Science of the Total Environment 955 (2024) 177193

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Chronological assessment of heuristic data driven approaches for soil water content simulation in subsurface drip irrigated rice

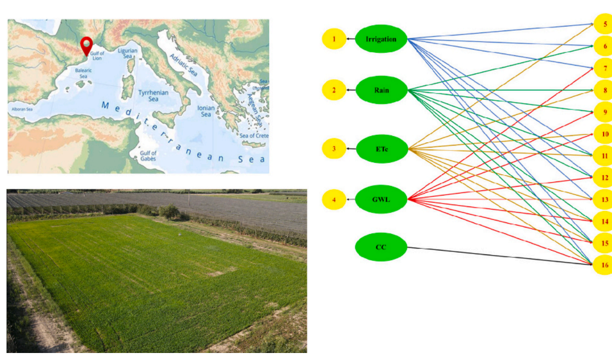
Jalal Shiri [a,b], Mohammad Hossein Kazemi [a], Sepideh Karimi [a,*], Sílvia Cufí [c], Francisco Ramírez de Cartagena [c], Jaume Pinsach [c], Gerard Arbat [c,*]

[a] Water Engineering Department, Faculty of Agriculture, University of Tabriz, Tabriz, Iran
[b] Water Engineering and Science Research Institute (WESRI), University of Tabriz, Tabriz, Iran
[c] Department of Chemical and Agricultural Engineering and Technology, University of Girona, Carrer de Maria Aurèlia Capmany, 61, 17003 Girona, Catalonia, Spain

HIGHLIGHTS

- Gene expression programming was used to predict soil moisture.
- The method was applied to subsurface drip irrigated rice.
- k-fold testing data considering two different chronologic strategies was used.
- Considering different growing stages of rice improved the predictions.
- Gene expression programming allowed to accurately predict soil moisture.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Charlotte Poschenrieder

Keywords:
K-fold testing
Machine learning
Meteorological data
Soil water content
Subsurface drip irrigation

ABSTRACT

Accurate estimation of soil water content (SWC) is essential for effective agriculture and water resources management. While various methods have been developed for in-situ SWC measurement, practical limitations and the need for comprehensive water sensor networks make their use complicated. To overcome these challenges, heuristic data-driven models may provide a suitable alternative to practical methods for SWC simulation under different cultivation conditions. In this paper, the application of gene expression programming (GEP) methodology was proposed to simulate SWC at three different depths in rice fields using information related to weather and groundwater. A modeling study was conducted that applied the robust k-fold testing data assignment method, considering two different chronologic strategies of “k” defining to evaluate both strategies. The first one was based on the definition of the “k” values based on yearly data partitioning, while the second one considered growing stages as the “k” definition criterion. Besides evaluating the models using error statistics, a further uncertainty analysis was also conducted to check stability and confidence. The obtained results revealed that selection of “k” based on growing stages produced more accurate and stable results. Among the target parameters, water content at the third layer was predicted with higher accuracy.

* Corresponding authors.
E-mail addresses: karimi_sepide@yahoo.com (S. Karimi), gerard.arbat@udg.edu (G. Arbat).

https://doi.org/10.1016/j.scitotenv.2024.177193
Received 22 July 2024; Received in revised form 22 October 2024; Accepted 22 October 2024
Available online 1 November 2024

1. Introduction

Surface and subsurface water are subjected to interaction between the earth and the atmosphere, with surface water playing a crucial role in providing the necessary water vapor required for evaporation and transpiration processes. When these processes are combined, water depletion occurs, disrupting the water balance in this layer (Corradini, 2014). Detecting accurate soil water content (SWC) is a challenging task, primarily due to the intricate relationships between various parameters, the variations of meteorological conditions, as well as its temporal and spatial variability. Hence, it is important to consider the effect of multiple parameters on earth-atmosphere interactions to obtain reliable SWC data (e.g. Todd and Mays, 2005; Das and Mohanty, 2006; Rab et al., 2011; Shiri et al., 2017; Li et al., 2023).

Nevertheless, accurate information on SWC quantity and variations is crucial for various activities, including hydrological processes analysis such as rainfall-runoff modeling, agricultural practices such as rice field irrigation management, infiltration modeling, evapotranspiration estimation, drainage systems design, groundwater recharge, and pollution. Given the global decline in available freshwater resources, precision agriculture and in particular precision irrigation seeks to enhance food production without exacerbating the strain on these resources, through the adoption of different strategies seeking to increase water use efficiency (WUE) (Lakhiar et al., 2024). Precision irrigation implied a design and management to maintain soil moisture at optimal values to fully satisfy crop water demand, eliminating runoff and minimizing deep percolation (Daccache et al., 2015). The intrinsic spatial heterogeneity of soil characteristics needs to be assessed in precision irrigation management, therefore a precise division of the sub-field areas with similar characteristics, defining site-specific management zones (SSMZ), is crucial for effective irrigation management (De Lara et al., 2018; Reyes et al., 2019); being SWC the key parameter to estimate soil water availability (Fleming et al., 1999; Lakhiar et al., 2024). Under this context, the meticulous monitoring of soil moisture variability over space and time is essential for site-specific application of irrigation used in precise irrigation (Hedley and Yule, 2009).

Both surface and subsurface drip irrigation systems have been found to be effective in promoting aerobic rice production. These techniques have proven to be viable alternatives to traditional continuous flooding irrigation, as they reduce irrigation water requirements while maintaining or slightly reducing rice production. Additionally, drip irrigation increases water productivity, as demonstrated by various studies (e.g. Mallareddy et al., 2023; Arbat et al., 2020). However, the irrigation management process of aerobic rice requires soil water status to remain above a certain threshold to avoid yield reduction, as noted by Bouman et al. (2007). Therefore, it is critical to monitor SWC in sub-layers to effectively manage drip irrigation. This requires regular monitoring of soil water content as described by Zotarelli et al. (2011) and Miller et al. (2014). SWC values are obtained through direct and indirect methods. Direct methods involve in-situ measurements of SWC using various techniques, while indirect methods estimate SWC based on other easily measured soil parameters (Shiri et al., 2017). The former methods are widely used in research and practical applications, while the latter ones have gained attention in recent years due to their cost-effectiveness and ease of use. Gravimetric approach, neutron probes, Gamma-ray and tensiometers are among direct methods that can be applied for point measurement of SWC (Todd and Mays, 2005; Rani et al., 2022) and, therefore, require the set-up of sensors network to obtain representative SWC values capturing the spatial variability of the soil hydraulic characteristics that allow to apply variable irrigation rates in each SSMZ of the field (Yari et al., 2017). Beside the practical issues, the set-up and maintenance of such networks is also expensive. Among soil water sensors to allow irrigation management, capacitance sensors are one of the most reliable options at a reasonable price (Domínguez-Niño et al., 2019), even though they are still too expensive to be extensively used in commercial rice fields. Indirect methods, on the other hand, obtain the SWC values by using other corresponding parameters; approaches including optical, thermal and microwave remote sensing are examples of those. (e.g. Liu et al., 2003; Maltese et al., 2013; Ahmed et al., 2011). In line with developing machine learning applications in various disciplines, numerous studies have used machine learning/data driven approaches for simulating SWC using soil parameters. Among others, Coopersmith et al. (2016) applied machine learning (ML) techniques for estimating near surface soil moisture values. Karandish and Šimůnek (2016) used three ML techniques for SWC simulation and stated that the neuro-fuzzy (NF) and support vector machine (SVM) techniques produced accurate results under water stress conditions. Matei et al. (2017) employed SVM and neural networks (NN) for real time prediction of SWC and proved the capabilities of these models. Prasad et al. (2018a) developed an extreme learning machine integrated with ensemble empirical mode decomposition for prediction of SWC and confirmed its ability in this regard. Prasad et al. (2018b) introduced ensemble committee machine learning approach to forecast upper and lower layer soil moisture values. Achieng (2019) simulated water retention curves of sandy and loamy soils by using SVM, NN and deep learning methods and found that SVM with radial basis kernel function outperformed the rest of the applied models. Pekel (2020) applied decision tree regression to simulate SWC. He et al. (2022) applied SVM-chaotic whale optimization algorithm to estimate soil moisture in maize crop and confirmed the hybrid model superiority over the SVM model. Zounemat Kermani et al. (2022) compared individual ML techniques with ensemble Bayesian model averaging method to simulate SWC and confirmed the superiority of the ensemble model in both simulating and forecasting SWC. Majumdar et al. (2023) applied hybrid SVM-grey wolf optimization algorithm to model SWC in a rice field and confirmed its ability to predict SWC. Other formerly published studies have also used NN, multivariable linear regression (MLR), SVM, neuro-fuzzy and genetic programming (GP) methods to simulate SWC in different soil types (e.g. Borgesen and Schaap, 2005; Merdun et al., 2006; Shiri et al., 2017). The inputs used to simulate SWC in these studies were various soil parameters (e.g. bulk density, texture, mean particle-size diameter, geometric standard deviation of soil particles, organic matter) as well as some meteorological parameters (e.g. air temperature).

The literature review revealed that various methods have yielded distinct outcomes across specific soil and climate conditions, leading to variable results for both model types and study areas. Meanwhile, application of gene expression programming (GEP) for SWC is scarce in literature. Despite this, there is scarce research addressing SWC simulation in rice crops. Cultivation of rice requires great amounts of water, representing the largest volume of water used in agriculture at a global scale. In response to the growing challenges posed by water freshwater scarcity and increased competition for its use, innovative approaches are being developed to optimize irrigation systems and enhance crop management techniques. To deal with the limited water availability, subsurface drip irrigation emerges as a promising solution. This system not only contributes to irrigation water saving compared to traditional flooding irrigation but also facilitates the expansion of rice cultivation into fields with reduced water availability compared to the ones within traditional rice cultivation areas. Hence, accurate information about SWC under rice cultivation equipped by subsurface irrigation systems would contribute to maximize its benefits. The objective of the present study is to evaluate the efficacy of the heuristic data-driven gene-expression programming (GEP) approach to simulate soil water content at three distinct soil depths by implementing various data management strategies. Genetic programming proves to be particularly viable in scenarios where the relationships among relevant variables are not well understood, theoretical analysis is hindered by assumptions, rendering

Map source: https://www.freeworldmaps.net/europe/mediterranean/physical.html

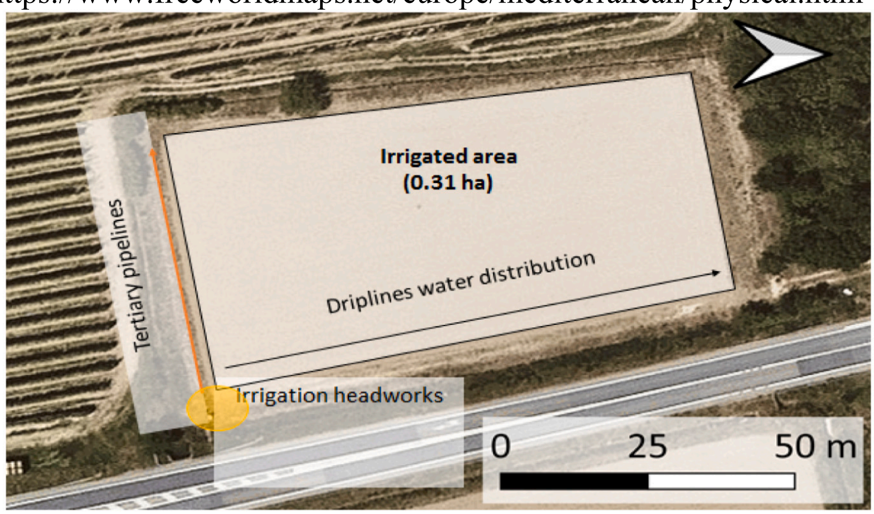


Fig. 1. Localization of the study area.

