Computers and Electronics in Agriculture 218 (2024) 108746

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Estimation of pore-water electrical conductivity in soilless tomatoes cultivation using an interpretable machine learning model

Mirko Sodini [a,b,*], Sonia Cacini [a], Alejandra Navarro [c], Silvia Traversari [d,e], Daniele Massa [a,c]

[a] CREA Research Centre for Vegetables and Ornamental Crops, Council for Agricultural Research and Economics, Via dei Fiori 8, 51017 Pescia, Italy
[b] Department of Agricultural, Food, Environmental and Animal Sciences, University of Udine, Via delle Scienze 206, 33100 Udine, Italy
[c] CREA Research Centre for Vegetables and Ornamental Crops, Council for Agricultural Research and Economics, Via Cavalleggeri 25, 84098 Pontecagnano, Italy
[d] Research Institute on Terrestrial Ecosystems (IRET), National Research Council (CNR), Via Giuseppe Moruzzi 1, 56124 Pisa, Italy
[e] National Biodiversity Future Center (NBFC), 90133 Palermo, Italy

ARTICLE INFO

Keywords:
Generalized additive model
Irrigation
bulk EC conversion
SHapley Additive exPlanations
XGBoost model

ABSTRACT

Soilless culture is widely adopted for improving produce quality and yield and increasing input efficiency. Most of the benefits potentially achievable in soilless systems are possible through precise and continuous management and adjustment of plant nutrition. Under operational conditions, the electrical conductivity (EC) is the main driving parameter leading fertigation strategies, but its measure in the drainage water can be not completely representative of the root zone in the growing medium. Nowadays low-cost sensors can be adopted to measure bulk EC (EC_b) in the substrate. The Hilhorst equation is commonly used to convert the EC_b into pore-water EC (EC_w). This equation is widely calibrated for soil cultivation, but unable to perform properly for soilless substrate with high moisture content and water permittivity. In this work, two cultivation cycles of cherry tomato, managed in a closed-loop soilless system, were used to calibrate and validate two alternative models to the above equation (i.e., generalized additive model - GAM, and extreme gradient boost model - XGBoost). The models predicted EC_w from the EC_b recorded by substrate sensors. Plants were grown in rockwool using two different strategies for nutrient solution refill achieving different EC_w trends during the cultivation. The Hilhorst equation confirmed its unsuitability for EC_w prediction in soilless systems. EC_w prediction through GAM was not satisfying at low and high EC_w values. XGBoost was the most suitable model for EC_w estimation, particularly at extreme EC values.

1. Introduction

A continuous monitoring of the electrical conductivity (EC) of nutrient solution is crucial to adequately manage open- and closed-loop soilless systems (Ahn and Son, 2022, Massa et al., 2020). In fact, low EC in the root zone may limit nutrient absorption and plant growth, reducing the quality of many vegetable species. Indeed, a moderate salinity of substrate has shown positive effects on the organoleptic quality and nutritional parameters of fruit-bearing crops, also reducing the accumulation of undesired molecules such as nitrates in leaf vegetables (Germano et al., 2022, Rouphael et al., 2018). However, high concentrations of salts at root level may cause physiological disorders in most of cultivated plants such as stomata closure, early flowering, water status imbalance, and qualitative and quantitative alterations. Furthermore, the presence of saline ions in the root zone can also cause nutrient deficiency impairing the normal absorption of other ions (Massa and Melito, 2019, Munns and Tester, 2008, Navarro et al., 2007).

The EC is influenced by both nutrient and non-nutrient ions. Therefore, to ensure a proper balance between the concentration of nutrients in the fertigation solution and their actual uptake rate for optimal plant growth and development, it would be necessary to know exactly the ionic composition of the nutrient solution in the root zone. However, EC is the primary parameter monitored for managing the fertigation system under operational (commercial) conditions (Massa et al., 2020, Ahn and Son, 2019, Incrocci and Massa, 2017). In fact, EC monitoring allows the early detection of the accumulation or depletion of soluble salts in the recirculating nutrient solution, due to fluctuations in nutrient absorption related to crop evolution (Ahn and Son, 2022, Massa et al., 2020). The refill nutrient solution can then be regulated by adjustments of the dilution factor of the concentrated stock solution

* Corresponding author at: Department of Agricultural, Food, Environmental and Animal Sciences, University of Udine, Via delle Scienze 206, 33100 Udine, Italy.
E-mail address: mirko.sodini@uniud.it (M. Sodini).

https://doi.org/10.1016/j.compag.2024.108746
Received 21 June 2023; Received in revised form 7 February 2024; Accepted 12 February 2024
Available online 16 February 2024

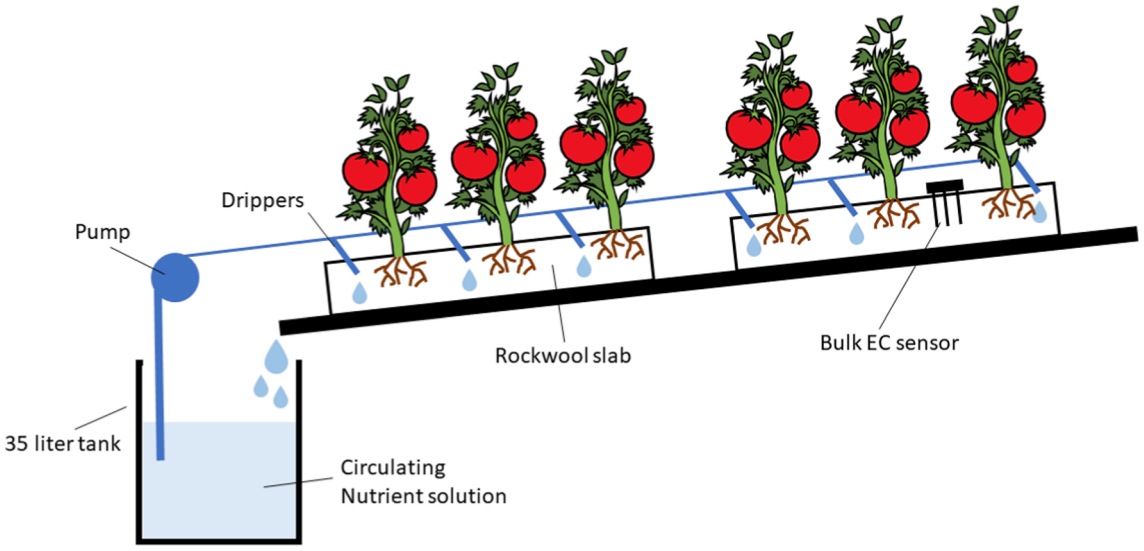

Fig. 1. Schematic representation of the cultivation bench system.

through fertigation devices (Massa et al., 2011). Notwithstanding, the tuning of fertilizer injection cannot prevent the accumulation of some undesired elements (e.g., saline ions) naturally present in the irrigation water at concentrations higher than actual plant needs. Massa et al. (2020) reported two different EC-based strategies for managing the nutrient solution in closed-loop systems: i) the refill nutrient solution is opportunely modified to achieve a target EC level in the root zone; in this case, the continuous dilution of added fertilizers, combined with the parallel accumulation of undesired ions, will lead over time to a depletion of nutrients up to the minimum concentration tolerated by the crop; ii) the refill nutrient solution has always a constant EC and composition; this strategy will lead to an EC increase in the root zone until the maximum EC value tolerated by the crop. In both cases, when the critical condition occurs, the circulating solution is discharged and replaced, completely or partly, with a newly prepared nutrient solution. The choice between the two strategies mainly depends on the type of crop and the irrigation water quality (Massa et al., 2020, Ahn and Son, 2019, Incrocci et al., 2017).

