

Monthly Groundwater Level Simulation and Forecasting in a Snow-Influenced Aquifer (Heby, Sweden) Using LSTM and Sparrow Search- and PSO-Optimized Support Vector Regression

Nasrin Khajeh Mahmoudi^{1*}, Abuzar Nekoei²

¹ MSc in Geography and Urban Planning, University of Zabol, Iran (nasrinkhajehmahmoudi@gmail.com)

² Department of civil engineering, ST.C., Islamic Azad University, Tehran, Iran.

Received: 07 May 2025

Accepted: 17 June 2025

Available online: 25 June 2025

Abstract: Groundwater levels in cold climates respond to recharge that is delayed by snow accumulation and melt, which makes data-driven modelling more difficult than in rain-dominated regions. This study models monthly groundwater levels (GWL) of an observation well near Heby, Sweden (1981–2020), using precipitation, potential evaporation and temperature from the E-OBS dataset. A degree-day snow routine was used to convert precipitation into rain and snowmelt, and the resulting effective water input was used as the main predictor. Multiple linear regression (MLR), support vector regression (SVR), random forest (RF), long short-term memory (LSTM) and two hybrid models in which SVR was tuned by the sparrow search algorithm (SVR-SSA) and particle swarm optimization (SVR-PSO) were evaluated in two scenarios: simulation from meteorological inputs only (S1) and one-month-ahead forecasting with lagged GWL (S2). The models were trained on 1981–2008, validated on 2009–2013 and tested on 2014–2020. In S1 the LSTM clearly outperformed all other models (test RMSE = 12.5 cm, NSE = 0.68), while the hybrid SVRs (NSE = 0.45) generalized worse than the default SVR (NSE = 0.64) despite lower validation errors. In S2 SVR-SSA was the best model (RMSE = 7.6 cm, NSE = 0.88), improving on persistence (RMSE = 11.0 cm) by 30%. The six-month cumulative effective water input was the most important meteorological predictor. The results show that sequence learning is most valuable when no past groundwater observations are available, and that metaheuristic tuning is beneficial mainly in the autoregressive forecasting setting.

Keywords: Groundwater level, LSTM, Sparrow search algorithm, Particle swarm optimization, Snowmelt, Sweden

1. Introduction

Groundwater is the largest accessible store of fresh water and supplies drinking water to a large part of the world's population. Its levels respond to climate variability and change, to land-use change and to abstraction (Taylor et al., 2013). Reliable models of groundwater level (GWL) dynamics are needed to manage water supply, to assess drought impacts on aquifers and to fill gaps in monitoring records. Physically based groundwater models require detailed hydrogeological information that is often unavailable, so time-series and data-driven models have become attractive alternatives (Bakker and Schaars, 2019; Rajaee et al., 2019).

Machine learning has been applied very widely to GWL prediction. Recent reviews show that artificial neural networks, support vector machines, tree ensembles and, increasingly, deep learning are the dominant approaches (Boo et al., 2024; Jesse et al., 2025). Wunsch et al. (2021) compared LSTM, convolutional neural networks and NARX models for German wells and showed that sequence models can simulate GWL from meteorological inputs alone. Later work used such models to project groundwater levels under climate change (Wunsch et al., 2022). Pham et al. (2022) compared seven machine learning models for wells in Bangladesh and found bagging-based tree ensembles to perform best.

Hybrid models that tune the hyperparameters or weights of a learner with metaheuristic algorithms are among the most frequently proposed approaches. Examples include SVR optimized by particle swarm optimization (PSO) for the Zanzan aquifer (Mozaffari et al., 2022), six evolutionary algorithms coupled with ANFIS, SVM and ANN in the Ardebil plain (Seifi et al., 2020), metaheuristic ANFIS models for the Urmia aquifer (Kayhomayoon et al., 2022), least-squares SVR tuned by the interior search algorithm in the Karaj plain (Moravej et al., 2020), and extreme learning machines optimized by hybrid PSO–grey wolf algorithms (Samantaray et al., 2024). Most of these studies were carried out in semi-arid, rain-fed aquifers. They also mostly include lagged GWL among the inputs, which makes the problem largely autoregressive.

In cold climates, recharge is controlled by snow accumulation during winter and melt in spring, which introduces a non-linear, temperature-dependent delay between precipitation and groundwater response. Such aquifers are a demanding test for data-driven models. The objectives of this study were (i) to develop monthly GWL models for a snow-influenced aquifer near Heby, Sweden, using physically motivated inputs derived from a degree-day snow routine, (ii) to compare MLR, SVR, RF, LSTM and two hybrid SVR models tuned by the sparrow search algorithm (SSA; Xue and Shen, 2020) and PSO (Kennedy and Eberhart, 1995), and (iii) to assess the models in a simulation scenario without past

GWL observations and in a one-month-ahead forecasting scenario.