Table 1
General soil characteristics of the soil in the study area.

Soil classification: typic calcixerept (soil taxonomy 1975–1999)
Characteristics of the soil horizons • 0–26 cm; Ap; Silty-clay-loam texture (6.7 % sand, 57.4 % silt, 35.9 % clay); pH (1:2.5) = 8.1; EC (1:5) = 0.3 dS/m; Organic matter = 3.8 %; Water retention at −10 kPa = 40 %; Water retention at -33 kPa = 36 % (vol.); Water retention at −1500 kPa = 20 % (vol.); bulk density 1.35 (±0.08 kg m^{-3}). • 26–62 cm; Bwkn1; Silty-clay-loam texture (3.8 % sand, 61.8 % silt, 34.4 % clay); pH (1:2.5) = 8.2; EC (1:5) = 0.7 dS/m; Organic matter = 1.3 %; Water retention at −10 kPa = 39 %; Water retention at -33 kPa = 31 % (vol.); Water retention at −1500 kPa = 18 % (vol.); bulk density 1.55 (±0.06 kg m^{-3}). • 62–90 cm; Bwkn2; Silty-loam texture (4.3 % sand, 71.8 % silt, 23.9 % clay); pH (1:2.5) = 8.4; EC (1:5) = 0.7 dS/m; Organic matter = 0.8 %; Water retention at -33 kPa = 29 % (vol.); Water retention at −1500 kPa = 13 % (vol.).

their solutions of limited utility, minor enhancements in performance are routinely quantifiable, and the extensive volume of computer-readable data necessitates laborious processing (Banzhaf et al., 1998). As mentioned, interactions between various parameters in soil medium are complex, so GEP would be a suitable method for mapping such complex relations between SWC and other parameters. To evaluate the performance of the developed GEP models, the robust k-fold testing data assignment method is employed, determining the value of "k" considering several factors. Finally, groundwater and meteorological data was applied for the first time to simulate SWC at three different soil depths in a subsurface drip irrigated rice field.

2. Materials and methods

2.1. Experimental site

The Baix Ter area includes a total surface area devoted to rice of approximately 1000 ha. Located in northeast Spain, the river Ter suffers from a strong competition for water use between different economic sectors. On the other hand, the production of rice plays a pivotal role in the cultivation of the renowned "Arròs de Pals" quality label, recognized for its exceptional culinary attributes. The subsurface drip irrigation experimental plot of 0.38 ha, located at the borders of the traditional rice cultivation area (42° 0′ 12″ N, 3° 8′ 46″ E). The irrigation water source was from an irrigation channel network, the same one as the one used to irrigate flooded fields (Fig. 1). The soil texture of the plot was silty clay with 6.5 % of sand, 53.4 % of silt and 40.1 % of clay. Table 1 presents the general soil characteristics of the study site. The irrigation configuration consisted in driplines buried at 0.15 m and spaced 0.75 m. Medium-walled driplines (1.00 mm), 14.20 mm internal diameter with integrated pressure-compensating and anti-suction emitters of 1 L h^{-1} spaced 0.3 m apart, were used. Temperature, rainfall, atmospheric pressure, air humidity, global radiation, and wind speed were measured with an automated meteorological station, located at few meters of the filed. Reference evapotranspiration (ET_o) was determined using FAO-56 method.

Volumetric soil water contents were measured at 0.10, 0.20 and 0.30 m soil depth in the mid-way between driplines, replicated in two separate field locations 9 m apart. During 2020 and 2021 moisture contents were measured every hour using 10HS (Meter Group, Pullman, United States), and during 2023 every 30 min using Drill & Drop (Sentek, Stepney, Australia) capacity probes. A total of 27,744 volumetric soil water contents were measured; the average value for each day, soil depth, and position relative to the drip line was considered in the study.

The volumetric soil water content at each of these 3 depths will be indicated as SWC1, SWC2 and SWC3, respectively. The three soil depths were selected because, based on experimental observations, most of the rice root system was located between 0.00 and 0.20 m depth, therefore, soil moisture at 0.10 and 0.20 m was useful for irrigation management, and an increase in soil moisture at 0.30 m might indicate water losses due to deep drainage or a rise in the water table above this depth.

The 10HS capacitance probes were calibrated for 0.00–0.10, 0.10–0.20 and 0.20–0.30 m soil depths. Following laboratory calibration method described by (Parvin and Degre, 2016) and Drill & Drop probes following the procedure described by (Paltineanu and Starr, 1997), the two probes showed an accuracy of at least ±3 % vol. in all soil depth ranges. Additionally, the soil volumetric water content measured with both probes were compared showing differences of less than 2 %.

The general irrigation criterion was based on maintaining soil water potential around −10 kPa at 0.00–0.10 m depth during the irrigation season; this potential is considered field capacity for rice. For the particular soil conditions this would equivale to maintaining soil water contents approximately over 0.40 m^3 m^{-3} during the irrigation season, except when the machinery needed to enter in the field, that irrigation was cut off during some days to prevent soil compaction, or at the end of the season, when the grain is already developed and the rice does not need more irrigation. Groundwater level was measured on a weekly basis during the irrigation campaign using a manual measuring tape with an accuracy of ±5 mm. During the cropping season the water table level ranged from 0.31 to 0.95 m in 2020; 0.37 to 1.05 m in 2021 and 0.34 to 1.18 m in 2023.

Rice development is characterized by different growing stages that can be defined using Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie (BBCH) scale (Yuzugullu et al., 2017). Using direct field observations and based on BBCH scale the following stages were distinguished: germination (GERM), tillering (TILL), panicle formation (PAN), flowering (FLOW), milk grain (MILK) and maturity (MAT). The length of each period slightly changed in the different growing seasons, being the number of observations for the years 2020, 2021 and 2023, the following: GERM (26, 26 and 36); TILL (35, 45 and 30); PAN (25, 25 and 15); FLOW (15, 10 and 15); MILK (10, 15 and 15); mat (44, 25 and 27).

2.2. Gene expression programming (GEP)

Genetic Programming (GP) (Koza, 1992) applies the principles of the Genetic Algorithm (GA) (Goldberg, 1989) by using a "parse tree" structure to search for solutions. Gene Expression Programming (GEP) is similar to GAs and GP in that it selects the parse trees based on fitness values and introduces genetic variation through one or more genetic operators (Ferreira, 2006). Two major advantages of GEP, compared to GP, are the simplicity of the chromosomes (they are linear, compact, small in size and easy to manipulate genetically) and the ability of the expression trees to be modified according to fitness values. This is because the expression trees are the representations of their related chromosomes.

The process of applying Gene Expression Programming (GEP) to estimate SWC values involves five major steps. Firstly, the root mean square error (RMSE) with parsimony pressure is chosen as the fitness function, based on trial and error process. Next, the terminal and function sets are defined. The terminal set includes input variables which are used to feed the GEP models for estimating SWC values. The function set includes four basic operators (+,-,*,/) as well as ln(x),

e^x, $\sqrt{}$, $\sqrt[3]{}$, x^2 and x^3 as advised by Shiri et al. (2017). In the third step, the chromosomal architecture is defined, with a length of head = 8 and a number of genes per chromosome = 3. The head size plays a pivotal role in determining the complexity of the initial set of solutions generated by the learning algorithm of GEP. More specifically, the head size refers to the number of nodes in the first layer of the model, which generates the possible solutions to the problem at hand. Each of these nodes comprises a distinct combination of functions and input terminals that correspond to the variables and constants in the data. Subsequently, the learning algorithm explores various combinations of functions and terminals in an iterative process, refining the model based on the accuracy of each solution. Thus, the head size parameter significantly impacts the speed and accuracy of the learning process. Further, the number of genes per chromosome is a crucial factor to consider in GEP. This is because the number of genes determines the number of terms in your model, with each gene coding for a different sub-expression tree. Although one could use a single large gene to develop intricate models, breaking the chromosome into smaller, more manageable units gives an advantage to the learning process. By using multi-genic chromosomes, the learning process becomes more efficient, and it becomes easier to discover elegant and efficient models. Information on these values selection is given in the next section. The fourth step involves selecting the linking function. In GEP, when the number of genes is more than one, it's important to select an appropriate linking function that can connect the mathematical terms encoded in each gene. Addition linking functions are typically used in such studies based on previous research. Finally, the basic GEP operators are fixed per run to establish the GEP models. GEP with random numerical constants was employed in this study that uses an additional gene domain to encode random numerical constants, allowing for direct manipulation of these constants during learning procedure. As a result, this algorithm is equipped with additional genetic operators, such as RNC mutation, Dc mutation, Dc inversion, and Dc-IS transpositions, which are particularly designed to deal with random numerical constants. The GEP-RNC algorithm's learning process requires a careful selection and adjustment of numerical constants, particularly if one is dealing with constants directly. Regarding the number of constants per gene parameter, it is advisable to utilize a small set of 10 distinct constants per gene. This approach provides ample diversity for most problems without substantially increasing structural complexity (Ferreira, 2006).

