AqC 6.3

Valdrich Fernandes[1], **Perry de Louw**[1,2], **Coen Ritsema**[1], **Ruud Bartholomeus**[1,3]

[1]Soil Physics and Land Management group, Wageningen University & Research, Wageningen, the Netherlands
[2]Department of Soil and Groundwater, Deltares, Utrecht, the Netherlands
[3]KWR Water Research Institute, Nieuwegein, the Netherlands

Corresponding author: Valdrich Fernandes, `valdrich.fernandes@wur.nl`

Abstract

Managed Aquifer Recharge (MAR) is widely applied to enhance groundwater storage and promote the sustainable use of this essential resource. Techniques such as Machine Learning (ML) as surrogate for computationally intensive numerical models, are increasingly considered for identifying suitable locations for MAR. While ML has been demonstrated to be suitable for steady-state simulations, its application for transient modelling is much more challenging. However, understanding how water recharged during wet periods is retained and remains available during dry periods is critical, highlighting the importance of transient responses. In this study, we therefore employed ML to mimic the transient effect of MAR by decomposing the time series of groundwater storage when MAR stopped into two components—the MAR-response and the decay coefficient—based on the observation that the storage decays following an exponential curve. This decomposition provides a simplified yet effective representation of groundwater storage changes over time. Using U-Net and XGBoost, we demonstrated that ML can accurately capture these dynamics for the Baakse Beek catchment, located in the drought sensitive sandy soils of the Netherlands. The ML models achieved an R^2 of more than 0.85 in predicting the two components of the long-term dynamics of the stored water. Additionally, using explainable AI techniques, specifically SHapley Aditive exPlanation values, we identified the site management decisions and the properties of the surface water network near the recharge site most significantly impact the effectiveness of the MAR in the region. This focus on model interpretability ensures transparency in the predictions, improving on the models' generalizability and fosters trust among hydrologists and stakeholders.

1 Introduction

Seasonal water availability has increasingly become a concern worldwide. Even countries with typically robust water systems have recently faced periods of water stress during extended periods of droughts. The 2018-2020 drought, for instance, affected most of continental Europe, limiting surface water availability in major international rivers (Rakovec et al., 2022). Similar extended periods of droughts are expected to become more frequent in future climatic conditions (Aalbers et al., 2023; Gu et al., 2023; Hari et al., 2020; van der Wiel et al., 2021). The 2018 drought saw a general deterioration of surface water quality, further exacerbating the stress on water users dependent on these sources (Wolff & van Vliet, 2021). This combination of poor water quality and scarcity is expected to worsen in Western Europe, particularly between June and October (Jones et al., 2024). Capturing excess water during the seasonal surplus could help mitigate some of the demand during the dry season.

Addressing this vulnerability requires prioritizing water management solutions that increase summer water availability (Bartholomeus et al., 2023), while still accounting for risks of floods and waterlogging. Efficient groundwater management could address some of the vulnerability by serving as a buffer, capturing water during surplus that supports demands during dryer periods. While much of this recharge occurs naturally, it can be further enhanced by recharging surplus water into the groundwater aquifers using Managed Aquifer Recharge (MAR) techniques (Dillon et al., 2019, 2020; Hartog & Stuyfzand, 2017).

In a previous study, we demonstrated U-Net's potential to estimate the steady-state groundwater response to MAR (Fernandes et al., 2024) through a technique called surrogate modelling. The surrogate model is a computationally simplified model trained on data generated using the computationally expensive model. However, the steady-state MAR response typically does not describe the transient behaviour of MAR, as water is recharged during periods with a water surplus, to buffer the water availability using the subsurface. The first goal of this paper therefore is to investigate a potential technique for ML to capture the transient dynamics of groundwater response to MAR capturing both immediate effects and longer-term dynamics during periods without recharge.

In addition to capturing the transient behaviour of MAR systems in ML, we focus on model interpretability. Dai et al. (2024) identified interpreting the ML models to be complex and cumbersome. However, we hypothesize that by using explainable AI techniques such as SHapley Aditive exPlanation (SHAP) on ML models, the relations identified by the ML model can be explained, further building trust in the ML model.

SHAP is a powerful model interpretation technique that quantifies the contribution of each input feature to a model's predictions, offering detailed insights into the relations identified by the model. By quantifying the contribution of each feature, SHAP not only indicates the importance but also the direction of the effect of that feature on the model's prediction. Furthermore, by evaluating feature contributions across an entire dataset, SHAP uncovers the generalized relations that underpin the model's predictions, unlike traditional local sensitivity analysis that focuses on features at an individual observation. This dataset-wide perspective allows for identifying consistent, generalized effects of the inputs on the outputs, which could serve as a foundation for other Multi-Critera Decision Analysis (MCDA).