The EC can be forecasted in closed loops by prediction models (Moon et al., 2018) and/or measured by water sensors in the drainage tank or substrate sensors directly placed in the root zone (Massa et al., 2020; Incrocci et al., 2017). Under operational conditions, the measurements made in the drainage tank could indeed differ from the root zone EC. This discrepancy can be attributed to various factors, with the type of substrate and the irrigation system and scheduling being particularly relevant (Venezia et al., 2022, Incrocci et al., 2006). Substrate sensors, which contemporarily measure EC, water content (namely water permittivity), and temperature, are nowadays widely used in agriculture (Hardie, 2020). However, these sensors do not directly measure EC of the circulating solution in the root zone but rather the bulk EC of the volume portion where the sensor is placed (EC_b). This measurement is influenced by water content, salinity, and temperature (Kargas and Kerkides, 2010). Generally, EC_b is representative of the circulating solution EC *in situ* and can be used to have an indication of the root zone salinity. However, when salinity increases and/or the water content varies significantly, the difference between the two measures may significantly increase (Lim et al., 2017) causing uncertainty in the operational fertigation management of soilless systems (Massa et al., 2020; Ahn and Son, 2019).

Electrical conductivity of the solution circulating in the pores of the substrate is defined pore-water electrical conductivity (EC_w) and can be obtained by soil elution (Corwin and Yemoto, 2020; Wright, 1986) or direct soil water extraction (Lim et al., 2017). Alternatively, EC_w can be estimated by mathematical models; the most used is the Hilhorst equation (Hilhorst, 2000). Substrate temperature (T_s), permittivity (ε_w), and EC_b are the main driving variables of this equation. Many commercial sensors, e.g., GS3 (METER Group Inc., Pullman, WA, USA) or WET (Delta-T Devices Ltd, Cambridge, UK), measure the parameters used in the Hilhorst equation to estimate EC_w. However, these measures are not always reliable and depend on the environmental conditions such as soil/substrate moisture, temperature, and salinity (Bañón et al., 2021, Kocárek and Kodešová, 2012, Kargas and Kerkides, 2012, Rosenbaum et al., 2011).

The aim of this study was to identify the most performing model for the EC_w prediction from the GS3 and later models (such as the TEROS 12 sensor, compatible with GS3) sensor outputs (Ferrarezi et al., 2020) in a soilless system, comparing the Hilhorst model with two models well known to capture complex non-linear relationships in data: generalized additive model (GAM), and extreme gradient boost model (XGBoost). Two greenhouse trials with different EC management strategies were established using tomato plants to test and calibrate the models. Cherry tomato was chosen for its tolerance to EC variation and level in the root zone and its commercial relevance among other soilless crops. Indeed, soilless tomato is normally grown with nutrient solution EC between 2.5 and 3.5 dS m^{-1} and cherry tomato performs well even at higher EC level (3–7 dS m^{-1}) showing positive effect on produced quality without any reduction in yield (Lu et al., 2022, Petersen et al., 2015). For example, EC above 4.5 dS m^{-1} has been found to increase sensory characteristics and sugar content (Cliff et al., 2012), lycopene and other antioxidant molecules (De Pascale et al., 2001, Wu and Kubota 2008).

2. Material and methods

2.1. Greenhouse experiment

2.1.1. Plant material and growing conditions

This study was carried out in a greenhouse located at the Research Centre for Vegetable and Ornamental Crops of the Council for Agricultural Research and Economics (Pescia, Tuscany, Italy, latitude 43.54 N, longitude 10.42E) under typical Mediterranean climate conditions. The experiment was repeated in two different experimental trials: spring-summer (2021) and autumn–winter (2021–2022). Greenhouse temperature was kept between 13 and 34 °C by a cooling system during summer, and a heating system during winter. Climatic parameters, i.e., air temperature, air relative humidity (RH), and global radiation were monitored with an environmental station (OPI system, EVJA srl., NA,

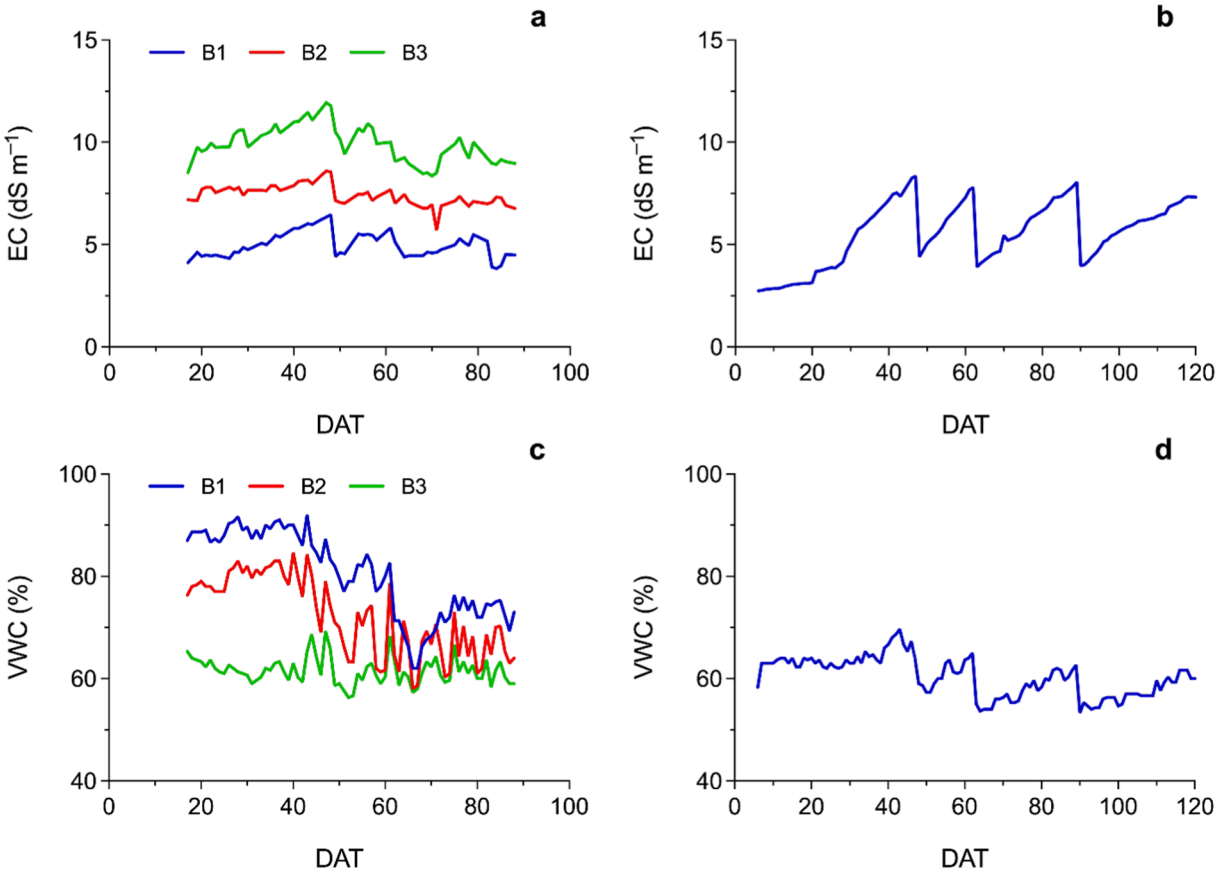

Fig. 2. Electrical conductivity (EC) values measured daily in the recirculating nutrient solution through a portable conductimetry (i.e., pore-water EC) and volumetric water content (VWC) values measured through the GS3 sensor. Data correspond to the single EC and WC values for the B1, B2, and B3 treatments (4.2, 7.5, and 9.0 dS m^{-1}) in the first experimental trial (a, c) or to the average value of the three benches in the second experimental trial (b, d). DAT = days after transplant.

