Review Paper

Machine learning for faster estimates of groundwater response to artificial aquifer recharge

Fernandes, V. J., de Louw, P. G., Bartholomeus, R. P., & Ritsema, C. J. (2024). *Journal of Hydrology*, *637*, 131418.

Presenter: Vo Thi Kim Huong

Advisor: Prof. Jui-Sheng Chen

Date: 2024/11/27

OUTLINE

1 INTRODUCTION

Motivation and Objective

2 METHODOLOGY

How were Machine Learning models trained?

3 RESULTS & DISCUSSION

Model Performance

4 CONCLUSION

Review original Objective

RESEARCH MOTIVATION

Expansion of surface drainage network

Increased exploitation of groundwater

Impact of drought

MANAGED ARTIFICIAL RECHARGE (MAR)

MAR: infiltration, direct injection, and filtration techniques.

⬇

identify the optimal location, recharge rate and combination of the recharge sites

Challenge: Traditional numerical groundwater models **too slow** for decision-making in MAR site optimization.

Machine Learning (ML) models: capture interaction between variables without run detail simulation
→ predict groundwater response to recharge **quickly** and **efficiently**.

OBJECTIVE

Provide **faster** and **more efficient** estimates of **groundwater response** to artificial aquifer recharge

- Develop Faster & Efficiency Predictive Models
- Balance Between Speed and Accuracy
- Determine the Required Training Data
- Determine the Physical Characteristics

METHOD

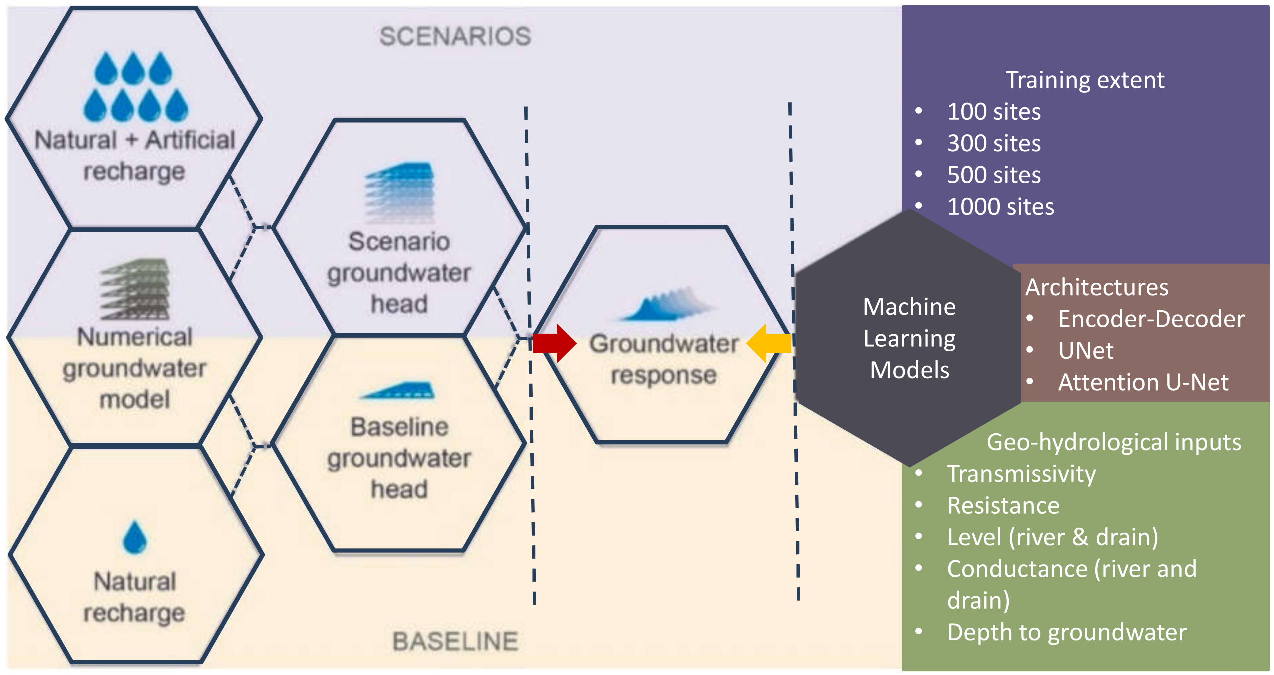


Fig. 1 Two main parts of methodology

Data Generation: Recharge rates (5–25 mm/day) and training area (0.01–1 km²) were selected using **random function** (Latin Hypercube Sampling and Orthogonal Array Latin Hypercube Sampling)

NUMERICAL MODEL (AMIGO – Actueel Model Instrument Gelderland Oost)

Simulate the groundwater system in the Baakse Beek catchment

Study area:

Baakse Beek catchment (Netherlands).