In this study, we thus present ML surrogate models to efficiently and accurately predict the transient dynamics of groundwater response to MAR, capturing the recharging process of MAR during periods of water surplus and the decay or discharge of the stored water volume during periods without recharge. By calculating the SHAP values of the surrogate model, we identify and quantify the influence of each geohydrological input on the groundwater response, providing explainable insights into the model's prediction. These insights not only enhance our understanding of the system but also lay the groundwork for integrating MCDA to identify potential locations for MAR. Through this approach, we aim to develop a comprehensive understanding of the key criteria influencing water storage due to MAR in the short to medium-term period.

2 Methodology

2.1 General overview

The methodology employed in this study borrows from Fernandes et al. (2024), where the groundwater response to hypothetical MAR sites are simulated with a numerical groundwater model, which makes up the training and testing data for the ML surrogate models. The groundwater response is calculated as the difference in the groundwater heads between scenarios with MAR and a scenario with only natural recharge. The location, size, and recharge rate are selected randomly to represent the entire range of possibilities within the demonstration region. As a continuation from Fernandes et al. (2024), we demonstrate the applicability of the surrogate modelling technique for the Baakse Beek catchment, which is located in the eastern sandy soils of the Netherlands. This region is characterised by highly transmissive and freely draining soils with a dense surface water network of ditches and streams that drain the groundwater reservoir. This network has been significantly intensified over time, increasing the region's vulnerability to water shortages during extended dry periods. In this study, we study how the water recharged by MAR decays to the natural groundwater levels during dry periods, periods without MAR.

2.2 Groundwater models

The artificial recharge is simulated using a groundwater model, AMIGO (Vreugdenhil, 2021), which represents the subsurface in the part of Gelderland province that is under the jurisdiction of the water board Rijn en IJssel in the Netherlands. AMIGO is a coupled unsaturated-saturated zone model similar to those used in

Querner et al. (2016), and van Walsum & Veldhuizen (2011) where the saturated zone is represented using MODFLOW-2005 (Harbaugh, 2005) while the unsaturated zone is represented using MetaSWAP. Gridded meteorological data from the AMIGO dataset, including precipitation and Makkink reference crop evaporation, is applied with the MetaSWAP that calculates the effective groundwater recharge, which is then applied to the saturated zone model. The saturated zone represents a 200 m thick sequence of Pleistocene sands, which is underlain by the clayey Breda formation. This formation represents the model basement, represented in the model by a no-flow boundary condition. Further details of the meteorological conditions and the model are available in Vreugdenhil (2021) and Fernandes et al. (2024). The model extent and the boundary conditions used to represent the Baakse Beek catchment were the same as in Fernandes et al. (2024), except for the outer boundary conditions and the river levels for the IJssel River to the west, where temporal data was available. The time-varying groundwater heads at the outer boundary were derived from a long-term simulation that is detailed in Vreugdenhil (2021). The initial groundwater head is also derived from this long-term simulation and is detailed in Section **??**. Furthermore, monthly river stage data is used for the IJssel River to the east.

In the simulated scenarios, the artificial recharge is simulated using the Recharge (RCH) package in MODFLOW as in Fernandes et al. (2024). This recharge is applied directly to the saturated zone, bypassing the unsaturated zone. It represents the effective recharge rate that reaches the saturated zone and is analogous to aquifer recharge applied below the surface, such as through sub-surface irrigation. As a result, the recharged water does not influence the evapotranspiration, except due to the increase in the groundwater level. While this is a simplification of reality, this assumption has minimum impact on the system, especially away from the recharge site. At the recharge site, we underestimate the evaporation and consequently overestimate the effective recharge site.

2.3 MAR Scenarios

In the scenarios considered, MAR is only applied in months when precipitation exceeds the evaporation demand for a year with average meteorological conditions. For this analysis, the meteorological data from the Hupsel station, data made available by the Royal Netherlands Meteorological Institute (KNMI), is analyzed as it is located within the study area. Complete data is available from October 1993 onwards based on which the annual cumulative evaporation excess precipitation is calculated. Based on the 31 years of available data, 2012 was selected as a representative year for average meteorological conditions, thas the precipitation excess fell within the 25th to 75th percentile for most of the year (see Figure **??**). The cumulative precipitation surplus at Hupsel increased between October 1, 2011, and March 1, 2012 (blue shaded region in Figure **??**), indicating higher water availability for artificial recharge. In the scenarios considered, artificial recharge was simulated during this period of water surplus and the progression of the groundwater response during the period without artificial recharge was analyzed until October 1, 2012.

