Journal of Industrial Intelligence
https://www.acadlore.com/journals/JII

Deployment Assessment of Ensemble Soft Sensors Under Input-Channel and Operating-Range Constraints: A Greenhouse Monitoring Case Study

Fucheng Wang[1], Jinqi Zhao[1,2*], Zhongyu Fu[1], Wanying Song[1], Jinze Ning[1]

[1] College of Engineering, Heilongjiang Bayi Agricultural University, 163319 Daqing, China
[2] Harbin Institute of Petroleum, 150000 Harbin, China

* Correspondence: Jinqi Zhao (1053796200@qq.com)

Received: 06-25-2026 **Revised:** 08-10-2026 **Accepted:** 08-25-2026

Citation: F. C. Wang, J. Q. Zhao, Z. Y. Fu, W. Y. Song, and J. Z. Ning, "Deployment assessment of ensemble soft sensors under input-channel and operating-range constraints: A greenhouse monitoring case study," *J. Ind Intell.*, vol. 4, no. 3, pp. 262–281, 2026. https://doi.org/10.56578/jii040303.

© 2026 by the author(s). Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: Soft sensors are widely used to estimate process variables that are hard to measure online, but a low prediction error says little about whether an estimator stays usable after an input channel is lost or operating conditions shift. This study examines how input-channel requirements, calibration coverage and local error limit the deployment of ensemble soft sensors. An assessment procedure combining validation-based model selection, channel ablation, componentwise range checks and error-tolerance curves was applied to a laboratory-scale solar greenhouse. Two campaigns in 2024 yielded 1,466 hourly records, of which 1,446 passed screening. Five ensemble regressors and two baselines were used to estimate same-hour soil temperature and probe-reported moisture from air temperature, relative humidity, illuminance and soil electrical conductivity (EC). With all four inputs, validation selected AdaBoost with support vector regression (AdaBoost–SVR, ABS) for temperature and extremely randomized trees (ET) for moisture, which gave heldout root mean squared errors (RMSEs) of 0.872 °C and 2.088 percentage points on 290 records. Without EC, these errors rose to 1.016 °C and 6.335 percentage points. Although 99.3% of the test inputs lay within the calibration envelope, winter moisture R^2 was negative. Under winter-only calibration, 98.6% of spring records fell outside the four-input envelope, and all six regressors gave moisture RMSEs above 17 percentage points. In this case, air-side inputs supported soil-temperature reconstruction, whereas the tested moisture models depended strongly on probe EC. High marginal range coverage did not ensure adequate local tracking. Reporting channel availability, range coverage and task-specific error separately gives engineers a documented basis for reviewing soft-sensor configurations before supervisory use, provided that models and limits are calibrated locally.

Keywords: Soft sensor; Process monitoring; Deployment assessment; Input-channel availability; Calibration envelope; Ensemble regression; Greenhouse

1 Introduction

Soft sensors estimate process variables that are expensive, slow or difficult to measure directly, and they are now routine in chemical, polymer, food and water processes. Recent work has automated soft-sensor design [1], exploited local spatiotemporal structure in deep models [2] and built model-based estimators for units such as sugar crystallizers [3]. Adaptive and multimode schemes handle process relationships that drift over time [4, 5]. Two application studies are especially relevant here. In in-line viscosity sensing, the accuracy of the soft sensor was bound up with the design of the acquisition hardware [6]. In industrial wastewater, the choice of reference variable determined what the model could be trusted to deliver [7]. Both studies expose a gap in how soft sensors are usually reported: papers state how accurate a model is but say much less about the conditions under which that accuracy holds once the estimator is part of a monitoring system.

Greenhouse monitoring makes this gap easy to see. Heating, ventilation, lighting and irrigation decisions all rely on sensors whose operating ranges shift with season and management. Machine- and deep-learning models have been used to predict greenhouse microclimate [8], to compare data-driven predictors with Computational Fluid Dynamics (CFD) across seasons [9] and to attach time-varying uncertainty to forecasts [10]. Other studies have extended the target list to evaporation [11], air temperature in edible-fungi houses [12] and indoor humidity [13].

For soil variables, Belouz and Zereg [14] estimated soil temperature from air temperature alone, and Baumberger et al. [15] mapped soil temperature and moisture with interpretable models. When such results are compared, one difference is often overlooked: some models forecast a future state while others reconstruct the current one, and they assume very different instrument sets. A model that needs a soil probe at run time cannot act as a backup when that probe fails.

Soil moisture illustrates the difficulty. Internet of Things (IoT) measurements [16], integrated climate, soil and vegetation data [17] and transfer learning [18] have all been used to predict moisture, often with conductivity among the predictors. In low-cost installations, conductivity, temperature and moisture may come from one integrated probe with shared electronics and compensation. A regression that uses the probe's conductivity output can then capture relationships introduced by the instrument as well as those associated with the soil. Commercial moisture sensors respond differently across conductivity levels and soil types [19, 20], apparent conductivity can be calibrated against water content [21], and some instruments measure water content, temperature and conductivity through a common sensing element [22]. Studies of dual-frequency measurement [23], joint changes in salinity, temperature and moisture [24] and soil-specific calibration [25] all underline the need to define the reference scale. Without a water-content calibration, an error on a probe's percentage scale cannot be converted directly into an irrigation volume.

The evaluation design matters as much as the model. Leakage through preprocessing and model selection inflates reported accuracy [26, 27], and a systematic review of maize-yield studies found that agricultural prediction scores are frequently over-interpreted [28]. For records collected in time order, random cross-validation mostly measures interpolation between neighbouring observations; blocked or temporal schemes are needed to test prediction under new conditions [29–32]. Hourly greenhouse data are a typical case, because adjacent records share weather, management and sensor state.

Errors in an estimate do not stay inside the model once it feeds a controller. Greenhouse studies now couple measurements with particle filtering and deep learning [33], with reinforcement learning and model predictive control [34–36] and with cooperative regulation of moisture and light [37]. Digital-twin platforms and greenhouse energy models link sensor streams to operating decisions in a similar way [38–41]. In each of these settings, the designer needs to know three things about a soft sensor: which channels it depends on, whether the current operating point was covered during calibration, and how large an error the downstream decision can tolerate. In the studies reviewed above, these three items are seldom reported together.

This paper sets out a deployment assessment that reports the three items separately and tests it on two monitoring campaigns from a scaled solar greenhouse. Same-hour soil temperature and probe-reported moisture were estimated from air temperature, air relative humidity, illuminance and soil electrical conductivity (EC). We compared a four-input configuration with an EC-free configuration that uses air-side sensors only, froze model selection on validation data, and measured heldout error both within the represented campaigns and after transfer from winter to spring. A componentwise calibration envelope and error-tolerance curves were then added to show what each check contributes to a monitoring decision.

The study makes three contributions. First, it quantifies how strongly each target depends on a predictor that may share hardware with the reference: removing EC increased temperature root mean squared error (RMSE) by 16.5% and roughly tripled moisture RMSE. Second, the measured data show that near-complete marginal range coverage can coexist with negative local R^2, so a passed range check is no guarantee of adequate tracking. Third, the procedure brings the channel set, reference definition, selected model, calibration bounds and heldout error profile together in one record for engineering review. The greenhouse supplies the experimental case. To apply the same checks elsewhere, one needs a defined input–reference relationship and local data for fitting the model, building the range bounds and testing it against the intended monitoring requirement.

2 Experimental Platform and Monitoring Data

2.1 Greenhouse and Acquisition System

The experimental platform was a scaled solar greenhouse in Tai'an, Shandong Province, China, with a span of 1.5 m, a height of 0.9 m and a length of 0.8 m. The enclosure had rock-wool composite side and rear walls, an aluminium-alloy frame and a transparent plastic front roof. The integrated soil probe was installed in a soil bed at a nominal measurement depth of 10 cm. All measurements come from this laboratory-scale enclosure and do not describe the spatial soil conditions of a commercial greenhouse.

A LabVIEW host program acquired the sensor channels once per hour over an RS485 bus and exported time-aligned records. Air temperature and relative humidity were measured with a TS-FTM01 sensor, and illuminance with an SM9561B sensor. The TS-FTM01 had a nominal temperature range of -20 to 80 °C, accuracy of ±0.3 °C (mean over the full range) and resolution of 0.1 °C. Its nominal humidity range was 0–99.9% relative humidity (RH), with an accuracy of ±3% RH, including nonlinearity, hysteresis and repeatability, and a resolution of 0.1% RH. The SM9561B had a nominal illuminance range of 0–200,000 lux, resolution of 1 lux, repeatability of ±5%

and response time of 0.1 s. Its stated accuracy was ±7% of full scale under the specified test conditions of 600 nm, 65% RH and 25 °C. Both sensor types used RS485 communication with the Modbus-RTU protocol. These nominal specifications do not quantify calibration uncertainty under the study conditions. An integrated multiparameter soil sensor with an RS485 interface provided soil temperature, moisture and EC readings. The four-input configuration used air temperature, air relative humidity, illuminance and soil EC as predictors. The air-side configuration left out EC and so needed no soil-probe input at inference. Both configurations were developed against the recorded soil temperature and moisture channels. Figure 1 outlines acquisition, model development and the offline assessment checks.