2.3. Modeling protocol

The first step with establishing the SWC models is defining the appropriate inputs matrix. Different approaches can be used for defining the input variables of models, e.g. regression analysis, Gamma test, etc., from which analyzing the physical interactions of the target parameter and governing factors would be efficient due to reflecting the real-world mechanisms controlling the studied phenomenon. As mentioned,

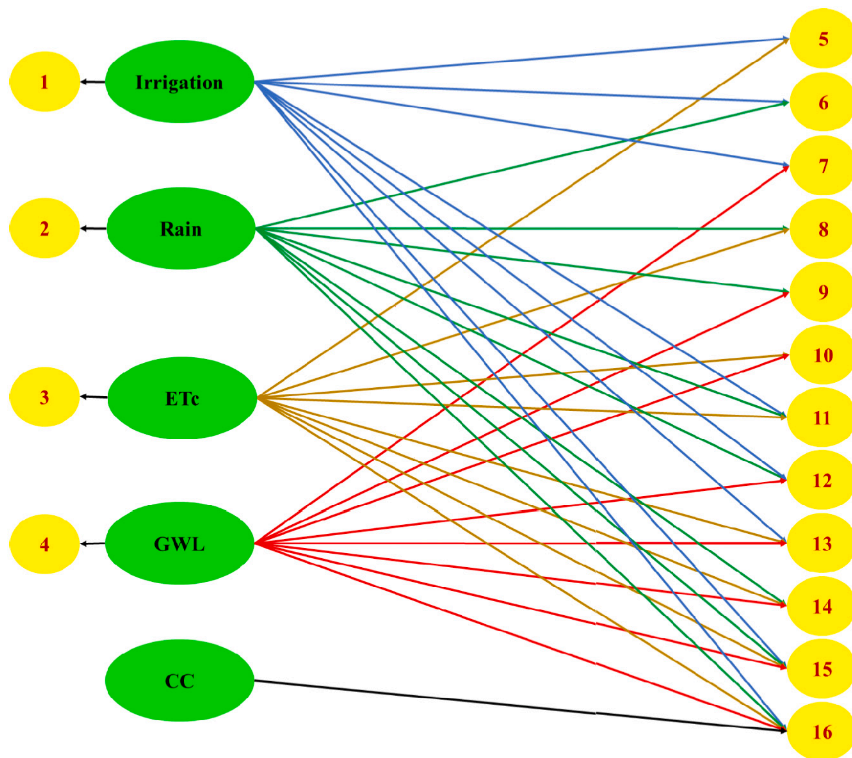


Fig. 2. Adopted input configurations (green points: input variables; yellow points: input configuration number)

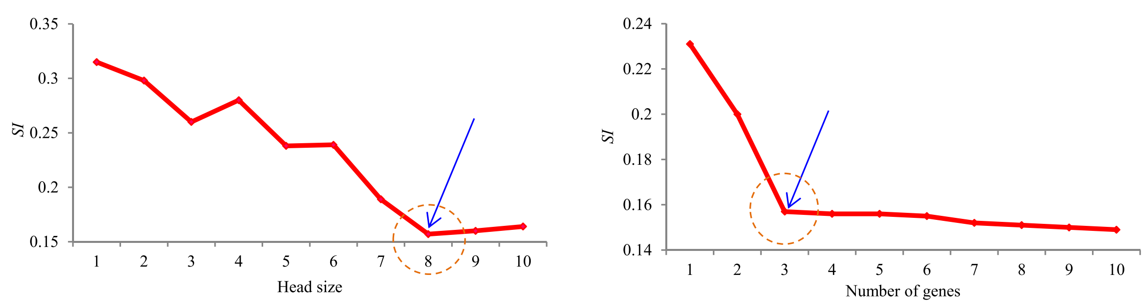


Fig. 3. Evaluation of the effect of head size and genes number on model accuracy (example case of GEP2 for SWC1 estimation).

previous studies have mostly used soil properties and hydraulic parameters to simulate SWC, although measuring some of hydraulic parameters for irrigation practitioners is quite expensive and time consuming. The present study aimed to simulate soil water content (SWC) based on agrometeorological data, crop development and groundwater level (GWL). To achieve this, several input variables were considered, namely, irrigation water depth (Ir), rainfall (R), actual evapotranspiration (ETc), GWL, and canopy cover (CC). Using these input variables, 16 input combinations were defined to simulate SWC at three soil depths (SWC1 for 0–12.5 cm; SWC2 for 12.5-25 cm and SWC3 for 25–30 cm), as can be seen in Fig. 2. Further, four input combinations were considered comprising SWC values (inputs 17–20), where SWC1 was used as input for SWC2 and SWC3 values (inputs 17&18, respectively). Then, SWC2 values was used as input for SWC3 estimation (input19) and finally, SWC1 and SWC2 values were simultaneously used for SWC3 simulation in input 20.

Once the input variables were identified, the second step would be adopting a proper strategy for assigning data matrix to GEP models. The most robust k-fold testing approach was utilized here by defining "k" based on temporal scales. By using k-fold testing, all available patters are divided by "k" parts, then one portion is reserved as testing patterns each time and the remaining patterns are used for training the models. The process is repeated "k" times so that all available events of input matrix are incorporated in both the training and testing stages. This is a robust and through data assigning method because all data are used as "unseen data" one time, so the model learn all the data (Karimi et al., 2018). A key point with applying this method is through defining the "k". Usually, this value is defined based on the specification of each study, so temporal or spatial criteria might be used to consider proper "k" value under different conditions. Temporal criterion was used here for "k" definition due to temporal base of the gathered data, and the available patterns were divided into three parts ($k = 3$) at the first step, considering the years of study period (2020, 2021 & 2023), i.e. strategy1. In the second strategy, the growing stages were used for partitioning the data, so the patterns were divided into six parts ($k = 6$) comprising growing stages of all three years (Germination (GERM), Tillage (TILL), Panicle formation (PAN), Flowering (FLOW), Milk grain (MILK) & Maturity (MAT) stages). Considering 20 input combinations of SWC modeling at three distinct depths, total 468 training-testing procedures were carried out in the present study. Two statistical indices, namely, the correlation coefficient

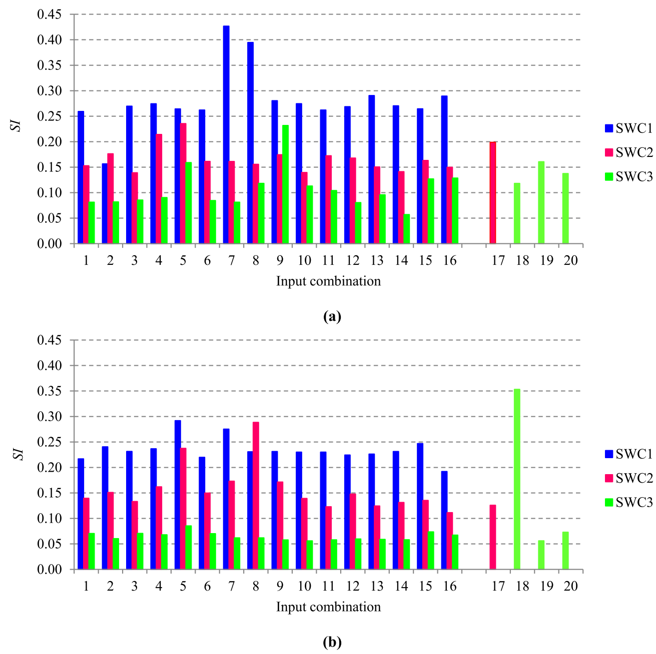


Fig. 4. Global *SI* values of the applied input combinations: a) strategy1, b) strategy2 (for input combinations 1–16, the highest SI values of the models are relatively higher for strategy1)

(R), and the weighted *RMSE* i.e. scatter index (*SI*) were used for evaluating and comparing the established models. These indices were computed for each test stage as well as entire period:

$$R = \left[\frac{\sum_{i=1}^{n}(SWC_m - \overline{SWC}_m)(SWC_o - \overline{SWC}_o)}{\sqrt{\sum_{i=1}^{n}(SWC_m - \overline{SWC}_m)^2 \sum_{i=1}^{n}(SWC_o - \overline{SWC}_o)^2}}\right] \quad (1)$$

$$SI = \frac{RMSE}{\overline{SWC}_o} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(SWC_m - SWC_o)^2}}{\overline{SWC}_o} \quad (2)$$

Where, SWC_m and SWC_o stand for the simulated and measured soil water content, respectively. $\overline{SWC}_m$ and $\overline{SWC}_o$ show the mean simulated and observed SWC values. N is the patterns number (data samples). Higher *R* and lower *SI* values show the better model performance and vice-versa.

3. Results and discussions

3.1. General assessment

GEP establishing steps were fixed on the basis of the outcomes of the previous studies as well as the trial and error process as mentioned in Section 2.2. As an example, the trial and error procedures for obtaining the optimum head size (length of head) and genes number is illustrated in Fig. 3 for SWC1 simulation with GEP2. From the figure it is revealed that the lowest *SI* values were obtained when the head size was fixed at 8 and the genes number was limited to 3. Even though increasing the genes number beyond 3 resulted in improved performance accuracy, the improvement was found to be insignificant. Therefore, in order to minimize the computational cost and model complexity, the genes number was ultimately selected as 3. Based on trial and error processes for sample models, these values were fixed per run of each GEP model in all studies cases.

Fig. 4 shows the global *SI* values of the applied models for both the

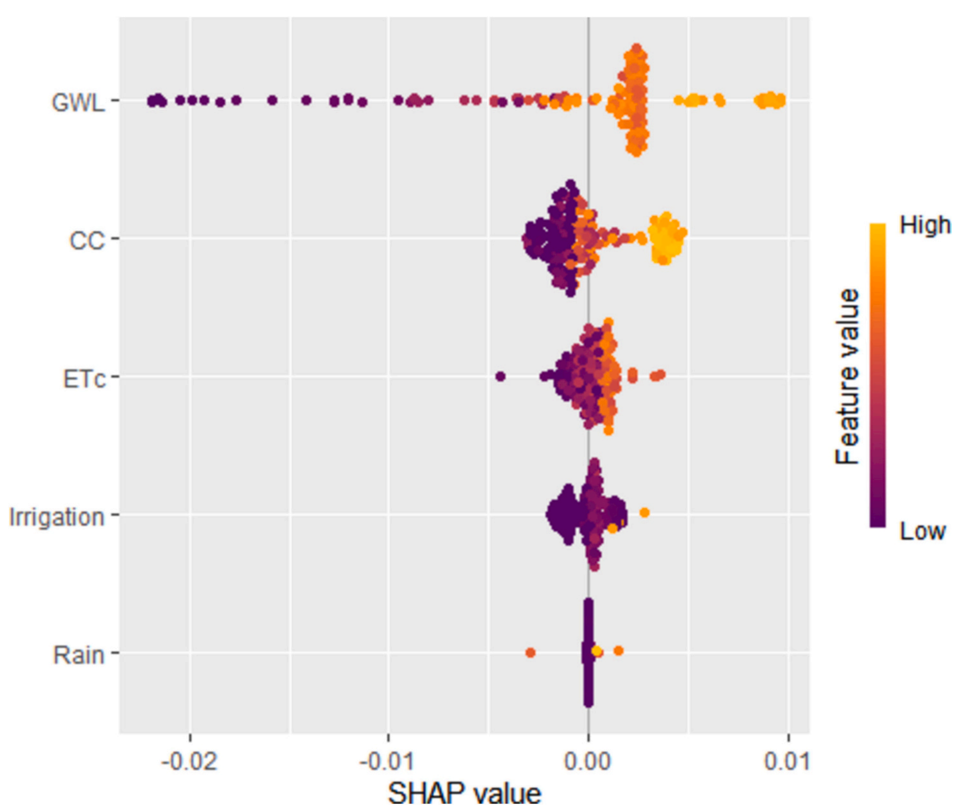


Fig. 5. SHAP values of the utilized input variables for SWC3 simulation with strategy1 (orange color shows direct positive effect and violet color shows direct adverse effect).

