Predicting Infiltration Rates in Soil Aquifer Treatment: A Comparative Study of Machine Learning Models Using Operational and Weather Data

Christian Reichhardt[1], Patrick Baier[1], Roy Elkayam[2,3], and Ovadia Lev[2]

[1] Karlsruhe University of Applied Sciences, Germany
christian.reichhardt@h-ka.de patrick.baier@h-ka.de
[2] The Hebrew University of Jerusalem, Israel
[3] Mekorot Water Co., Israel

Abstract. We present a comparative benchmark of machine learning models with static and temporal sequence inputs for predicting infiltration rates in Soil Aquifer Treatment systems, trained on ten years of operational and weather records. Using expanding-window cross-validation, nonlinear models achieved accurate basin-level predictions without explicit water-quality or soil-property measurements. Gradient Boosting performed best among cycle-aggregated tabular input models using current-cycle information and basin identity. LSTM achieved comparable performance for temporal 365-day inputs. Feature ablation shows that adding weather variables to models using basin identity and operational conditions did not yield a clear improvement in our setting, with LSTM achieving 0.654 cm/h MAE without weather information. Integrated Gradients showed strongest per-day attributions for recent inputs, with weaker, model-dependent long-range contributions. Our findings suggest that data-driven models based on recent, routinely available operational records can support SAT simulation, basin-level planning, and future control strategies.

Keywords: Soil Aquifer Treatment; SAT; Managed Aquifer Recharge; MAR; Infiltration Rate Prediction; Machine Learning in Hydrology; Time-Series Regression; Explainable AI; XAI

1 Introduction

Managed Aquifer Recharge (MAR) is increasingly vital for securing water supplies, with Soil Aquifer Treatment (SAT) being a widely implemented approach. In SAT, treated wastewater is directed into surface basins where it infiltrates through the soil into underlying aquifers. On its way through the soil, the water is purified by physical, chemical, and biological processes. Systems typically operate in alternating wetting and drying cycles, with the drying phases promoting soil regeneration and mitigation of basin clogging. During some of the drying phases, tillage is used to break the surface layer to further reduce clogging. Accurately predicting water infiltration rates under varying operational and environmental conditions is critical for optimizing SAT performance and increasing groundwater recharge. Many SAT systems are manually operated due to limited process understanding and a scarcity of long-term field-scale data. This can lead to suboptimal infiltration rates, reduced groundwater recharge, and inefficiencies in agricultural water management. Laboratory and column studies provide valuable insight, but often fail to

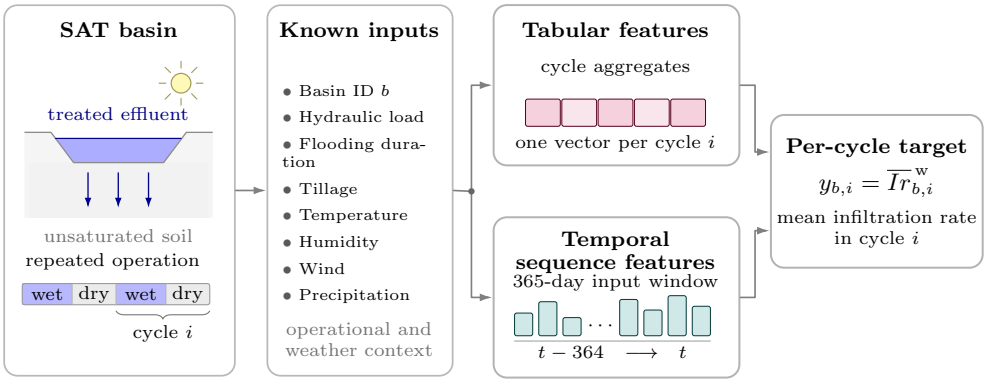


Fig. 1. Conceptual formulation of per-cycle SAT infiltration rate prediction. The same operational and meteorological inputs are represented either as one aggregated tabular vector per wetting-drying cycle or as a sequence of daily vectors over a 365-day window ending at the target cycle endpoint. Both model families predict the average infiltration rate for the current cycle

capture field-scale heterogeneity and sensor limitations, highlighting the need for data-driven, basin-level prediction frameworks based on routinely available monitoring data. In this study, we formulate infiltration rate estimation as a regression problem using operational and meteorological data from a large real-world SAT dataset. Specifically, we ask: *What is the infiltration rate under known operational and weather conditions?* This differs from classical forecasting because previous infiltration rates are not used as input. Instead, the objective is to learn a mapping from basin characteristics, operational decisions, and environmental conditions to infiltration rate. This setup is relevant for simulation and control applications, where candidate operational conditions may be known in advance, environmental inputs can be represented by forecasts, but future infiltration rates remain unknown.