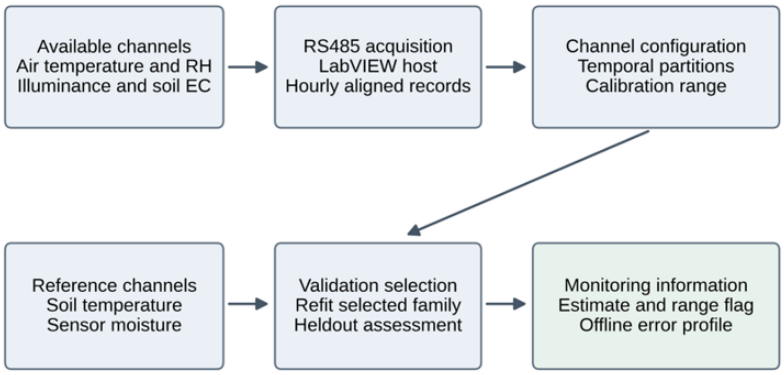


Figure 1. Workflow for soft-sensor development and deployment assessment

Note: RH = relative humidity; EC = electrical conductivity.

2.2 Monitoring Campaign

The monitoring data were collected in 2024. Campaign 1 extended from 25 January at 11:00 to 25 February at 11:00 and contained 745 hourly records. Campaign 2 extended from 28 April at 09:00 to 28 May at 09:00 and contained 721 records. The two campaigns were acquired on the same platform at one-hour intervals and spanned 61 days in total (31 and 30 days, respectively). Monitoring was not continuous between the two windows. The two campaigns comprise 1,466 records, each containing four inputs and two targets, giving 8,796 values across the six analysed channels before screening.

2.3 Data Screening and Measurement Definitions

Screening used predictor values only, independent of the targets and of model residuals. Seventeen records contained simultaneous zeros in illuminance, air relative humidity and air temperature (eleven in Campaign 1 and six in Campaign 2). We treated this combination as an acquisition anomaly; its device-level cause cannot be determined from the records. Three further Campaign 1 records exceeded a gross illuminance cutoff of 200,000 lux, with readings of 479,413.35, 1,917,415.34 and 7,669,601.57 lux. These 20 records were excluded without imputation. Records with zero illuminance alone, or with subzero soil temperature, were kept. The 200,000-lux cutoff coincides with the nominal upper measurement limit of the SM9561B. Readings above this limit were treated as out of range; their device-level cause was not inferred from the exported values.

After screening, Campaigns 1 and 2 contained 731 and 715 records, respectively. All six analysed channels were numeric and nonmissing, and no duplicate six-variable vectors were found. Table 1 reports record accounting and temporal partitions; Table 2 summarizes the retained operating ranges. Supplementary Data S1 contains the source records and the exclusion log.

Moisture was recorded as a percentage on the probe's output scale. Because no independent gravimetric or volumetric calibration was available, we refer to the target as sensor-reported moisture and express its errors in percentage points. In Campaign 2, one record equalled 40.00% and one was at or below 0.10% (minimum 0.02%). These endpoints are isolated readings; they neither define the probe's measurement limits nor demonstrate saturation. Both were retained under the predictor-only screening protocol. The other channels are expressed in °C for soil and air temperature, % for air relative humidity, lux for illuminance and μS/cm for EC.

We could not confirm that the probe's conductivity, temperature and moisture measurement chains are independent. The four-input analysis is therefore a reconstruction of recorded probe channels, while the EC-free analysis evaluates a configuration with no soil-probe predictor at inference. Neither configuration amounts to an independently calibrated soil-water measurement. The same distinction between available predictors and reference output arises in industrial soft sensing [7].

Table 1. Acquisition records, screening and ordered partitions

Item	Campaign 1	Campaign 2	Total
Recorded window in 2024	25 Jan–25 Feb	28 Apr–28 May	Two campaigns
Source records	745	721	1466
Simultaneous-zero exclusions	11	6	17
Gross illuminance exclusions	3	0	3
Retained records	731	715	1446
Training records	438	429	867
Validation records	146	143	289
Test records	147	143	290

Table 2. Retained channel distributions by monitoring campaign

Channel	C1 Mean ± SD	C1 Range	C2 Mean ± SD	C2 Range
Soil temperature (°C)	9.55 ± 7.74	-0.36 to 35.57	24.76 ± 5.86	12.46 to 36.06
Sensor moisture (%)	32.65 ± 2.51	19.04 to 37.96	14.93 ± 6.20	0.02 to 40.00
Illuminance (lux)	2164.80 ± 3661.43	15.00 to 14974.80	5978.96 ± 8289.96	14.70 to 40257.13
Conductivity (μS/cm)	422.46 ± 89.82	246.67 to 684.33	193.04 ± 85.19	2.54 to 500.55
Air relative humidity (%)	81.66 ± 5.66	60.96 to 92.95	60.54 ± 14.06	27.69 to 85.15
Air temperature (°C)	9.57 ± 6.82	0.56 to 31.72	26.42 ± 7.39	12.20 to 46.65

Note: C1 and C2 denote Campaigns 1 and 2; SD denotes sample standard deviation.

2.4 Operating-Condition Variation

Figure 2 shows daily temperature cycles in both campaigns and quite different moisture trajectories, so the data cover changing operating conditions rather than repeated visits to one state. The horizontal axis is the retained-record index divided by 24, which approximates days. Because removed records shorten the axis slightly, it preserves order but not exact elapsed time.

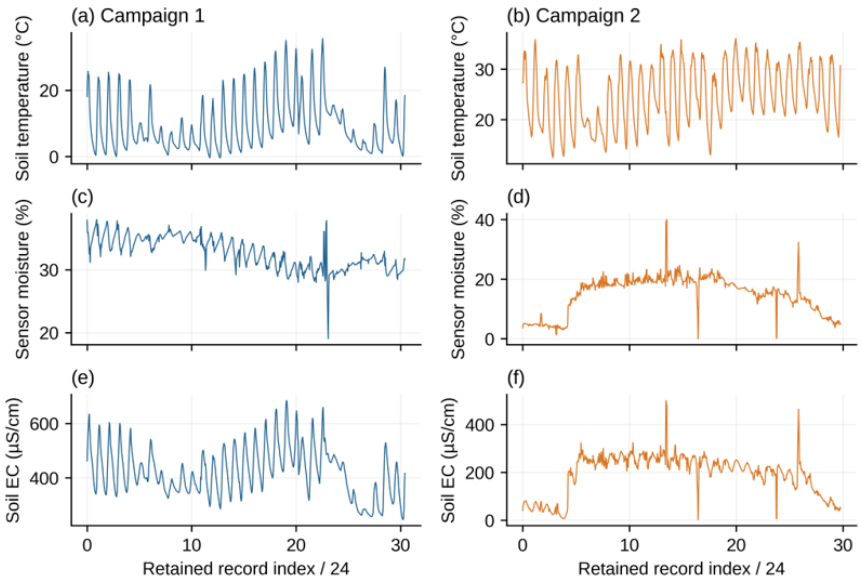


Figure 2. Soil temperature, sensor moisture and conductivity sequences in the two campaigns

Figure 3 and Table 2 show how strongly the inputs shifted between campaigns. From Campaign 1 to Campaign 2, mean air temperature rose from 9.57 to 26.42 °C, mean air relative humidity fell from 81.66% to 60.54%, mean illuminance rose from about 2,165 to 5,979 lux, and mean soil EC dropped from 422.46 to 193.04 μS/cm. Spring conditions were warmer, brighter and less humid, with a very different conductivity distribution. The data cannot isolate the physical causes of these changes, but they motivate two evaluation settings. Calibration on both campaigns tests reconstruction inside a represented domain, and calibration on Campaign 1 alone tests how the models cope with conditions they have not seen. The EC-free comparison adds a further question: can temperature or moisture

still be estimated once the shifting soil-probe input is removed?

