# Moving away from deterministic solutions: A probabilistic machine learning approach to account for geological model uncertainty in groundwater modelling

**Mathias Busk Dahl**[1,2], Troels Norvin Vilhelmsen[2], Rasmus Bødker Madsen[3], and Thomas Mejer Hansen[1]

[1]Department of Geoscience, Aarhus University, Aarhus Denmark (dahl@geo.au.dk)

[2]NIRAS, Ceres Allé 3, Aarhus, Denmark

[3]Near Surface Land and Marine Geology, The Geological Survey of Denmark and Greenland (GEUS), Aarhus, Denmark

- **Decision-making related to groundwater management often relies on results from a deterministic groundwater model representing one 'optimal' solution. However, such a single deterministic model lacks representation of subsurface uncertainties. The simplicity of such a model is appealing, as typically only one is needed, but comes with the risk of overlooking critical scenarios and possible adverse environmental effects. Instead, we argue, that groundwater management should be based on a probabilistic model that incorporates the uncertainty of the subsurface structures to the extent that it is known. If such a probabilistic model exists, it is, in principle, simple to propagate the uncertainties of the model parameter using multiple numerical simulations, to allow a quantitative and probabilistic base for decision-makers. However, in practice, such an approach can become computationally intractable. Thus, there is a need for quantifying and propagating the uncertainty numerical simulations and presenting outcomes without losing the speed of the deterministic approach.**

- **This presentation provides a probabilistic approach to the specific groundwater modelling task of determining well recharge areas that accounts for the geological uncertainty associated with the model using a deep neural network. The results of such a task are often part of an investigation for new abstraction well locations and should, therefore, present all possible outcomes to give informative decision support. To overcome the significant increase in computation time, we argue that this problem can be solved using a probabilistic neural network trained on examples of model outputs. We present a way of training such a network. Ultimately, this presentation aims to contribute with a method for incorporating model uncertainty in groundwater modelling without compromising the speed of the deterministic models.**

## BACKGROUND

A deterministic numerical groundwater model of the Egebjerg catchment in Denmark is used to manage the groundwater supply of the municipality. A stochastic hydrostratigraphic model of the area is available. From the model we can draw different realizations of the subsurface with slight variations in the geology (Fig. 1). Applying different realizations as the geological setting in the groundwater model affects the estimated recharge areas of wells in the groundwater model (Fig. 2). We need the full probabilistic recharge area to calculate the potential risks of establishing a new well.

## QUESTIONS:

- **How can we use the full stochastic hydrostratigraphic model to estimate recharge areas?**
- **Can we speed this up?**
- **What is required for this to work in a decision-support tool?**