2. Materials and Methods

2.1. Study Site and Data

The observation well is located near Heby in Uppsala County, east-central Sweden (approximately 59.9° N, 16.9° E; Fig. 1). The region has a humid continental climate with cold winters, a snow cover that usually lasts several months and a mean annual temperature of about 6–7 °C. The groundwater level record and the corresponding meteorological series were obtained from the example datasets distributed with the open-source groundwater time-series package Pastas (Collenteur et al., 2019; github.com/pastas/pastas). The dataset is documented there as head observations from a well near Heby; in Sweden such observations are collected within the national groundwater monitoring programme of the Geological Survey of Sweden (SGU). According to the same documentation, the daily precipitation, mean temperature and potential evaporation series for the site were derived from the E-OBS gridded dataset (Cornes et al., 2018).

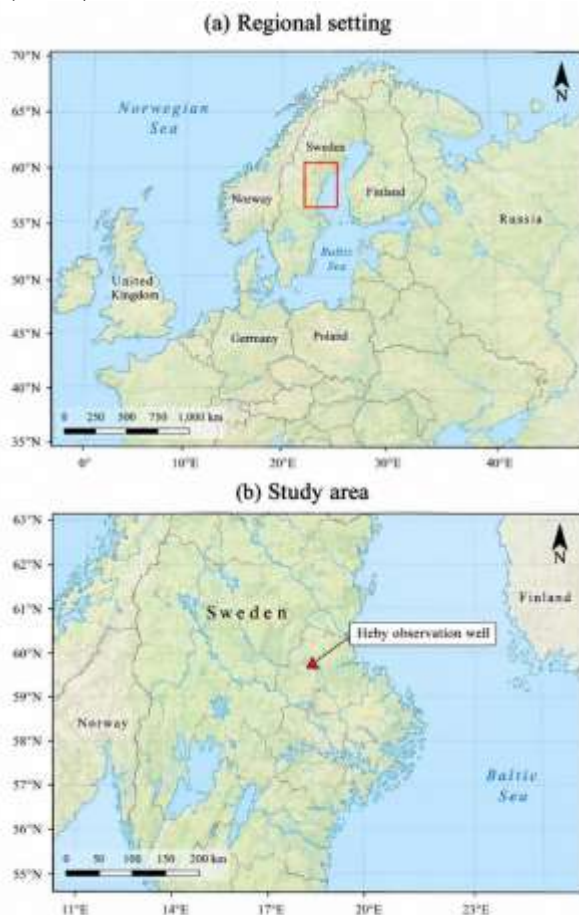


Figure 1. Location of the Heby observation well in east-central Sweden

The head record (January 1980 – November 2020) contains 3911 measurements at irregular intervals, from roughly twice a month in the early years to daily in later

years. The heads were averaged to monthly values. 478 of the 486 months between January 1980 and June 2020 contain at least one observation. For the model inputs, the few missing months (at most three consecutive) were linearly interpolated, but performance was evaluated only on observed months. Daily meteorological series were aggregated to monthly sums (precipitation and potential evaporation) and means (temperature). Mean annual precipitation was 577 mm and mean annual potential evaporation 497 mm. Descriptive statistics are given in Table 1, and the monthly series are shown in Fig. 2.

Table 1. Descriptive statistics of the monthly variables at Heby (1980–2020)

Variable	Min	Max	Mean	SD	Skew	Kurt
GWL (m a.s.l.)	78.27	79.28	78.81	0.18	-0.30	-0.11
P (mm)	0.00	189.90	48.67	29.70	1.15	1.83
PET (mm)	1.04	135.22	41.91	37.19	0.50	-1.15
T (°C)	-12.84	22.00	6.72	7.57	-0.02	-1.06
Snowfall (mm)	0.00	66.00	6.79	11.98	2.22	5.22
Melt (mm)	0.00	106.10	6.79	13.35	2.97	11.64

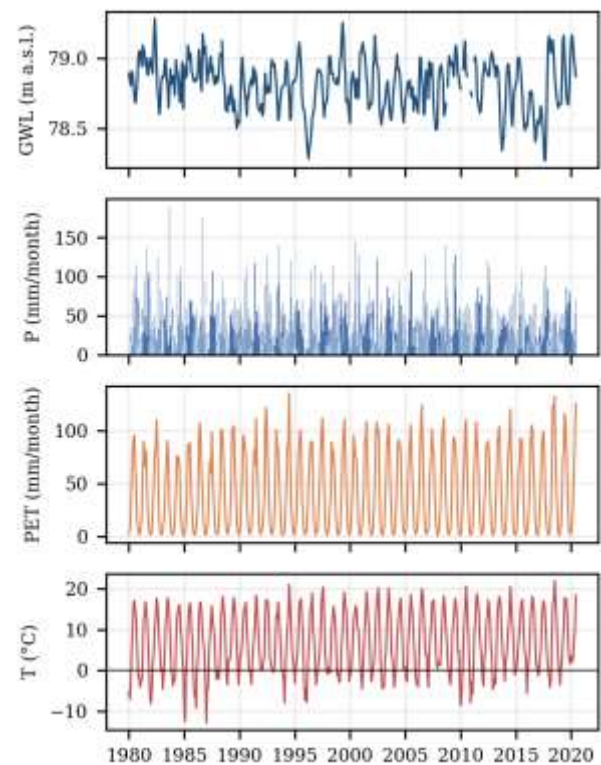


Figure 2. Monthly groundwater level, precipitation (light blue: snowfall), potential evaporation and air temperature at Heby