MODFLOW-2005: Tile drainage (**DRN package**), ditches/streams (**RIV package**) ➡ the surface water network drains the groundwater.

Boundary: maintain at distance of three times the leakage factor ➡ ensure not influence

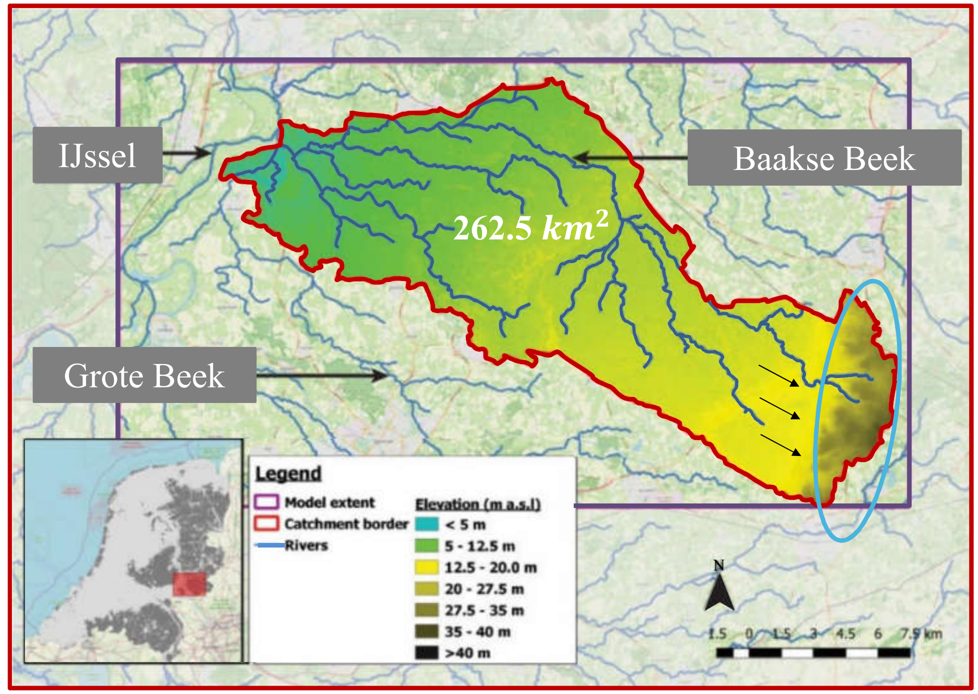


Fig. 2 Baakse Beek catchment (Netherland)

In steady-state simulations, the storage coefficient is zero and not used in the numerical or ML models

Geology: Pleistocene sands, 200 m thick; highly transmissivity → enabling groundwater flow

MACHINE LEARNING MODEL: HOW DO THEY WORK?

Predict steady-state groundwater response to artificial recharge.