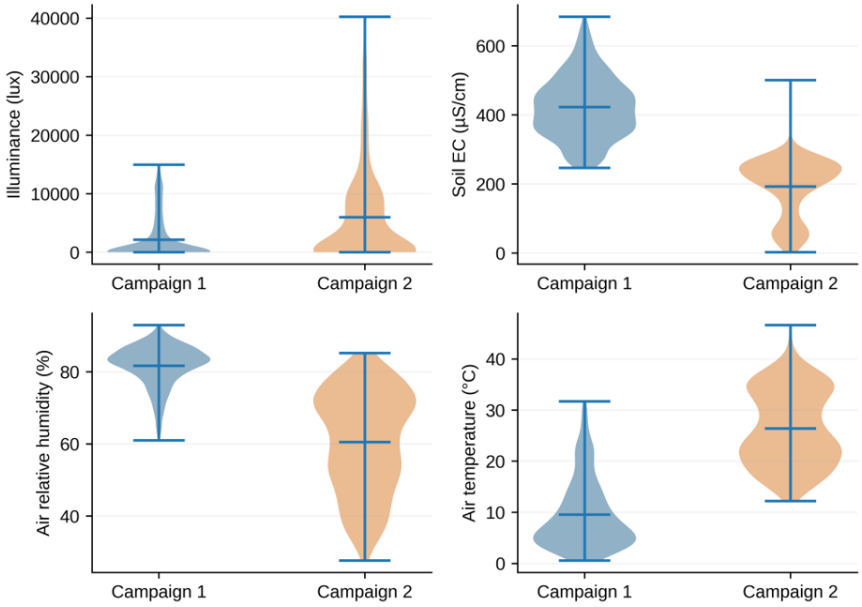


Figure 3. Distributions of the four predictor channels by campaign

3 Soft-Sensor Models and Evaluation Protocol

3.1 Same-Hour Estimation Task

For each retained hour, the predictor vector contains illuminance, soil EC, air relative humidity and air temperature, and separate single-output regressors are fitted for soil temperature and sensor-reported moisture (Eq. (1)). All quantities refer to the same hour. Campaign identity, timestamp, record index and lagged target values are excluded from the inputs, and reference measurements serve only for fitting, validation and evaluation. The task is thus current-state reconstruction. This matters for the air-side estimator in particular: if it is to be considered as a probe backup, it cannot rely on recent readings from that probe. The comparison was restricted to contemporaneous regressors. Recurrent models using histories of air-side predictors could also satisfy the input-channel constraint, but their potential contribution was not evaluated in this study.

$$\hat{y}_{k,t} = f_k\left(L_t, EC_t, RH_{a,t}, T_{a,t}\right), k \in \{T_s, M_s\} \tag{1}$$

where, L denotes illuminance, EC denotes soil electrical conductivity, RH_a denotes air relative humidity and T_a denotes air temperature; the subscript t identifies the same observation hour. The output index k identifies soil temperature, T_s, or sensor-reported moisture, M_s.

Two input configurations are compared: all four recorded channels, and the three air-side channels without EC. The gap between them measures what is lost by giving up a channel that may share hardware with the reference. Reference measurements are needed during development but never enter the inputs at inference.

3.2 Baselines and Ensemble Estimators

The mean baseline predicts the development-set average of each target. Ridge regression [42] provides a stronger, linear baseline on standardized inputs, with α selected from 0.1, 1, 10 and 100. Standardization is fitted on the active training data only and is wrapped inside the estimator pipeline. Together, the two baselines show how much of each target is already explained by a constant level or by a linear relationship.

Random forest (RF) averages regression trees fitted to bootstrap samples [43]. Each candidate used 200 trees, all available predictors as split candidates, and minimum leaf sizes of 1, 3 or 8. The four-channel and air-channel configurations contained four and three predictors, respectively. Regression splits minimized squared error. Extremely randomized trees (ET) also averaged 200 trees but used randomized thresholds and the complete training set for each tree, without bootstrap resampling; the same leaf-size candidates were evaluated [44].

Gradient boosting (GB) fitted 200 successive regression trees with a learning rate of 0.03, squared-error loss and a minimum leaf size of five [45]; the maximum tree depth was selected from 1, 2, and 3.

AdaBoost with support vector regression (AdaBoost–SVR, ABS) used support vector regression (SVR) as the base estimator in AdaBoost.R2 [46, 47]. The inputs were standardized before boosting. The radial-basis-function

SVR used $\varepsilon = 0.1$, γ = scale and C selected from 1, 10 and 100. AdaBoost used a maximum of 30 estimators, learning rate 0.5 and linear regression loss. Sample reweighting and weighted-median prediction followed the scikit-learn implementation. The averaging ensemble (VR) combined the separately selected RF, GB, and ET estimators with equal weights. Its composition was fixed before the heldout comparison.

All models were implemented in Python 3.12.14 using scikit-learn 1.8.0, NumPy 2.3.5 and pandas 2.2.3 [48]. The principal random seed was 17. Table 3 lists the candidate settings and the selected configurations. Each target was tuned separately; temperature and moisture were not required to share the same hyperparameters. No target scaling was applied.

Table 3. Candidate settings and validation-selected configurations for the principal four-input comparison

Model	Candidate Settings	Temperature Selection	Moisture Selection
Ridge	α: 0.1, 1, 10, 100	α = 10	α = 100
RF	200 trees; leaf: 1, 3, 8	Leaf = 1	Leaf = 8
GB	200 trees; rate 0.03; depth: 1, 2, 3	Depth = 3	Depth = 3
ABS	30 estimators; rate 0.5; SVR C: 1, 10, 100	C = 10	C = 1
ET	200 trees; leaf: 1, 3, 8	Leaf = 1	Leaf = 8
VR	Equal RF, GB and ET weights	Selected components	Selected components

Note: RF = random forest; GB = gradient boosting; ABS = AdaBoost with support vector regression (AdaBoost–SVR); SVR = support vector regression; ET = extremely randomized trees; VR = equal-weight averaging ensemble.

3.3 Ordered Development and Heldout Evaluation

The retained records of each campaign were divided, in their existing order, into the first 60% for training, the next 20% for validation and the final 20% for testing. Integer boundaries were the floor of $0.6n$ and $0.8n$. This produced 867 training records, 289 validation records and 290 heldout records. Campaign 1 contributed 438, 146 and 147 records, and Campaign 2 contributed 429, 143 and 143 records. The same boundaries were used for every model and both targets.

Candidate hyperparameters were selected by the pooled validation mean squared error. After selection, each estimator was refitted on the combined training and validation portions, comprising 1,156 records. Standardization was then refitted on those development records alone. The heldout portions were not used to choose hyperparameters or preprocessing. VR inherited the component configurations selected by the same procedure; no additional ensemble-weight optimization was performed.

The principal evaluation measures reconstruction within operating periods that are represented in development. Within each campaign, training precedes testing, although development data from the later campaign also contribute when the earlier campaign is evaluated. Transfer to an unrepresented period is tested separately in the Campaign 1-to-2 experiment of Section 3.6 [31, 32].

3.4 Error Metrics and Uncertainty Summaries

RMSE, mean absolute error (MAE) and the coefficient of determination (R^2) were calculated from the heldout reference–estimate pairs using Eqs. (2)–(4). R^2 was calculated from the residual and total sums of squares. Pearson correlation was computed separately in the numerical results. The corresponding mean squared error equals RMSE squared before rounding.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{2}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{4}$$

In Eqs. (2)–(4), y_i and $\hat{y}_i$ are the reference and estimated values for observation i, n is the number of evaluated observations, and $\bar{y}$ is the mean reference value in the corresponding evaluation set.

Pooled and campaign-specific metrics separate broad operating-level reconstruction from local tracking. A negative R^2 means that squared error exceeds the error of the evaluation-set mean. Differences between campaigns can dominate pooled variance, so R^2 is interpreted alongside error magnitude and baseline performance [49].

A moving-block bootstrap summarized fixed-model sampling variation in heldout RMSE [50]. In each of 1,000 replicates, blocks of 24 retained records were sampled with replacement separately within each campaign, concatenated and trimmed to the original test count. The 2.5th and 97.5th percentiles defined the reported limits. Because blocks follow retained-record order, excluded timestamps can make a block longer than 24 clock hours. The limits apply to aggregate RMSE for the observed campaigns; calibrating prediction intervals is a separate task [51, 52].

3.5 Input Ablation and Sensitivity Checks

Input ablation compared all four inputs, the three inputs without conductivity, and air temperature alone. Ridge, GB, ABS, and ET were refitted with the hyperparameters selected in the main analysis. Holding these settings fixed isolates the effect of removing inputs; re-optimizing the reduced-input models was outside the scope of this comparison.

Permutation importance for GB and ET was measured as the increase in heldout MAE over 20 shuffles of each predictor within its campaign, which keeps campaign membership intact. The diagnostic reflects how much a fitted model relies on each input, and its magnitude can be affected by correlated predictors.

Sensitivity to estimator randomization was evaluated at seeds 17, 29, 43, 71 and 101, with the selected hyperparameters and the same development-record order as in the principal analysis. For the embargo analysis, the last 24 development records before each campaign-specific test interval were removed and the remaining records kept their relative order. The hyperparameters from the principal analysis were retained, so this is a fixed-configuration sensitivity check and involves no re-selection on purged validation data. The test observations were unchanged, and models were refitted for each seed and embargo condition.