Our work makes three key contributions: (1) a benchmark of static and temporal machine learning architectures for infiltration rate prediction across 50 infiltration basins, (2) a comparison between current-cycle tabular features and 365-day temporal input sequences, and (3) feature ablation and interpretable attributions that clarify the influence of operational and climatic drivers. This comparative framework provides a foundation for digital SAT simulation, optimal control strategies, and adaptive management under varying environmental conditions.

2 State of the Art

SAT systems are a form of MAR in which treated effluent infiltrates from surface basins through an unsaturated aquifer medium, where microorganisms are substantially removed during subsurface passage [1]. Infiltration dynamics are affected by operational conditions, clogging processes, soil properties, water quality and environmental factors. Due to the scarcity of long-term real-world data, studies on infiltration often rely on column experiments [2–4]. While these studies provide mechanistic insight, their transferability to heterogeneous field-scale basins is limited. In parallel, several physical models have been proposed to describe infiltration into soil, such as the Green-Ampt model [5] and related approaches [6, 7]. These models often relate infiltration to hydraulic head, meaning the ponded water depth or pressure potential above the soil surface. However,

recent research by Elkayam and Lev [8] suggested that the infiltration rate in SAT may be largely head-independent under certain field conditions. Other studies indicate that clogging can be a driving factor for infiltration decline [3, 9]. The Shafdan SAT system has also been investigated with respect to pretreatment of the effluent [10]; biogeochemical processes [11] and the effects of longer drying phases on oxidation [12]. Lin et al. [13] reported a correlation between infiltration rate and effluent temperature in the Shafdan system. However, effluent temperature is not equivalent to ambient meteorological temperature, and the predictive value of routinely available weather variables remains unclear.

Machine learning provides an alternative approach for learning infiltration rate relationships directly from data records. Abdalrahman et al. [14] applied machine learning to predict infiltration rates in nine real-world wastewater infiltration basins over seven months, focusing on effluent water quality features. However, long-term field-scale prediction across many basins using routinely available operational and meteorological data remains largely unexplored. For sequential environmental data, machine learning models range from static tabular regressors to temporal architectures that process multivariate input sequences. Classical recurrent models such as Long Short-Term Memory networks (LSTM) [15], temporal convolution-based approaches such as TimesNet [16], and Transformer-based models [17, 18] can model temporal dependencies over extended input windows. These architectures are therefore relevant for testing whether long historical sequences provide additional predictive value over current-cycle operational features.

To the best of our knowledge, large-scale prediction of SAT infiltration rates using long-term operational and meteorological records across many infiltration basins has not yet been systematically studied. In particular, it remains unclear which routinely available operational and meteorological information is sufficient for accurate basin-level infiltration rate prediction, and whether longer historical input windows provide additional predictive value beyond current-cycle conditions.

3 Dataset

Our dataset originates from the Shafdan SAT facility in Israel, operated by Mekorot Water Co. The facility receives approximately 140 million cubic meters of treated wastewater per year from the Shafdan Wastewater Treatment Plant and distributes it across multiple fields, each containing several infiltration basins. Each basin operates independently with its own flooding and resting schedule, creating a diverse range of operational conditions within a single facility. Our dataset spans ten years of preprocessed records from 50 basins and combines for each basin: (1) Operational data. Per-basin monitoring records include high-frequency measurements of flooding duration, tillage events, hydraulic loading, and computed infiltration rates. (2) Meteorological data. Meteorological data originate from the Israel Meteorological Service (IMS)[4] and include temperature, relative humidity, wind speed, and precipitation. We assign weather observations to SAT fields by selecting the nearest available IMS weather station using haversine distance, excluding coastal stations that may not be representative of inland fields. Preprocessing includes invalid-value filtering, weather interpolation, and rolling-mean water-level smoothing. In total, the preprocessed dataset encompasses 57,689 cycles. We frame infiltration prediction as a per-cycle regression task, where a cycle represents a complete wet-dry phase from the start of flooding until the next resting period and lasts on average 3.21 days. We construct two feature representations corresponding to our model families:

[4] https://ims.gov.il/en/ObservationDataAPI.