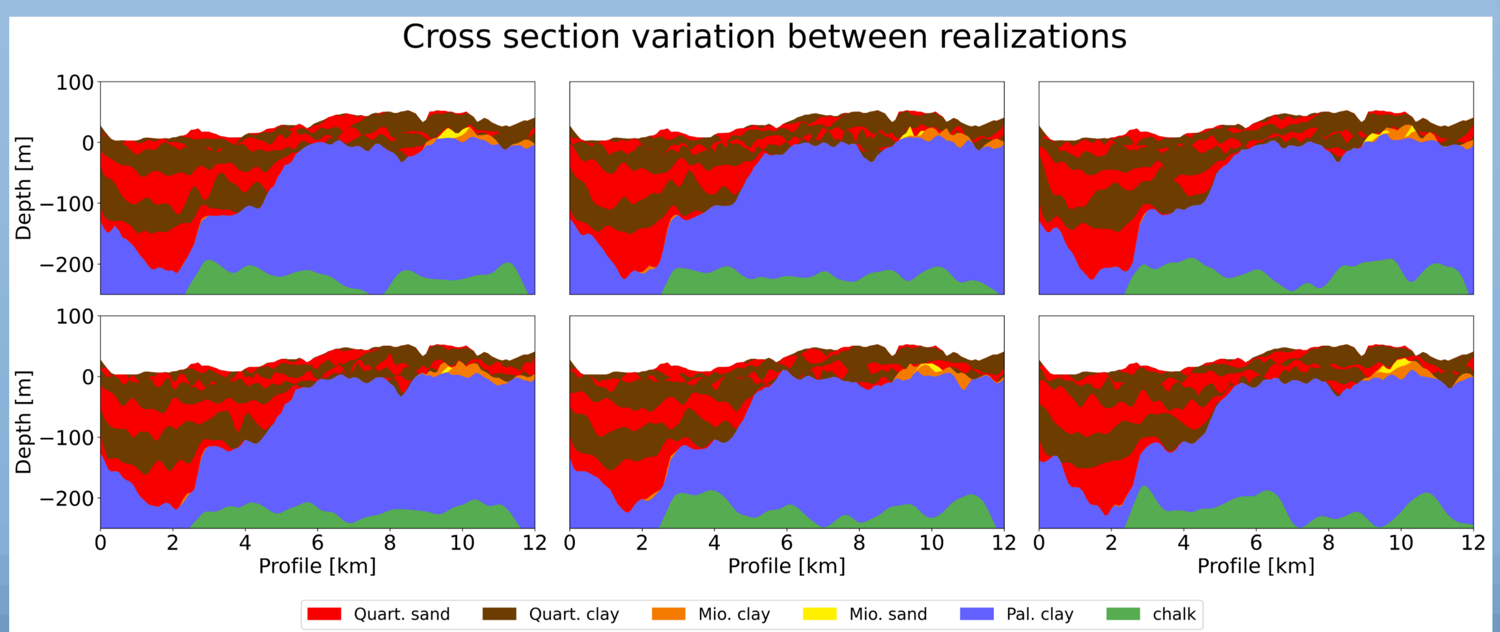
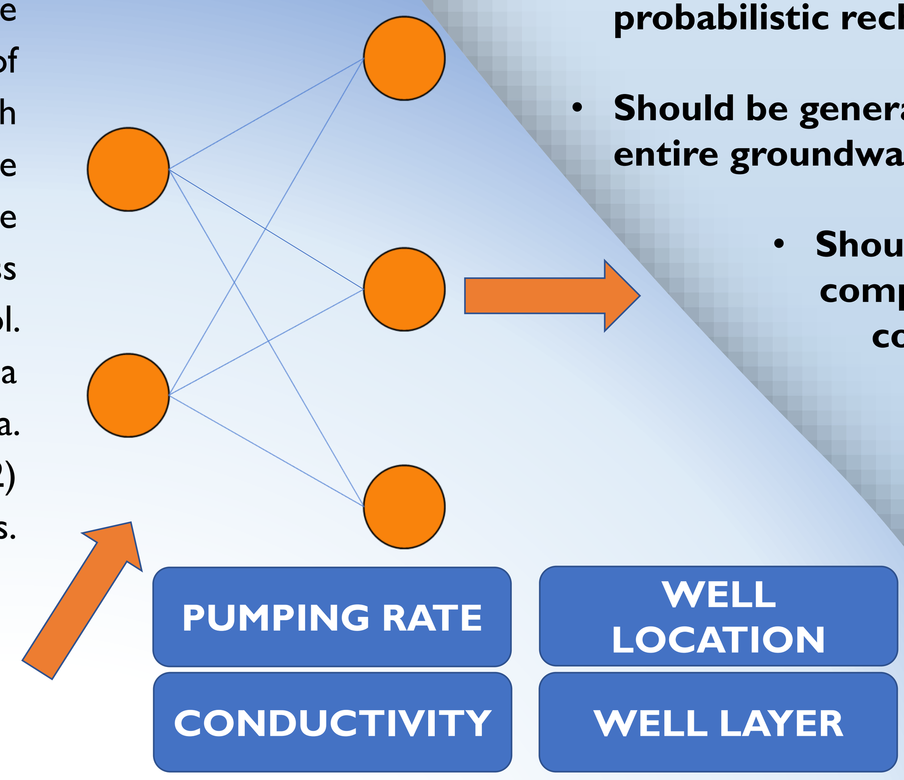



Fig 1: Geological cross sections of different realizations from the hydrostratigraphic model. The structural variations affects the outcome of the numerical simulations.

## METHODS

By averaging multiple MODPATH simulations of the same well set-up with different geological realizations we can obtain a probabilistic recharge area. However, this is a slow process not feasible in a decision-support tool. We overcome this, by training a neural network to predict such a recharge area. The network is trained with simulation results (Fig. 2) and selected input features (Fig. 3) related to the results.