Italy) Cherry tomato seedlings (*cv*. 'Genio', HM. Clause SAS) with 4/5 expanded leaves were transplanted in rockwool cubes (10 × 10 cm^2) and irrigated with half-strength nutrient solution (the composition of the nutrient solution in mM was: 12.00 N-NO_3^-, 1.00 N-NH_4^+, 1.50 P^{2-}, 10.00 K^+, 5.00 Ca^{2+}, 2.50 Mg^{2+}, 0.65Na^+, 5.66 S-SO_4^{2-}, 0.50Cl^- for macronutrient, and in μM: 15.0 Fe^{2+}, 20.0B^-, 1.0 Cu^{2+}, 5.0 Zn^{2+}, 10.0 Mn^{2+}, 1.0 Mo^{2+} for micronutrients; the EC was 2.72 dS m^{-1}) for one week to stimulate root penetration. Then, rockwool cubes were inserted in rockwool slabs (1 m × 0.15 m × 0.75 m), previously fully soaked with nutrient solution. The greenhouse consisted of six cultivation benches each organized in a separate closed loop soilless system. Each bench had 10 slabs and a drainage tank for collecting the excess of nutrient solution drained out from the substrate. The volume of the tank was 35 L over a total volume (tank plus slabs) of 110 L for each closed system. Three plants were transplanted on each slab for a final plant density of 3 plants m^{-2}. The nutrient solution absorbed by plants was continuously replenished in the drainage tank through water level sensors to keep the total nutrient solution volume as constant as possible in each cultivation bench (Fig. 1). Daily irrigation was provided by a drip irrigation system (10 drippers per slab) from the drainage tank, frequency and duration were scheduled to have a high leaching fraction (30–40 % depending on day time), however water was recirculated before sampling 3 times the volume of the drainage tank (which roughly corresponded to 100 % the total water volume of each bench), ensuring greater homogeneity between the composition of nutrient solution in the slabs and tank.

Plants were grown until the harvest of the sixth truss of fruits in both the cycles; the spring summer cycle lasted 90 d after transplant while the autumn–winter cycle lasted 148 d after transplant.

2.1.2. *Electrical conductivity management*

The EC of the nutrient solution was managed differently in the two cultivation cycles. In the spring-summer cycle, EC had three different salinity levels (three benches) kept fairly constant during the cultivation cycle. The initial nominal EC levels, i.e., 4.2, 7.5, and 9.0 dS m^{-1} (B1, B2, and B3 treatments, respectively), were obtained adding 15, 45, and 75 mmol L^{-1} of NaCl to the previously described nutrient solution. The nutrient solution absorbed by plants was reintegrated in the drainage tank with fresh nutrient solution with a full amount of nutrients: corresponding to the crop nutrient uptake concentration. No additional NaCl was added, maintaining a concentration equivalent to that found in the irrigation water (0.65 mM). This expedient was adopted to prevent NaCl accumulation and nutrient accumulation or depletion in the recirculating nutrient solution thus resulting in EC levels very close, on average, to the initial values throughout the cultivation cycle.(Fig. 2a). In the autumn–winter cycle, no NaCl was added to the initial nutrient solution. The nutrient solution absorbed by plants was replaced in the drainage tank by fresh nutrient solution with 10 mM (on average). Such an addition of NaCl to the irrigation water was causing a progressive increase in EC up to a maximum value (roughly 8 dS m^{-1}) that was then restored to the initial minimum value (roughly 4 dS m^{-1}) after flushing the nutrient solution (Fig. 2b). This treatment was applied in triplicate and, in terms of management techniques for the refill of the nutrient solution, was representing an operational (commercial) management (Massa et al., 2020). However, in the Mediterranean commercial greenhouses, salt-tolerant cherry tomato varieties are commonly grown hydroponically with EC reaching 7.0–7.5 dS m^{-1} in specific stages of the cultivation cycle to obtain high quality berries. As matter of fact, the tested variety did not show difference between 4.5 and 7.5 dS m^{-1} in previous experiments (unpublished data).

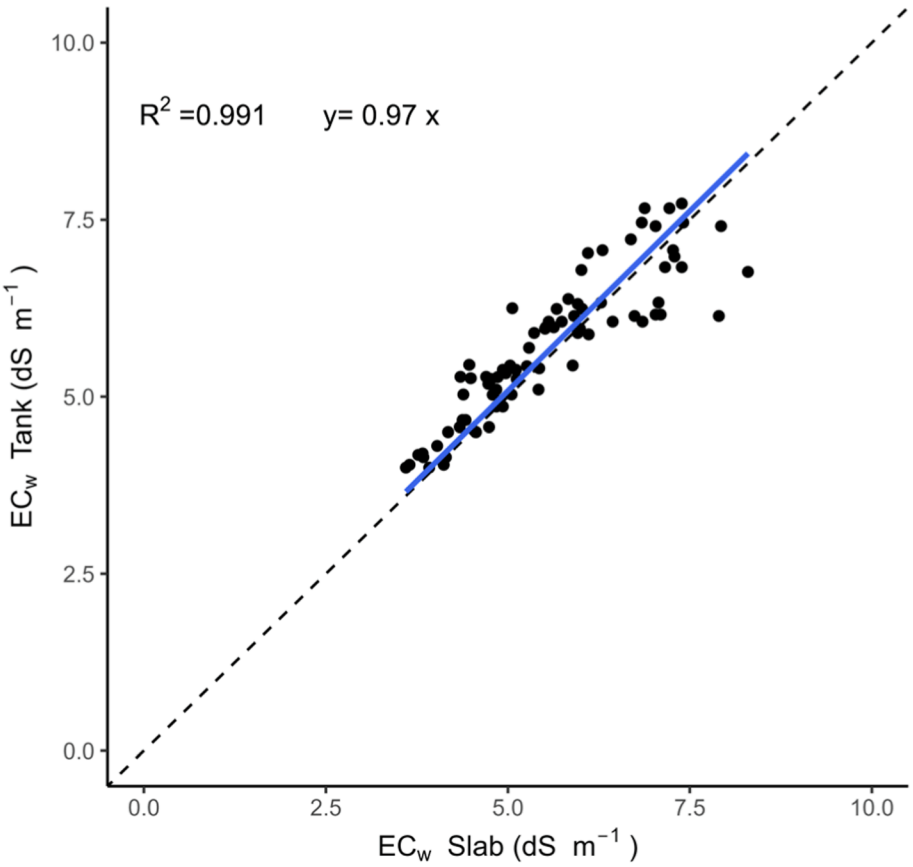

Fig. 3. Regression plot between the pore-water electrical conductivity (EC_w, dS m^{-1}) measured in the slab and the drainage tank (blue line) during the two tomato experimental trials. Data represent a limited number of measurements taken during the experimental trials (roughly 25 % of total). The identity line is reported in the graph (black dashed line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.2. Data collection and sensors

GS3 soil moisture sensors (METER Group Inc., Pullman, WA, USA) were used in this work to measure EC_b, T_s, and volumetric water content (VWC, Fig. 2c, d) in the growing medium. The volumetric water content was estimated through the measure of permittivity of a magnetic field emitted by the GS3 sensor (Amente et al., 2000, Hilhorst, 2000). The calibration of permittivity for the soilless substrates was done in accordance with the sensor manufacturer procedures. The GS3 sensor was placed in the side of the rockwool slabs, in a middle position between two tomato plants as indicated in Fig. 1, data of the sensors were monitored and recorded using an environmental station (OPI system, EVJA srl., NA, Italy).

The EC_w was measured directly in the tank of circulating nutrient solution with a portable conductivity-meter (Hanna Instruments, Woonsocket, RI, USA) every day at 9:00 am throughout the two cultivation trials. At the same time the output of the GS3 sensor was retrieved. The nutrient solution was recirculated for 10 min before the measurements to assure that circulating nutrient solution was homogenized with the drainage tank. The EC_w was also measured in the slab where the GS3 sensor was placed using the portable conductivity-meter, collecting the nutrient solution with a syringe (50 mL sample). The measure was repeated at different day time during the experiment.