Tabular features (per-cycle aggregation). For the static model family, features are aggregated over each flooding cycle: average relative humidity (%), average temperature (°C), average wind speed (m/s), precipitation intensity (mm/h), time since tillage (s), hydraulic load (m), and flooding duration (s). In order to allow the models to learn basin-specific characteristics such as general soil properties, we also one-hot encode the unique basin identifier.

Temporal sequence features (rolling 365-day window). For the temporal sequence input model family, we compile 365 daily intervals ending at the target cycle endpoint, including observations from the target cycle itself. Features at each daily timestep are the same as for per-cycle aggregation but aggregated per day. Instead of time since tillage, we use a binary tillage indicator per day.

To capture cycle-level variability within basins, we define the regression target as the average wetting-period infiltration rate for each cycle i in basin b. The target can be computed as described by Elkayam and Lev [8] from

$$\bar{Ir}^{\mathrm{w}}_{b,i} = \frac{1}{T^{\mathrm{w}}_{b,i}} \frac{1}{A_b} \int_{t=0}^{T^{\mathrm{f}}_{b,i}} Q_{b,i,t} dt = \frac{1}{T^{\mathrm{w}}_{b,i}} p_{b,i}, \tag{1}$$

where $T^{\mathrm{w}}_{b,i}$ is the total wetting duration in hours (sum of flooding and drainage times). $T^{\mathrm{f}}_{b,i}$ is the total flooding duration during which the valve is open and effluent inflow $Q_{b,i,t} > 0$. A_b corresponds to the area of the basin in m^2. Since hydraulic load $p_{b,i}$ is also used as an input feature, the learning task can be interpreted as estimating the effective wetting duration or drainage behavior associated with a given operational cycle, rather than predicting infiltration from completely independent variables. For convenience, the target is converted to cm/h. The mean infiltration rate over the whole dataset is 3.25 cm/h with a standard deviation of 1.82 cm/h. We also evaluated an extended infiltration rate definition that accounts for the integral of precipitation and evaporation, calculating evaporation according to Valiantzas [19]. Because the extended and standard infiltration rates are almost linearly related in our dataset ($R^2 = 0.998$), we use the simpler target definition throughout this study. This assumption may not hold in regions with substantially higher precipitation or evaporation variability.

4 Methodology

We compare established tabular and sequence-based machine learning models against a naive day-of-year baseline using data from the Shafdan SAT facility. We evaluate two model families trained on different feature representations:

Cycle-aggregated tabular features. This model family receives one per-cycle feature vector, described in section 3, and produces a single infiltration rate prediction. We evaluate four architectures: (1) *Linear Regression (LR)*, (2) *Decision Tree Regressor (DT)*, (3) *Random Forest Regression (RF)*, (4) *Gradient Boosting Regressor (GB)* [20]. All tabular models use the *scikit-learn* [21] implementation.

Temporal sequence features. Sequence-based models receive a 365-day daily feature sequence and produce a single infiltration-rate prediction. These models can learn short-term effects and longer-term patterns such as seasonal components. We evaluate: (1) an *LSTM* [15] regression model with configurable hidden state size and layer depth, projecting the final hidden state to a scalar output via a linear layer, (2) *iTransformer* as described by Liu et al. [17], (3) *Non-stationary Transformer* (NSTransformer) [18], and (4) *TimesNet* [16].

We use the *Ray Tune* framework [22] for hyperparameter optimization. For each model, we draw 40 random configurations and evaluate each configuration across all three folds. We benchmark all models against a day-of-year baseline. For each basin, this baseline computes the historical mean infiltration rate from the training data grouped by day of year and uses the grouped values as predictions. Thus, it captures seasonal patterns as described by Lin et al. [13] but has no explicit information about operational conditions.