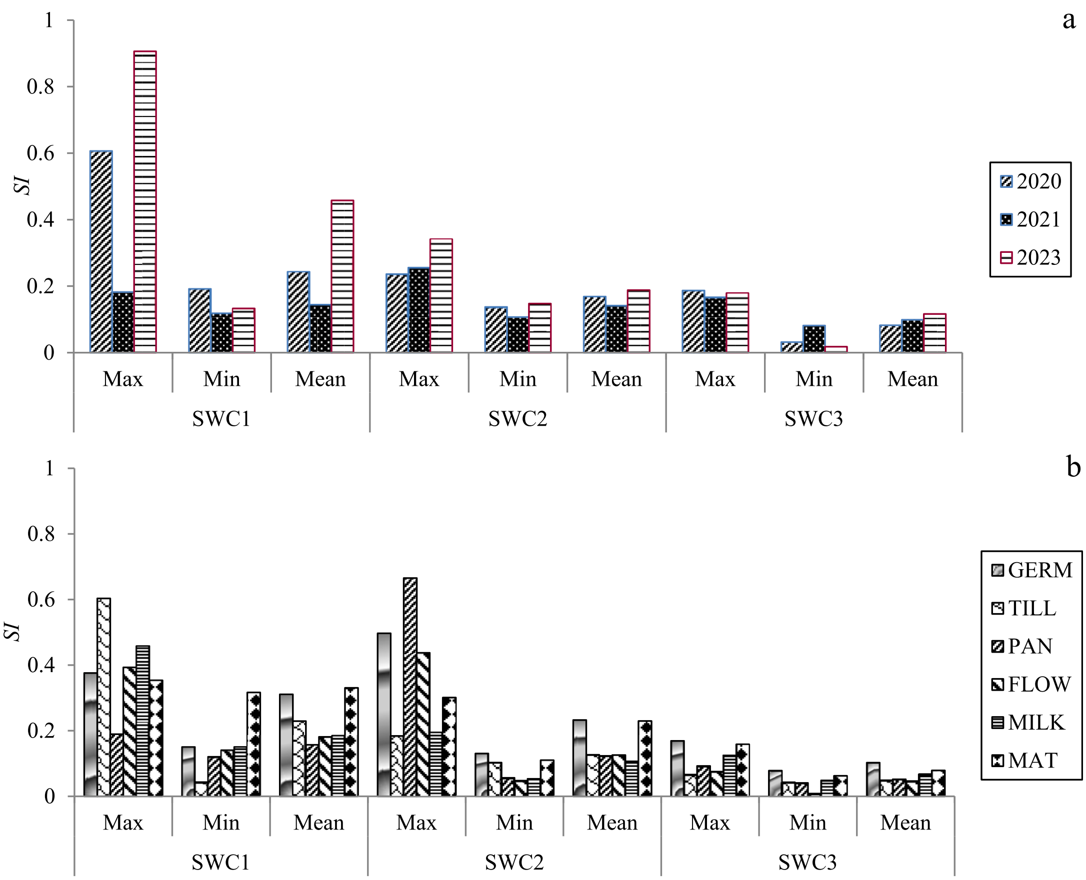


Fig. 6. Comparing *SI* variation of GEP models per test stage: a) stratgey1, b) strategy2 (SWC1 and SWC3 simulation models presented the highest and lowest *SI* values, respectively.

adopted strategies of "k" defining. These values have been obtained through merging all simulated-observed values of the target parameters (SWC1, SWC2 and SWC3) of all test stages in a unique matrix. Comparing the adopted strategies, the average *SI* values (average of all used input combinations) were obtained as 0.256, 0.166 and 0.107 for SWC1, SWC2 and SWC3 simulation when the first strategy was applied, while they were 0.218, 0.157 and 0.065 for the strategy2 (defining "k" based on the stages of growing season). A same trend was also observed for the *R* and maximum and minimum *SI* values (not presented here) and the differences between the strategies were more obvious for maximum *SI* values. Although the average *SI* differences between the strategies are low, adopting strategy2 has produced more accurate results for all three target parameters. As mentioned, strategy2 has ability to align the data partitioning process with the physiological growth stages of the rice crop, which may ensures that the utilized data partitions for training and testing phase are more representative of the available data matrix in real world applications. Nonetheless, SWC3 simulation models produced better results than the other two models.

Comparing the global performance of the GEP models per adopted strategy, GEP2 and GEP14 produced the most accurate results for SWC1 and SWC3 simulation in strategy1, respectively, while GEP3 and GEP10 were best models for SWC2 modeling. The better performance of GEP2 in SWC1 simulation might be linked to direct effect of rainfall on surface layer soil moisture. Further, as the second soil layer is subjected to the effect of GWL rising as well as crop root water uptake (in term of water evapotranspiration), better performance of GEP3 and GEP10 models might be justified (Fig. 4). Finally, the models comprising irrigation water volume, groundwater level and rainfall as the soil moisture sources, produced the better results than the rest of the models for SWC2 and SWC3, among which GEP14 had slightly better performance. Nevertheless, owing to lower error values, the single-input GEP4 model that uses only GWL records can be considered as a simple model of SWC3 simulation. It should be however noted that these outcomes have been obtained based on global performance analysis of the models, so per test year analysis might revise these concluding remarks.

For the second strategy, however, the model ranking was not similar to the previous cases and GEP16 presented the highest performance accuracy for SWC1 and SWC2, while the best SWC3 simulations were obtained by GEP10 and GEP19 models. Better performance of GEP16 in SWC1 and SWC2 estimation might be attributed to inclusion of all input parameters in model development. For SWC3, although GEP10 presented the best performance among the models using soil and weather parameters, a range of models might be considered as optimal models as *SI* differences between them was very low. GEP19 that uses SWC2 values for SWC3 estimation also gave better results that may show the strong relation between the moisture content of two successive soil layers. This would be a great step forward because application of these models would allow reducing the number of soil water sensors.

Digging to GEP performance comparison for three targets simulation, it is seen that the models developed for SWC3 estimation presented the lowest *SI* values compared to SWC1 and SWC2 models. Better performance of GEP at the third layer might be due to the inherent stability and reduced susceptibility to environmental noise at deeper layer. Nonetheless, SWC at deeper layers are more influenced by seasonal variations of other parameters, so their variations at diurnal scale are low, facilitating their simulation.

Fig. 5 presented the SHAP values of the inputs for SWC3 simulation. The values have been computed for strategy1 as an illustrative example. SHAP is a powerful tool for visual description of the variable importance. It is seen that GWL has made the highest impact on SWC3, followed by CC, ETc, irrigation water and rainfall. Negative values of SHAP push the simulated SWC values to lower magnitudes, while positive

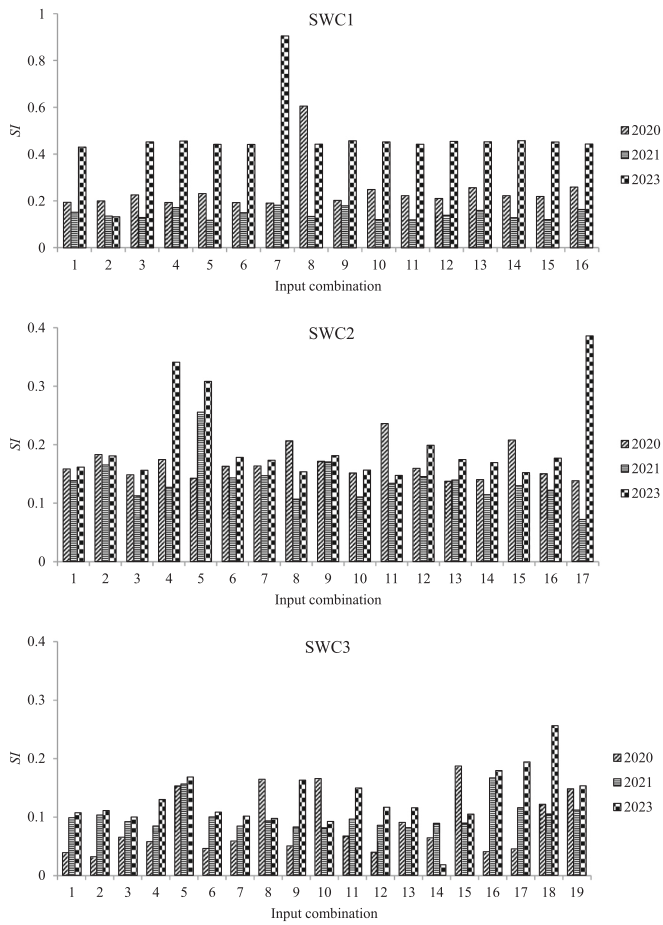


Fig. 7. Temporal variations of *SI* values of the GEP models per test stage (year)-strategy1 (Obvious variations can be seen for test years in most of the input combinations)

SHAP values push them into higher magnitudes. For GWL, the higher SHAP values are generally positive (orange color) denoting that it has direct positive effect on SWC3. For CC, irrigation and rainfall parameters, the lower values (violet points) have both direct and adverse effect on it. Analysis of the correlation values between the target and input parameters showed that crop evapotranspiration and canopy cover presented the highest correlation value with the SWC2. Similar trend was also observed for irrigation water volume (higher correlation with SWC2) that can be due to water infiltration into soil layers, and the higher water extraction rate between 0.10 and 0.20 m.

Finally, as could be anticipated, groundwater level showed the higher correlation with SWC3 that is related to rising/falling regime of water table, which occurred in the soil third sub-layer. Further, analysis of some statistical characteristics of the SWC data showed that besides lower values of SWC3 compared to SWC1 and SWC2, it has the lowest standard deviation and coefficient of variation (CV) values (lower range of soil water change at 30 cm depth) that can explain lower *SI* values of GEP models for SWC3 simulation. In cases where water table is deep, the lower moisture content of deep layer might be attributed to lower organic matter in this layer, as can be seen from Table 1 (Zheng et al., 2015), nevertheless in the present study water table was very shallow and had a great influence in the deeper layer, being its moisture contents higher than in shallower layers during most of the irrigation season. On the other hand, minimum values of CV were observed for the deeper layer (20-30 cm) in all cases, confirming the outcomes reported by Wang et al. (2014). This might be expected because less water changes were observed in the deeper soil layers due to the shallow groundwater level and shallow root system of the rice as well as buffering effect, soil texture impact (deeper layer may contain finer particles that hold water more efficiently) and diminished compaction effects. It is seen that the topsoil layer exhibited a significant variation in soil water content (SWC) due to alternate wetting and drying conditions, indicating a strong interaction between soil, atmosphere and irrigation. However, as the depth of SWC measurements increased, the influence of weather gradually diminished, resulting in a smaller range of SWC variation. Therefore, the depth of observations plays a crucial role in determining the extent of SWC variation and modeling performance, as discussed by Li et al. (2021) and Shaikh and Sahoo (2024). Finally, higher variations of soil moisture content at the first layer may be due to keeping a constant matric potential at 10 cm soil depth as irrigation managing criterion by using sub-surface irrigation system at the studied field and also that the shallow layer of the soil is more exposed to evaporation, root water extraction and rain. This means that, in contrary to the traditional methods where irrigation is managed through maintaining a water depth depending on the rice growing stage, the goal by the present experiments was keeping a constant matric potential by using subsurface irrigation method. So, various amounts of water should be supplied by irrigation system regardless of the rice growth stage that makes higher variances for soil moisture values at the first layer (SWC1).