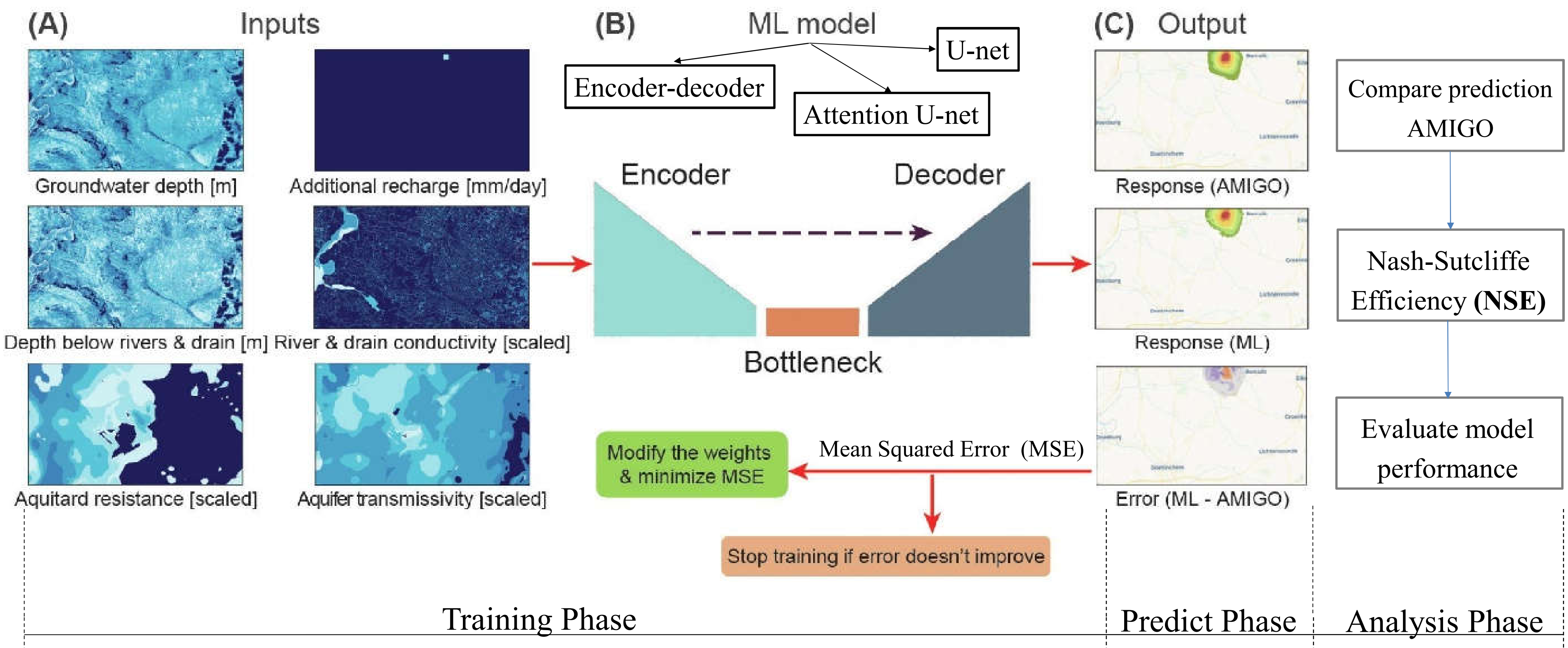


Fig. 3 Training process of ML models

KEY CHARACTERISTICS

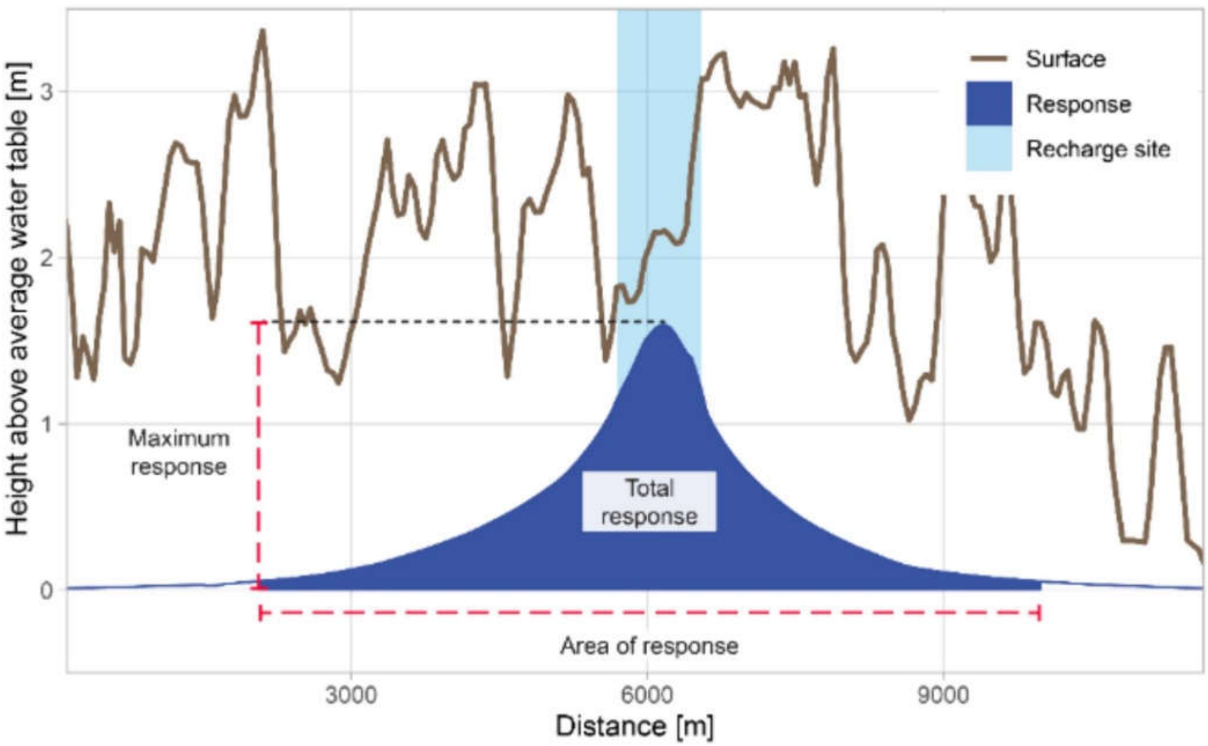


Fig. 4. Cross-sectional view of how groundwater response to artificial recharge

Maximum response

Highest increase in groundwater head

Area of response

The spatial extent where the groundwater head increases by more than 1 cm.

Total response

Total volume affected

together quantify performance of ML models in predict groundwater system's reaction to artificial recharge.

