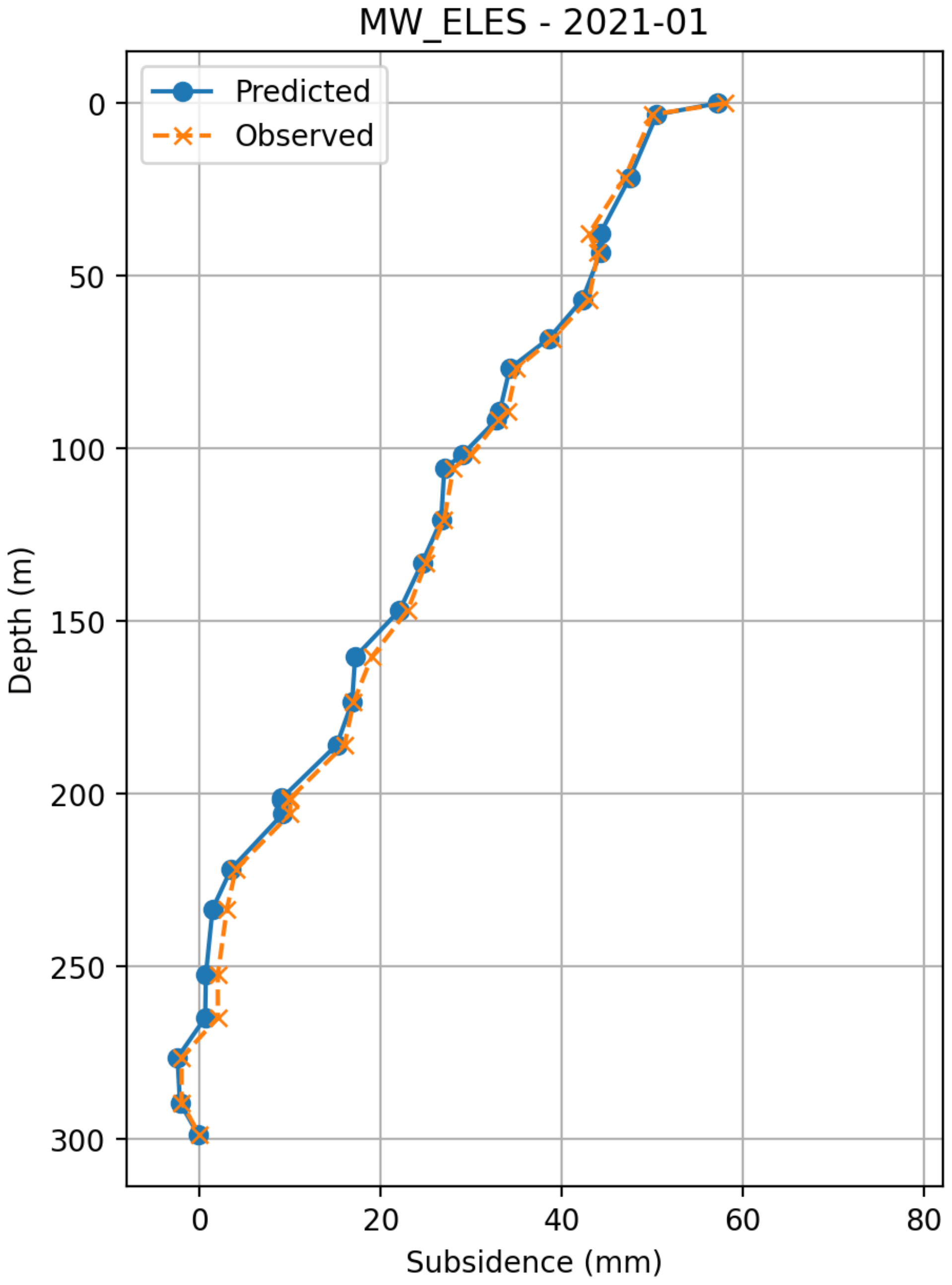


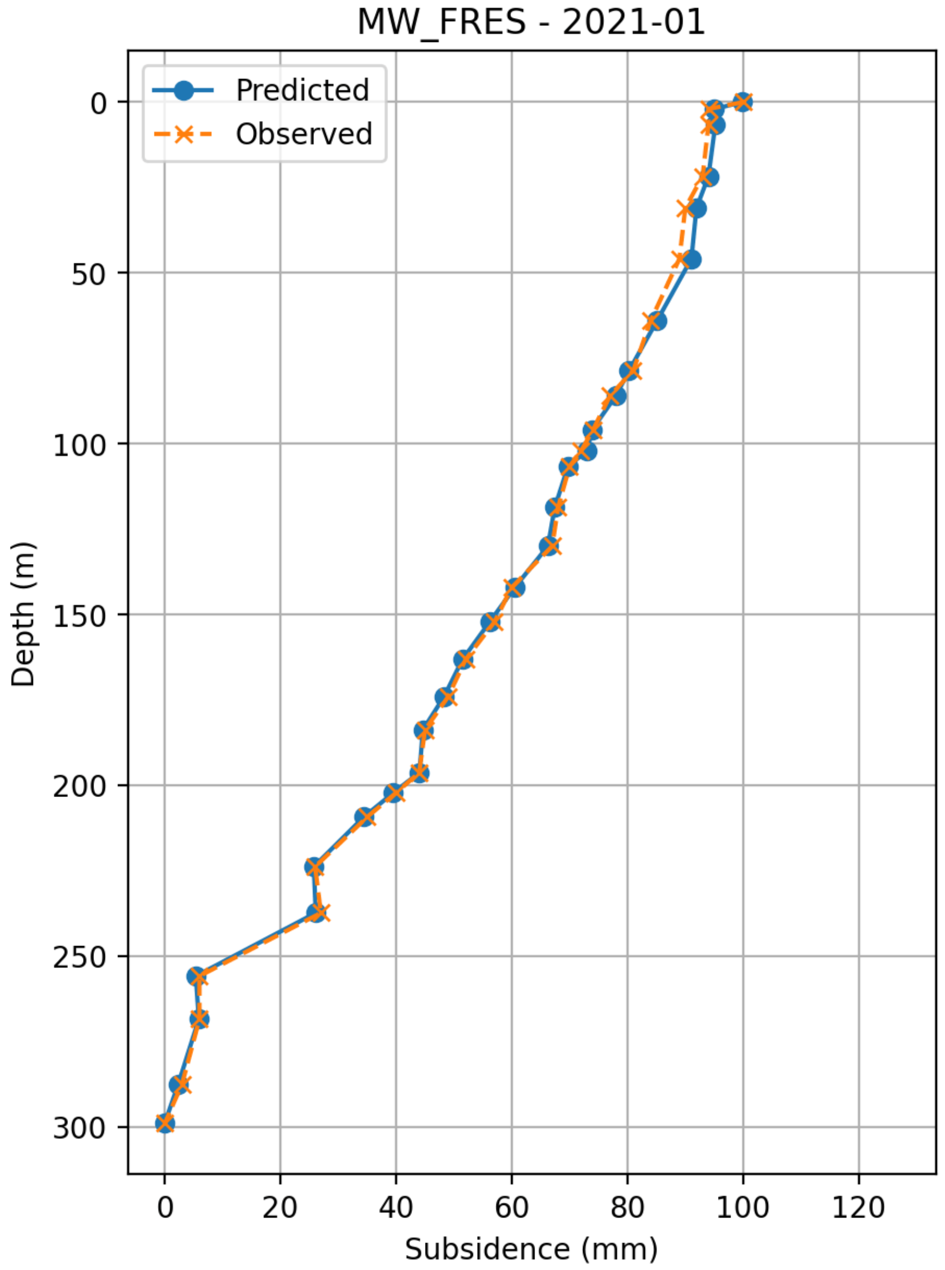
Spatial prediction



Temporal prediction examples



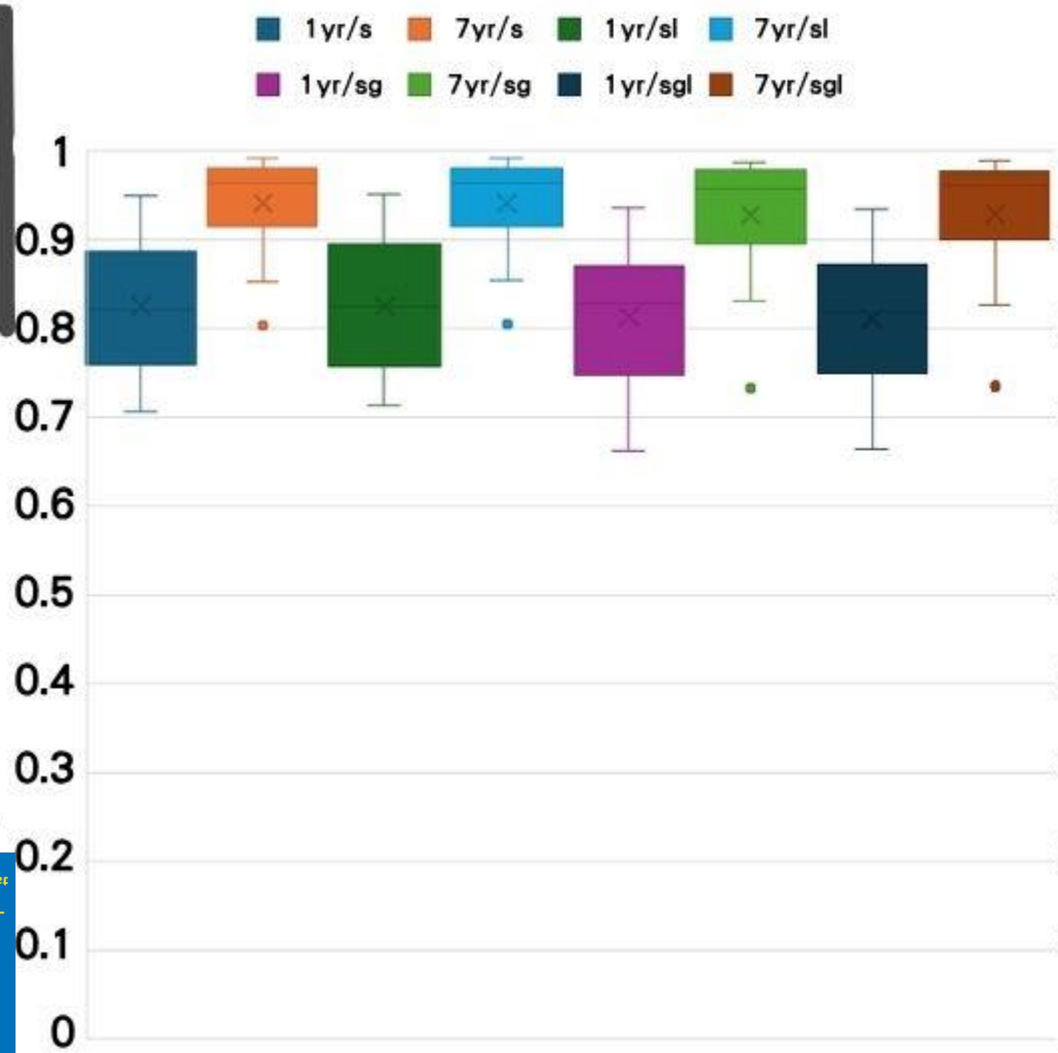