All models use three chronological expanding-window folds with approximately one-year held-out test periods, separated from training by an approximately one-year gap to prevent 365-day test windows from overlapping the training period. Records without complete preceding 365-day history or outside a fold's train/test periods are excluded. For cycle-level models, the folds contain 19,007/4,590, 28,327/5,029 and 38,316/5,048 train/test cycles. Temporal models use closely aligned splits but slightly different sample counts due to sequence construction; because training sets overlap, fold counts are not expected to sum to the 57,689-cycle dataset.

For hyperparameter tuning (including finding the number of training epochs for deep regression networks), we use the chronologically last 15% of each training partition as validation data. After selecting the best hyperparameters across all folds, we retrain the final model on the complete training partition, including the validation data, and evaluate on the withheld test partition. Hyperparameter trials are not pruned before completion across all folds, because poor performance on a single fold does not necessarily imply poor average performance. All numerical features are standardized using statistics computed from the respective training partition only. Deep regression networks are trained with a maximum of 200 epochs, with early stopping if the validation loss does not improve for 10 consecutive epochs. Neural models are trained by minimizing the Mean Squared Error (MSE) using the Adam [23] optimizer.

We report Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Symmetric Mean Absolute Percentage Error (SMAPE), and coefficient of determination (R^2). SMAPE is used for basin-wise comparisons because it is scale-independent, symmetric, and bounded.

5 Experimental Results

All results are reported as mean and standard deviation across the three cross-validation folds for both model families. Integrated Gradients [24] is used to obtain interpretable feature attributions for selected temporal models.

Table 1 summarizes the prediction results for both model families. Among the tabular models, Gradient Boosting performs best overall, achieving the lowest MAE, lowest RMSE, and highest R^2. Decision Tree obtains the lowest SMAPE, while Linear Regression performs substantially worse and shows an unstable coefficient of determination. For temporal sequence models, LSTM achieves the best performance across all reported metrics, with NSTransformer ranking second among the temporal models. Overall, the best cycle-aggregated tabular model performs competitively with, and in several metrics outperforms, the temporal sequence models. For Gradient Boosting, basin-wise SMAPE analysis (not shown) indicates better robustness for basins in the fields *Soreq 2*, *Yavne 2*, *Yavne 3*, and *Yavne 1K*, while basins in *Yavne 4* and *Yavne 1* show larger errors.

Integrated Gradients approximates feature attributions by integrating gradients of the model output with respect to the input along a straight path from a predefined zero baseline to the observed inputs [24]. Figure 2 shows the last 30 days for LSTM and

Table 1. Prediction results reported as mean and standard deviation across cross-validation folds. Best mean values are highlighted separately for tabular and temporal groups

Model	MAE (cm/h) ↓	RMSE (cm/h) ↓	SMAPE (%) ↓	R^2 ↑
Baseline	1.045 ± 0.207	1.372 ± 0.235	34.10 ± 6.35	0.290 ± 0.178
Tabular Feature Models				
DT	0.790 ± 0.110	1.138 ± 0.139	**26.10 ± 2.25**	0.574 ± 0.059
GB	**0.690 ± 0.113**	**0.977 ± 0.125**	28.08 ± 3.91	**0.684 ± 0.055**
LR	1.034 ± 0.158	1.514 ± 0.357	37.42 ± 5.05	0.238 ± 0.257
RF	0.742 ± 0.120	1.032 ± 0.156	28.24 ± 3.74	0.648 ± 0.069
Temporal Sequence Feature Models				
iTransformer	0.972 ± 0.139	1.327 ± 0.164	34.22 ± 5.84	0.408 ± 0.147
LSTM	**0.710 ± 0.082**	**1.076 ± 0.128**	**25.91 ± 3.28**	**0.619 ± 0.051**
NSTransformer	0.745 ± 0.126	1.116 ± 0.172	26.79 ± 3.63	0.591 ± 0.073
TimesNet	0.758 ± 0.123	1.143 ± 0.216	28.26 ± 0.65	0.574 ± 0.094