2.2. Degree-Day Snow Routine and Effective Water Input

To represent the delay of recharge by snow storage, a simple degree-day snow model (Hock, 2003) was run at daily time steps. Precipitation on days with mean

temperature $T \leq 0$ °C was added to the snowpack S as snowfall, and on warmer days melt was computed as

$$M_d = \min(S, DDF \cdot T_d), T_d > 0 \quad (1)$$

with a degree-day factor $DDF = 3 \text{ mm } ^\circ\text{C}^{-1} \text{ d}^{-1}$, a typical value for open terrain. The monthly effective water input was defined as

$$W = P_{\text{rain}} + M - \text{PET} \quad (2)$$

where P_{rain} is rainfall, M is snowmelt and PET is potential evaporation. On average, 14% of the annual precipitation fell as snow. The mean seasonal regime (Fig. 3) shows that the groundwater level rises in March–May, when snowmelt and spring rain coincide with low evaporation, and reaches its minimum in September–October at the end of the evaporation season. GWL was most strongly correlated with the six-month sum of W lagged by one to two months ($r = 0.66$), whereas the correlation with same-month W was weak ($r = -0.20$). The Mann–Kendall test indicated a statistically significant but very small decreasing trend in the monthly GWL ($Z = -3.02$, $p = 0.003$, Sen's slope $= -0.2 \text{ cm per century}$).

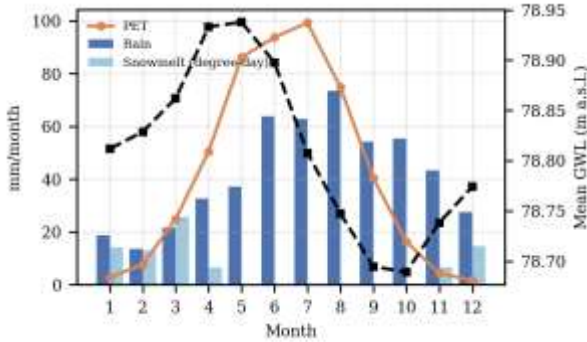


Figure 3. Mean monthly rain, degree-day snowmelt, potential evaporation and groundwater level at Heby

2.3. Modelling Scenarios and Inputs

Two scenarios were defined. In scenario S1 (simulation), $\text{GWL}(t)$ was estimated only from meteorological information: W at lags 0–3, cumulative W over 3, 6 and 12 months, P , PET , T , melt and snowfall of month t , and the sine and cosine of the month (14 inputs). This setting corresponds to gap filling or to climate-impact studies in which no GWL observations are available (Wunsch et al., 2021). In scenario S2 (forecasting), $\text{GWL}(t-1)$ and $\text{GWL}(t-2)$ were added (16 inputs), so that the model produces a one-month-ahead forecast. In both scenarios the LSTM received sequences of the previous 12 months of $[P, P_{\text{rain}}, M, \text{PET}, T, \sin, \cos]$, with $\text{GWL}(t-1)$ added in S2.

The data were split chronologically into training (1981–2008, $n = 334$ observed months), validation (2009–2013, $n = 54$) and test (2014–June 2020, $n = 78$) periods. Inputs were standardized with training-set statistics. The validation set was used for early stopping of the LSTM and as the objective of the metaheuristic search. In S2 a persistence forecast ($\text{GWL}(t) = \text{GWL}(t-1)$) served as the benchmark.

2.4. Models

MLR was fitted by ordinary least squares. **SVR** (Drucker et al., 1997) with a radial basis function kernel was used with default settings ($C = 1$, $\epsilon = 0.1$, $\gamma = \text{'scale'}$) as the standalone model. **RF** (Breiman, 2001) used 500 trees with a minimum leaf size of 2. Permutation importance of the RF on the test set was used to rank the predictors.

The **LSTM** (Hochreiter and Schmidhuber, 1997; Kratzert et al., 2018) consisted of one layer of 32 units, dropout of 0.2, a 16-unit ReLU dense layer and a linear output. It was trained with Adam (learning rate 0.002, batch size 16, mean squared error) for up to 400 epochs with early stopping (patience 40). Five networks with different random seeds were averaged.