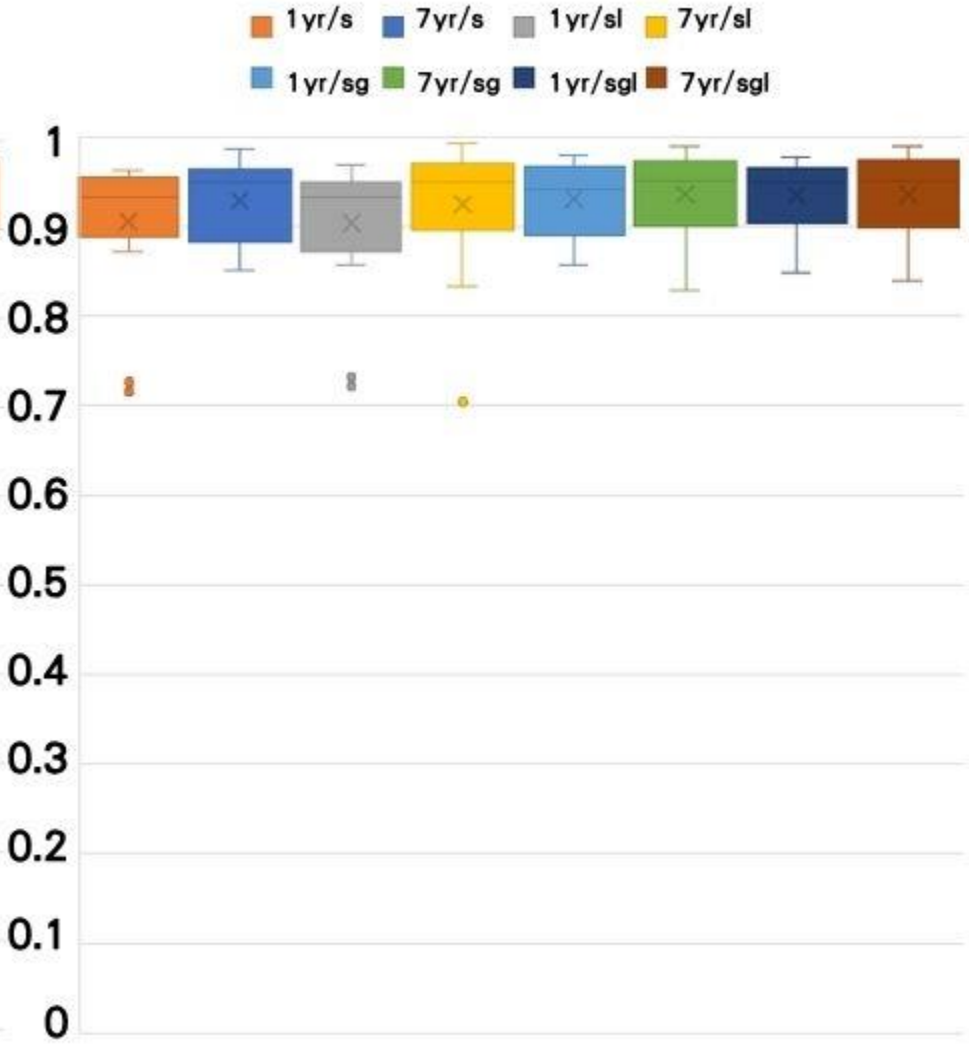
GIF：https://drive.google.com/drive/folders/1-5XmfX8jcO1hUL4ls1Jvr8kLWRMt1w9X?usp=sharing

Temporal Prediction—Model Comparison

LSTM
R^2 distribution



XGBoost
R^2 distribution



s: land subsidence, g: groundwater-level variation, l: lithology

LSTM

Rank	Training period/Inputs	RMSE(mm)	R^2
Best	7yr / LS, LI	8.41	0.94
Worst	1yr / LS, GWL, LI	17.92	0.81