3.6 Transfer Between Monitoring Campaigns

A separate experiment trained only on Campaign 1 and evaluated every retained record of Campaign 2. Candidate configurations were selected on the first 60% and the following 20% of Campaign 1. The selected model was refitted on the first 80% of Campaign 1 and then applied to Campaign 2, which contributed no fitting observations, standardization statistics or hyperparameter-selection information. VR used equal weights and components chosen by the same Campaign 1 procedure. The design tests transfer to a different observed operating range without recalibration on the target campaign.

The transfer test uses all retained Campaign 2 records, whereas the principal test uses the final portions of both campaigns, so the two results refer to different calibration settings and evaluation populations. Changes in input range, target range, management and probe behaviour may all contribute to transfer error. The experiment measures performance without recalibration on the target campaign; it neither estimates how much recalibration data would be needed nor isolates the cause of the error.

3.7 Deployment Assessment Procedure

The deployment assessment fixed a model for each target and input configuration before any test evaluation. Ridge, RF, GB, ABS, and ET were tuned by validation mean squared error over the grids in Section 3.2, and the fixed equal-weight RF–GB–ET voter was scored on the same validation records. The family with the lowest validation RMSE was refitted on the development records in the principal record order and evaluated on the unchanged test portions. Four-input and three-input configurations were tuned independently with identical candidate grids. The result is a validation-selected comparison that can be read alongside the retrospective test rankings in Table 4.

For each configuration, the development-set minimum and maximum of every predictor defined a componentwise calibration envelope. A test record counted as inside the envelope only when all its predictors lay within these bounds, and error summaries were reported separately for inside and outside records. In the transfer analysis, model selection and envelope construction both used Campaign 1 alone before evaluation on Campaign 2. The check detects marginal extrapolation only; a record inside every bound may still have an unfamiliar combination of inputs.

Error-tolerance curves give the fraction of heldout absolute errors within 0.5, 1, 2, and 3 °C for temperature and within 1, 2, 3, and 5 percentage points for sensor moisture. The bands were fixed before testing and serve as illustrations. They are neither manufacturer accuracy limits nor validated irrigation thresholds.

The assessment record lists the required channels, reference definition, validation-selected model, development bounds and heldout error summaries. At run time, the available inputs yield an estimate and a range flag without any reference measurement; errors and tolerance exceedances can be computed only when a reference is available. Under the proposed handling sequence, a configuration is withheld when a required channel is missing, and out-of-range inputs are flagged for reference review. We assessed this sequence offline only and did not test its operational responses in the field. Appendix Table A2 reports fitting and batch prediction times.

Table 4. Combined heldout performance on 290 observations

Target	Model	RMSE	MAE	R^2	RMSE 95% Limits
Temperature	Mean	11.339	10.365	−0.0086	—
Temperature	Ridge	1.467	1.096	0.9831	1.105–1.926
Temperature	RF	0.814	0.576	0.9948	0.696–0.959
Temperature	GB	0.799	0.579	0.9950	0.670–0.967
Temperature	ABS	0.872	0.670	0.9940	0.781–0.951
Temperature	ET	0.771	0.563	0.9953	0.630–0.878
Temperature	VR	0.759	0.548	0.9955	0.631–0.900
Moisture	Mean	10.886	9.818	−0.1149	—
Moisture	Ridge	3.095	2.620	0.9099	2.526–3.590
Moisture	RF	4.843	3.599	0.7793	3.853–6.027
Moisture	GB	3.651	2.858	0.8746	3.132–4.384
Moisture	ABS	1.834	1.374	0.9684	1.275–2.253
Moisture	ET	2.088	1.623	0.9590	1.596–2.514
Moisture	VR	3.159	2.563	0.9061	2.886–3.630

Note: RMSE = root mean squared error; MAE = mean absolute error; RF = random forest; GB = gradient boosting; ABS = AdaBoost with support vector regression (AdaBoost–SVR); ET = extremely randomized trees; VR = equal-weight averaging ensemble. Temperature errors are in °C; moisture errors are in sensor-scale percentage points. Limits are fixed-model moving-block bootstrap percentiles. An em dash (—) denotes not applicable.

4 Results

4.1 Data Accounting and Operating Ranges

Screening retained 98.6% of the acquired records (Table 1). Temperature and moisture followed different temporal patterns (Figure 2), which justifies fitting and evaluating them separately. The environmental variables were observed rather than manipulated, so the relationships reported below are predictive associations and carry no causal interpretation.

4.2 Combined Heldout Performance

Table 4 lists the principal heldout results with moving-block bootstrap limits. For temperature, VR gave the lowest RMSE (0.759 °C), followed by ET (0.771 °C), GB (0.799 °C), RF (0.814 °C) and ABS (0.872 °C). Ridge regression reached 1.467 °C, which means VR reduced the error by 48.2%. The four tree-based ensembles lie within 0.06 °C of each other and their bootstrap limits overlap (Figure 4), so little weight should be placed on their exact order.

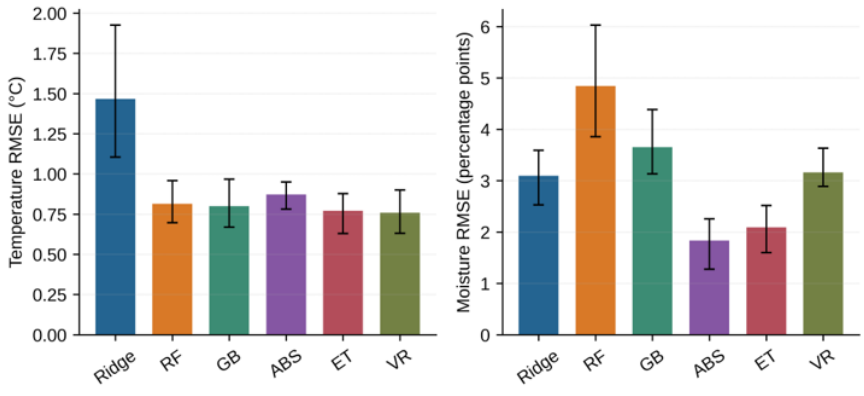


Figure 4. Combined test RMSE with 95% moving-block bootstrap limits

Note: RMSE = root mean squared error; RF = random forest; GB = gradient boosting; ABS = AdaBoost with support vector regression (AdaBoost–SVR); ET = extremely randomized trees; VR = equal-weight averaging ensemble.

For moisture, ABS had the lowest RMSE (1.834 percentage points; MAE 1.374), followed by ET (2.088), with ridge, VR, GB and RF further behind. ABS improved on ridge by 40.8%. The mean baselines were far worse for both targets, showing that the models capture more than a constant operating level. The best family differed between targets, and averaging three tree models (VR) helped for temperature but not for moisture.

Validation and test rankings did not agree. Among the individually tuned models, ABS had the lowest validation RMSE for temperature (0.982 °C) and ET for moisture (2.859 percentage points), whereas VR and ABS were best on the test set. This is why Section 4.7 reports validation-frozen choices rather than the test-best models.

4.3 Campaign-Specific Behaviour

Table 5 breaks the test results down by campaign. In the combined parity plots (Figure 5), temperature follows a single relationship, whereas moisture splits into campaign-specific clusters. Figure 6 keeps the heldout records in time order and reveals what the scatter plot hides: the temperature estimates follow the observed daily cycles, but the moisture estimates in Campaign 1 do not follow the small local changes.

Table 5. Campaign-specific heldout RMSE and prediction R^2

Target	Model	C1 RMSE	C1 R^2	C2 RMSE	C2 R^2
Temperature	Ridge	1.132	0.953	1.745	0.866
Temperature	RF	0.663	0.984	0.945	0.961
Temperature	GB	0.653	0.985	0.926	0.962
Temperature	ABS	0.702	0.982	1.017	0.954
Temperature	ET	0.645	0.985	0.881	0.966
Temperature	VR	0.627	0.986	0.874	0.966
Moisture	Ridge	2.927	−8.466	3.259	0.477
Moisture	RF	6.612	−47.314	1.620	0.871
Moisture	GB	4.966	−26.250	1.301	0.917
Moisture	ABS	2.181	−4.259	1.389	0.905
Moisture	ET	2.633	−6.661	1.309	0.916
Moisture	VR	4.267	−19.122	1.234	0.925

Note: RMSE = root mean squared error; RF = random forest; GB = gradient boosting; ABS = AdaBoost with support vector regression (AdaBoost–SVR); ET = extremely randomized trees; VR = equal-weight averaging ensemble. C1 and C2 denote Campaigns 1 and 2. Campaign 1: $n = 147$; Campaign 2: $n = 143$. Temperature RMSE is in °C and moisture RMSE is in sensor-scale percentage points.