To ensure representative antecedent conditions for all scenarios, a two-year start-up run was simulated from 1st October 2009 to 30th September 2011. The initial groundwater levels for this start-up run were derived from a simulation starting in April 2004, which was conducted outside the scope of this study. Since initial soil moisture conditions for MetaSWAP were unavailable for this earlier run, conditions that would ensure an average recharge flux were used as the starting conditions for the start-up run.

We simulated a total of 720 artificial recharge sites, evenly spaced across the model domain. The recharge rate at these sites ranged from 5 to 25 mm/day, and the site sizes varied between 0.1 and 1 km^2. The recharge rate and the size of each site were selected randomly using a Latin hypercube sampling technique, similar to the methodology employed by Fernandes et al. (2024). Six recharge sites were simulated simultaneously in each scenario to reduce the number of groundwater model simulations needed. The interaction between the recharge sites was minimized by maintaining a separation distance of six times the leakage factor between the sites. This distance ensures that groundwater levels return to their natural state between adjacent recharge sites, hence minimizing the interaction between the sites.

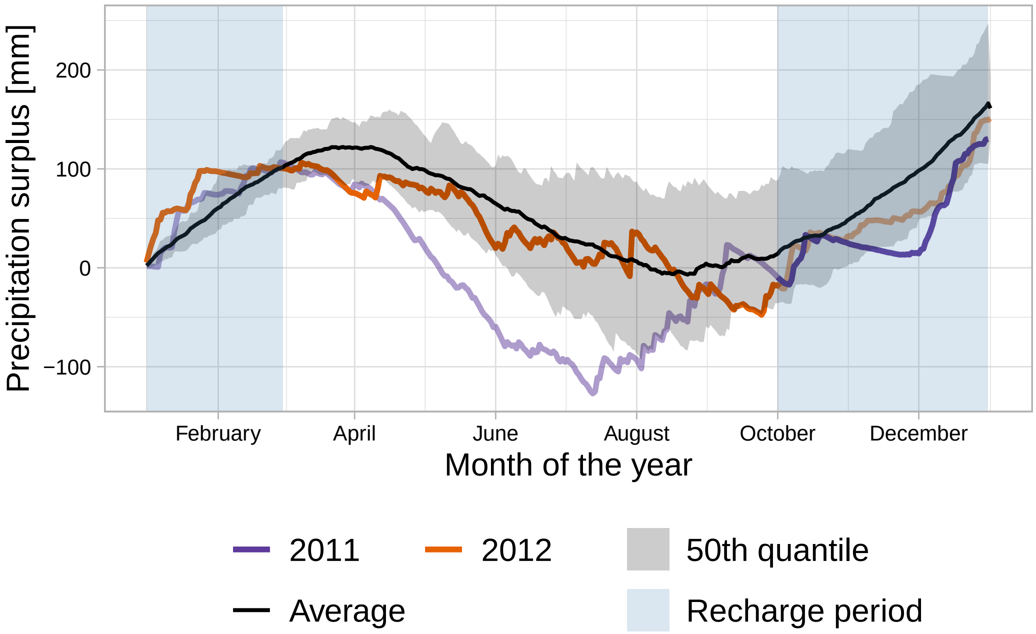

Figure 1: Cumulative potential precipitation surplus (i.e. precipitation minus reference evaporation according to Makkink (1957)) in the simulated years, 2011 and 2012 as compared to the average surplus between 1993 and 2024. Artificial recharge is applied between 1st October 2011 till 1st March 2012, when precipitation exceeds the reference evaporation.

2.4 Progression of the storage during the summer season

In this study, we aim to predict the progression of groundwater storage during periods without MAR. Groundwater storage is defined as the total increase in groundwater heads, analogous to the volume of the aquifer saturated due to MAR. Rather than predicting the complete time series of the groundwater storage in the period without MAR, we decompose it into two parameters of the exponential decay function. The response peaks during the recharge period and then gradually returns to natural groundwater levels once recharge stops. Simulation results (see Figure **??**) suggest that this decay follows an exponential curve, a pattern commonly observed in systems where the rate of change is proportional to the current quantity. This behaviour is also found in the discharge from a linear reservoir in hydrology, which decays exponentially after precipitation ends. In such reservoirs, the discharge Q_t at time t is directly proportional to the water storage S_t within the catchment, governed by the reservoir constant k (Dingman, 2015; Pedersen et al., 1980; Wittenberg, 1994). This relationship is captured in Equation **??**, with the change in discharge over time described by Equation **??**. Assuming no external fluxes like recharge or extraction, the storage dynamics can be derived as shown in Equation **??**.