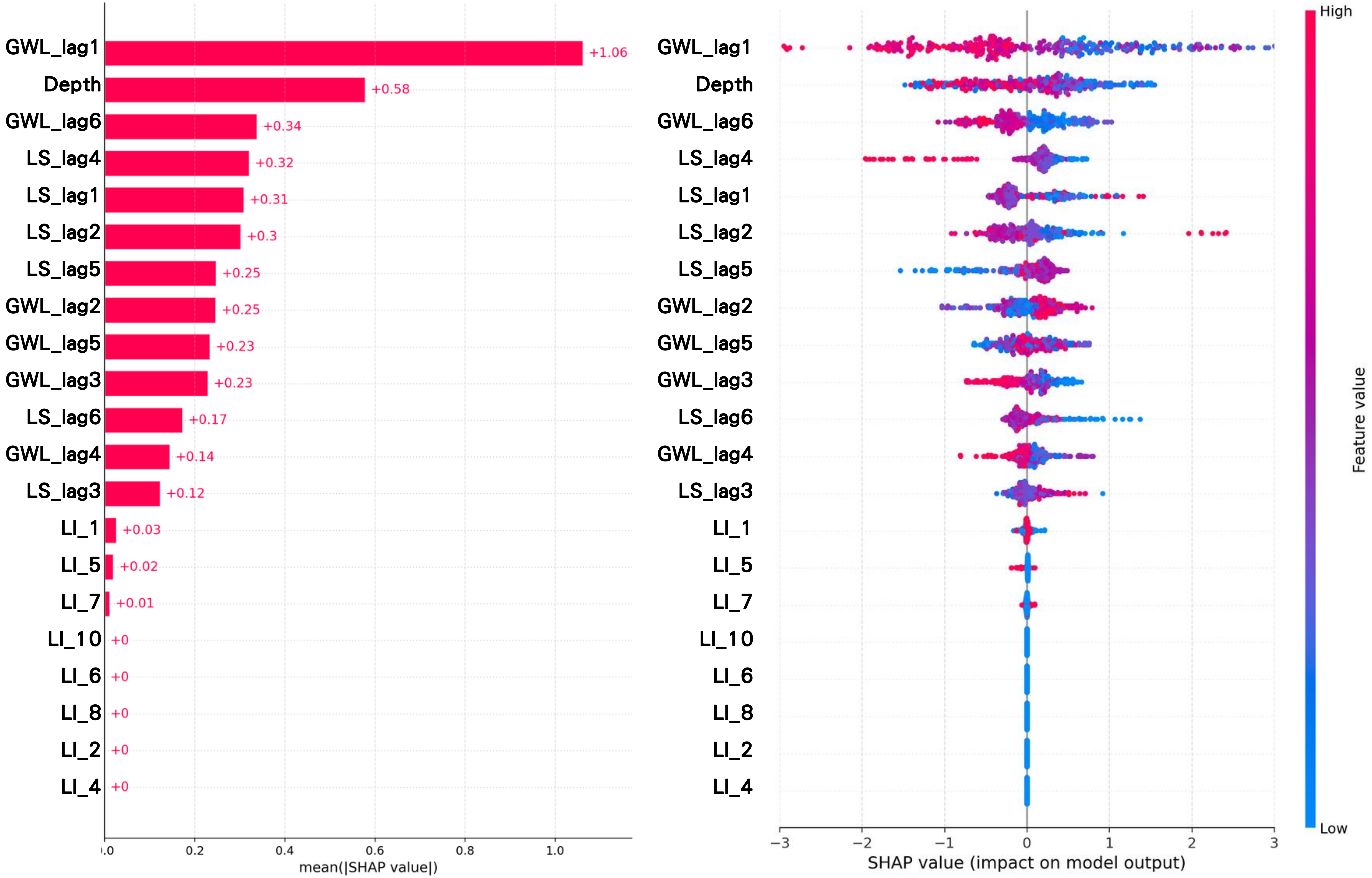
XGBoost

Rank	Training period/Inputs	RMSE(mm)	R^2
Best	7yr / LS, GWL, LI	8.70	0.94
Worst	1yr / LS	10.88	0.91