MODEL PERFORMANCE: FACTOR EFFECT

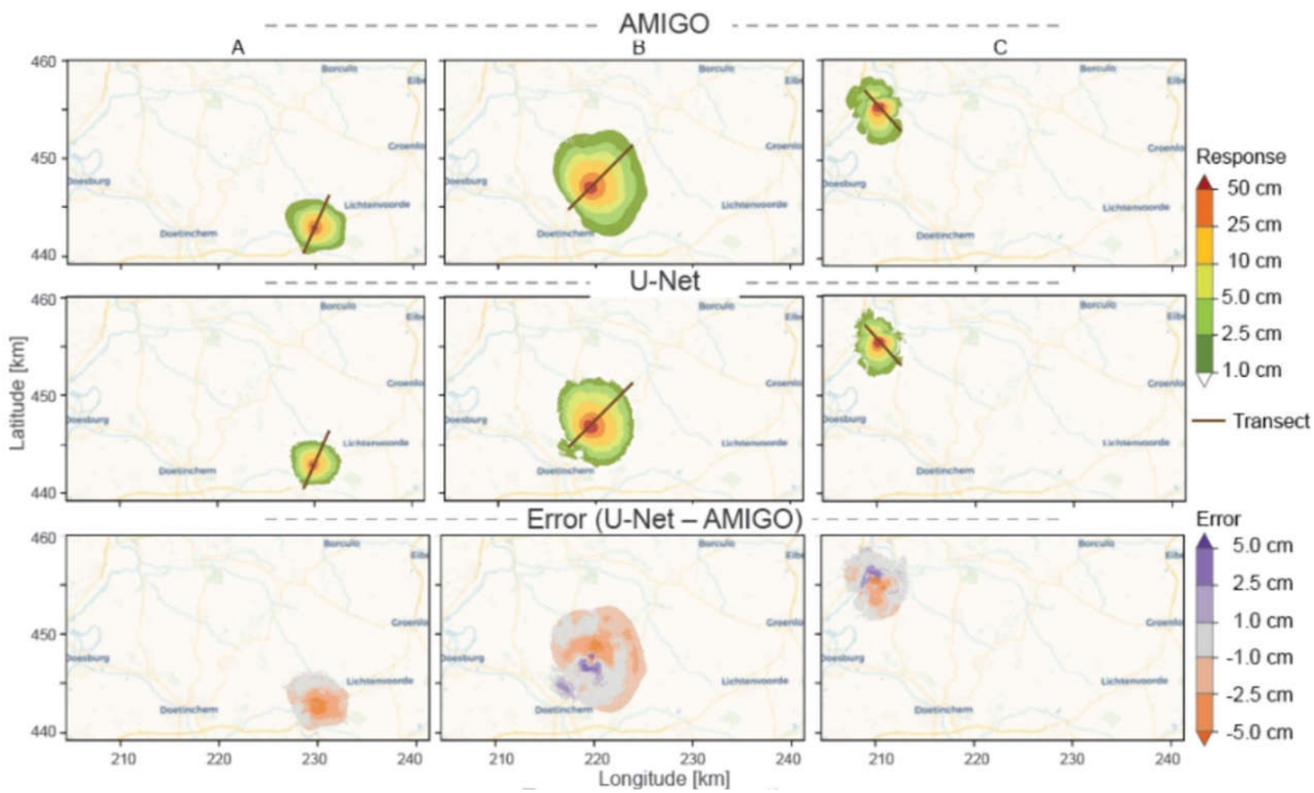


Fig. 5 Map view of response for 3 recharge sites

The recharge sites selected for their **asymmetric response** caused by **the interaction** between **the groundwater and the