3.2. Analysis per test stage

3.2.1. Strategy1

With applying this strategy, the "k" value of k-fold testing was selected based on the number of available years (k = 3), so the models were trained and tested by leaving each time the available patterns of one year for model testing. Fig. 6 illustrates the maximum, minimum and mean *SI* values of GEP models for each test year (2020, 2021 and 2023). It is seen that the GEP model provided the most accurate results for SWC3 simulation as discussed before. Nonetheless, the mean and maximum *SI* values of the models were higher for 2023 test year than 2020 and 2021, summarizing that the GEP models presented more accurate results in these two years for all three target parameters. Differences among *SI* values are higher for SWC1 models. Analyzing the statistical characteristics of soil moisture data showed that SWC1 had the highest standard deviation (=12.7) and CV (=0.385) values for 2023 data, which were about double- and triple the amounts of these values for the previous years. Such scattering in data quantity might include difficulties for interpolating the target values through machine learning techniques (Kazemi et al., 2020).

A temporal analysis of *SI* variations per test year of each input combination was additionally performed as can be seen in Fig. 7. First, obvious variations of model performance is seen among test years for all input combinations presenting that the modeling accuracy is fluctuated among test years and emphasizing on the necessity of a through data scanning, i.e. k-fold testing adaptation for model assessment. With very minor exceptions, error statistics were higher for test year 2023. Further, with minor exceptions, *R* values of the models for 2023 were lower than 2020 and 2021. Next, at the first look, no general conclusion might be obtained for deciding on the most accurate model based on yearly performance indicator. For instance, with simulation of SWC1, while GEP7 presented the lowest *SI* value in test year 2020, GEP6 and GEP3 gave the most accurate results for 2021 and 2023 test stages, respectively. A similar trend can be observed for SWC2 and SWC3 models and *R* values (not presented here), too. However, differences among performance indicators of the best models at each test year are generally low so that the average *SI* differences among best models of each target parameter were 0.048, 0.026 and 0.042 for SWC1, SWC2 and SWC3 models, respectively. The average *R* differences for the same cases were 0.091, 0.102, and 0.115, respectively.

Comparing per test year results with global model assessment (Section 3.1) shows some agreement on best model selection: While GEP2 is ranked as the first best model of SWC1 simulation based on global assessment, GEP7, GEP6 and GEP2 models were selected as most

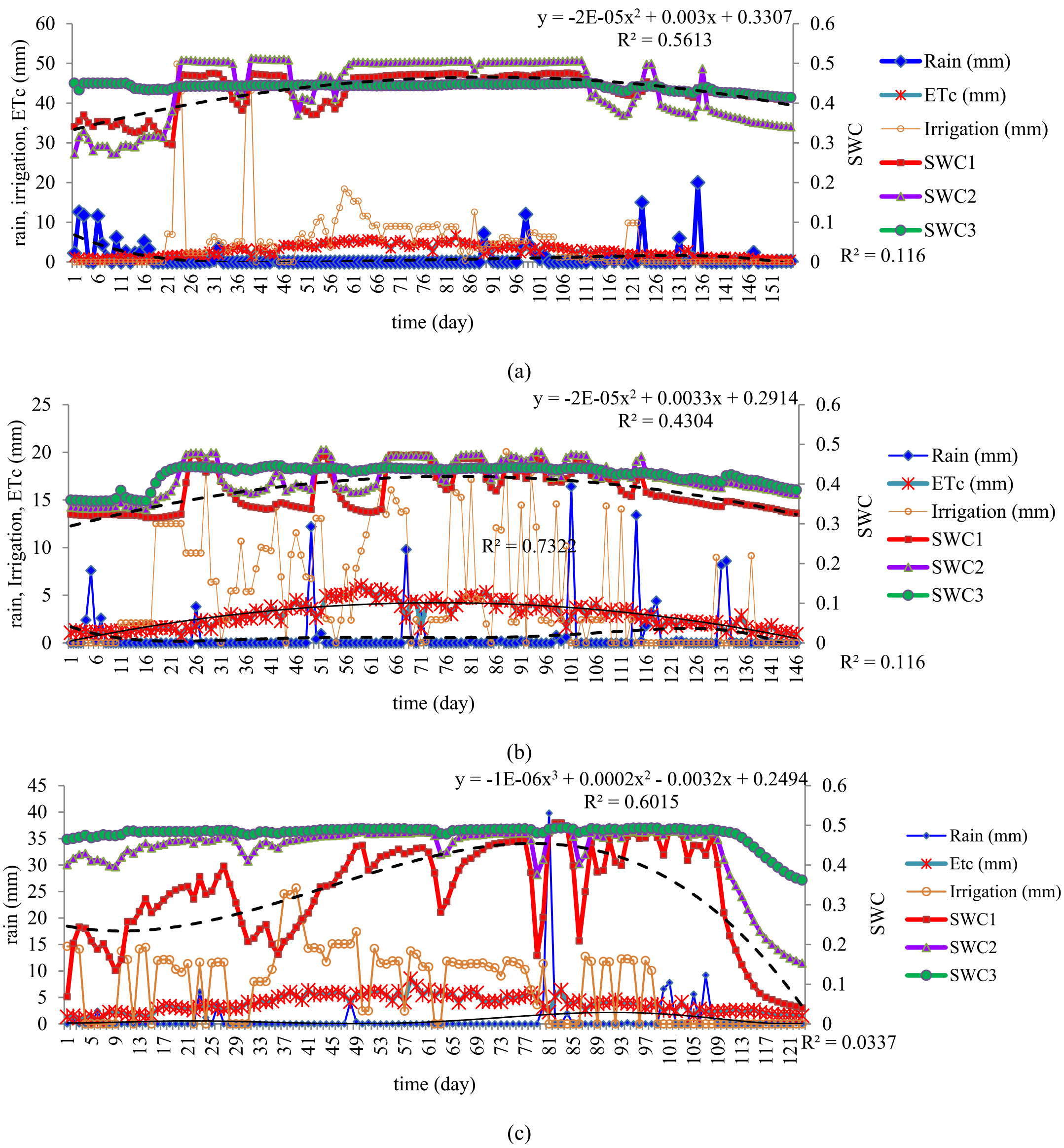


Fig. 8. Soil water content and rainfall changes at different soil depths (SWC1, SWC2 and SWC3) during 2020 (a), 2021 (b) and 2023 (c) rice growing seasons: SWC3 values presented the lowest variations in all periods among three studied targets. SWC1 and SWC2 variations are affected by the changes in irrigation water, rainfall amount, while SWC3 is not obviously affected.

Table 2
Key uncertainty parameters of the best GEP models.

Target	Model	e	SD	WUB	PEI	
					Lower bound	Upper bound
Stratgey1						
SWC1	GEP2	−0.0317	0.0592	0.1161	−0.1477	0.0842
SWC2	GEP3	0.0034	0.0603	0.1182	−0.1148	0.1216
SWC2	GEP10	0.0012	0.0608	0.1191	−0.1179	0.1203
SWC3	GEP14	0.0080	0.0271	0.0532	−0.0452	0.0612
Strategy2						
SWC1	GEP16	0.0002	0.0802	0.1572	−0.1569	0.1575
SWC2	GEP16	−0.0009	0.0505	0.0990	−0.1080	0.0900
SWC3	GEP10	−0.0010	0.0270	0.0524	−0.0532	0.0516
	GEP19	0.0022	0.0258	0.0506	−0.0484	0.0528

Note: e: average error (simulation-observation); SD: standard deviation of error; WUB: width of uncertainty bound, PEI: 95 % interval prediction error.

accurate models for test years 2020, 2021 and 2023, respectively. It is noticeable that rain was included in all these three models. Besides 2023 best models ranking that is in complete agreement with global model assessment (Section 3.1.), differences between model performances were low for the other cases, so it might be stated that both analysis (global and per test stage) confirm each other for SWC1 simulation. Similar conclusions can be given for the rest of target parameters, too. In order to obtain a thorough and accurate assessment of a soil moisture simulation model's capability, it is crucial to consider the differences between overall and temporal performance accuracy. Some models exhibit significant disparities between these two metrics, indicating that a k-fold testing data assignment approach should be adopted. This will ensure that the model's performance is evaluated comprehensively across different time periods, and that any potential weaknesses or strengths are identified and addressed accordingly. By taking this approach, researchers and practitioners can confidently assess the suitability of different soil moisture simulation models for their specific needs, and make informed decisions about which models to use for their applications. However, in order to effectively model SWC in rice field soil, it is crucial to determine the most appropriate method for defining the “k” through the process of assigning data. This topic will be explored in greater detail in the upcoming section.

3.2.2. Strategy2

The “k” of k-fold testing was defined based on different growing stages of the rice and each time the available patterns of one stage were used for model testing. As the amount of moisture fluctuates throughout the various growing stages throughout the year, its variance is not consistent. Hence, examining the applied models through scanning the available patterns based on growing stages (as test set) would be necessary and relevant. Fig. 6 compares *SI* ranges of the GEP models adopted by strategy2. Again, the applied models provided the highest performance accuracy for SWC3 simulation among three considered targets. The *SI* ranges presented in Fig. 4 show that models' performance accuracy varied among different testing patterns (growing stages), especially for SWC1 and SWC2 simulations. While the lowest average *SI* (and the highest average *R*) was obtained for panicle test stage in SWC1 GEP models, milk and flowering stages presented the lowest average *SI* for SWC2 and SWC3 models, respectively. Nonetheless, the highest maximum, minimum and mean *SI* values were reported for germination test stage in all cases. Analyzing per growing stage of SWC variance