2.2.1. Pore-water EC prediction with the Hilhorst equation

The EC_w (dS m^{-1}) was estimated with the following equation (Hilhorst, 2000):

$$EC_w = \frac{\varepsilon_w \bullet EC_b}{\varepsilon_b - \varepsilon_{b=0}} \quad (1)$$

where EC_b is the bulk electrical conductivity as output of the GS3 sensor, $\varepsilon_{b=0}$ is the permittivity when EC_b is equal to 0, ε_b is the dielectric permittivity of the bulk soil, ε_w is the permittivity of the soil pore-water. The last quantity can be calculated as follows:

$$\varepsilon_w = 80.3 - 0.37 \times (T_s - 20) \quad (2)$$

where T_s represents the substrate temperature measured by the sensor. The dielectric permittivity of the bulk soil ε_b was calculated as described by the manufacturing instructions, using the following equation:

$$\varepsilon_b = \frac{\varepsilon_{raw}}{50} \quad (3)$$

where ε_{raw} is the raw data of dielectric permittivity provided by the instrument.

2.2.2. Pore-water EC prediction with a GAM

Different types of linear and nonlinear regressions were tested to estimate the EC_w. Since the relationship between EC_w, temperature, ε_b, and EC_b was not linear, a GAM with EC_b as the nonlinear parameter was selected using the package "mgcv" in R. The GAMs allow to create smooth functions using additive base functions; in our study, the cubic spline was selected as smooth type with the "shrinkage" function (bs = "cs" in mgcv). The model was built using the same variables of the Hilhorst equation:

$$\mathrm{GAM}\varepsilon_w = EC_w \times EC_b \times T_s(EC_b \text{smoothed with cubic spline})$$

2.2.3. Pore-water EC prediction with a machine learning model

Data were divided in two datasets: "Training" and "Test" with the

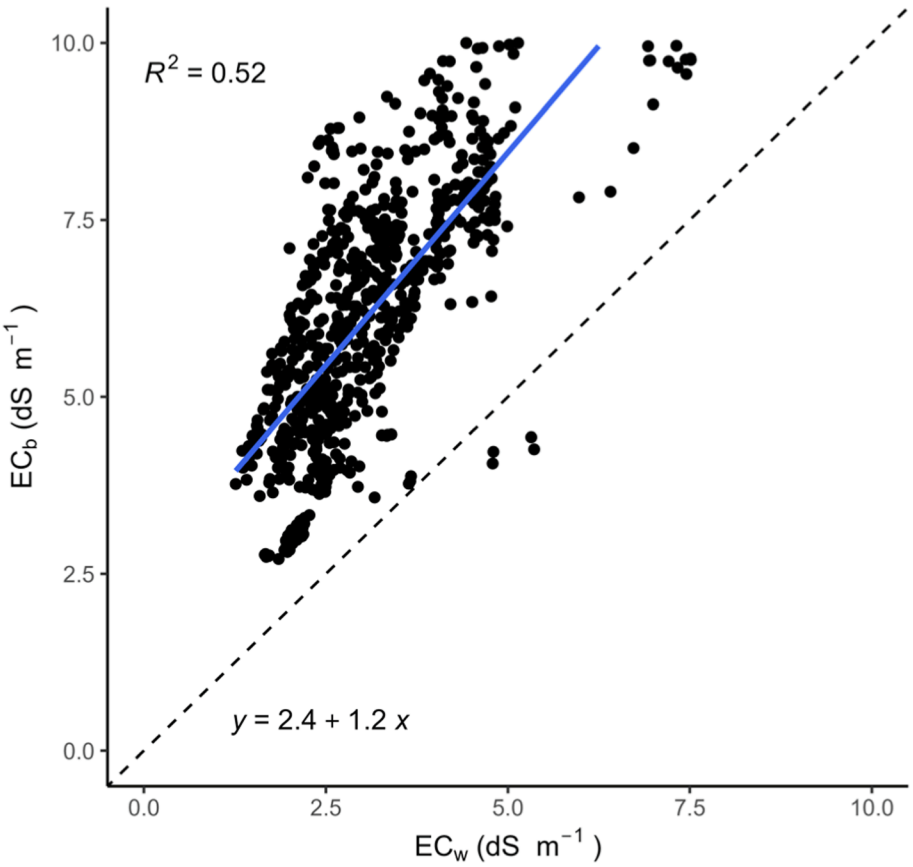

Fig. 4. Regression plot between bulk electrical conductivity (EC_b, dS m^{-1}) and pore-water electrical conductivity (EC_w, dS m^{-1}) measured during the two tomato experimental trials (blue line). The identity line is reported in the graph (black dashed line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

proportion of 80 % and 20 %, respectively, and scaled to values from 0 to 3 to avoid problem of variance homogeneity between the variables. The selected models were trained with the Training dataset and tested with the Test dataset to increase the generalization capacity. A machine learning model based on the XGBoost was built with the same parameters used in the GAM and the Hilhorst equation. The XGBoost algorithm is a decision tree boosting system usable for classification and regression models (Chen and Guestrin, 2016). The XGBoost model was trained with the Training dataset using the package "xgboost" in R, identifying the model parameters that better optimized the algorithm through a turning process. The selected model was then used with the Test dataset to predict the EC_w. The SHAP (SHapley Additive exPlanations) values were used to explain how the independent variables ε_w, EC_b, and T_s were able to predict the dependent variable EC_w. The SHAP approach offers interpretability for assessing machine learning models. It relies on Shapley values, a concept derived from game theory, to measure the contribution of individual features in a predictive model (Lundberg et al., 2020). The SHAP values were applied with the package "SHAPforxgboost" in R.

2.2.4. Evaluation of model accuracy

The accuracy of the tested models was evaluated with different metrics: 1) MAE (Mean Absolute Error) that represents the distance between the measured and predicted values obtained by averaging the absolute difference; 2) MSE (Mean Squared Error) that represents the average quadratic distance between the measured and predicted values; 3) RMSE (Root Mean Squared Error) that is derivate by the square root of MSE; 4) R^2 (Coefficient of determination) that is a measure of how well the values fit compared to the original values, measured by the ratio of the residuals variance over the total variance; and 5) Equation coefficients, intercept and the slope of the regression line between predicted and measured data (in a perfect model would tend to 0 and 1, respectively).

3. Results and discussion

The scatter plot between the EC_w values measured in the nutrient solution, sampled from the rockwool slabs, and the EC_w values measured in the drainage tank, recorded during the two tomato experimental trials, is reported in Fig. 3. The close dispersion around the identity line (1:1) and the determination coefficient, explaining the 88 % of experimental variability, underlined a good agreement between the EC_w measured in substrate and in drainage tank. However, this high level of homogeneity was the result of the experimental conditions. The high leaching fraction imposed at each irrigation events and the nutrient solution recycling before the samplings allowed to obtain an equilibrium between EC_w in the cultivation substrate and the drainage tank, as already assessed in a previous work by Massa et al. (2011). This experimental expedient allowed the calibration of the proposed models using a dataset obtained from the combination of EC_w measurements from both rockwool slabs and drainage tanks.

Under operational conditions, the above conditions are unusual. The EC_w measured in the substrate can in fact differ sometimes from that measured in the drainage tank depending on irrigation scheduling, climate, plant growth stage, cultivation system, and substrate conditions (Massa et al., 2020, Venezia et al., 2022, Heinen, 2001). Therefore, a continuous EC monitoring at root zone level might improve the application of precision fertigation management, but the simple measurement EC_b can be misleading due to the poor correlation with EC_w (Fig. 4).