## TRAINING DATA SET

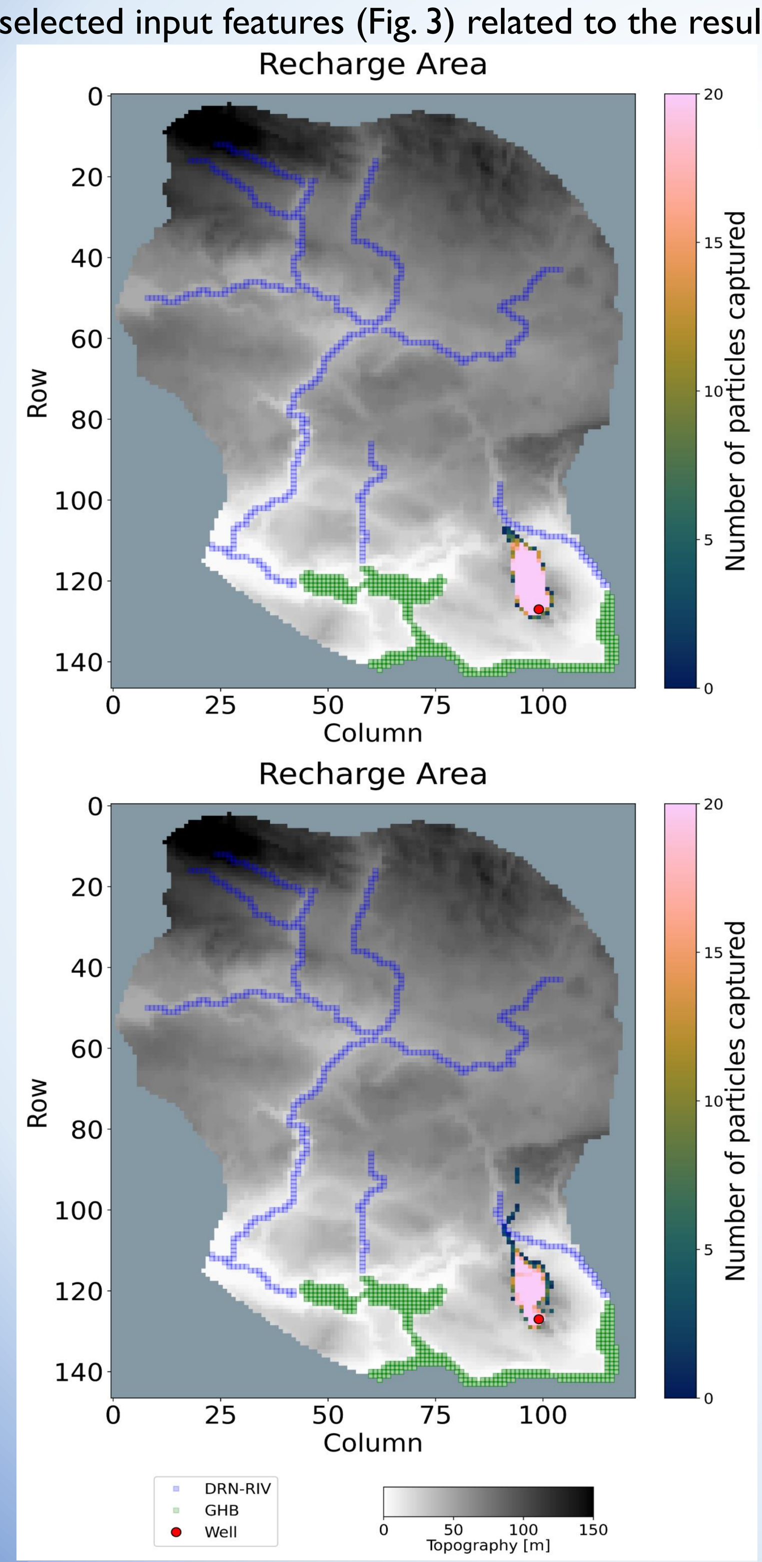



Fig 2: Two different well recharge areas from the same well position. Different realizations from the geological model causes drastically different results in the groundwater model. Recharge areas are calculated with MODPATH forward-tracking.