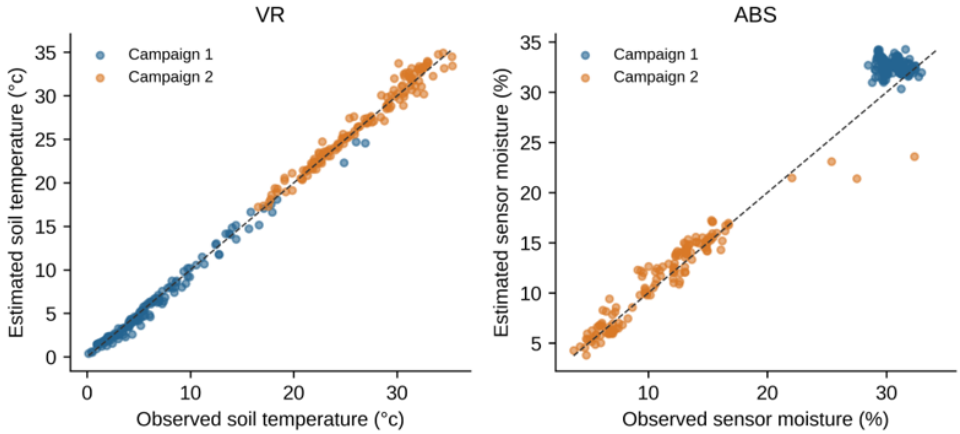


Figure 5. Observed–estimated pairs for VR temperature and ABS moisture predictions

Note: VR = equal-weight averaging ensemble; ABS = AdaBoost with support vector regression (AdaBoost–SVR).

In the Campaign 1 test interval, sensor moisture varied only between 28.45% and 32.93% (standard deviation 0.954 percentage points), and all five ensembles had negative R^2. Even the best of them, ABS, had an RMSE of 2.181 percentage points, more than twice the spread of the data around their mean. In Campaign 2, where moisture varied far more, the same models explained most of the within-campaign variance (R^2 = 0.87–0.93). The pooled moisture score thus gives a much more favourable picture than the Campaign 1 result alone.

The pooled moisture fit captures much of the level difference between campaigns, but the Campaign 1 errors are too large to show that the model responds to small local changes. Any irrigation application would need threshold-specific evaluation against a physically calibrated water-content reference.

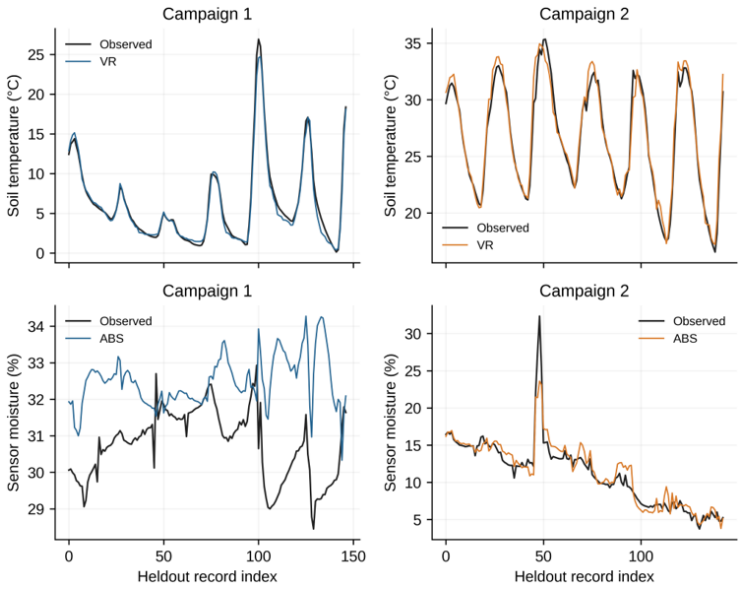


Figure 6. Campaign-specific test sequences for VR temperature and ABS moisture predictions

Note: VR = equal-weight averaging ensemble; ABS = AdaBoost with support vector regression (AdaBoost–SVR).

4.4 Dependence on Conductivity and Other Inputs

Table 6 and Figure 7 compare the input ablations. Temperature remained predictable from air-side measurements: removing EC raised the RMSE of GB, ABS, and ET by 0.10–0.14 °C. Moisture behaved very differently. Without EC, the RMSE of the three nonlinear models rose from 1.8–3.7 to 6.3–6.4 percentage points, and air temperature alone did no better. Channel requirements are target-specific, and a single sensor-importance ranking for both outputs would be misleading.

Table 6. RMSE after refitting with reduced input availability

Target	Model	All Inputs	No EC	Air Temperature Only
Temperature	Ridge	1.467	1.533	1.772
Temperature	GB	0.799	0.900	0.995
Temperature	ABS	0.872	1.016	0.962
Temperature	ET	0.771	0.884	1.403
Moisture	Ridge	3.095	7.328	7.417
Moisture	GB	3.651	6.429	7.120
Moisture	ABS	1.834	6.400	6.966
Moisture	ET	2.088	6.335	7.167

Note: RMSE = root mean squared error; EC = electrical conductivity; GB = gradient boosting; ABS = AdaBoost with support vector regression (AdaBoost–SVR); ET = extremely randomized trees. Temperature errors are in °C and moisture errors are in sensor-scale percentage points. Hyperparameters were not retuned after input removal.

Within-campaign permutation importance (Appendix Table A1) points the same way. Air temperature dominates the temperature models and EC dominates the moisture models. The two analyses ask different questions: ablation lets the model compensate after refitting, whereas permutation disrupts a channel in a model that has already been fitted. That they agree strengthens the case that the moisture estimates depend on EC, although neither analysis shows that EC independently measures soil water content.

4.5 Transfer to an Unrepresented Campaign

Table 7 reports the winter-to-spring transfer experiment. Temperature errors increased and the model ranking changed; GB and VR kept R^2 above 0.90, whereas ABS failed ($R^2 = -1.30$). Moisture transfer failed for every model, with RMSE between 17.8 and 22.2 percentage points and negative R^2 throughout. Figure 8 compares these values with the setting in which both campaigns contributed development data. A moisture model calibrated on Campaign 1 should not be used unchanged in the second operating regime.

This result bears directly on repeated use of one monitoring platform. Temperature, moisture and conductivity distributions all changed between campaigns, and several estimators had to extrapolate beyond the Campaign 1

range. Recalibration will need reference measurements collected under the new conditions, and comparing adaptive methods would require more labelled data.

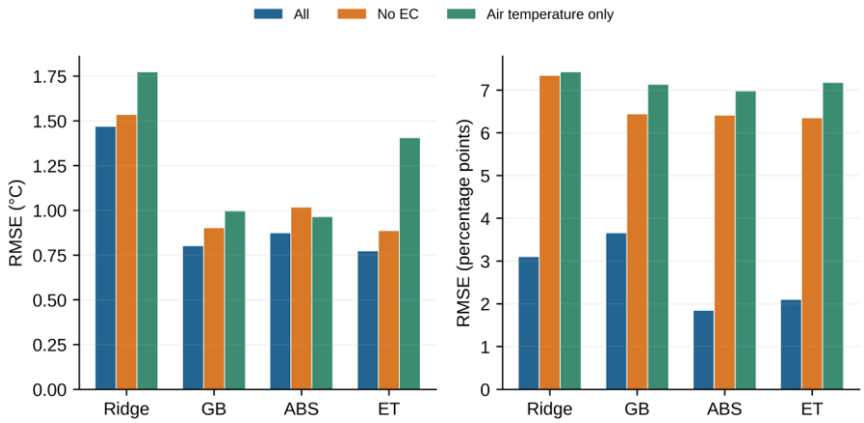


Figure 7. Input ablation with fixed main-analysis hyperparameters

Note: RMSE = root mean squared error; EC = electrical conductivity; GB = gradient boosting; ABS = AdaBoost with support vector regression (AdaBoost–SVR); ET = extremely randomized trees.

Table 7. Transfer from Campaign 1 development data to all 715 retained Campaign 2 observations

Model	Temp. RMSE	Temp. MAE	Temp. R^2	Moist. RMSE	Moist. MAE	Moist. R^2
Ridge	4.903	4.280	0.299	22.173	20.851	−11.824
RF	3.205	2.601	0.700	19.590	18.652	−9.010
GB	1.829	1.362	0.902	18.135	17.080	−7.578
ABS	8.885	7.284	−1.304	17.839	16.789	−7.300
ET	2.111	1.686	0.870	18.196	17.183	−7.637
VR	1.791	1.274	0.906	18.633	17.638	−8.057

Note: RMSE = root mean squared error; MAE = mean absolute error; RF = random forest; GB = gradient boosting; ABS = AdaBoost with support vector regression (AdaBoost–SVR); ET = extremely randomized trees; VR = equal-weight averaging ensemble; Temp. = temperature; Moist. = moisture. Temperature errors are in °C; moisture errors are in sensor-scale percentage points. Tuning and fitting used Campaign 1 alone.