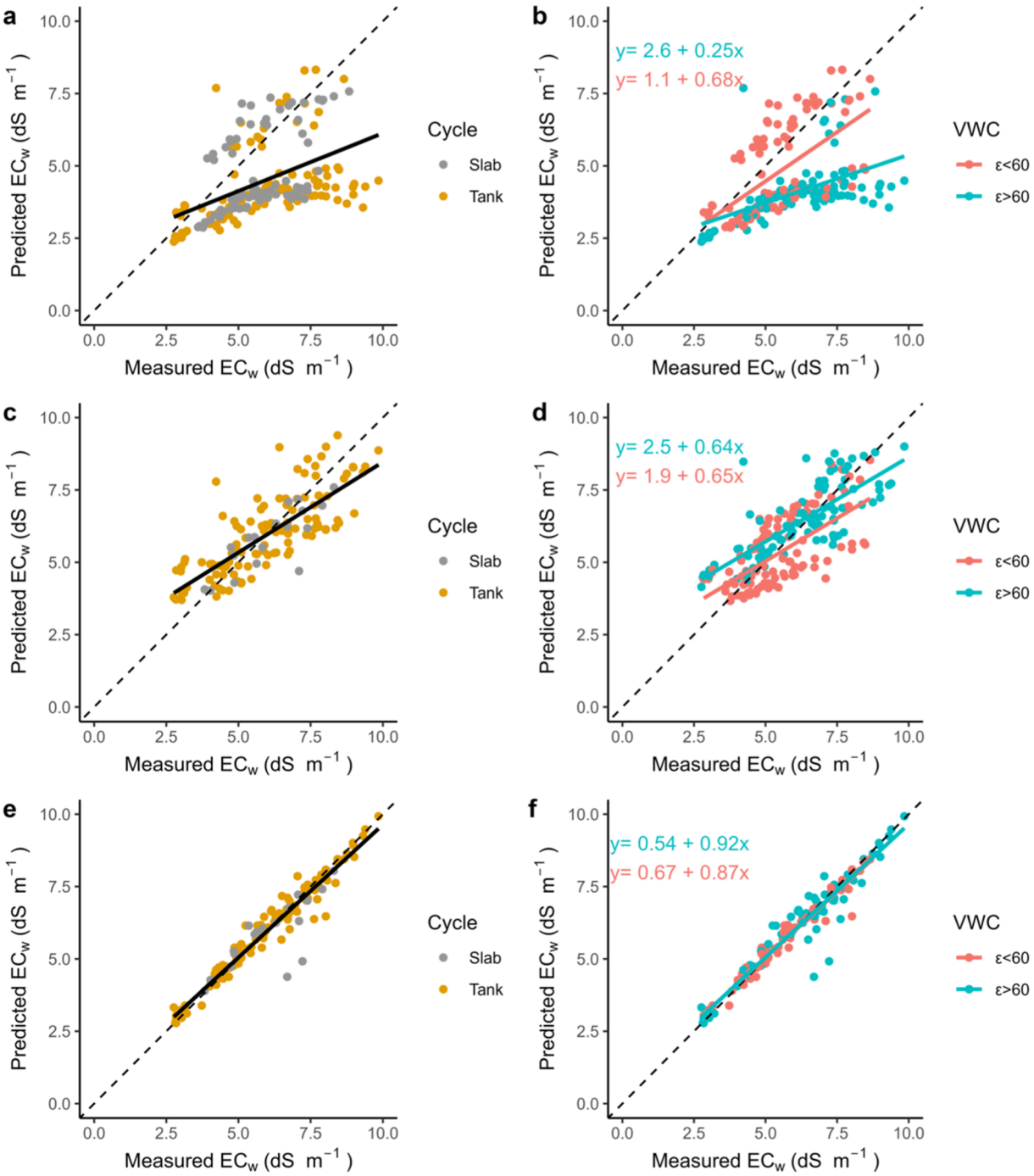

Fig. 5. Comparison between the predicted and the measured pore-water electrical conductivity (EC_w) values (dS m^{-1}) using different models and the "Test" dataset. The linear regression plots show predicted and measured values in the slabs (grey dots) and the tanks (brown dots) using: Hilhorst equation (a), GAM (c), and XGBoost (e). The interactions using two levels of ε_w (<60 in red and > 60 in blue) are showed for: Hilhorst equation (b), GAM (d), and XGBoost (f). The equations are reported in Table 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Indeed, previous studies reported not clear relationships between the above variables directly measured in the substrate through GS3 sensors under salinity treatments (Lim et al., 2017). In agreement, under our experimental conditions, EC_w increased 1.2 folds as a function of EC_b at higher salinity levels; the scatter plot between the EC_w and EC_b values measured in the rockwool slabs during the two tomato trials is reported in Fig. 4. The unclear pattern and the low determination coefficient highlighted the unsuitability of a simple linear model to describe the relation between these two values. Therefore, results underline the necessity of a model to convert EC_b directly measured in the substrate into EC_w for the correct management of the nutrient solution in soilless systems.

The EC_w values predicted with the Hilhorst equation and those measured in the cultivation system were poorly correlated (Fig. 5a), as indicated by the very low value of the determination coefficient (R^2 = 0.13) and high values of the main other error coefficients (Table 1). In particular, the EC_w predicted by this equation was more accurate at low salinity level while a strong underestimation of data was highlighted at higher EC_w values also showing different trends with poor prediction accuracy (Fig. 6). Similar unsatisfactory results were obtained by Bañón

Table 1
Comparison among the accuracy metrics of the three different prediction models of pore-water electrical conductivity (EC_w). Data are plotted in Fig. 5. Mean absolute error (MAE), root mean square error (RMSE), mean squared error (MSE), determination coefficient (R^2), linear regression between predicted and measured values.

Model	MAE (dS m^{-1})	RMSE (dS m^{-1})	MSE (dS m^{-1})	R^2	Predictions *vs* measurements equation
Hilhorst equation	1.94	2.43	5.88	0.132	$y = 2.1 + 0.40x$
GAM	0.826	1.027	1.06	0.668	$y = 2.5 + 0.59x$
XGBoost	0.405	0.623	0.389	0.876	$y = 0.8 + 0.87x$

et al. (2021) who argued that this variability could mainly be related to different levels of ε_w in the root zone. To better identify the role of ε_w in this poor correlation, this parameter was grouped in two levels: below and above 60 % (v/v), a threshold limit always proposed by Bañón et al. (2021). For ε_w above 60 % the slope of the linear regression between predicted and measured values decreased, while for ε_w below 60 % the slope coefficient increased although a very high degree of data spread was observed at low salinity levels. Our results confirmed that the Hilhorst equation is a poor predictor when ε_w is not constant and/or higher than 60 % as previously reported by other authors (Bañón et al., 2021, Scoggins and van Iersel, 2006).

Since the relationship between EC_b and ε_w data was non-linear, GAMs could be appropriate for the simulation of EC_w allowing the creation of models with non-linearity and smooth of the prediction. GAMs have been already broadly adopted to estimate EC in water basin (Maloney et al., 2012, Morton and Henderson, 2008) and lands (Jordán et al., 2004). In our study case, the variable EC_b was set as non-linear. The presence of ε_w or T_s as smoothed function did not increase the model precision (data not shown). Therefore, a simpler model with only EC_b as smoothed variable was selected. This model increased the accuracy of EC_w predictions compared with Hilhorst equation (Fig. 5c) as highlighted by the higher determination coefficients ($R^2 = 0.66$), and lower MAE, RMSE, and MSE (Table 1). Indeed, data simulated by the GAM followed more precisely the time course of measured data compared with the simulations obtained with the Hilhorst equation (Fig. 6). The GAM was not influenced by the level of ε_w, as indicated by