In this analogy, the groundwater response to artificial recharge can be compared to the storage S in a linear reservoir. The initial storage S_0 represents the MAR-response, or the groundwater response at the end of the recharge period, while the reservoir constant k serves as the decay coefficientwhich defines the rate at which the groundwater storage is drained to the surface water network. By estimating these two parameters, the progression of groundwater storage after the recharge period can be effectively described, providing a clear understanding of how the system returns to equilibrium. These two parameters, the MAR-response and the decay coefficient, are the target variables for the ML models. These terms are used throughout the paper to identify the two separate phases of MAR.

$$Q_t = k \cdot S_t \tag{1}$$

$$Q_t = Q_0 \cdot e^{-k \times t} \tag{2}$$

$$S_t = S_0 \cdot e^{-k \times t} \tag{3}$$

2.5 ML procedure

Table 1: Models used in this study and the purpose for which they are used. The effect of the inputs on the MAR-response and the decay coefficient are estimated based on the SHAP values of the trained XGBoost models

Response characteristic	Model	Scenario type[a]	SHAP i
MAR-response	Num. groundwater model - Training data	T	-
	Num. groundwater model	S	No
	U-Net	T	No
	XGBoost	T	Yes
Decay coefficient	Num. groundwater model - Training data	T	-
	XGBoost	T	Yes

[a]T - Transient scenarios; S - Steady-state scenarios

2.5.1 ML models used

Two ML models were used to analyze the response to artificial recharge, U-Net and XGBoost. U-Net has previously been shown to accurately reproduce the steady-state response to artificial recharge (Fernandes et al., 2024). In this study, it is trained to estimate the spatial distribution of the MAR-response, at the end of the recharge period, see Table **??**. The response is a spatially distributed result of the simulations, calculated for each cell around a recharge site. In contrast, the decay coefficient characterizes the overall response, describing it with a single value for each recharge site, even when the site spans multiple cells. Similarly, the total MAR-response can be calculated for each recharge site based on the spatial distribution of the MAR response. XGBoost has been shown to accurately represent nonlinear relations (Chen & Guestrin, 2016), lending itself to estimate each recharge site's aggregated MAR-response and the decay coefficient, as shown in Table **??**.

2.5.2 Data preparation

U-Net is trained on two-dimensional representations of the geo-hydrological properties from the numerical groundwater model, AMIGO. Following Fernandes et al. (2024), six inputs were shortlisted to capture the interactions within the system, applied recharge rate, aquifer transmissivity, vertical resistance below the aquifer, steady state groundwater depth, drain and river conductance, and groundwater depth below the drain level and river stage. Among these inputs, the aquifer properties were further pre-processed as the 15 model layers were often discontinuous, represented by constant values (Fernandes et al., 2024; Vreugdenhil, 2021). The model layers were combined to represent an aquifer if the vertical resistance between the layers was less than 200 days leaving us with seven aquifers. Of the seven aquifers, the properties of the first aquifer were fed as an input to the model as the artificial recharge was applied to the phreatic aquifer. Further details on the inputs to the U-Net model are described in Fernandes et al. (2024).

XGBoost requires tabular inputs rather than the spatial distribution of the geo-hydrological properties used to train U-Net. However, the inputs need to represent the properties around the recharge site that could impact the groundwater response at the site. To satisfy this condition, the geo-hydrological properties around the site are summarized using a Gaussian filter. This method of feature extraction represents a distance-weighted average of the inputs. The weights are calculated based on the standard deviation of the Gaussian. After comparing multiple standard deviations from 250 m to 1500 m, a standard deviation of 500 m trained the model with the highest R^2. Based on this, a filter with a standard deviation of 500 m is used for all the inputs. Some inputs are not defined everywhere, especially the surface water network. As an example, the river conductance is defined only at river streams, where the RIV package of MODFLOW is active. Locations where the RIV package is inactive, are represented by zero conductance before the Gaussian filter is applied.