surface** water network (Groote Beek River and IJssel River).

MODEL PERFORMANCE: ACCURACY AND SPEED

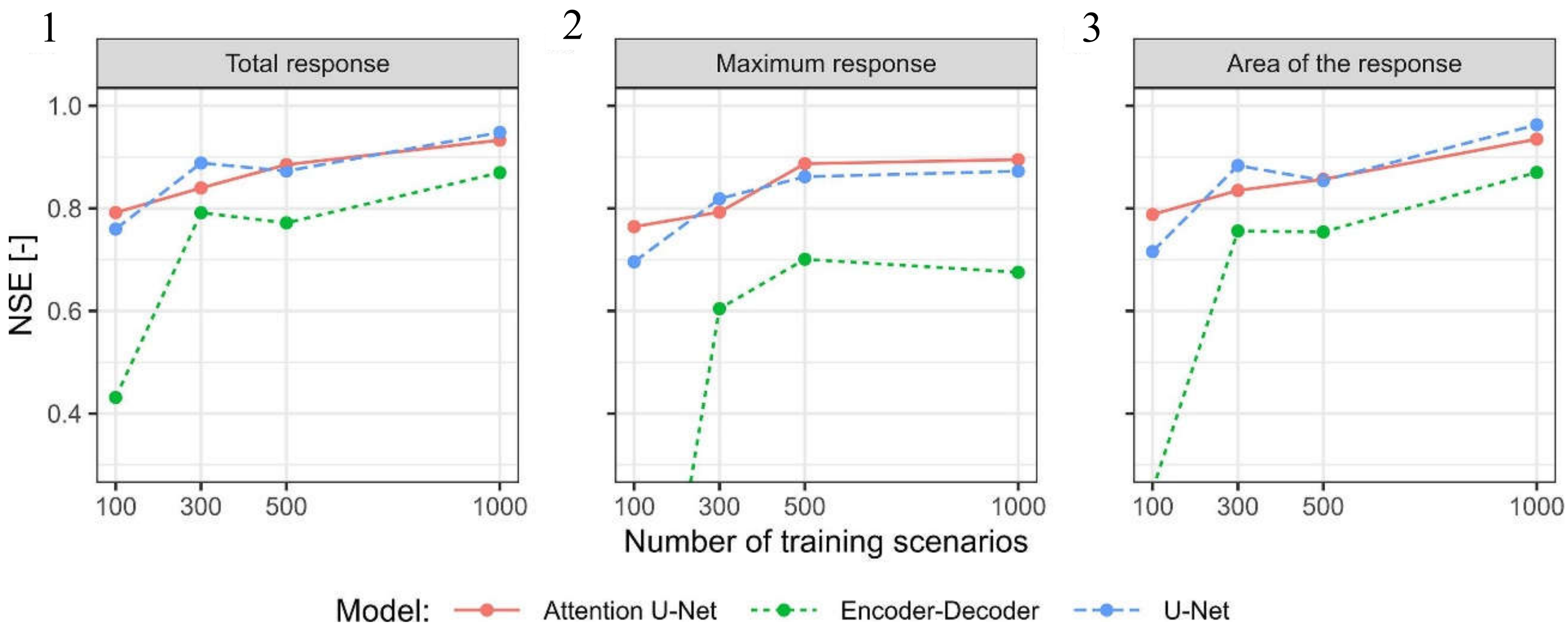


Fig. 7 NSE evaluate performance of 3 models

U-Net and Attention U-Net outperform the Encoder-Decoder model.