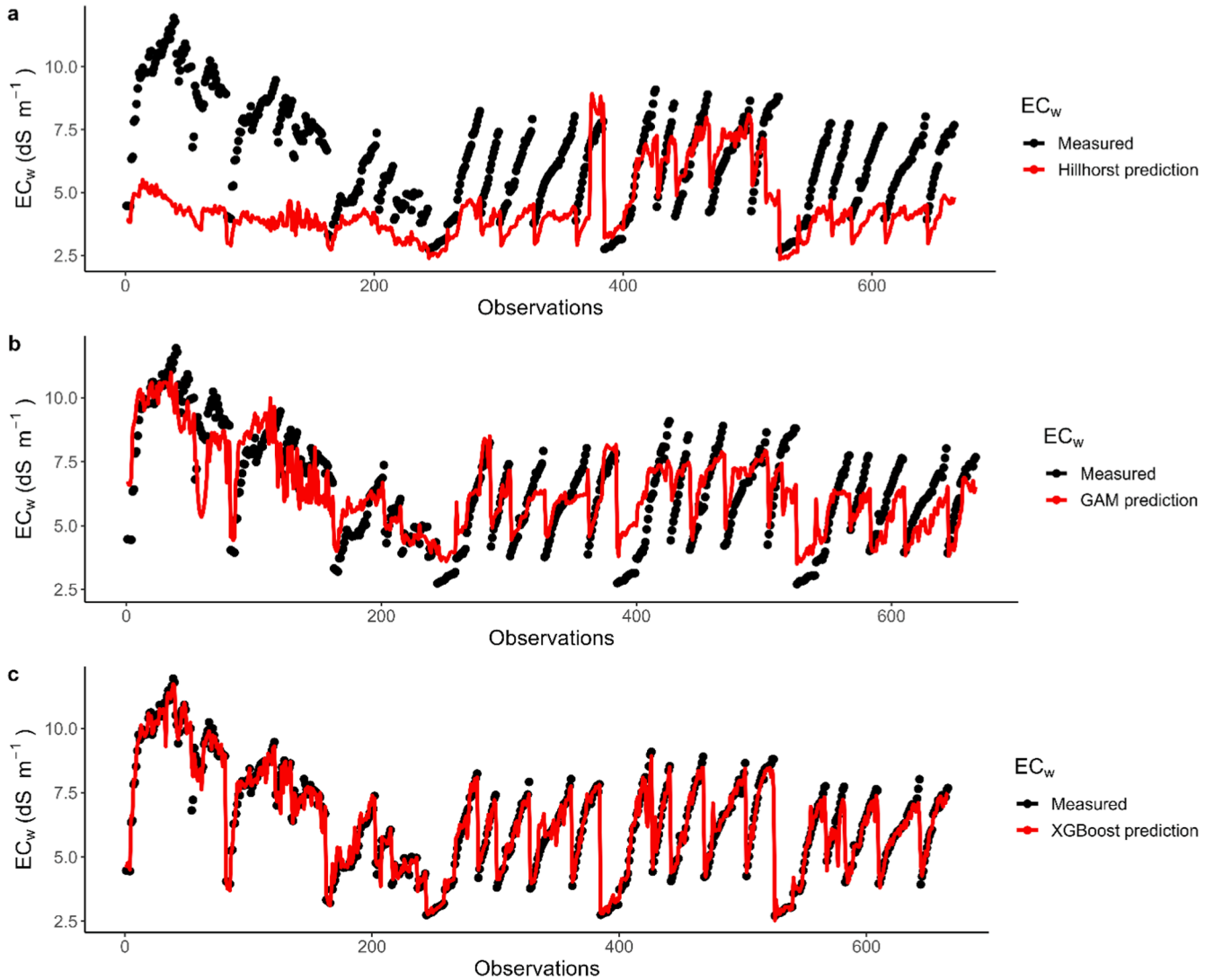

Fig. 6. Values of pore-water electrical conductivity (EC_w, dS m^{-1}) measured (black dots) and predicted with the different models (red lines): Hilhorst equation (a), GAM (b), and XGBoost (c). The observations represent all datasets (Training + Test). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

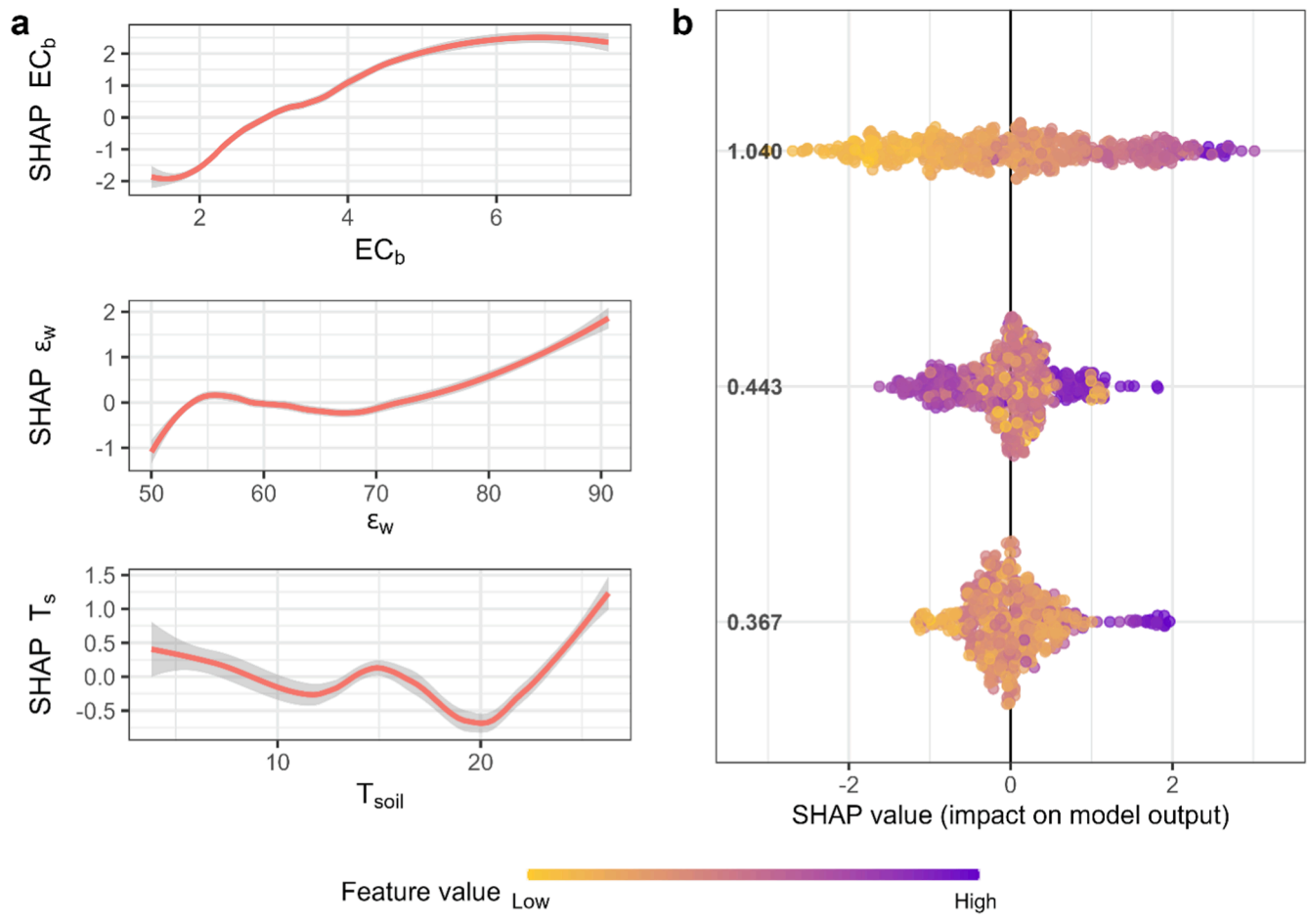

Fig. 7. Graphical representation of SHAP values for the XGBoost model. In the SHAP summary plot (a) each point represents an observation for three different variables, the position on the X axes indicates the impact of that observation on the output value (EC_w) and the value on the rows represents the total impact on model output. The dots color indicates the value for a particular variable, purple for high value and yellow for low value. The SHAP dependence plot (b) shows the contribution of each variable to the model output in a XY graph. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the slope coefficient in Fig. 5e. However, the linear regression analysis between predicted and measured data showed an intercept of 2.5 dS m^{-1} and a slope of 0.59, indicating a trend of overestimation for low values of EC_w and an underestimation for high values of EC_w (Fig. 5c) with a general poor performance in the simulation of the extreme values (Fig. 6).