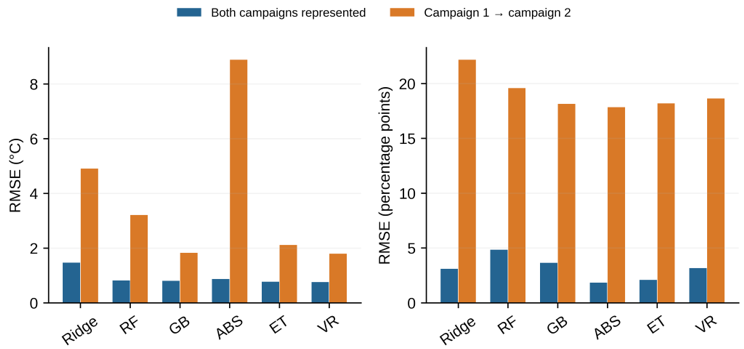


Figure 8. RMSE under represented-period calibration and Campaign 1-to-2 transfer

Note: RMSE = root mean squared error; MAE = mean absolute error; RF = random forest; GB = gradient boosting; ABS = AdaBoost with support vector regression (AdaBoost–SVR); ET = extremely randomized trees. VR = equal-weight averaging ensemble.

4.6 Randomization and Boundary Sensitivity

Table 8 and Figure 9 summarize the five-seed and embargo results. Removing the 24 development records nearest each test interval changed mean RMSE by at most 0.03 °C for temperature and 0.28 percentage points for moisture. ABS and ET remained more accurate than RF and GB for moisture, and the temperature ensembles remained closely grouped. The small changes under this embargo suggest that immediate boundary adjacency alone does not account for the main performance pattern.

Table 8. Five-seed sensitivity with and without a 24-record development embargo

Target	Model	No Embargo RMSE	24-Record Embargo RMSE
Temperature	RF	0.8184 ± 0.0045	0.8422 ± 0.0109
Temperature	GB	0.7993 ± 0.0000	0.8130 ± 0.0000
Temperature	ABS	0.8994 ± 0.0251	0.9246 ± 0.0147
Temperature	ET	0.7815 ± 0.0145	0.7969 ± 0.0116
Moisture	RF	4.8748 ± 0.0407	4.8964 ± 0.0423
Moisture	GB	3.6512 ± 0.0004	3.3692 ± 0.0003
Moisture	ABS	1.8166 ± 0.0198	1.8210 ± 0.0140
Moisture	ET	2.1119 ± 0.0422	2.1570 ± 0.0332

Note: RMSE = root mean squared error; RF = random forest; GB = gradient boosting; ABS = AdaBoost with support vector regression (AdaBoost–SVR); ET = extremely randomized trees. Entries are mean ± sample standard deviation across five seeds, rounded to four decimal places. Temperature errors are in °C and moisture errors are in sensor-scale percentage points. Development records preserve the relative order used in the principal analysis.

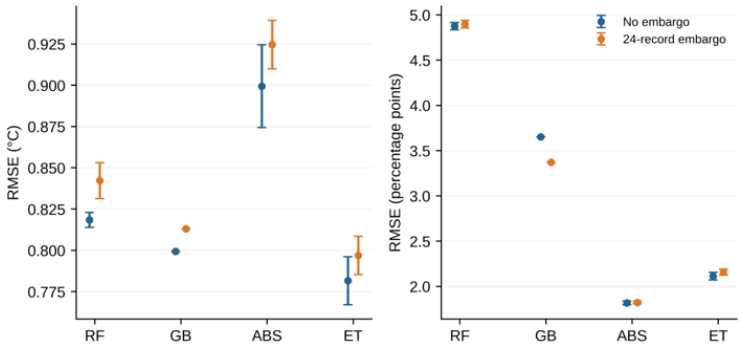


Figure 9. RMSE across five seeds with and without a 24-record embargo (mean ± SD)

Note: RMSE = root mean squared error; RF = random forest; GB = gradient boosting; ABS = AdaBoost with support vector regression (AdaBoost–SVR); ET = extremely randomized trees. SD denotes sample standard deviation.

These checks cover estimator randomness and boundary adjacency. Serial dependence remains within each test interval, and two campaigns span a limited range of environmental conditions. The larger shift to an unrepresented period is examined separately in Section 4.5.

4.7 Assessment Under Channel and Operating Range Constraints

Validation selected ABS for temperature and ET for moisture in both input configurations (Table 9). With four inputs, their test RMSEs were 0.872 °C and 2.088 percentage points; without EC, they were 1.016 °C and 6.335 percentage points, increases of 16.5% and 203.4%, respectively (Figure 10). Unlike the retrospective rankings in Table 4, these values come from choices fixed before the test data were examined.

Table 9. Validation-frozen estimator selection under two input configurations

Inputs	Target	Model	Validation RMSE	Test RMSE	C1 Test R^2	C2 Test R^2
Four channels	Temperature	ABS	0.982	0.872	0.982	0.954
Four channels	Moisture	ET	2.859	2.088	−6.661	0.916
Air channels	Temperature	ABS	1.355	1.016	0.980	0.932
Air channels	Moisture	ET	5.136	6.335	−11.159	−2.449

Note: RMSE = root mean squared error; ABS = AdaBoost with support vector regression (AdaBoost–SVR); ET = extremely randomized trees. C1 and C2 denote Campaigns 1 and 2. Temperature RMSE is in °C; moisture RMSE is in sensor-scale percentage points. Air channels comprise illuminance, air temperature and air relative humidity. Family selection used validation records alone.

The componentwise range check placed 288 of the 290 principal test records inside the development envelope for both configurations (Table 10). Two outside records are too few for a reliable estimate of out-of-range error. More telling is that this near-complete coverage coexisted with negative Campaign 1 moisture R^2: the inputs passed the range check while local moisture changes went untracked.

With Campaign 1 as the only calibration source, the four-input envelope contained just 10 of the 715 Campaign

2 records (98.6% outside), whereas the air-side envelope contained 358 (49.9% outside). Dropping EC removes one constraint from the componentwise test, so the higher coverage reveals nothing about transferability on its own. Moisture RMSE was in fact still 15.9 percentage points for records inside the air-side envelope.

Figure 11 shows, for each campaign and configuration, the share of estimates within each error band. A designer can compare these curves with a tolerance set independently for the intended decision. The curves do not define an irrigation tolerance, and an inside-range flag should not be read as the probability that an estimate is accurate.

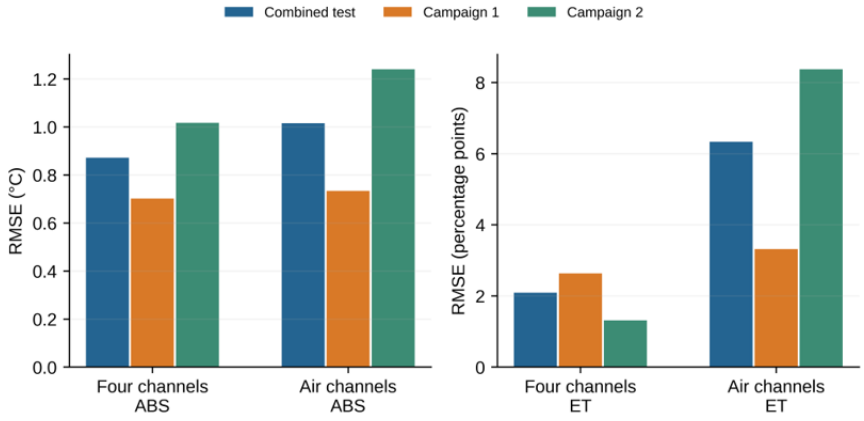


Figure 10. Test RMSE of validation-selected models under four-input and air-side configurations

Note: RMSE = root mean squared error; ABS = AdaBoost with support vector regression (AdaBoost–SVR); ET = extremely randomized trees. Temperature RMSE is in °C; moisture RMSE is in sensor-scale percentage points. Air channels comprise illuminance, air temperature and air relative humidity. Family selection used validation records alone.

Table 10. Calibration-envelope coverage and conditional errors of validation-selected models

Calibration setting	Inputs	Target	Inside n	Outside n	Inside RMSE	Outside RMSE
Both periods	Four channels	Temperature	288	2	0.870	1.134
Both periods	Four channels	Moisture	288	2	2.073	3.653
Both periods	Air channels	Temperature	288	2	1.014	1.278
Both periods	Air channels	Moisture	288	2	6.246	14.200
Period 1 → 2	Four channels	Temperature	10	705	3.540	4.920
Period 1 → 2	Four channels	Moisture	10	705	11.313	19.682
Period 1 → 2	Air channels	Temperature	358	357	2.092	7.116
Period 1 → 2	Air channels	Moisture	358	357	15.936	16.891

Note: RMSE = root mean squared error. Both periods refers to the pooled development calibration and 290 heldout records. Period 1 → 2 uses Campaign 1 for tuning and fitting and all 715 Campaign 2 records for evaluation. Temperature errors are in °C and moisture errors are in percentage points. Campaign 1-only validation selected Ridge for temperature and random forest for moisture in both configurations. The principal outside-envelope group contains only two records.