2.5. Hybrid SVR Models

The **sparrow search algorithm** (Xue and Shen, 2020) divides the population into producers, which search for food, and scroungers, which follow the producers. A fraction of the birds act as scouts that react to danger. The position of a producer is updated as

$$X_i^{t+1} = X_i^t \cdot \exp(-i / (\alpha \cdot \text{iter}_{\text{max}})) \text{ if } R_2 < \text{ST}; X_i^{t+1} = X_i^t + Q \cdot L \text{ otherwise} \quad (3)$$

where α is a random number in $(0, 1]$, R_2 is the alarm value, ST is the safety threshold, Q is a normally distributed random number and L is a vector of ones. Scroungers move towards the best producer, and scouts jump towards the global best or away from the worst position. **PSO** (Kennedy and Eberhart, 1995) updates the velocity v_i and position x_i of each particle from its personal best p_i and the global best g :

$$v_i^{t+1} = w v_i^t + c_1 r_1 (p_i - x_i^t) + c_2 r_2 (g - x_i^t) \quad (4)$$

In both hybrids a candidate solution encoded $\log_{10}C \in [-1, 3]$, $\log_{10}\gamma \in [-4, 0]$ and $\log_{10}\epsilon \in [-3, -0.5]$. The fitness was the RMSE of the standardized validation predictions of an SVR trained on the training period. SSA and PSO were run with 20 agents for 40 iterations (MEALPY; Van Thieu and Mirjalili, 2023).

2.6. Performance Criteria

The models were evaluated with R^2 , RMSE, MAE, the Nash–Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970), the Kling–Gupta efficiency (KGE; Gupta et al., 2009) and Willmott's index (Willmott, 1981):

$$\text{NSE} = 1 - \Sigma(O_i - P_i)^2 / \Sigma(O_i - \bar{O})^2 \quad (5)$$

$$\text{KGE} = 1 - \sqrt{[(r-1)^2 + (\sigma_P/\sigma_O - 1)^2 + (\mu_P/\mu_O - 1)^2]} \quad (6)$$

where r is the linear correlation and σ and μ are the standard deviations and means of predictions (P) and observations (O). Because the mean head ($\approx 78.8 \text{ m a.s.l.}$) is large compared with its variability, the bias term of KGE is close to zero and KGE mainly reflects correlation and variability.

3. Results

3.1. Hyperparameter Optimization

Table 2. Optimized SVR hyperparameters, validation fitness and run time

Scenario	Model	C	γ	ϵ	Fitness	Time (s)
S1	SVR-SSA	112.9	0.00026	0.0017	0.5822	13.4
S1	SVR-PSO	162.0	0.00020	0.0010	0.5803	5.3
S2	SVR-SSA	998.9	0.00015	0.0859	0.3296	15.6
S2	SVR-PSO	917.6	0.00010	0.0340	0.3304	13.8

SSA and PSO converged to very similar solutions (Table 2; Fig. 4). In both scenarios they selected a large C and a very small γ , i.e. a smooth, nearly linear kernel with weak regularization. The two algorithms reached practically the same validation fitness. PSO reached a marginally lower fitness in S1 and SSA in S2. The searches took 5–16 s.

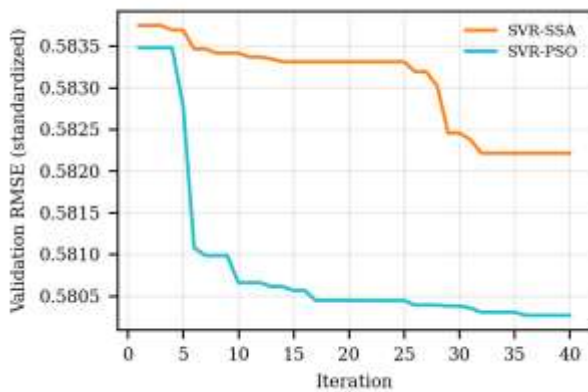


Figure 4a. Convergence of SSA and PSO in scenario S1

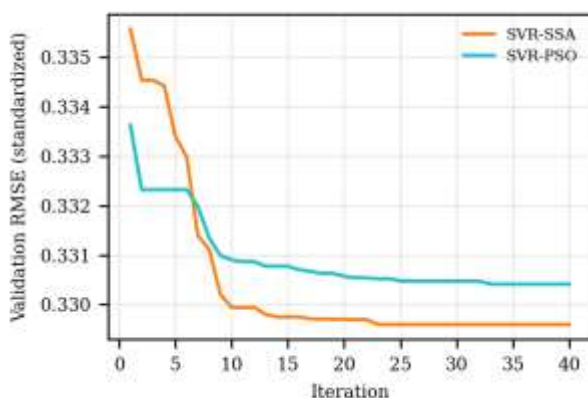


Figure 4b. Convergence of SSA and PSO in scenario S2

3.2. Scenario S1: Simulation from Meteorological Inputs

Table 3. Performance in scenario S1 (test period 2014–2020 unless indicated)

Model	R ²	RMSE (m)	NSE	KGE	R ² (val)	RMSE (val)
MLR	0.502	0.179	0.348	0.615	0.831	0.103

SVR	0.718	0.134	0.636	0.586	0.776	0.108
RF	0.521	0.175	0.382	0.570	0.804	0.109
LSTM	0.752	0.125	0.683	0.704	0.834	0.078
SVR-SSA	0.509	0.165	0.451	0.583	0.838	0.095
SVR-PSO	0.509	0.165	0.451	0.583	0.839	0.095

The results of scenario S1 are summarized in Table 3. The LSTM was clearly the best model in the test period, with $R^2 = 0.752$, $RMSE = 0.125$ m, $NSE = 0.683$ and $KGE = 0.704$. The default SVR ranked second ($NSE = 0.636$), while MLR, RF and both hybrid SVRs reached NSE values of only 0.35–0.45. The RF fitted the training period best ($NSE = 0.92$) but generalized poorly, which shows its tendency to overfit a few hundred monthly samples.