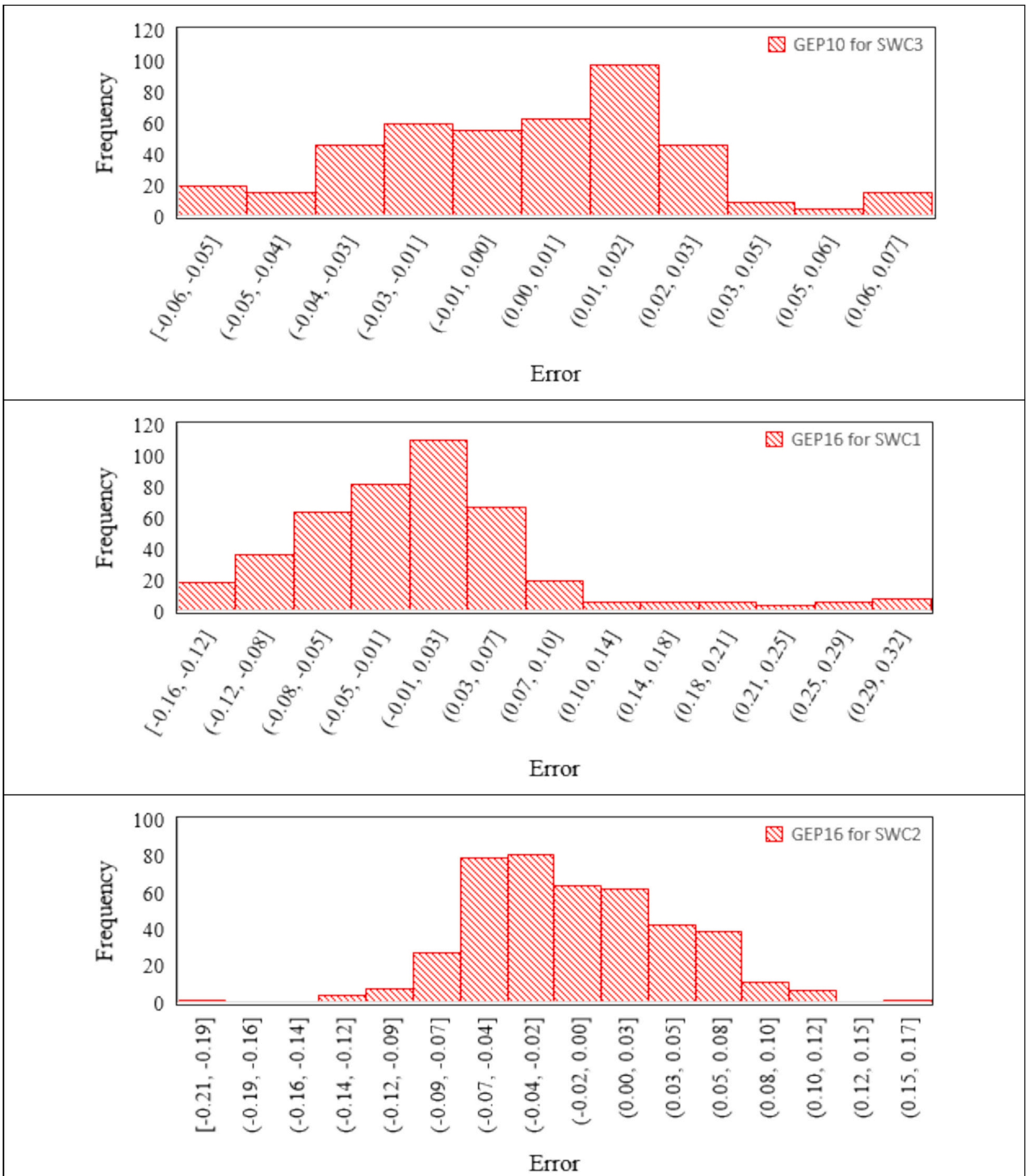


Fig. 9. Error histograms of the best gene expression programming (GEP) models.

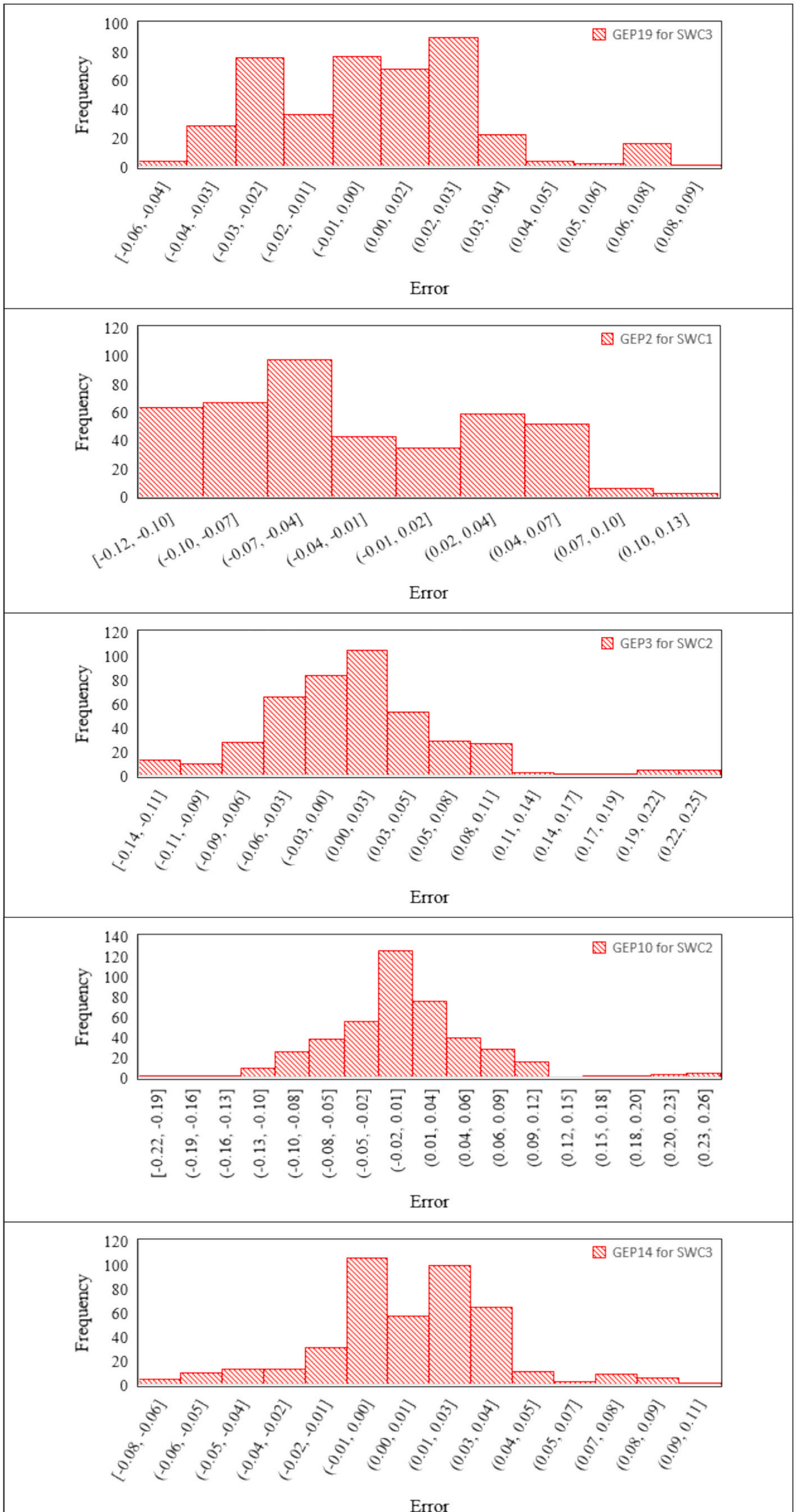


Fig. 9. (*continued*).

showed that the highest CV belonged to germination stage at all three studied depths, with the maximum correlation values with the recharge amount (irrigation+rainfall). An additional analysis was also performed to assess the performance of the GEP models per test stage. Consequently, double-input GEP5 (comprising irrigation and ETc as input variables) presented the most accurate results for germination, panicle initiation and milk stages in SWC1 simulation, while quintuple-input GEP16 gave the best results for tillering, flowering and milk stages in the same case. In SWC2 and SWC3 simulations, however, no general conclusion might be obtained for selecting a best model and each testing stage belonged to different input combination when the best results were assessed.

3.3. Further discussions

The study aimed to evaluate the effectiveness of the GEP model in estimating the SWC of rice fields, using various combinations of input parameters. The outcomes of the study indicated that the GEP method was able to capture the intricate nonlinear relationships between the input variables and accurately simulate SWC values of a *Typic Calcixerept* soil.

One of the key findings of the study was that rainfall had a significant impact on the modeling of SWC1 when the first strategy was applied. The GEP2 model that relied exclusively on rainfall values produced the most accurate results. This was expected because the first soil layer has the greatest interactions with rainfall. However, when testing the second strategy, which uses the growing season as test stages, it cannot solely simulate the SWC at the first top layer. Dai et al. (2022) argued that multiple small precipitation events may not have the same impact on soil moisture as a single large precipitation event that produces the same total rainfall. To support this argument, a visual analysis of temporal SWC and rainfall variations is depicted in Fig. 8. The figure illustrates that SWC3 is not affected by the amount of rainfall, while SWC1 and SWC2 show variations according to the variations in rainfall amount. However, these variations do not completely conform to each other, suggesting that there may be other factors at play that affect SWC1 and SWC2 in addition to rainfall. Upon scrutinizing the data for the example year of 2021, it can be observed that SWC1 experiences minimal fluctuations between days 1–24, with a variation range of less than 0.009. During this period, five rainfall events were recorded, with the maximum value being 7.6 mm. From day 25 onwards, SWC1 starts to increase despite the absence of any rainfall events. The ETc values presented no sharp variations during this period. However, it should be noted that a substantial amount of irrigation water (12.5 mm) was added to the soil on day 18 (germination period), which could have contributed to the increase in SWC1. On day 49 (tillering period with higher ETc values), a single rainfall event occurred with a considerable amount of 12.20 mm, resulting in a noticeable increase in SWC1. Similarly, on day 101, a rainfall event with a magnitude of 16.40 mm was recorded. This increase in SWC1 was not due to any irrigation water input, hence indicating that rainfall was the primary factor behind the increase. Similar statements can be given for the years 2020 and 2023. These observations confirm the statements given by Dai et al. (2022). Further, the lags between the rainfall occurrence and SWC1 increase might be due to the crop interception and the buffering effect of soil cover (rice) as discussed by Niu et al. (2015).

Attending to SWC variations with relation to ETc, it is seen from Fig. 8 that ETc values follow a second order polynomial with relatively higher correlation values in all three studies years. Among target parameters, SWC3 variations do not follow ETc variations, which is consistent with the fact that rice has a very shallow root system. According to experimental observations, rice roots rarely extend beyond 0.20 m, meaning soil moisture below this depth was practically unaffected by transpiration. The significant dampening of transpiration and evaporation in the third layer, along with the influence of the water table at this depth, maintained stable moisture levels throughout the observed periods (Fig. 8). This greater stability in moisture levels at the third layer may explain the higher accuracy of the models for this depth, as indicated by the SI ranges presented in Fig. 4.

SWC1 and SWC2 presented similar variations with ETc with the higher similarity for SWC2. So, it is evident that variations of SWC with respect to ETc are diminished by increasing the target soil depth that is in line with the conclusions obtained by Shaikh and Sahoo (2024). On the other hand, Xi et al. (2023) stated that the effect of ETc on SWC variations is small for high frequencies (weekly, monthly or seasonal), while it shows higher impact for lower frequency (seasonal to annual) that is in cope with the present statements. In case of the irrigation water amount variations, the highest similarity was found with SWC2, which might be due to the depth of emitter installation that has the most interact with the second sub layer.