Table 2. Model results for different input feature sets: basin features, operational features, and weather features. Best mean values are highlighted per model

Features	MAE (cm/h) ↓	RMSE (cm/h) ↓	SMAPE (%) ↓	R^2 ↑
GB				
basin only	1.149 ± 0.149	1.515 ± 0.177	38.39 ± 4.06	0.243 ± 0.109
+ operational	**0.684 ± 0.123**	0.982 ± 0.138	**27.68 ± 3.59**	0.684 ± 0.048
+ weather	0.690 ± 0.113	**0.977 ± 0.125**	28.08 ± 3.91	**0.684 ± 0.055**
LSTM				
basin only	1.159 ± 0.171	1.524 ± 0.197	39.25 ± 5.32	0.234 ± 0.125
+ operational	**0.654 ± 0.077**	**1.029 ± 0.127**	**24.19 ± 2.80**	**0.654 ± 0.038**
+ weather	0.710 ± 0.082	1.076 ± 0.128	25.91 ± 3.28	0.6194 ± 0.0508

NSTransformer, omitting static basin identifiers. Across the seven displayed features and three folds, the remaining 335 days account for about 11% of the absolute IG sum for LSTM and 38% for NSTransformer, but with much smaller per-day attributions than the final 30 days. For LSTM, all plotted earlier-day mean attributions remain below 0.01; for NSTransformer, two values exceed it. Thus, temporal models rely most strongly on recent inputs, although NSTransformer retains more distributed long-range attribution.

We perform a feature ablation by retraining the best GB as well as the best LSTM configurations with different feature sets. We distinguish between basin, operational, and weather features. Basin features correspond to one-hot encoded basin identifiers; operational features include flooding duration, tillage information, and hydraulic load; weather features include humidity, temperature, wind speed, and precipitation. Results are presented in Table 2. Basin identity alone provides only limited predictive accuracy for both models. Adding weather variables on top of basin and operational features does not yield a clear benefit. The MAE of GB remains nearly unchanged (0.684 vs. 0.690 cm/h), while the LSTM performs even better without weather information.

6 Discussion

The results suggest that nonlinear models can achieve accurate basin-level infiltration predictions without explicit water-quality or soil-composition measurements. The competitive performance of cycle-aggregated tabular models indicates that much of the pre-

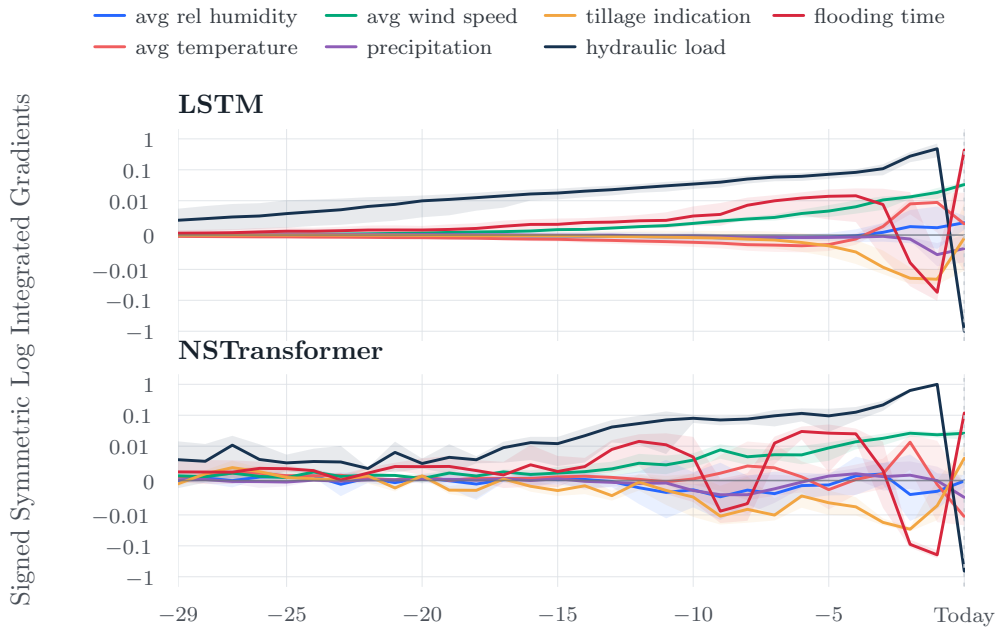


Fig. 2. Integrated Gradients feature attributions of final LSTM and NSTransformer models over the final 30 days of the input window, summed over test samples and normalized per fold by the largest absolute attribution (range -1 to 1). Shaded bands indicate standard deviation

dictive information in this dataset is contained in current-cycle operational conditions and basin identity, making such models attractive for practical deployment.