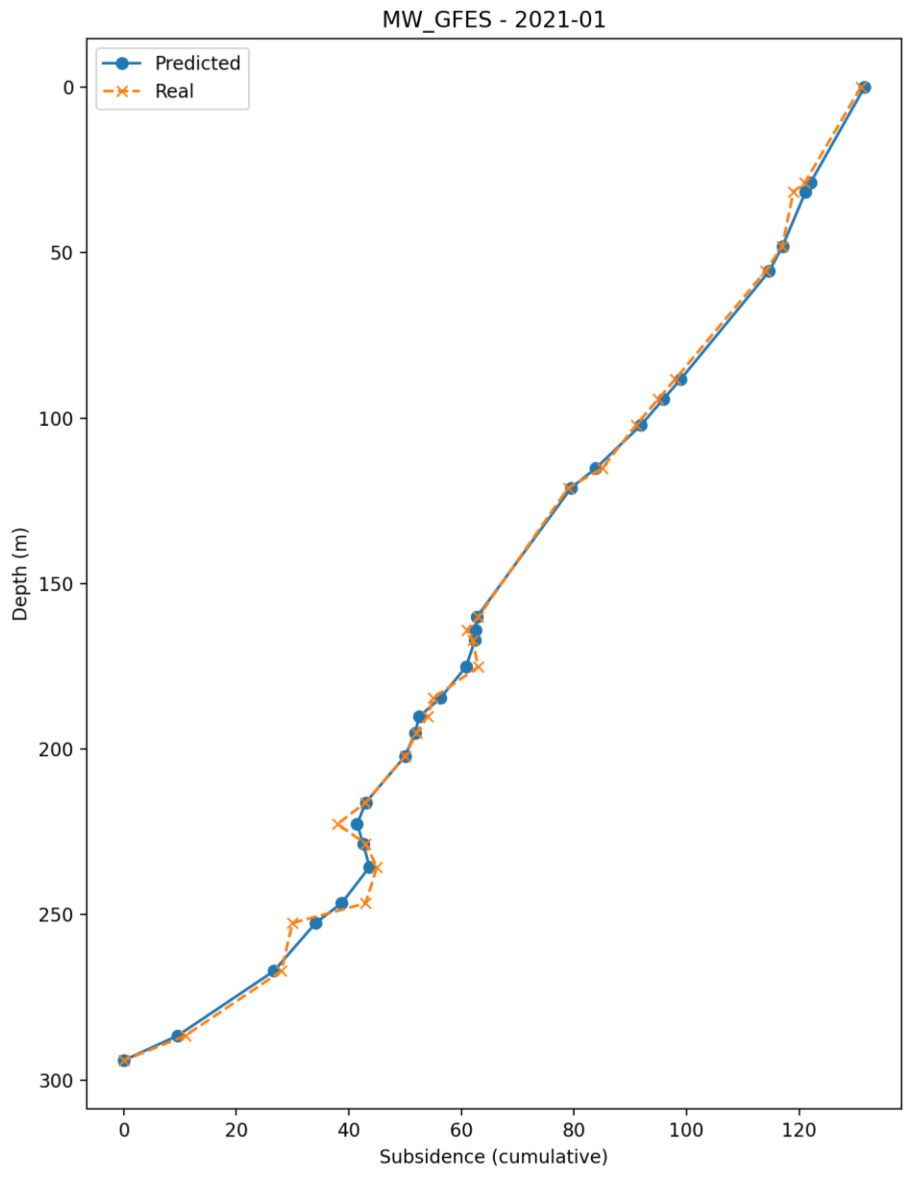
LS: land subsidence, GWL: groundwater-level variation, LI: lithology

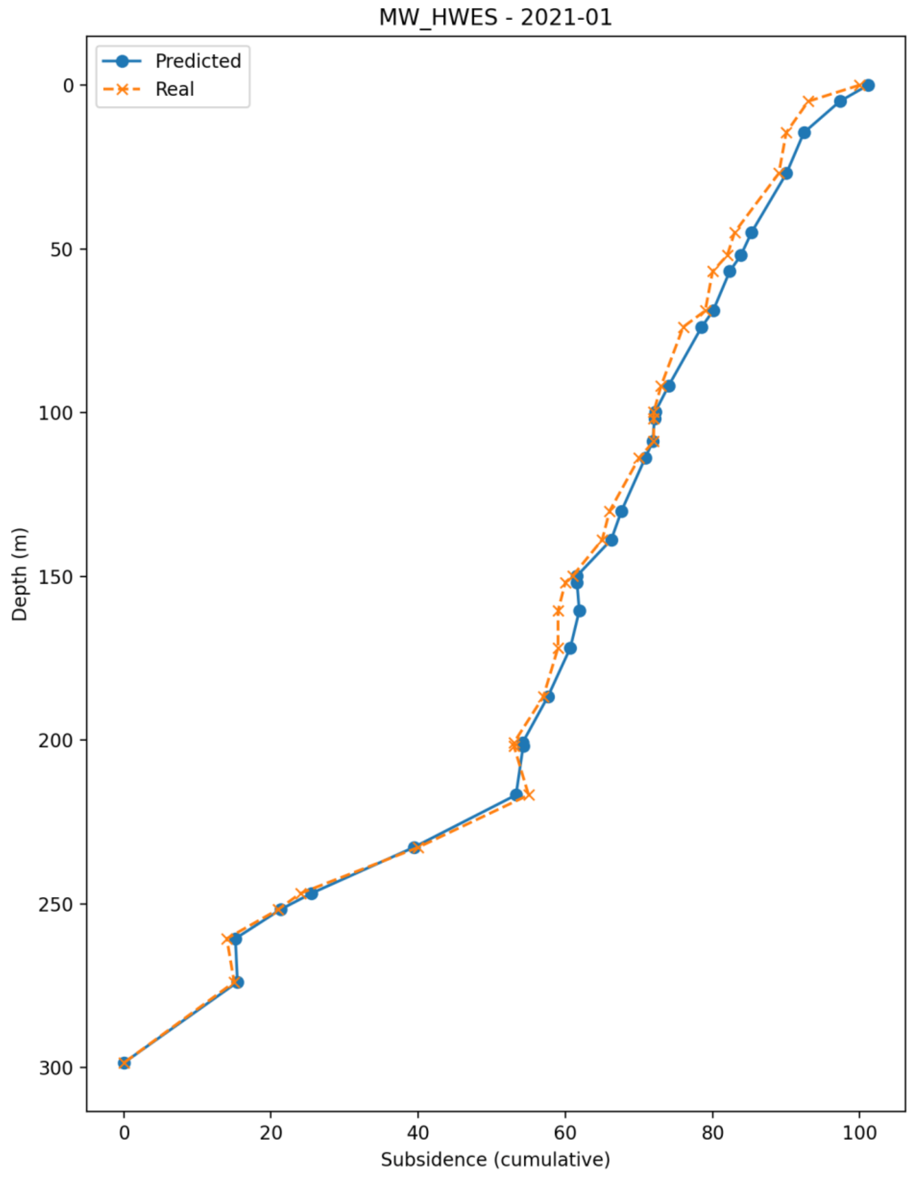
LSTM showed wider R^2 variation, while XGBoost maintained more stable performance across settings.