It is worth noting that, while single-input models yielded optimal results for SWC1 and SWC2 simulations under strategy1, multi-input models proved more effective when strategy2 was employed. This could be attributed to the temporal changes in the parameters used. As illustrated in Fig. 8, SWC values exhibit a polynomial trend throughout the year. For example, for SWC1 in 2021, a second-order polynomial trend line was fitted. Similarly, fifth-order and second-order polynomial trend lines could be fitted to the rainfall and ET_c time series, respectively. Therefore, when testing the models using a full year of data, single-input models may yield satisfactory results. It is important to note that the behavior of other input parameters e.g. irrigation water content and GWL exhibit polynomial trend lines with higher correlation values over a year. It may be assumed that the utilization of these variables as the sole input parameter could lead to better predictions for SWC1. However, as these variables exhibit lower interactions with soil water at the first layer, the single input model consisting of these variables fails to surpass rain-based models for SWC1 estimation with the first strategy. Nevertheless, these parameters exhibit higher linear correlation values with SWC1 than rainfall amount, but the correlation analysis only captures linear relations and ignores some essential features of nonlinear relations. Summarizing, although lower linear correlation values were observed between rain and SWC1, it is seen that the surface soil water content is highly affected by precipitation amount, confirming the conclusions reported by Yao et al. (2013), Duan et al. (2016) and Cheng et al. (2020).

Upon comparing the simulated values of three distinct soil depths, it was observed that the estimated soil water content was lower than the measured values in the surface soil layer (SWC1). One potential explanation for this discrepancy is that we did not account for the capillary rise of water from the second layer, which could be substantial when the topsoil becomes dry due to lack of water input, as noted by Jiang et al. (2008).

Digging to the adopted data management strategies, it is imperative to adopt k-fold testing data assignment and its significance should not be underestimated. This is due to the detection of apparent variations in the error statistics per test stage of k-fold testing with both the adopted strategies. Therefore, the adoption of k-fold testing data assignment is crucial in ensuring the accuracy of the statistical analysis outcomes. When comparing the performance of two different strategies for modeling, it was found that adopting the second strategy produced more accurate models for all three targets. This can be attributed to the fact that the second strategy involves better partitioning available patterns, which leads to better performance accuracy than using a single testing data matrix that includes all patterns from the entire year. This is because SWC variations among different growing stages are considerable throughout the year, and using yearly testing data matrix can reduce the accuracy of the trained GEP model.

Further, as soil water (its state, movement and quantity) involve complex and uncertain processes, it's essential to conduct an uncertainty analysis to ensure the reliability and applicability of models used for SWC estimation. So, all measured-simulated pairs of all test stages per adopted strategy were combined to make a unique matrix per strategy, then the uncertainty analysis was conducted for this global matrix. In this study, some key parameters, i.e. average prediction error (e), standard deviation of error (SD), 95 % confidence bound (WUB), and 95 % interval prediction error (PIE) of the selected models for SWC estimation. This analysis provides critical insights into the accuracy and limitations of the models. Table 2 sums up the key uncertainty parameters of the best GEP models for all targets. Analyzing the values presented in the table, it is observed that GEP has generally overestimated the SWC values when strategy1 was adopted, except for GEP2. The magnitudes of over/underestimations are obviously lower for strategy2 for SWC1 estimation, where the over/underestimations for SWC1 and SWC2 with strategy1 were approximately 15 times and 4 times greater than those with strategy2.

Comparing between the strategies it is seen that adopting second strategy is better than strategy1, confirming the statements given in the previous sections. Fig. 9 presents the error histograms of the discussed models. The figure clearly shows that error values followed the Gaussian distribution that illustrates the stability of these models in SWC estimation (Bilali et al., 2021).

It is noted that sources of uncertainty of SWC include various parameters, including soil texture (homogeneity or heterogeneity), weather data errors, characteristics of irrigation inflows/outflow and differences of infiltration rates (Chaubey et al., 1999; Feki et al., 2018; Kelly et al., 2021). For the present case, all three sources of uncertainty would be influential in making the overall SWC monitoring/estimation uncertainty. The potential for inaccuracies in soil-texture or water-flux measurements poses a risk to the precision of estimating soil water content. This, in turn, could lead to suboptimal irrigation practices by farmers, possibly impacting crop yields and the overall profitability of agricultural operations as discussed by Kelly et al. (2021). Despite increasing the irrigation cost by over-irrigation, the risk of waterlogging and drainage-related issues with the cultivated lands are general consequences of SWC underestimations, while forced deficit irrigation that leads to the reduction of the crop yield would be a result of SWC overestimations. Thus, precise irrigation or emerging machine learning-based models focusing on reducing the SWC monitoring/estimation errors are suitable approaches for providing accurate information on SWC for irrigation scheduling. Again, the values presented in Table 2 are low for both the strategies (although the second is obviously better) confirming the suitability of the proposed model for accurate and applicable SWC estimation, with lower risk of over-irrigation or deficit irrigation.

The SWC generated by the GEP model can be integrated into existing precision irrigation systems, enhancing the monitoring of soil moisture across different SSMZ in the field. This approach reduces the need for multiple measurement points. It can be applied across various irrigation systems, leveraging simulation model results to control solenoid valves when target moisture levels are reached at specific locations (Kumar et al., 2023), or by enabling site-specific irrigation based on soil water availability at different soil depth intervals, this method supports variable-rate irrigation using center pivot or lateral move systems (McCarthy et al., 2023).

4. Conclusions

Soil water content (SWC) simulation using gene expression programming (GEP) was carried out in rice cropped using subsurface drip irrigation, through. K-fold testing was applied for training and testing the models by adopting two different strategies for “k” definition, either based on complete one-year patterns (stragey1) or on growing season patterns (strategy2). Results revealed that both strategies lead to promising results in SWC simulation, with models obtained with the second strategy showing a slightly better accuracy. It was observed that rainfall is an important factor for SWC estimation in topsoil layer. The amount and intensity of rainfall were two key elements impacting SWC, confirming previously published results obtained with other methods. An uncertainty analysis confirmed the capability and stability of the developed models. The study conducted is the first to attempt to simulate soil water content (SWC) of a rice field under subsurface drip irrigation system in existing literature. While numerous studies have been carried out to simulate SWC, it was challenging to compare the findings of the present research with those of the previously published studies in terms of data handling, applied strategies, irrigation method and field characteristics. Further, it is noted that the present study focused on SWC simulation of rice filed under subsurface drip irrigation system and the conclusions might be valid for similar configurations. Further studies might be needed through evaluation of SWC under different soil cover types and irrigation systems. Nonetheless, integrating remote sensing data or other environmental datasets to improve model accuracy as well as investigation into the seasonal variability of SWC and how it correlates with crop yield would also be valuable study subjects for future.

CRediT authorship contribution statement

Jalal Shiri: Writing – original draft, Validation, Supervision, Formal analysis, Conceptualization. **Mohammad Hossein Kazemi:** Writing – review & editing, Visualization, Validation, Methodology, Formal analysis. **Sepideh Karimi:** Writing – review & editing, Methodology, Formal analysis, Data curation. **Sílvia Cufí:** Writing – review & editing, Methodology, Investigation. **Francisco Ramírez de Cartagena:** Writing – review & editing, Supervision, Conceptualization. **Jaume Pinsach:** Writing – review & editing, Project administration, Funding acquisition. **Gerard Arbat:** Writing – review & editing, Methodology, Investigation, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to express their gratitude for funding “MEDWATERICE - Towards a sustainable water use in Mediterranean rice-based agro-ecosystems” (PCI2019-103738) and “PROMEDRICE - Effective farming practices to protect water resources in rice-based agroecosystems” (PCI2023-143435) to MICIU/AEI/10.13039/501100011033 and the cofinancing of the European Union, being both projects part of the PRIMA Programme supported by the European Union. Contributions of co-author Sílvia Cufí were funded by Ministerio de Ciencia, Innovación y Universidades of Spain through an FPU predoctoral grant (FPU20/01123).

Data availability

Data will be made available on request.

References

Achieng, K.O., 2019. Modelling of soil moisture retention curve using machine learning techniques: artificial and deep neural networks vs support vector regression models. Comput. Geosci. 133, 104320.
Ahmed, A., Zhang, Y., Nichols, S., 2011. Review and evaluation of remote sensing methods for soil-moisture estimation. SPIE Rev. 2 (1), 028001.
Arbat, G., Sílvia, C., Miquel, D.R., Jaume, P., Jaume, P.B., Joan, P., de Cartagena, Francisco Ramírez, 2020. Modeling approaches for determining dripline depth and irrigation frequency of subsurface drip irrigated rice on different soil textures. Water (Switzerland) 12 (6). https://doi.org/10.3390/W12061724.
Banzhaf, W., Nordin, P., Keller, P.E., Francone, F.D., 1998. Genetic Programming. Morgan Kaufmann, San Francisco, CA (512 pp).
Bilali, A.E., Taleb, A., Brouzine, Y., 2021. Comparing four machine learning model performances in forecasting the alluvial aquifer level in a semi-arid region. J. Afr. Earth Sci. 181, 104244.
Borgesen, C.D., Schaap, M.G., 2005. Point and parameter pedotransfer functions for water retention predictions for Danish soils. Geoderma 127 (1–2), 154–167.