How can the role of different features be interpreted? Hydraulic load is both an operational input and part of the target definition. Consequently, the large gain from adding operational features in the feature ablation, as well as the high attribution of hydraulic load, should not be interpreted as direct physical feature importance alone. Rather, the model partly learns how a specified loading condition relates to the resulting effective wetting duration and infiltration rate. Negative tillage attributions may reflect reactive management after reduced infiltration. Adding weather variables on top of basin identity and operational features did not provide a clear predictive benefit in this experimental setup. This should not be interpreted as evidence that weather has no physical influence on infiltration, but rather as a lack of additional predictive value given the available predictors. A correlation analysis (not shown) provided limited evidence that operational inputs directly encode weather. The strongest pooled weather-operation association was weak, and month explained little variance in operational features. Moreover, ambient meteorological variables may only imperfectly represent relevant drivers such as effluent temperature or local basin-surface conditions. The relatively low performance of the day-of-year baseline further suggests that seasonal information alone is insufficient to explain infiltration variability at this site. Integrated Gradients further indicated that temporal sequence models assign the strongest per-day attributions to recent inputs, while longer-range contributions were smaller per day and more model-dependent.

How can these findings be leveraged for control strategies? The results motivate basin-level decision-support models that condition on recent basin state and candidate opera-

tional records. Because cycle-aggregated models performed competitively, SAT simulation and planning tools may not require extensive year-long input sequences in this setting. Given proposed operational conditions, such models could estimate expected infiltration rates, support basin availability assessment, and provide a basis for simulation-based optimization or reinforcement-learning control.

7 Limitations and Future Work

Although the dataset spans ten years, the number of observed years allows only a small number of chronological cross-validation folds. The data is also subject to sensor noise, basin-scale heterogeneity, manual operational records, and imperfect assignment of weather observations from nearby meteorological stations. The findings are site-specific to the Shafdan SAT system, its climate, and its historical management regime. Predictions should therefore be interpreted primarily within the range of operational and weather conditions observed during training.

Further limitations arise from the target definition and feature representation. Since infiltration rate is defined as $p_{b,i}/T^{\mathrm{w}}_{b,i}$ and hydraulic load $p_{b,i}$ is used as an input feature, the task partly estimates the effective wetting duration associated with a given loading condition. Future work should evaluate alternative target formulations, generalization to held-out basins, and prospective decision-support settings using prescribed operational scenarios and forecast environmental inputs.

8 Conclusion

In summary, our findings indicate that data-driven models based on routinely available operational records can predict SAT infiltration rates with useful accuracy. Cycle-aggregated Gradient Boosting performed strongest among tabular models and was competitive with temporal sequence models, while the ablation study showed that the lowest MAE was achieved by the LSTM using basin identity and operational features without weather inputs. Thus, weather variables should be interpreted as evaluated candidate predictors rather than necessary components of the best-performing model in this dataset. Integrated Gradients showed strongest per-day relevance for recent inputs, with weaker and model-dependent long-range effects. These results support compact data-driven models for SAT simulation, basin-level planning, and future control strategies, while emphasizing that conclusions are site-specific and should be validated under different climatic, operational, and soil conditions.

Disclosure of AI Assistance

AI-assisted tools were used solely as supportive instruments during manuscript preparation and code development. They contributed to wording suggestions, text refinement, code refinement, code review and editorial improvements. The conception of the research, methodological decisions, analysis, interpretation of findings, and all final content remain the responsibility of the authors.

Funding

This work was funded by the Federal Ministry for Research, Technology and Space (BMFTR), Germany, under grant number 02WIL1733.