The model obtained with the XGBoost approach performed better than the other two models in terms of error coefficients (Table 1) and explained the 83 % of the experimental variability (R^2 = 0.83). The linear regression between predicted and measured values produced an intercept and a slope very close to 0 and 1, respectively (Fig. 5e and Table 1). The XGBoost model was also able to adequately predict the EC_w in the slabs even at extreme values of EC and ε_w (Fig. 5e and Fig. 6). Machine learning models have already been applied to predict water EC (Mestanza et al., 2022, Al-Mukhtar and Al-Yaseen, 2019), and the decision tree-based models, such as XGBoost, are among the most used (Ali Khan et al., 2022, Guan et al., 2022). Compared with other tree-based machine learning approaches, XGBoost models have the capacity to exploit complex nonlinear relationships between driving variables thereby preventing model overfitting and leading to more generalizable models (Jia et al., 2019).

The SHAP values for the XGBoost model were calculated allowing to observe the influence of each variable in every sample of our dataset (Fig. 7). For high values of EC_b and ε_w (color tending to purple), high correspondent outputs were observed. On the contrary, a drop in the output values around 20 °C was observed for T_s. Despite SHAP values are capable to explain complex model such as XGBoost, this result did not imply that T_s directly increased or decreased the EC_w, but rather that the probability of finding EC_w with low values was higher when T_s was close to 20 °C. In other words, the SHAP coefficient showed correlation but not causation (Han et al., 2022, Kaur et al., 2020). In a recent work, Mestanza et al. (2022) found that machine learning models, such as XGBoost, were the best models to predict extreme soil EC in term of RMSE and MAE, while GAMs were accurate only at low EC levels (0–4 dS m^{-1}). Our results in soilless were consistent with their findings in soil.

Cherry tomatoes are confirmed an optimal test species to calibrate EC models for soilless systems due to their good tolerance to high substrate water content and EC_b that in our experimental conditions were even higher than those previously described in the available literature (Bañón et al., 2021, Kargas and Lim et al., 2017, Kerkides, 2012).

4. Conclusions

To the best of our knowledge, this is the first calibration study in substrate (rockwool) soilless system for the determination of EC_w from EC_b, using XGBoost model, using a very large and variable range of EC in the root zone. This work demonstrates the strength of machine learning models for the EC_w prediction in soilless growing condition allowing to overcome the low accuracy of Hilhorst equation at high ε_w. Results highlighted the possibility of monitoring continuously and with high

accuracy the EC_w status at root level with a bulk, low cost, EC sensor such as GS3. However, machine learning models such as XGBoost tend to be very precise in the training conditions but their applicability under different agronomic conditions such as a different substrate would require further validation and investigations.

Funding sources

This work was supported by the PRIMA programme under grant agreement No. 1916, project iGUESS-MED. The PRIMA programme is supported by the European Union.

CRediT authorship contribution statement

Mirko Sodini: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Sonia Cacini:** Investigation, Supervision, Writing – review & editing. **Alejandra Navarro:** Funding acquisition, Visualization, Writing – review & editing. **Silvia Traversari:** Investigation, Visualization, Writing – review & editing. **Daniele Massa:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

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.

Data availability

Data will be made available on request.

Acknowledgement

This work has been supported by the project iGUESS-MED – "Innovative Greenhouse Support System in the Mediterranean Region: efficient fertigation and pest management through IoT based climate control", grant agreement 1916; financed by the European Commission, in the PRIMA program SECTION 1 IA Sustainability and competitiveness of Mediterranean greenhouse and intensive horticulture – CALL 2019.

References

Ahn, T.I., Son, J.E., 2019. Theoretical and experimental analysis of nutrient variations in electrical conductivity-based closed-loop soilless culture systems by nutrient replenishment method. Agronomy 9 (10), 649. https://doi.org/10.3390/agronomy9100649.

Ahn, T.I., Son, J.E., 2022. Application of an Alternative Nutrient Replenishment Method to Electrical Conductivity-Based Closed-Loop Soilless Cultures of Sweet Peppers. Horticulturae 8, 295. https://doi.org/10.3390/horticulturae8040295.

Ali Khan, M., Izhar Shah, M., Faisal Javed, M., Ijaz Khan, M., Rasheed, S., El-Shorbagy, M.A., Roshdy El-Zahar, E., Malik, M.Y., 2022. Application of random forest for modelling of surface water salinity. Ain Shams Eng. J. 13 (4), 101635 https://doi.org/10.1016/j.asej.2021.11.004.

Al-Mukhtar, M., Al-Yaseen, F., 2019. Modeling water quality parameters using data-driven models, a case study Abu-Ziriq Marsh in South of Iraq. Hydrology. 6 (1), 24. https://doi.org/10.3390/hydrology6010024.

Amente, G., Baker, J.M., Reece, C.F., 2000. Estimation of Soil Solution Electrical Conductivity from Bulk Soil Electrical Conductivity in Sandy Soils. Soil Sci. Soc. Am. J. 64 (6), 1931–1939. https://doi.org/10.2136/sssaj2000.6461931x.

Bañón, S., Álvarez, S., Bañón, D., Ortuño, M.F., Sánchez-Blanco, M.J., 2021. Assessment of soil salinity indexes using electrical conductivity sensors. Sci. Hort. 285, 110171 https://doi.org/10.1016/j.scienta.2021.110171.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining pp. 785–794. Doi: 10.1145/2939672.2939785.

Cliff, M.A., Li, J.B., Toivonen, P.M.A., Ehret, D.L., 2012. Effects of nutrient solution electrical conductivity on the compositional and sensory characteristics of greenhouse tomato fruit. Postharvest Biol. Technol. 74, 132–140. https://doi.org/10.1016/j.postharvbio.2011.12.007.

Corwin, D.L., Yemoto, K., 2020. Salinity: Electrical conductivity and total dissolved solids. Soil Sci. Soc. Am. J. 84 (5), 1442–1461. https://doi.org/10.1002/saj2.20154.

De Pascale, S., Maggio, A., Fogliano, V., Ambrosino, P., Ritieni, A., 2001. Irrigation with saline water improves carotenoids content and antioxidant activity of tomato. J. Hortic. Sci. Biotechnol. 76 (4), 447–453. https://doi.org/10.1080/14620316.2001.11511392.

Ferrarezi, R.S., Nogueira, T.A.R., Zepeda, S.G.C., 2020. Performance of Soil Moisture Sensors in Florida Sandy Soils. Water 12, 358. https://doi.org/10.3390/w12020358.

Germano, R.P., Melito, S., Cacini, S., Carmassi, G., Leoni, F., Maggini, R., Montesano, F. F., Pardossi, A., Massa, D., 2022. Sweet basil can be grown hydroponically at low phosphorus and high sodium chloride concentration: Effect on plant and nutrient solution management. Sci. Hortic. 304, 111324 https://doi.org/10.1016/j.scienta.2022.111324.

Guan, Y., Grote, K., Schott, J., Leverett, K., 2022. Prediction of Soil Water Content and Electrical Conductivity Using Random Forest Methods with UAV Multispectral and Ground-Coupled Geophysical Data. Remote Sens. 14 (4), 1023. https://doi.org/10.3390/rs14041023.