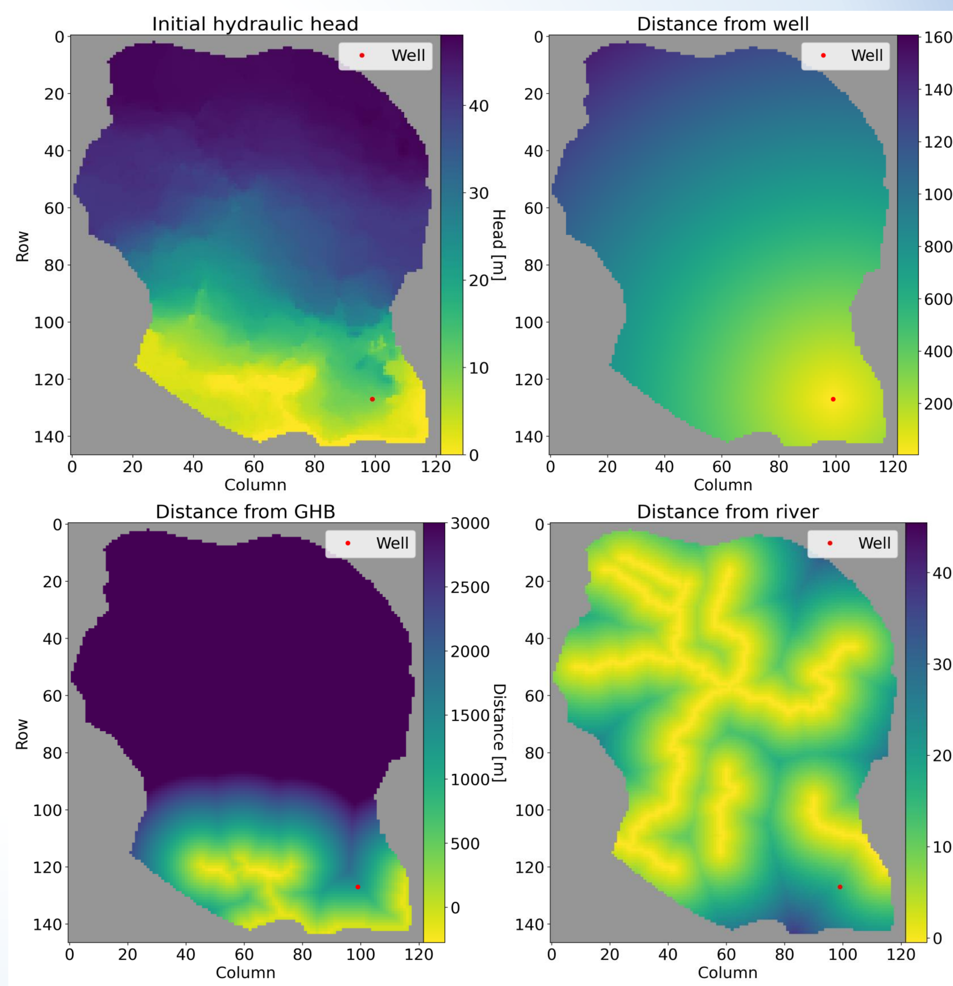



Fig 3: Input features used to train the neural network. We select specific features from the groundwater model and train the network to understand the mapping between these and the desired probabilistic recharge area. Since some of these features vary for the same well set-up, we also use statistics of these features like mean, minimum, maximum, and standard deviation. Shown here are the input features initial hydraulic head of the top layer in the model, distance from the well, distance from river modules, and distance from a General Head Boundary (lake and fjord).
Listed above the figure are some of the additional input features such as pumping rate, hydraulic conductivity, well location, and the model layer the well is located in.

## RESULTS

The training data set is constructed to hold simulation results (Target data) and input features from well set-ups across the entire model, such that the neural network predicts accurately in different locations with different geological settings.

### The trained neural network:

- **Should be able to reproduce the probabilistic recharge area in Fig. 4.**
- **Should be generally applicable in the entire groundwater model.**
- **Should drastically reduce the computational requirements compared to MODFLOW and MODPATH.**
- **Should produce results within seconds and not hours including input feature creation.**