References

1. Dillon, P.: Managed aquifer recharge: an introduction. National Water Commission (2009)
2. Gorski, G., Dailey, H., Fisher, A.T., Schrad, N., Saltikov, C.: Denitrification during infiltration for managed aquifer recharge: Infiltration rate controls and microbial response. Science of the Total Environment **727** (2020) 138642
3. Pavelic, P., Dillon, P.J., Mucha, M., Nakai, T., Barry, K.E., Bestland, E.: Laboratory assessment of factors affecting soil clogging of soil aquifer treatment systems. Water Research **45**(10) (2011) 3153–3163
4. Cha, W., Kim, J., Choi, H.: Evaluation of steel slag for organic and inorganic removals in soil aquifer treatment. Water Research **40**(5) (2006) 1034–1042
5. Green, W.H., Ampt, G.A.: Studies on Soil Physics. The Journal of Agricultural Science **4**(1) (1911) 1–24
6. Horton, R.E.: An approach toward a physical interpretation of infiltration-capacity. Soil Science Society of America Journal **5**(C) (1940) 399–417
7. Philip, J.R.: Theory of Infiltration. Advances in Hydroscience **5** (1969) 215–296
8. Elkayam, R., Lev, O.: Head-Independent Infiltration Rate in Aquifer Recharge with Treated Municipal Wastewater. ACS ES&T Water **4**(12) (2024) 5678–5688
9. Bouwer, H., Rice, R.C.: Effect of Water Depth in Groundwater Recharge Basins on Infiltration. Journal of Irrigation and Drainage Engineering **115**(4) (1989) 556–567
10. Zucker, I., Mamane, H., Cikurel, H., Jekel, M., Hübner, U., Avisar, D.: A hybrid process of biofiltration of secondary effluent followed by ozonation and short soil aquifer treatment for water reuse. Water Research **84** (2015) 315–322
11. Goren, O., Burg, A., Gavrieli, I., Negev, I., Guttman, J., Kraitzer, T., Kloppmann, W., Lazar, B.: Biogeochemical processes in infiltration basins and their impact on the recharging effluent, the soil aquifer treatment (SAT) system of the Shafdan plant, Israel. Applied Geochemistry **48** (2014) 58–69
12. Ben Moshe, S., Weisbrod, N., Barquero, F., Sallwey, J., Orgad, O., Furman, A.: On the role of operational dynamics in biogeochemical efficiency of a soil aquifer treatment system. Hydrology and Earth System Sciences **24**(1) (2020) 417–426
13. Lin, C., Greenwald, D., Banin, A.: Temperature dependence of infiltration rate during large scale water recharge into soils. Soil Science Society of America Journal **67**(2) (2003) 487–493
14. Abdalrahman, G., Lai, S.H., Kumar, P., Ahmed, A.N., Sherif, M., Sefelnasr, A., Chau, K.W., Elshafie, A.: Modeling the infiltration rate of wastewater infiltration basins considering water quality parameters using different artificial neural network techniques. Engineering Applications of Computational Fluid Mechanics **16**(1) (2022) 397–421
15. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9** (1997) 1735–1780
16. Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., Long, M.: TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. In: Proceedings of the 11th International Conference on Learning Representations. (2023)
17. Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., Long, M.: iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. In: Proceedings of the 12th International Conference on Learning Representations. (2024)
18. Liu, Y., Wu, H., Wang, J., Long, M.: Non-stationary Transformers: Exploring the Stationarity in Time Series Forecasting. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. (2022)
19. Valiantzas, J.D.: Simplified versions for the Penman evaporation equation using routine weather data. Journal of Hydrology **331**(3–4) (2006) 690–702
20. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Annals of statistics **29**(5) (2001) 1189–1232
21. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É.: Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research **12** (2011) 2825–2830

22. Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J.E., Stoica, I.: Tune: A Research Platform for Distributed Model Selection and Training. arXiv preprint arXiv:1807.05118 (2018)
23. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations. (2015)
24. Sundararajan, M., Taly, A., Yan, Q.: Axiomatic Attribution for Deep Networks. In: Proceedings of the 34th International Conference on Machine Learning. (2017) 3319–3328