Han, L., Yang, G., Yang, X., Song, X., Xu, B., Li, Z., Wu, J., Yang, H., Wu, J., 2022. An explainable XGBoost model improved by SMOTE-ENN technique for maize lodging detection based on multi-source unmanned aerial vehicle images. Comput. Electron. Agric. 194, 106804 https://doi.org/10.1016/j.compag.2022.106804.

Hardie, M., 2020. Review of Novel and Emerging Proximal Soil Moisture Sensors for Use in Agriculture. Sensors 20 (23), 23. https://doi.org/10.3390/s20236934.

Heinen, M., 2001. FUSSIM2: brief description of the simulation model and application to fertigation scenarios. Agronomie 21, 285–296. https://doi.org/10.1051/agro:2001124.

Hilhorst, M.A., 2000. A Pore Water Conductivity Sensor. Soil Sci. Soc. Am. J. 64 (6), 1922–1925. https://doi.org/10.2136/sssaj2000.6461922x.

Incrocci, L., Malorgio, F., Della Bartola, A., Pardossi, A., 2006. The influence of drip irrigation or subirrigation on tomato grown in closed-loop substrate culture with saline water. Sci. Hortic. 107 (4), 365–372. https://doi.org/10.1016/j.scienta.2005.12.001.

Incrocci, L., Massa, D., Pardossi, A., 2017. New trends in the fertigation management of irrigated vegetable crops. Horticulturae 3, 37. https://doi.org/10.3390/horticulturae3020037.

Jia, Y., Jin, S., Savi, P., Gao, Y., Tang, J., Chen, Y., Li, W., 2019. GNSS-R Soil Moisture Retrieval Based on a Xgboost Machine Learning Aided Method: Performance and Validation. Remote Sens. 11 (14), 1655. https://doi.org/10.3390/rs11141655.

Jordán, M.M., Navarro-Pedreño, J., García-Sánchez, E., Mateu, J., Juan, P., 2004. Spatial dynamics of soil salinity under arid and semi-arid conditions: Geological and environmental implications. Environ. Geol. 45 (4), 448–456. https://doi.org/10.1007/s00254-003-0894-y.

Kargas, G., Kerkides, P., 2010. Evaluation of a Dielectric Sensor for Measurement of Soil-Water Electrical Conductivity. J. Irrig. Drain. Eng. 136, 553–558. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000218.

Kargas, G., Kerkides, P., 2012. Comparison of two models in predicting pore water electrical conductivity in different porous media. Geoderma 189–190, 563–573. https://doi.org/10.1016/j.geoderma.2012.06.024.

Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., Wortman Vaughan, J., 2020. Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 1–14. Doi: 10.1145/3313831.3376219.

Kocárek, M., Kodešová, R., 2012. Influence of temperature on soil water content measured by ECH2O-TE sensors. Int. Agrophys. 26 (3), 259–269. https://doi.org/10.2478/v10247-012-0038-2.

Lim, S.J., Shin, M.N., Son, J.K., Song, J.D., Cho, K.H., Lee, S.H., Ryu, J.H., Cho, J.Y., 2017. Evaluation of soil pore-water salinity using a Decagon GS3 sensor in saline-alkali reclaimed tidal lands. Comput. Electron. Agric. 132, 49–55. https://doi.org/10.1016/j.compag.2016.11.017.

Lu, T., Yu, H., Wang, T., Zhang, T., Shi, C., Jiang, W., 2022. Influence of the Electrical Conductivity of the Nutrient Solution in Different Phenological Stages on the Growth and Yield of Cherry Tomato. Horticulturae 8, 5. https://doi.org/10.3390/horticulturae8050378.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9.

Maloney, K.O., Schmid, M., Weller, D.E., 2012. Applying additive modelling and gradient boosting to assess the effects of watershed and reach characteristics on riverine assemblages. Methods Ecol. Evol. 3 (1), 116–128. https://doi.org/10.1111/j.2041-210X.2011.00124.x.

Massa, D., Melito, S., 2019. Signaling Molecules in Ecophysiological Response Mechanisms of Salt-Stressed Plants, in: Khan, M.I.R., Reddy, P.S., Ferrante, A., Khan, N.A. (Eds.), Plant Signaling Molecules. Woodhead Publishing, pp. 1–18. Doi: 10.1016/B978-0-12-816451-8.00001-0.

Massa, D., Incrocci, L., Maggini, R., Bibbiani, C., Carmassi, G., Malorgio, F., Pardossi, A., 2011. Simulation of crop water and mineral relations in greenhouse soilless culture. Environ. Model. Software 26, 711–722. https://doi.org/10.1016/j.envsoft.2011.01.004.

Massa, D., Magán, J.J., Montesano, F.F., Tzortzakis, N., 2020. Minimizing water and nutrient losses from soilless cropping in southern Europe. Agric. Water Manage. 241, 106395 https://doi.org/10.1016/j.agwat.2020.106395.

Mestanza, C., Chicchon, M., Gutiérrez, P., Hurtado, L., Beltrán, C., 2022. Prediction of Soil Saturated Electrical Conductivity by Statistical Learning, in: Lossio-Ventura, J.A., Valverde-Rebaza, J., Díaz, E., Muñante, D., Gavidia-Calderon, C., Valejo, A.D.

B., Alatrista-Salas, H. (Eds.), Information Management and Big Data, Communications in Computer and Information Science. Springer International Publishing, Cham, pp. 397–412. Doi: 10.1007/978-3-031-04447-2_27.

Moon, T., Ahn, T.I., Son, J.E., 2018. Forecasting Root-Zone Electrical Conductivity of Nutrient Solutions in Closed-Loop Soilless Cultures via a Recurrent Neural Network Using Environmental and Cultivation Information. Front. Plant Sci. 9, 859. https://doi.org/10.3389/fpls.2018.00859.

Morton, R., Henderson, B.L., 2008. Estimation of nonlinear trends in water quality: An improved approach using generalized additive models. Water Resour. Res. 44, 7. https://doi.org/10.1029/2007WR006191.

Navarro, A., Bañón, S., Olmos, E., Sánchez-Blanco, M.J., 2007. Effects of sodium chloride on water potential components, hydraulic conductivity, gas exchange and leaf ultrastructure of *Arbutus unedo* plants. Plant Sci. 172 (3), 473–480. https://doi.org/10.1016/j.plantsci.2006.10.006.

Petersen, K.K., Willumsen, J., Kaack, K., 2015. Composition and taste of tomatoes as affected by increased salinity and different salinity sources. J. Hortic. Sci. Biotechnol. https://doi.org/10.1080/14620316.1998.11510966.

Rosenbaum, U., Huisman, J.A., Vrba, J., Vereecken, H., Bogena, H.R., 2011. Correction of Temperature and Electrical Conductivity Effects on Dielectric Permittivity Measurements with ECH2O Sensors. Vadose Zone J. 10 (2), 582–593. https://doi.org/10.2136/vzj2010.0083.

Rouphael, Y., Petropoulos, S.A., Cardarelli, M., Colla, G., 2018. Salinity as eustressor for enhancing quality of vegetables. Sci. Hortic. 234, 361–369. https://doi.org/10.1016/j.scienta.2018.02.048.

Venezia, A., Colla, G., Di Cesare, C., Stipic, M., Massa, D., 2022. The effect of different fertigation strategies on salinity and nutrient dynamics of cherry tomato grown in a gutter subirrigation system. Agric. Water Manage. 262, 107408 https://doi.org/10.1016/j.agwat.2021.107408.

Wright, R.D., 1986. The pour-through nutrient extraction procedure. The Pour-through Nutrient Extraction Procedure 21 (2), 227–229.

Wu, M., Kubota, C., 2008. Effects of high electrical conductivity of nutrient solution and its application timing on lycopene, chlorophyll and sugar concentrations of hydroponic tomatoes during ripening. Sci. Hortic 116 (2), 122–129. https://doi.org/10.1016/j.scienta.2007.11.014.