The hybrid SVRs had the lowest validation $RMSE$ of all models (0.095 m, against 0.108 m for the default SVR), yet their test $RMSE$ was higher (0.165 vs 0.134 m). The configurations selected on the five-year validation period therefore did not transfer to the test period. This is a clear case of overfitting to the validation set, which is easily overlooked when a single split is used both for tuning and for model selection. Fig. 5 shows that the LSTM reproduced the timing and amplitude of the seasonal cycle, including the strong spring rises, better than the SVR models. All models underestimated the test-period levels slightly (mean error -4 to -9 cm), and the LSTM showed the smallest bias apart from the default SVR.

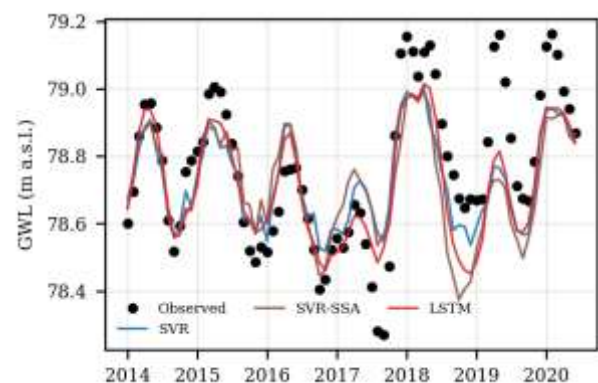


Figure 5. Observed and simulated GWL in the test period, scenario S1

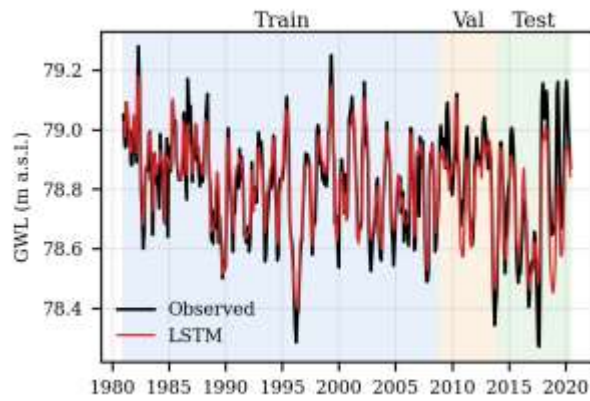


Figure 6. Observed and LSTM-simulated GWL over the full record, scenario S1

The full-record simulation (Fig. 6) shows that the LSTM captured the multi-year variability, including the low levels of dry years and the high levels of wet springs, without any groundwater input. The training and validation losses (Fig. 7) decreased smoothly, and early stopping prevented overfitting. The Taylor diagram (Fig. 8) confirms that the LSTM combined the highest correlation (0.87) with a realistic variance (normalized SD = 0.74). The default SVR had a similar correlation (0.85) but damped variability (0.61), and the remaining models had correlations of only 0.71–0.72.

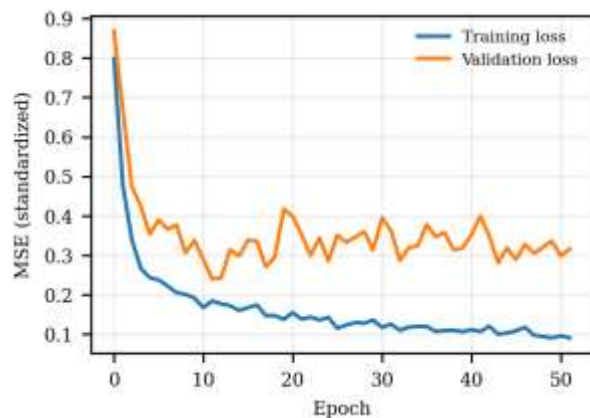


Figure 7. Training and validation loss of the LSTM (scenario S1, first ensemble member)

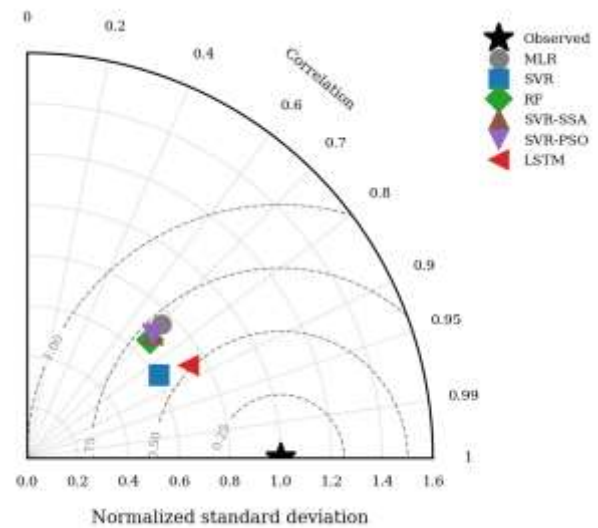


Figure 8. Taylor diagram of test-period simulations, scenario S1

The permutation importance of the RF (Fig. 9) shows that the six-month ($W6(t)$) and twelve-month ($W12(t)$) cumulative effective water input were by far the most important predictors, followed by $PET(t)$ and $W(t)$. This confirms that the aquifer integrates recharge over several months and that snowmelt-adjusted water input is a physically meaningful predictor.