The Gaussian filter is applied to all inputs to the XGBoost model, which we categorized into three groups, site properties, aquifer properties, and surface network properties. The site properties are represented in U-Net by a map of the applied recharge, which represents both the recharge rate and the size of the recharge site. These two inputs are provided separately to the XGBoost. For aquifer properties, the inputs include transmissivity, vertical resistance, and groundwater depth for the situation without MAR (baseline scenario). The transmissivity and the vertical resistance undergo the same pre-processing as those used for the inputs to U-Net. While groundwater depth is not strictly an aquifer property, it results from the interaction of various location specific factors, such as transmissivity and seepage flux to the surface water network. Due to the localized nature of these influences, groundwater depth is included along with other aquifer properties. The groundwater depth calculated from the baseline scenarios, with only natural recharge, is divided into two periods: winter and summer. These periods were defined to coincide with the periods with and without recharge in Section **??**. As the artificial recharge is applied during the winter months, the MAR-response is more strongly influenced by the winter conditions. On the contrary, the decay coefficient depends on summer conditions and is therefore predicted using the average summer groundwater depth. Among the surface water network properties, the river and drain conductance are a common input between the two models, XGBoost and U-Net. Additionally, river density is included as an input to XGBoost to quantify drainage intensity and is calculated based on the cells where the RIV package in MODFLOW is active. Ditch distance is a common alternative for capturing similar information, which is calculated from river conductance. However, as river density is less dependent on the length of the river within the cell, it exhibits a weaker correlation with river conductance. This reduced correlation and the fact that river density is bounded between 0 and 1 result in more stable model training makes river density the more robust parameter for quantifying drainage intensity, especially for ditches and streams that are more than 25 m (numerical model resolution) away from each other. Finally, the groundwater depth relative to the drain level and river stage was excluded from the inputs to XGBoost as it was highly correlated with groundwater depth. Strong correlation complicates input attribution without significantly improving the model performance.

While the MAR-response is the response at the end of the recharge period, the decay coefficient was calculated from exponential curves fitted to the time series of the numerical model simulation results. The curves were fitted using a robust linear regression using the 'rlm' function from the 'MASS' package in R. However, the distribution of the target variables presents a right-skewed distribution, see Figure **??**. This skew would make the trained ML model biased to predict low values as these values are more common. To minimize the bias and make the target variables more normally distributed, the MAR-response and decay coefficient, are log-transformed. However, the target variables described in the figures and results of this study are back-transformed to the original scale.

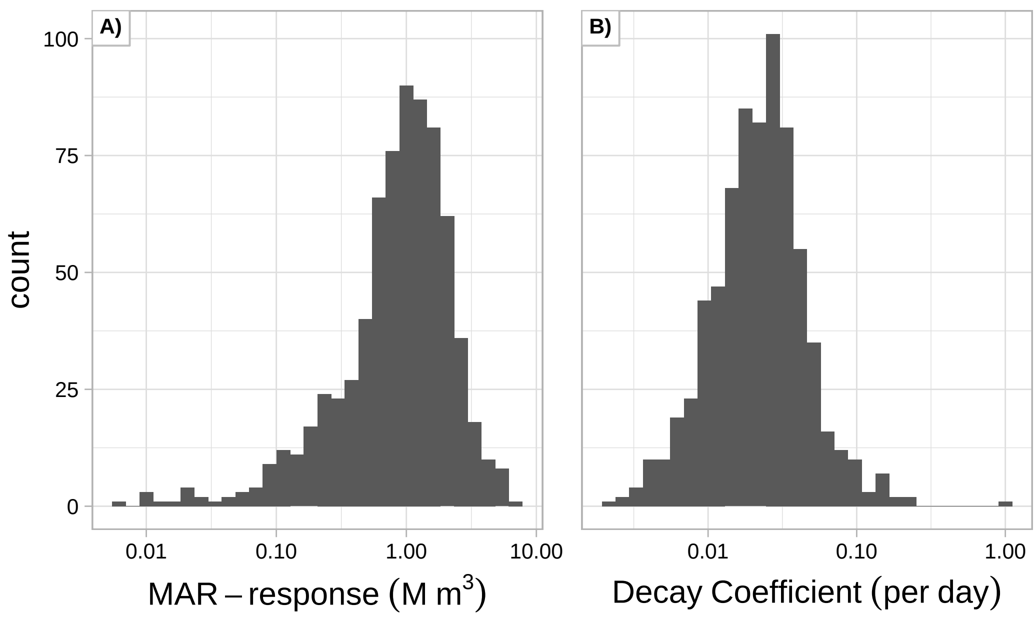

Figure 2: Distribution of the target variables, the MAR-response, A, and the decay coefficient, B, based on 720 hypothetical recharge sites.

2.5.3 Training process

U-Net is trained using the ADAM optimizer with a reducing learning rate schedule with a relatively small batch size of 12. These settings have been shown to effectively train the model (Fernandes et al., 2024). The model was trained on the results from 500 recharge sites and its performance was tracked during training using an unseen dataset, validation set, of 100 recharge sites. The training was concluded if the model performance didn't improve over ten epochs.

The training process for XGBoost follows a comprehensive approach aimed at maximizing the performance and generalizability of the trained model. The results of all the scenarios are divided into training (N = 500) and testing sets (N = 220), which