MODEL PERFORMANCE: ACCURACY AND SPEED

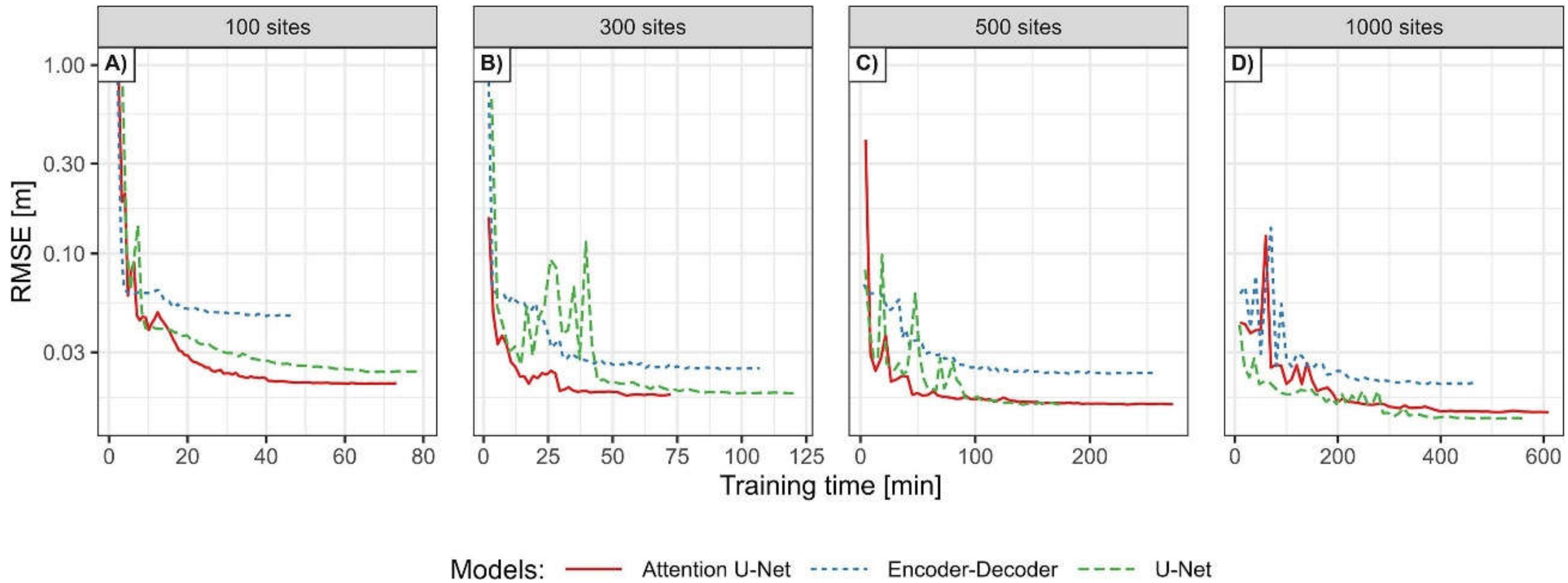


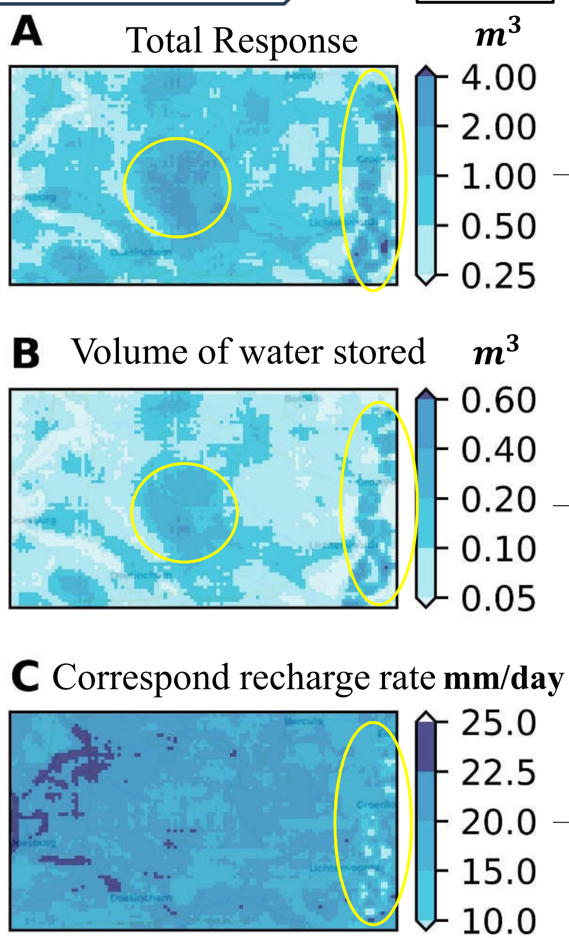
Fig. 8 Validation MSE

Add training sites improve the final results, increase the training time.

ML models: **0.06 to 0.43 seconds**
AMIGO: **1290 seconds (~21 mins) per run**.
U-Net: **3000 scenarios** in the time AMIGO runs one.

APPLICATIONS

Unit



eastern: 4.35 mil.
central: 3.0 mil.

eastern: 0.65 mil.
central: 0.45 mil.

eastern: 11 mm/day
central: 23 mm/day

- The eastern regions: **best result**

→ identify best recharge/stored water location

Compare 3 Key Characteristics of **720 recharge sites**

- ML: **144 seconds**
- AMIGO: **11 hours**

Compare multiple-location rapidly

Fig. 10. Results of ML on recharge rate of 5-25mm/day over 10 ha – 7.722 recharge sites

MODEL LIMITATION AND INSIGHTS

<table>
<tr><th>MODEL</th><th>EVALUATION TIMES (seconds)</th><th>INPUTS</th><th>NSE (TOTAL RESPONSE)</th><th>STRENGTH</th><th>LIMITATIONS</th></tr>
<tr><td>Encoder-Decoder</td><td>0.06 – 0.43</td><td rowspan="3">6 Inputs</td><td>0.75</td><td>Simpler, lower computational cost</td><td>Struggles with complex spatial details</td></tr>
<tr><td>U-Net</td><td>0.09 - 0.11</td><td>0.95 (best)</td><td>Best performance, skip connections capture spatial details</td><td>Higher memory requirement</td></tr>
<tr><td>Attention U-Net</td><td>0.09 - 0.11</td><td>Similar to U-Net</td><td>Focuses on important regions, can improve local accuracy</td><td>No significant improvement over U-Net, higher memory demand</td></tr>
<tr><td>AMIGO Model</td><td>1290</td><td>105 Inputs</td><td></td><td>Highly accurate, attention mechanism improves focus on important regions</td><td>Extremely slow, computationally expensive</td></tr>