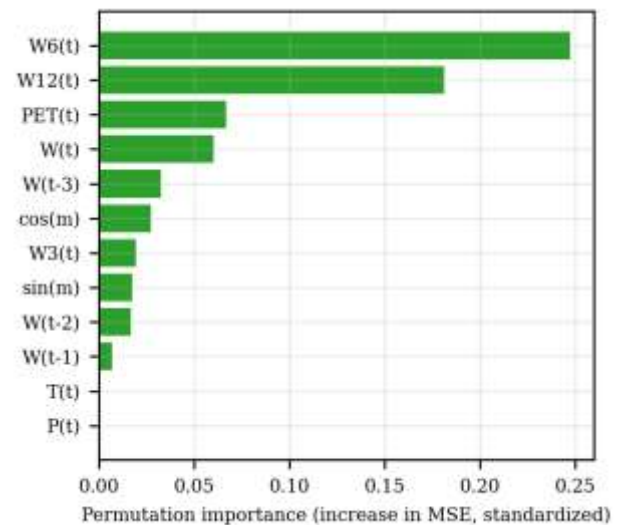


Figure 9. Permutation importance of the predictors (RF, scenario S1)

3.3. Scenario S2: One-Month-Ahead Forecasting

Table 4. Test-period performance in scenario S2 (one-month-ahead forecasts)

Model	R^2	RMSE (m)	MAE (m)	NSE	KGE	WI
Persistence	0.774	0.110	0.086	0.756	0.879	0.937
MLR	0.890	0.082	0.061	0.863	0.861	0.960
SVR	0.880	0.097	0.068	0.810	0.680	0.931
RF	0.896	0.081	0.062	0.868	0.809	0.959
LSTM	0.840	0.104	0.082	0.780	0.719	0.924

SVR-SSA	0.905	0.076	0.057	0.882	0.860	0.966
SVR-PSO	0.894	0.080	0.060	0.872	0.868	0.963

Adding the lagged GWL greatly improved all models (Table 4). Persistence already reached $NSE = 0.756$, and every learning model exceeded this value. SVR-SSA achieved the best scores ($R^2 = 0.905$, $RMSE = 0.076$ m, $NSE = 0.882$, $WI = 0.966$), followed by SVR-PSO ($RMSE = 0.080$ m), RF (0.081 m) and MLR (0.082 m). In this scenario the metaheuristic tuning reduced the test $RMSE$ of the default SVR by 21% (SSA) and 18% (PSO), and SVR-SSA improved on persistence by 30%. The LSTM, which was the best model in S1, ranked last among the learners in S2 ($RMSE = 0.104$ m). Its sequence input contained only the last observed level, while the tabular models used two lags together with explicitly aggregated water input.