Bouman, B.A.M., Humphreys, E., Tuong, T.P., Barker, R., 2007. Rice and water. In: Sparks, Donald L. (Ed.), Advances in Agronomy, vol. 92. Academic Press, pp. 187–237. https://doi.org/10.1016/S0065-2113(04)92004-4.
Chaubey, I., Haan, C.T., Grunwald, S., Salisbury, J.M., 1999. Uncertainty in the model parameters due to spatial variability of rainfall. J. Hydrol. 220, 48–61.
Cheng, R.R., Chen, Q., Zhang, J., Shi, W., Li, G., Du, S., 2020. Soil moisture variations in response to precipitation in different vegetation types: a multi-year study in the loess hilly region in China. Ecohydrology 13, e2196.
Coopersmith, Evan J., Cosh, Michael H., Bell, Jesse E., Boyles, Ryan, 2016. Using machine learning to produce near surface soil moisture estimates from deeper in situ Records at U.S. climate reference network (USCRN) locations: analysis and applications to AMSR-E satellite validation. Adv. Water. Res. 98, 122–131.
Corradini, C., 2014. Soil moisture in the development of hydrological processes and its determination at different spatial scales. J. Hydrol. 516, 1–5.
Daccache, A., Knox, J.W., Weatherhead, E.K., Daneshkhah, A., Hess, T.M., 2015. Implementing precision irrigation in a humid climate – recent experiences and on-going challenges. Agric Water Manag 147, 135–143. https://doi.org/10.1016/j.agwat.2014.05.018.
Dai, L., Fu, R., Guo, X., Du, Y., Zhang, F., Cao, G., 2022. Soil moisture variations in response to precipitation across different vegetation types on the northeastern Qinghai-Tibet plateau. Front. Plant Sci. 13, 854152.
Das, N.N., Mohanty, B.P., 2006. Root zone soil moisture assessment using remote sensing and vadose zone modeling. Vadose Zone J. 5, 296–307.
De Lara, A., Khosla, R., Longchamps, L., 2018. Characterizing spatial variability in soil water content for precision irrigation management. Agronomy 8 (5), 5. https://doi.org/10.3390/agronomy8050059.
Domínguez-Niño, J., Bogena, H., Huisman, J., Schilling, B., Casadesús, J., 2019. On the accuracy of factory-calibrated low-cost soil water content sensors. Sensors 19. https://doi.org/10.3390/s19143101.
Duan, L., Huang, M., Zhang, L., 2016. Differences in hydrological responses for different vegetation types on a steep slope on the Loess Plateau, China. J. Hydrol. 537, 356–366.
Feki, M., Ravazzani, G., Cepi, A., Mancini, M., 2018. Influence of soil hydraulic variability on soil moisture simulations and irrigation scheduling in a maize field. Agric Water Manag 202, 183–194.
Ferreira, C., 2006. Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence. Springer, Berling, Heidelberg New York (478 pp).
Fleming, K.L., Westfall, D.G., Wiens, D.W., Rothe, L.E., Cipra, J.E., Heermann, D.F., 1999. Evaluating farmer developed management zone maps for precision farming. In: Proceedings of the Fourth International Conference on Precision Agriculture. John Wiley & Sons, Ltd, pp. 335–343. https://doi.org/10.2134/1999.precisionagproc4.c29.
Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison –Wesley, ReadingMA (432 pp).
He, B., Jia, B., Zhao, Y., Wang, X., Wei, M., Dietzel, R., 2022. Estimate soil moisture of maize by combining support vector machine and chaotic whale optimization algorithm. Agric Water Manag 267, 107618.
Hedley, C.B., Yule, I.J., 2009. Soil water status mapping and two variable-rate irrigation scenarios. Precis. Agric. 10 (4), 342–355. https://doi.org/10.1007/s11119-009-9119-z.
Jiang, J., Zhang, Y., Wegehenkel, M., Yu, Q., Xia, J., 2008. Estimation of soil water content and evapotranspiration from irrigated cropland on the North China plain. J. Plant Nutr. Soil Sci. 171, 751–761.
Karandish, F., Šimůnek, J., 2016. A comparison of numerical and machine-learning modeling of soil water content with limited input data. J. Hydrol. 543, 892–909.
Karimi, S., Sadaraddini, A.A., Nazemi, A.H., Xu, T., Fakherifard, A., 2018. Generalizability of gene expression programming and random forest methodologies in estimating cropland and grassland leaf area index. Comput. Electron. Agric. 144, 232–240.
Kazemi, M.H., Shiri, J., Marti, P., Majnooni-Heris, A., 2020. Assessing temporal data partitioning scenarios for estimating reference evapotranspiration with machine learning techniques in arid regions. J. Hydrol. 590, 125252.
Kelly, T.D., Foster, T., Schultz, D.M., Mieno, T., 2021. The effect of soil-moisture uncertainty on irrigation water use and farm profits. Adv. Water Resour. 154, 103982.
Koza, J.R., 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. The MIT Press, Cambridge, MA (840 pp).
Kumar, S.V., Singh, C.D., Rao, K.V.R., Kumar, M., Rajwade, Y.A., Babu, B., Singh, K., 2023. Evaluation of IoT based smart drip irrigation and ETc based system for sweet corn. Smart Agric. Technol. 5, 100248. https://doi.org/10.1016/j.atech.2023.100248.
Lakhiar, I.A., Yan, H., Zhang, C., Wang, G., He, B., Hao, B., Han, Y., Wang, B., Bao, R., Syed, T.N., Chauhdary, J.N., Rakibuzzaman, M., 2024. A review of precision irrigation water-saving technology under changing climate for enhancing water use efficiency, crop yield, and environmental footprints. Agriculture 14 (7), 7. https://doi.org/10.3390/agriculture14071141.
Li, Y., Yu, Y., Sun, R., Shen, M., Zhang, J., 2021. Simulation of soil water dynamics in a black locust plantation on the loess plateau, western Shanxi Province, China. Water 13 (9), 1213.
Li, X., Xu, X., Wang, K., Li, X., 2023. Estimation of root zone soil moisture at point scale based on soil water measurements from cosmic-ray neutron sensing in a karst catchment. Agric Water Manag 289, 108511.
Liu, W., Baret, F., Gu, X., Zhang, B., Tong, Q., Zheng, L., 2003. Evaluation of methods for soil surface moisture estimation from reflectance data. Int. J. Remote Sens. 24 (10), 2069–2083.
Majumdar, P., Mitra, S., Bhattacharya, D., 2023. Soil moisture simulation of rice using optimized support vector machine for sustainable agricultural applications. Sustain. Comput. Inform. Syst. 40, 100924.
Mallareddy, M.R.T., Subramanian, P.B., Naseeruddin, R., Nithya, N., Arulanandam, M.N.E., Choudhary, A., Murugesan, D., Elangovan, S., Padmaja, B., Vijayakumar, S., 2023. Maximizing water use efficiency in rice farming: a comprehensive review of innovative irrigation management technologies. Water 15, 1802. https://doi.org/10.3390/w15101802.
Maltese, A., Capodici, F., Ciraolo, G., La Loggia, G., 2013. Mapping soil water content under sparse vegetation and changeable sky conditions: comparison of two thermal inertia approaches. J. Appl. Remote. Sens. 7 (1), 073548.
Matei, O., Rusu, T., Petrovan, A., Mihut, G., 2017. A data mining system for real time soil moisture prediction. Process. Eng. 181, 837–844.
McCarthy, A., Foley, J., Raedts, P., Hills, J., 2023. Field evaluation of automated site-specific irrigation for cotton and perennial ryegrass using soil-water sensors and model predictive control. Agric Water Manag 277, 108098. https://doi.org/10.1016/j.agwat.2022.108098.
Merdun, H., Cinar, O., Meral, R., Apan, M., 2006. Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. Soil Tillage Res. 90 (1–2), 108–116.
Miller, Gilbert, Farahani, Hamid J., Hassell, Richard L., Khalilian, Ahmad, Adelberg, Jeffrey, Wells, Christina E., 2014. Field evaluation and performance of capacitance probes for automated drip irrigation of watermelons. Agric Water Manag 131. https://doi.org/10.1016/j.agwat.2013.09.012.
Niu, C.Y., Musa, A., Liu, Y., 2015. Analysis of soil moisture condition under different land uses in the arid region of Horqin sandy land, northern China. Solid Earth 7, 1979–2009.
Paltineanu, I.C., Starr, J.L., 1997. Real-time soil water dynamics using multisensor capacitance probes: laboratory calibration. Soil Sci. Soc. Am. J. 61 (6), 1576–1585. https://doi.org/10.2136/sssaj1997.03615995006100060006x.
Parvin, N., Degre, A., 2016. Soil-specific calibration of capacitance sensors considering clay content and bulk density. Soil Res. 54. https://doi.org/10.1071/SR15036.
Pekel, E., 2020. Estimation of soil moisture using decision tree regression. Theor. Appl. Climatol. 139 (3), 1111–1119.
Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2018a. Soil moisture forecasting by a hybrid machine learning technique: ELM integrated with ensemble empirical mode decomposition. Geoderma 330, 136–161.
Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2018b. Ensemble committee-based data intelligent approach for generating soil moisture forecasts with multivariate hydro-meteorological predictors. Soil Tillage Res. 181, 63–81.
Rab, M.A., Chandra, S., Fisher, P.D., Robinson, N.J., Kitching, M., aumann, C.D., Imho, M., 2011. Modelling and prediction of soil water contents at field capacity and permanent wilting point of dryland cropping soils. Soil. Res. 49 (5), 389–407.
Rani, A., Kumar, N., Kumar, J., Kumar, J., Sinha, N.K., 2022. Machine Learning for Soil Moisture Assessment. Deep Learning for Sustainable Agriculture, Elsevier, pp. 143–168. Chapter6.
Reyes, J., Wendroth, O., Matocha, C., Zhu, J., 2019. Delineating site-specific management zones and evaluating soil water temporal dynamics in a Farmer's field in Kentucky. Vadose Zone J. 18 (1), 180143. https://doi.org/10.2136/vzj2018.07.0143.
Shaikh, J., Sahoo, S., 2024. Effect of evapotranspiration on soil moisture dynamics in top surface layer of a loamy land in climate change condition. Acta. Hort. Regitec. 27, 6–14.
Shiri, J., Keshavarzi, A., Kisi, O., Karimi, S., 2017. Using soil easily measured parameters for estimating soil water capacity: soft computing approaches. Comput. Electron. Agric. 141, 327–339.
Todd, D.K., Mays, L.W., 2005. Groundwater Hydrology, 3rd edition. John Wiley and sons Inc. (636P).
Wang, X.Y., Zhang, W.J., Wang, Z.Q., Liu, X.P., Wang, S.F., 2014. Soil moisture status under deep-rooted and shallow-rooted vegetation in the semiarid area of loess plateau in China. Pol. J. Environ. Stud. 23 (2), 511–520.
Xi, X., Zhuang, Q., Kim, S., Gentine, P., 2023. Evaluating the effects of precipitation and evapotranspiration on soil moisture variability within CMIP5 using SMAP and ERA5 data. Water Resour. Res. 59, e2022WR034225.
Yao, S.X., Zhao, C.C., Zhang, T.H., Liu, X.P., 2013. Response of the soil water content of mobile dunes to precipitation patterns in Inner Mongolia, northern China. J. Arid Environ. 97, 92–98.
Yari, A., Madramootoo, C., Woods, S., Adamchuk, V., Huang, H.-H., 2017. Assessment of field spatial and temporal variabilities to delineate site-specific management zones for variable-rate irrigation. J. Irrig. Drain. Eng. 143. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001222.
Yuzugullu, O., Marelli, S., Erten, E., Sudret, B., Hajnsek, I., 2017. Determining rice growth stage with X-Band SAR: a metamodel based inversion. Remote Sens. 9 (5), 5. https://doi.org/10.3390/rs9050460.
Zheng, H., Gao, J., Teng, Y., Feng, C., Tian, M., 2015. Temporal variations in soil moisture for three typical vegetation types in Inner Mongolia, Northern China. Plos One 10 (3), e0118964.
Zotarelli, L., Dukes, M.D., Scholberg, J.M.S., Femminella, K., Muñoz-Carpena, R., 2011. Irrigation scheduling for green bell peppers using capacitance soil moisture sensors. J. Irrig. Drain. Eng. 137 (2), 73–81.
Zounemat Kermani, M., Golestanni Kermani, S., Alizamir, M., Fadaee, M., 2022. Soil moisture simulation using individual versus ensemble soft computing models. Int. J. Environ. Sci. Technol. 19, 10089–10104.