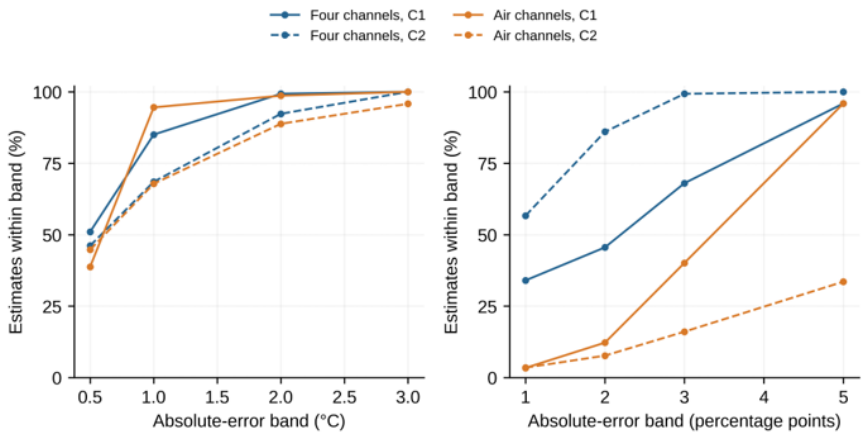


Figure 11. Heldout coverage within illustrative error bands by campaign and input configuration

Note: C1 and C2 denote Campaigns 1 and 2.

5 Discussion

5.1 Information Path and the Role of Each Estimator

The role a soft sensor can play depends on its information path. The four-input models take EC from the same soil probe that supplies the references, while the air-side models avoid this dependency. Their combined temperature RMSE stayed close to 1 °C, which makes an air-side backup worth evaluating further within the represented operating conditions. For moisture, removing EC roughly tripled the RMSE. The four-input moisture model was more accurate overall, yet its poor tracking in Campaign 1 limits where it can be used. Comparing the probe's moisture reading with an EC-based reconstruction might serve as a probe-consistency check, although shared measurement dependencies could also hide faults. Fault-labelled reference data would be needed to test such a role.

Using a different predictor set does not by itself demonstrate that an estimator can serve as a backup instrument. Shared power, communications or compensation can affect several channels at once, and the campaigns contain no labelled device failures. Studies of sensor fusion [53] and adaptive greenhouse control [54] both emphasize examining the complete acquisition chain. The EC-free analysis quantifies one predictor constraint; backup performance has to be assessed with reference measurements taken during the relevant failure conditions.

In an IoT monitoring system, each estimate can be displayed with its channel status, model identity and range status [55], so that operators can see what it is based on. The timings in Appendix A2 cover algorithmic workload only. Acquisition, communication reliability and end-to-end latency are properties of the deployed system and must be measured there. Direct measurements should remain in service until operational error limits and response procedures have been validated.

5.2 Separating Channel, Range and Error Evidence

The assessment breaks the deployment decision into three engineering questions. Required-channel status decides whether a configured model can run at all. If EC is unavailable, a separately evaluated air-side configuration is needed; our results give no support for feeding an imputed EC value into the four-input model. Development bounds flag marginal extrapolation, and heldout errors can then be compared with a tolerance derived from the intended use. Each answer belongs in the supervisory record, since none of them can stand in for another.

Our results show why this separation matters. Almost all principal test inputs passed the range check, yet moisture tracking in Campaign 1 was poor. Conversely, dropping EC raised range coverage during transfer simply because one bound disappeared, while moisture errors stayed large. Range status should prompt a calibration review; it says nothing definite about accuracy. The handling sequence we propose is to withhold any configuration with missing channels, attach a range flag to every estimate that can be produced, and request reference measurements when unfamiliar conditions appear.

5.3 Engineering Transfer to Production Monitoring

Table 11 links each result to a production-monitoring decision. What carries over to other installations is the link between channel definition, calibration evidence and an explicit monitoring requirement; the fitted models stay with this platform. A commercial greenhouse, for instance, could apply the sequence to one cultivation zone with a defined substrate, probe set and acquisition path. The operator would align the permitted predictors with calibrated references at the required depth, position and sampling interval, hold out whole production periods or zones for testing, and log irrigation, ventilation, crop stage and salinity changes to explain coverage gaps.

Before a backup role is assigned, a channel map should show which probes, compensation circuits, power supplies and communication links are shared. While the soil-probe inputs remain available, a reconstruction can be compared with its own reference channel. If the estimator is instead meant to provide an alternative information path after a probe fails, that configuration needs its own calibration and testing during such events. In our case, excluding EC cost little for temperature but a great deal for moisture, and a common replacement policy for both targets would have hidden this difference.

Operating limits are likewise tied to the target installation. Its own development records set the predictor bounds, and the reference uncertainty together with the intended decision sets the acceptable error against which heldout tolerance curves are compared. Where the estimate is to trigger alarms or automatic actions, event records are needed to quantify false alarms, missed events, delays and fallback behaviour.

Process plants raise the same questions about predictor availability, reference definition and changing operating conditions, for example in sugar crystallization [3] and in automated soft-sensor development [1]. An estimator that draws on a signal from the analyser it is meant to back up has a similar information dependency. The assessment sequence can organize the evidence needed in such cases, but the greenhouse models and numerical limits reported here cannot be carried over to those processes.

Table 11. Deployment decisions supported by the case study and requirements for production use

Engineering Decision	Evidence From This Case	Production-Site Requirement
Information role	Four-input reconstruction uses soil EC; the air-channel configuration excludes soil-probe predictors.	Map shared devices, compensation, power and communications before assigning a backup role.
Configuration choice	Validation-selected temperature RMSE: 0.872 to 1.016 °C without EC; moisture: 2.088 to 6.335 percentage points.	Calibrate each intended configuration with local references and test its required availability conditions.
Operating coverage	Principal coverage: 288/290 records. Transfer coverage: 10/715 with four inputs and 358/715 with air inputs.	Establish local bounds and evaluate whole withheld operating periods or zones.
Monitoring tolerance	Figure 11 reports empirical error-band fractions; high range coverage coexists with poor local moisture tracking.	Set tolerances from the intended decision and reference uncertainty, then assess heldout errors.
Exception response	The case supplies offline estimates and range flags, without labelled device-failure or alarm trials.	Validate reference checks, missed events, false alarms, delays and fallback responses before automated use.

Note: RMSE = root mean squared error; EC = electrical conductivity. Error values refer to the validation-selected models in Table 9. Production-site requirements specify local calibration and operational verification.

5.4 Why Pooled Accuracy and Model Rankings Can Mislead

The moisture results show how a high pooled R^2 can hide poor tracking within one campaign. Mean moisture differed by almost 18 percentage points between the campaigns, and reconstructing that difference contributes heavily to the pooled score even when the smaller Campaign 1 changes are missed. Irrigation thresholds would have to be evaluated within the relevant moisture range and on a physically calibrated scale.

For temperature, several ensembles performed almost equally well: VR, ET, GB, and RF were separated by less than 0.06 °C, with overlapping bootstrap limits. Because the rankings also changed between validation and testing, model choice should weigh consistency across monitoring periods as well as the lowest error on a single split.

EC clearly improves the reconstruction of sensor moisture, but it may encode a relationship specific to the probe and its surroundings. Work on knowledge-informed agricultural modelling has argued that fitted structure should be tied to the process it represents [56]. We regard the EC dependence as a property to be reported and managed, and we do not take it as evidence that the estimator measures water content independently. For the EC-free temperature task, most of the useful information comes from air temperature.

5.5 Measurement Calibration and Operating-Condition Shift

Bulk soil conductivity depends jointly on water content, temperature, salinity and texture, and a single uncalibrated curve generally cannot separate these factors [57]. The association between EC and the probe's moisture output may thus reflect both soil conditions and the probe's internal compensation. An independent water-content calibration is required before a percentage-point error can be translated into a water-storage error or an irrigation volume.

The transfer experiment puts an empirical limit on the fitted moisture relationship, since winter calibration did not give reliable reconstruction in spring. Changes in the moisture range, the EC–moisture relationship, management and sensor behaviour are all plausible contributors. Logging irrigation, ventilation, soil amendments and probe recalibration alongside the measurements would help to distinguish these effects.

Sensor placement also affects how representative the reference is; greenhouse experiments have shown its influence on recorded temperature distributions [58]. Our evaluation used a single probe position in one small platform. A multi-position reference array and measurements at larger structural scales would be needed to judge how well the estimates represent a production greenhouse.