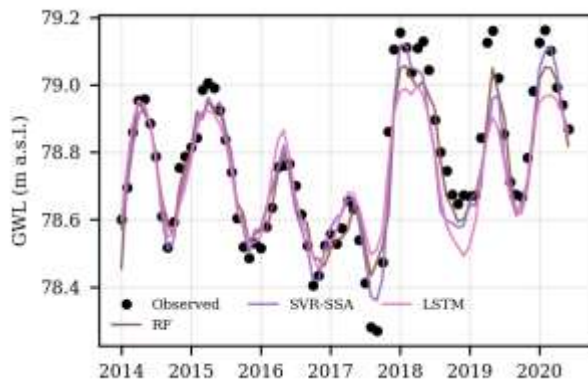


Figure 10. Observed and forecast GWL in the test period, scenario S2

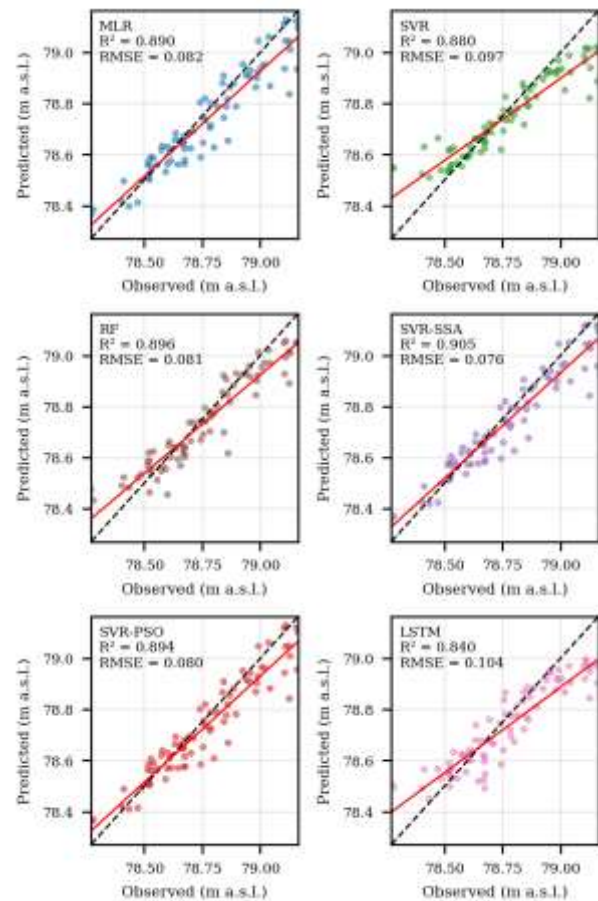


Figure 11. Observed versus forecast GWL in the test period, scenario S2

The forecasts (Figs. 10 and 11) follow the observations closely. The largest errors occur at the onset of the spring rise, when the change in level is driven by snowmelt that is not fully described by the monthly inputs. As expected, the RF importance in S2 (Fig. 12) is dominated by $GWL(t-1)$, followed by $W6(t)$ and $W(t-1)$. The error distributions (Fig. 13) show that SVR-SSA, SVR-PSO, RF and MLR have narrower error ranges than persistence and the LSTM.

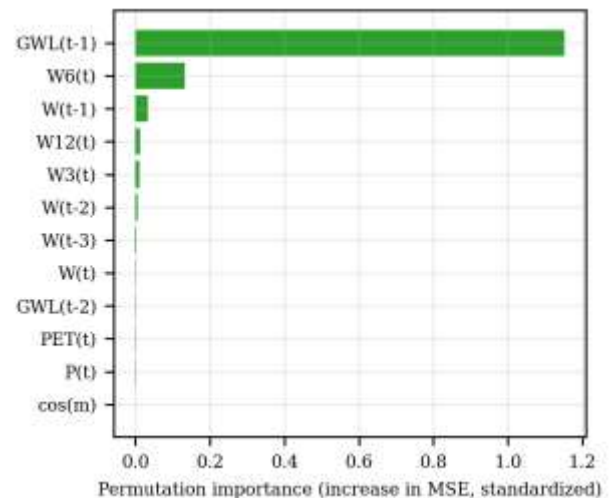


Figure 12. Permutation importance of the predictors (RF, scenario S2)

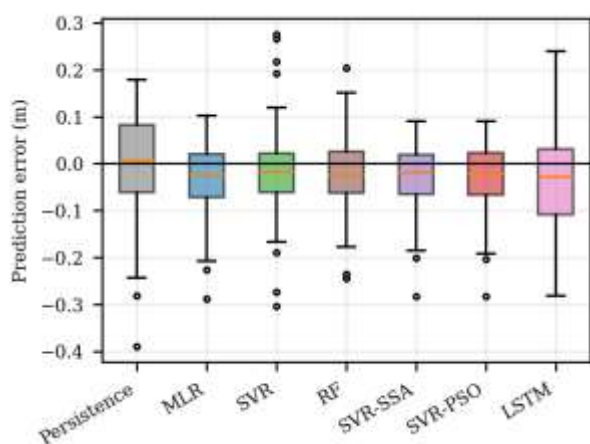


Figure 13. Distribution of forecast errors in the test period, scenario S2

4. Discussion

The two scenarios lead to different conclusions about which model to use. Without past groundwater information (S1), the model must reconstruct the aquifer's memory from the meteorological history. The LSTM, which learns this memory internally from 12-month sequences, was superior, in agreement with Wunsch et al. (2021). The tabular models had to rely on hand-crafted cumulative inputs and were less successful, even though these inputs were physically meaningful. With past groundwater levels available (S2), the aquifer memory is largely contained in $GWL(t-1)$, and simple, well-regularized models such as MLR and tuned SVR performed best. In practice, the choice between the approaches should therefore depend on the availability of recent observations.

The behaviour of the hybrid models deserves attention. In S2, SSA and PSO improved the SVR considerably and produced the best forecasts. In S1 the same procedure selected hyperparameters that minimized the validation error but degraded the test performance. The validation period (2009–2013) contained only 54 observed months and may not represent the wetter and drier conditions of the test period. Many hybrid-model studies report only the positive case. Our results suggest that metaheuristic tuning should be combined with time-series cross-validation or with a longer validation period, and that improvements should be confirmed on independent data. The close agreement between SSA and PSO also indicates that the choice of optimizer mattered much less than the choice of validation design.