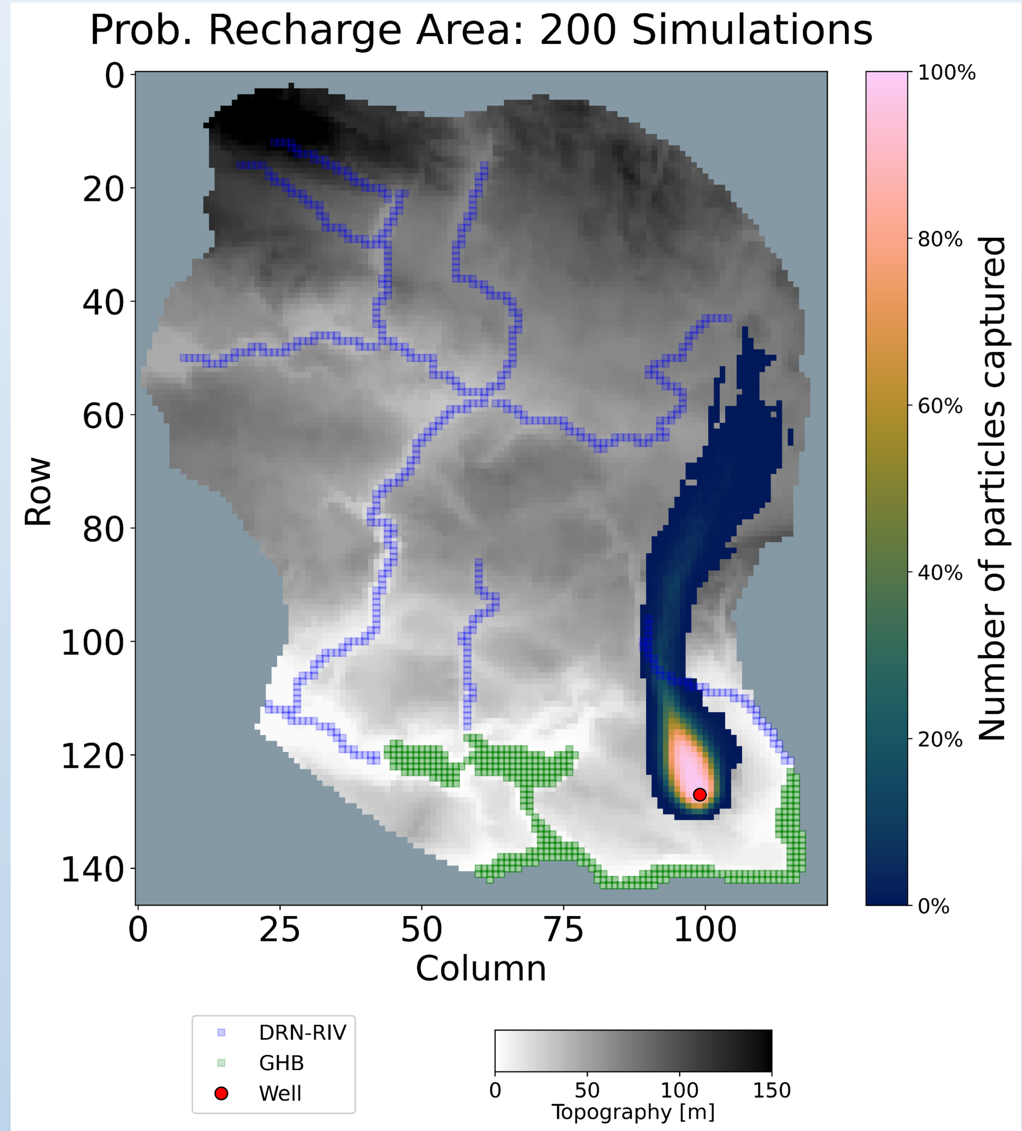



Fig 4: A probabilistic recharge area for the well marked with a red dot. By performing forward-tracking of 20 particles in each top cell of the model, we calculate, how many of these end up in the well. By doing this multiple times with varying geological setting, we obtain a probabilistic recharge area. The colorbar indicates the probability of capturing groundwater from an area in the model. Orange to purple is the most likely area to capture from.

## DISCUSSION

The addition of stochastic geological modelling and the expected speed up allows for new, not earlier feasible, investigations. Investigations that requires hours or days of computation time can be unaffordable for small businesses and rarely paid for by municipalities. This could be investigations such as searching for optimal locations for new wells with included risk assessments. By incorporating this neural network into a decision-support tool with a GUI makes it feasible for them to perform these investigations themselves in real-time.

We expect to put a lot of work into the training and input feature selection of this ongoing project. The input feature selection changes as we move away from the deterministic model since many of these are depending on geology. It would be interesting to investigate how different input feature combinations affect the training phase.

## CONCLUSIONS

- **We present a way of incorporating a stochastic hydrostratigraphic model into recharge area simulations in MODFLOW and MODPATH.**
- **To overcome the large increase in computation time, we suggest to train a neural network that takes in a few selected input features and predicts a probabilistic recharge area from these.**
- **The stochastic addition and increase in speed allows for new investigation possibilities and the trained neural network could be a great asset in a decision-support tool.**

AARHUS UNIVERSITY

NIRAS

Sharing is encouraged

EGU