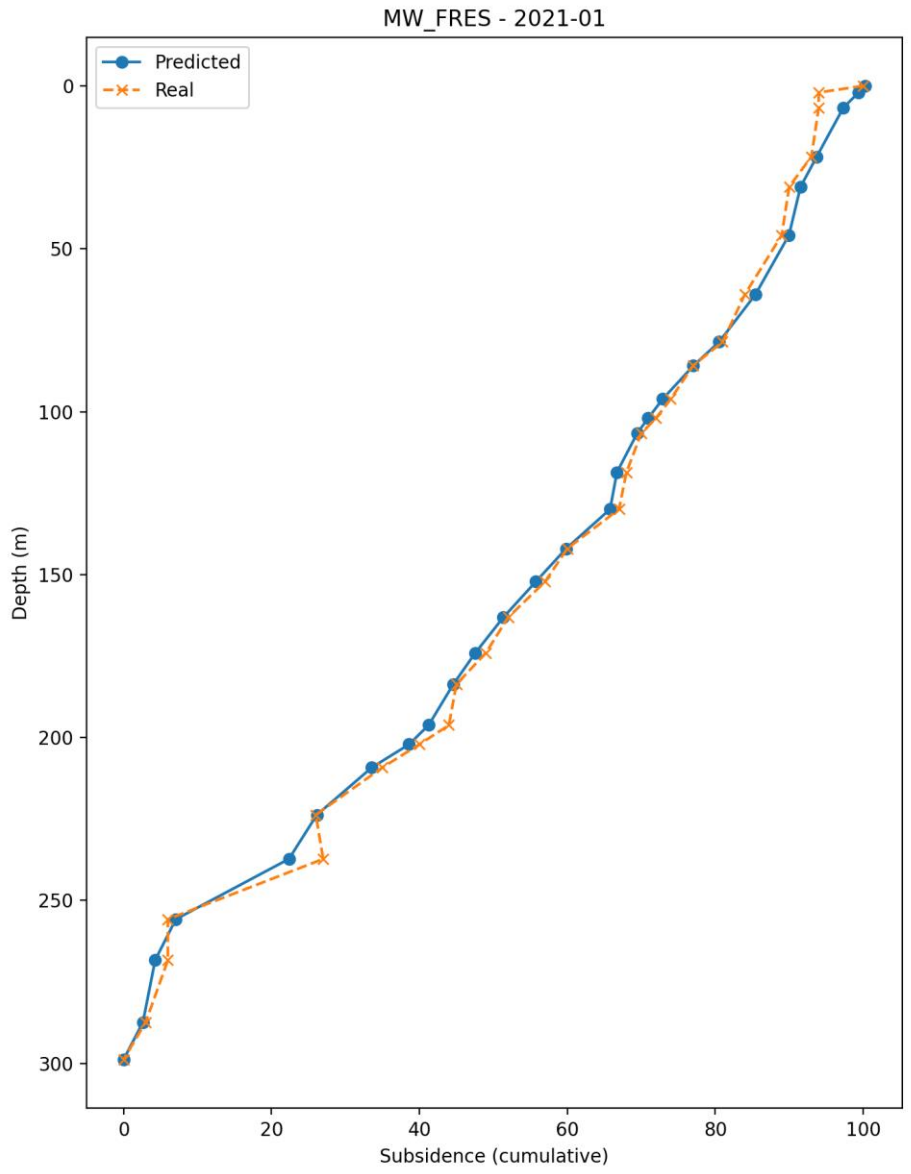
SHAP interpretation



Spatial prediction at unseen stations







GIF：https://drive.google.com/drive/folders/1-5XmfX8jcO1hUL4ls1Jvr8kLWRMt1w9X?usp=sharing