</table>

MODEL LIMITATION AND INSIGHTS

MODEL	EVALUATION TIMES (seconds)	INPUTS	NSE (TOTAL RESPONSE)	STRENGTH	LIMITATIONS
Encoder-Decoder	0.06 – 0.43		0.75	Simpler, lower computational cost	Struggles with complex spatial details
U-Net	0.09 - 0.11	6 Inputs	0.95 (best)	Best performance, skip connections capture spatial details	Higher memory requirement
Attention U-Net	0.09 - 0.11		Similar to U-Net	Focuses on important regions, can improve local accuracy	No significant improvement over U-Net, higher memory demand
AMIGO Model	1290	105 Inputs		Highly accurate, attention mechanism improves focus on important regions	Extremely slow, computationally expensive

⟶Allows more scenario evaluation, optimization in ground water management

HIGHLIGHT:

- **A first-step** in apply ML
- ML models: evaluate **thousands of recharge scenarios** in the time it takes the AMIGO model to **simulate one scenario**, demonstrate their potential for **real-time decision-making** in groundwater management and **optimization** of recharge strategies.
- **U-Net outperformed** other models with the best accuracy, evaluate **thousands of scenarios** with **high accuracy**
- Increasing training data improved accuracy, especially for area and total response.

LIMITATIONS:

- Steady-state conditions limits the model's applicability $\longrightarrow$ develop transient simulation to account for dynamic changes in groundwater system over time
- Consider impact of deeper aquifer
- **Higher river stages** could **reduce the river flux** and increase the response more than the model predicts

Thank You