The study has limitations. The degree-day factor was fixed and not calibrated, and potential evaporation was used instead of actual evaporation. Only one well was analysed, and no information on abstraction or land use was available. The significant trend in GWL is very small, but it contributes to the negative bias of all S1 simulations in the test period. Future work should test the approach on multiple wells in the Swedish

monitoring network, include calibrated snow routines or snow water equivalent products, and quantify predictive uncertainty.

5. Conclusions

Monthly groundwater levels in a snow-influenced aquifer near Heby, Sweden, were simulated and forecast with MLR, SVR, RF, LSTM and two metaheuristic-optimized SVR models (SVR-SSA and SVR-PSO), using effective water input from a degree-day snow routine. The main conclusions are: (1) in the simulation scenario, the LSTM clearly outperformed the other models (test NSE = 0.68); (2) in the one-month-ahead forecasting scenario, SVR-SSA was the most accurate model (NSE = 0.88, RMSE = 7.6 cm), closely followed by SVR-PSO; (3) metaheuristic tuning improved SVR by about 21% in S2 but overfitted the short validation period in S1; and (4) the six- and twelve-month cumulative effective water input, which accounts for snow storage, were the key meteorological drivers of groundwater levels. Sequence-learning models are therefore recommended for gap filling and climate-impact simulations, and tuned kernel models for operational short-term forecasting.

6. Data Availability

Groundwater level and E-OBS-based meteorological series for the Heby well were obtained from the example data of the Pastas package (github.com/pastas/pastas; Collenteur et al., 2019). The code used for the analysis is available from the authors on request.

7. References

- Bakker, M., & Schaars, F. (2019). Solving groundwater flow problems with time series analysis: You may not even need another model. *Groundwater*, 57(6), 826–833.
- Boo, K. B. W., et al. (2024). Groundwater level forecasting with machine learning models: A review. *Water Research*, 252, 121249.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Collenteur, R. A., Bakker, M., Caljé, R., Klop, S. A., & Schaars, F. (2019). Pastas: Open source software for the analysis of groundwater time series. *Groundwater*, 57(6), 877–885.
- Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., & Jones, P. D. (2018). An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123(17), 9391–9409.
- Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems*, 9, 155–161.
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving

- hydrological modelling. *Journal of Hydrology*, 377(1–2), 80–91.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Hock, R. (2003). Temperature index melt modelling in mountain areas. *Journal of Hydrology*, 282(1–4), 104–115.
- Jesse, G., et al. (2025). A systematic review of machine learning models for groundwater level prediction. *Applied Computing and Geosciences*.
- Kayhomayoon, Z., et al. (2022). A combination of metaheuristic optimization algorithms and machine learning methods improves the prediction of groundwater level. *Water*, 14(5), 751.
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95 – International Conference on Neural Networks* (Vol. 4, pp. 1942–1948). IEEE.
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22(11), 6005–6022.
- Moravej, M., et al. (2020). Groundwater level simulation and forecasting using interior search algorithm-least square support vector regression (ISA-LSSVR). *Groundwater for Sustainable Development*.
- Mozaffari, S., et al. (2022). Forecasting groundwater levels using a hybrid of support vector regression and particle swarm optimization. *Water Resources Management*, 36(6). <https://doi.org/10.1007/s11269-022-03118-z>
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282–290.
- Pham, Q. B., et al. (2022). Groundwater level prediction using machine learning algorithms in a drought-prone area. *Neural Computing and Applications*, 34. <https://doi.org/10.1007/s00521-022-07009-7>
- Rajaei, T., Ebrahimi, H., & Nourani, V. (2019). A review of the artificial intelligence methods in groundwater level modeling. *Journal of Hydrology*, 572, 336–351.
- Samantaray, S., et al. (2024). Groundwater level prediction using an improved ELM model integrated with hybrid particle swarm optimisation and grey wolf optimisation. *Groundwater for Sustainable Development*, 26, 101178.
- Seifi, A., Ehteram, M., Singh, V. P., & Mosavi, A. (2020). Modeling and uncertainty analysis of groundwater level using six evolutionary optimization algorithms hybridized with ANFIS, SVM, and ANN. *Sustainability*, 12(10), 4023.
- Taylor, R. G., Scanlon, B., Döll, P., et al. (2013). Ground water and climate change. *Nature Climate Change*, 3, 322–329.
- Van Thieu, N., & Mirjalili, S. (2023). MEALPY: An open-source library for latest meta-heuristic algorithms in Python. *Journal of Systems Architecture*, 139, 102871.
- Willmott, C. J. (1981). On the validation of models. *Physical Geography*, 2(2), 184–194.
- Wunsch, A., Liesch, T., & Broda, S. (2021). Groundwater level forecasting with artificial neural networks: A comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX). *Hydrology and Earth System Sciences*, 25(3), 1671–1687.
- Wunsch, A., Liesch, T., & Broda, S. (2022). Deep learning shows declining groundwater levels in Germany until 2100 due to climate change. *Nature Communications*, 13, 1221.
- Xue, J., & Shen, B. (2020). A novel swarm intelligence optimization approach: Sparrow search algorithm. *Systems Science & Control Engineering*, 8(1), 22–34.