Spatial rollout performance

LSTM

MW_HWES	Baseline R^2	Model R^2		Baseline R^2	Model R^2
2021-1	1.00	1.00	2021-7	0.84	0.88
2021-2	0.91	0.92	2021-8	0.87	0.89
2021-3	0.87	0.92	2021-9	0.88	0.87
2021-4	0.82	0.88	2021-10	0.85	0.84
2021-5	0.79	0.87	2021-11	0.82	0.82
2021-6	0.70	0.80	2021-12	0.83	0.82

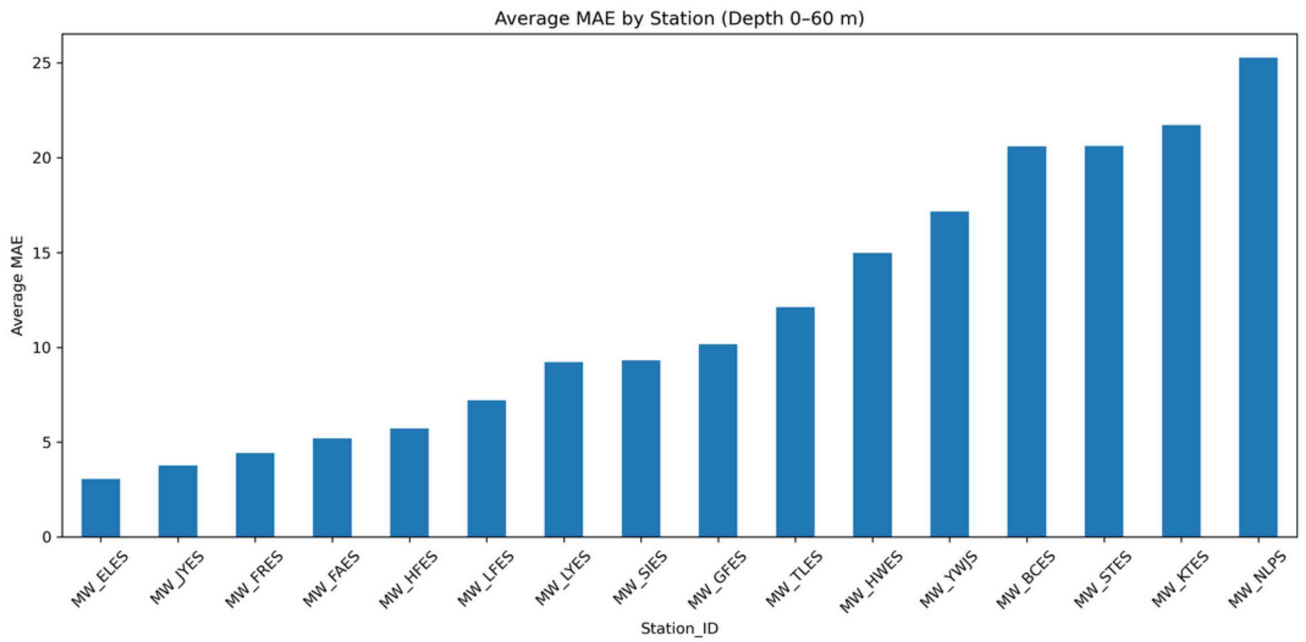
XGBoost

MW_HWES	Baseline R^2	Model R^2		Baseline R^2	Model R^2
2021-1	1.00	1.00	2021-7	0.84	0.98
2021-2	0.91	0.95	2021-8	0.87	0.96
2021-3	0.87	0.99	2021-9	0.88	0.93
2021-4	0.82	1.00	2021-10	0.85	0.97
2021-5	0.79	0.98	2021-11	0.82	0.98
2021-6	0.70	0.94	2021-12	0.83	0.98

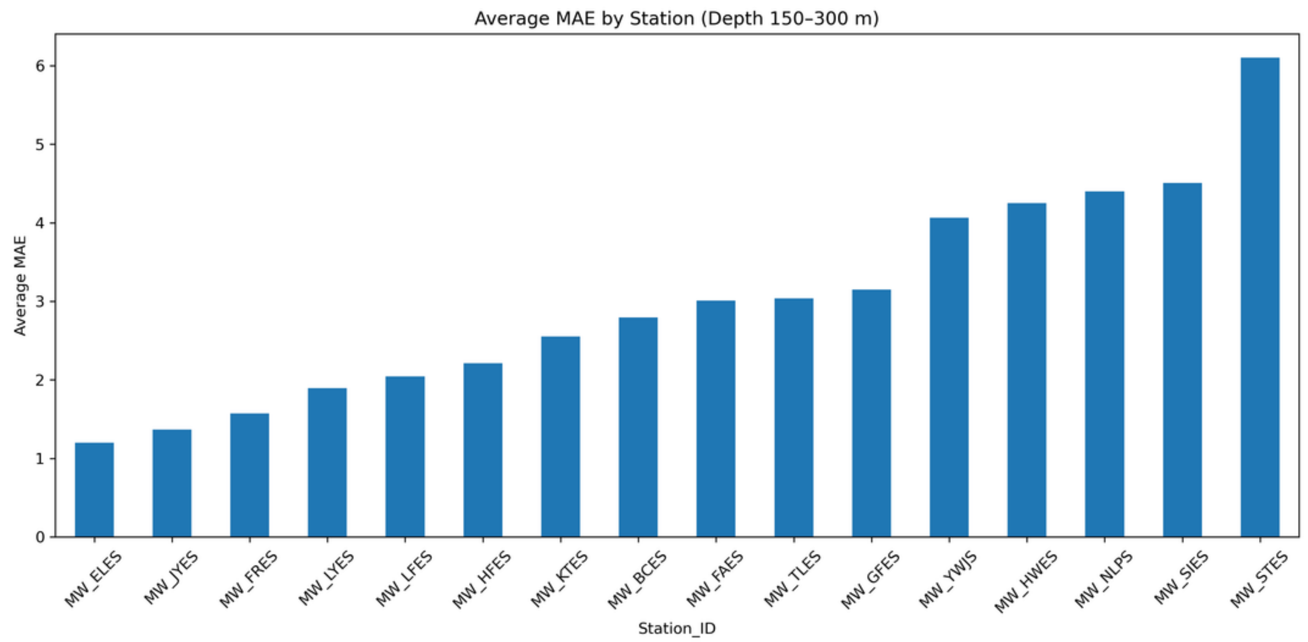
Period	Baseline R^2	Model R^2
6-month rollout	0.70 - 1.00	0.80~1.00
12-month rollout	0.70~1.00	0.80~1.00

Period	Baseline R^2	Model R^2
6-month rollout	0.70 - 1.00	0.94~1.00
12-month rollout	0.70~1.00	0.93~1.00

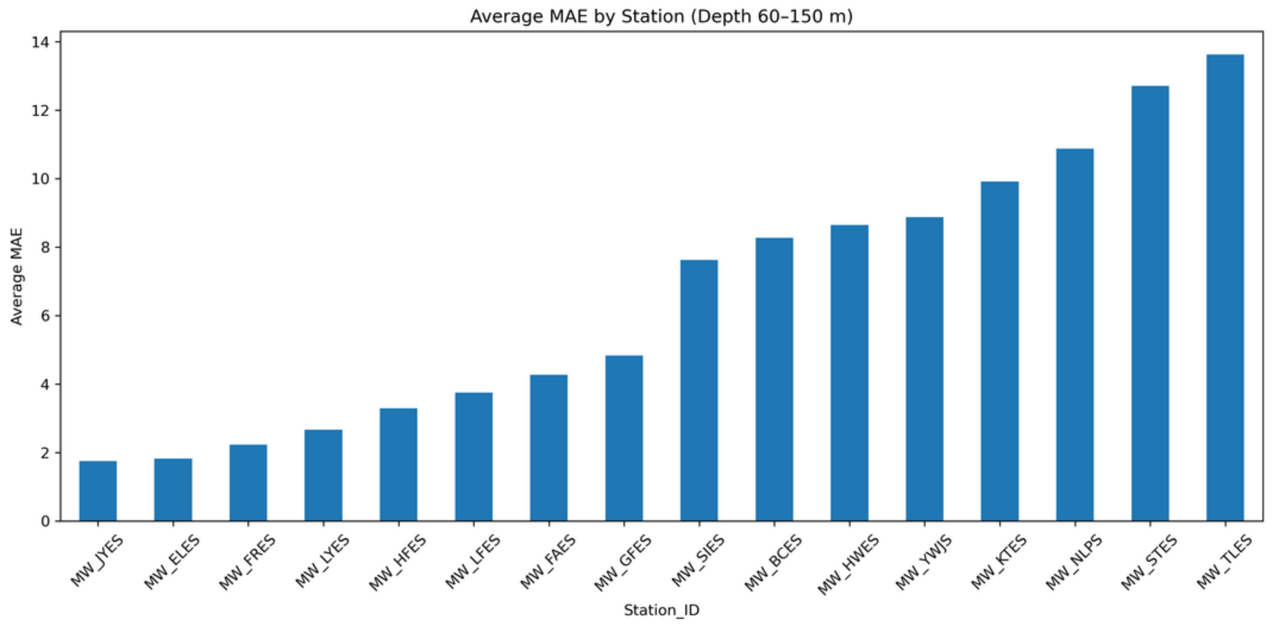
Depth-dependent error analysis



Shallow layer(0m~60m)



Deep layer(150m~300m)



Middle layer(60m~150m)

Depth	0m~60m	60m~150m	150m~300m
Median MAE(mm)	10	5	2

Results Summary

- Machine learning models effectively predicted cumulative land subsidence.
- XGBoost showed more stable performance, while LSTM was more sensitive to input settings.
- Nearby monitoring stations improved spatial prediction at unseen stations.
- Shallow layers showed larger prediction errors than deeper layers.

Conclusion

- CWGAIN-GP produced reliable groundwater-level imputation results.
- LSTM and XGBoost effectively predicted cumulative land subsidence in both temporal and spatial tasks.
- SHAP results showed that previous subsidence and groundwater-related variables were key predictors.
- Prediction accuracy varied with depth, with larger errors in shallow layers.