5.6 Uncertainty and the Boundary Between Estimation and Control

The bootstrap limits describe uncertainty in aggregate test RMSE, conditional on the fitted models and the observed sequences. Intervals for individual predictions need a separate calibration step. Conformal prediction [59] and prediction-interval soft sensors [60] are candidate approaches, although how well they stay calibrated under serial dependence and shifting conditions would itself have to be tested.

This study covered estimation only, without actuation. A control study would have to define irrigation or climate thresholds, delays, constraints and fallback behaviour, and then test the complete loop. Closed-loop experiments of

this kind must come before these estimates are used in autonomous irrigation or climate control.

5.7 Study Limitations and Reproducibility

The study is limited to two campaigns, one probe position and one laboratory-scale greenhouse. It includes no independent water-content calibration and no replicated comparison across soil types, and probe-specific calibration uncertainty was unavailable. Soil texture, crop status and irrigation or ventilation interventions were not recorded as explanatory variables, so their contributions to the differences between campaigns cannot be separated. The hourly sampling cannot resolve within-hour dynamics, and recurrent models using air-side histories were not evaluated. The conclusions therefore apply to the observed input–reference relationships; broader use would require more sites, probe types, operating records and calibrated references.

All evaluations were offline. The error bands are sensitivity values, and the records contain no labelled device failures or alarm events. Operational acceptance will require task-specific limits and event-based validation at the intended installation.

Supplementary Data S1 provides the source measurements, exclusions, ordered partitions, model configurations and saved predictions, together with executable analysis scripts. It includes an unscreened sensitivity analysis whose partition boundaries follow the unscreened record counts. No synthetic observations were added to either analysis.

6 Conclusions

This study evaluated a soft-sensor assessment procedure that reports required channels, marginal calibration coverage and heldout error separately. In the represented greenhouse periods, removing EC increased the validation-selected temperature RMSE from 0.872 to 1.016 °C (16.5%) and the moisture RMSE from 2.088 to 6.335 percentage points (203.4%). Air-side inputs retained useful information for temperature reconstruction, whereas the tested moisture models relied heavily on the soil-probe EC channel.

Near-complete range coverage did not ensure accurate local tracking. The principal calibration envelope contained 99.3% of the test inputs, yet winter moisture R^2 was negative, and winter-only calibration led to spring moisture RMSEs above 17 percentage points for all six regressors. Range flags are useful for identifying conditions that need a calibration review, but whether an estimator suits a monitoring task has to be judged from reference-based errors against a stated tolerance.

The assessment sequence can be applied at other installations, but the fitted relationships and numerical limits require local evaluation. Production use calls for local calibration, documented acquisition dependencies and testing under the operating changes relevant to the intended task. Further validation should involve multiple probes and sites with independent water-content references, followed by fault-event and closed-loop evaluation wherever a backup or automatic-control role is proposed.

Author Contributions

Conceptualization, F.C.W., J.Q.Z., and Z.Y.F.; methodology, W.Y.S. and J.Z.N.; software, F.C.W., J.Q.Z., and Z.Y.F.; validation, F.C.W., W.Y.S., and J.Z.N.; formal analysis, J.Q.Z. and Z.Y.F.; investigation, F.C.W. and W.Y.S.; resources, J.Q.Z. and J.Z.N.; data curation, Z.Y.F. and J.Z.N.; writing—original draft preparation, F.C.W., Z.Y.F., and J.Z.N.; writing—review and editing, J.Q.Z., Z.Y.F., and W.Y.S.; visualization, F.C.W. and W.Y.S.; supervision, J.Q.Z. and J.Z.N.; project administration, Z.Y.F. and W.Y.S.; funding acquisition, J.Q.Z. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Heilongjiang Provincial Natural Science Foundation Joint Fund Cultivation Project (Grant No.: PL2026E124), the Research Project on Ecological Environmental Protection in Heilongjiang Province (Grant No.: HST2025GF002), and the Doctoral Research Startup Fund (Grant No.: XDB202205).

Data Availability

The source measurements, screening logs, retained observations, ordered partitions, analysis and figure-generation scripts, model settings, heldout predictions and numerical results are supplied in Supplementary Data S1. Source time labels are preserved as exported. The supplement documents software dependencies, random seeds and the files underlying each manuscript table and figure.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] L. Cao, J. Su, E. Conde, L. C. Siang, Y. Cao, and B. Gopaluni, "A novel automated soft sensor design tool for industrial applications based on machine learning," *Control Eng. Pract.*, vol. 160, p. 106322, 2025. https://doi.org/10.1016/j.conengprac.2025.106322

[2] X. Wang, X. Qi, and Y. Zhang, "Deep learning with local spatiotemporal structure preserving for soft sensor development of complex industrial processes," *Appl. Soft Comput.*, vol. 164, p. 111974, 2024. https://doi.org/10.1016/j.asoc.2024.111974

[3] H. Morales, F. di Sciascio, E. Aguirre-Zapata, and A. N. Amicarelli, "A model-based supersaturation estimator (inferential or soft-sensor) for industrial sugar crystallization process," *J. Process Control*, vol. 129, p. 103065, 2023. https://doi.org/10.1016/j.jprocont.2023.103065

[4] M. Ni and S. Li, "Adaptive cascade enhancement broad learning system combined with stacked correlation information autoencoder for soft sensor modeling of industrial process," *Comput. Chem. Eng.*, vol. 177, p. 108324, 2023. https://doi.org/10.1016/j.compchemeng.2023.108324

[5] X. Zhang, C. Song, J. Zhao, and D. Xia, "Gaussian mixture continuously adaptive regression for multimode processes soft sensing under time-varying virtual drift," *J. Process Control*, vol. 124, pp. 1–13, 2023. https://doi.org/10.1016/j.jprocont.2023.02.003

[6] Z. H. Wang, Y. T. Li, and F. C. Wen, "A novel in-line polymer melt viscosity sensing system of integrated soft sensor and machine learning," *IEEE Sens. J.*, vol. 23, no. 11, pp. 12 181–12 189, 2023. https://doi.org/10.1109/jsen.2023.3267682

[7] M. Hassnain, S. M. W. Lee, and M. R. Azhar, "Soft sensing of biological oxygen demand in industrial wastewater using machine learning models," *J. Water Process Eng.*, vol. 78, p. 108699, 2025. https://doi.org/10.1016/j.jwpe.2025.108699

[8] W. Liu, Y. Li, M. Yang, K. Pang, Z. Xu, M. Yao, Y. Bai, and F. Zhang, "Microclimate prediction of solar greenhouse with pad–fan cooling systems using a machine and deep learning approach," *Agriculture*, vol. 15, no. 20, p. 2107, 2025. https://doi.org/10.3390/agriculture15202107

[9] M. E. Alaoui, L. O. Chahidi, M. Rougui, A. Mechaqrane, and S. Allal, "Evaluation of CFD and machine learning methods on predicting greenhouse microclimate parameters with the assessment of seasonality impact on machine learning performance," *Sci. Afr.*, vol. 19, p. e01578, 2023. https://doi.org/10.1016/j.sciaf.2023.e01578

[10] W. J. Choi and M. Yang, "Probabilistic deep learning framework for greenhouse microclimate prediction with time-varying uncertainty and covariance analysis," *Agriculture*, vol. 15, no. 23, p. 2461, 2025. https://doi.org/10.3390/agriculture15232461

[11] N. Averbuch and M. Moshelion, "Beyond VPD: multi-sensor, sub-hourly microclimate improves machine-learning prediction of greenhouse evaporation in a Mediterranean semi-arid greenhouse," *J. Hydrol. Reg. Stud.*, vol. 65, p. 103324, 2026. https://doi.org/10.1016/j.ejrh.2026.103324

[12] D. Tian, J. Shi, H. Zhang, X. Wei, A. Zhao, and J. Feng, "Regularized extreme learning machine-based temperature prediction for edible fungi greenhouse," *Appl. Eng. Agric.*, vol. 40, no. 4, pp. 483–499, 2024. https://doi.org/10.13031/aea.15767

[13] S. Rezaei Melal, M. Aminian, and S. M. Shekarian, "A machine learning method based on stacking heterogeneous ensemble learning for prediction of indoor humidity of greenhouse," *J. Agric. Food Res.*, vol. 16, p. 101107, 2024. https://doi.org/10.1016/j.jafr.2024.101107

[14] K. Belouz and S. Zereg, "Extreme learning machine for soil temperature prediction using only air temperature as input," *Environ. Monit. Assess.*, vol. 195, no. 8, p. 962, 2023. https://doi.org/10.1007/s10661-023-11566-2

[15] M. Baumberger, B. Haas, S. Sivakumar, M. Ludwig, N. Meyer, and H. Meyer, "High-resolution soil temperature and soil moisture patterns in space, depth and time: An interpretable machine learning modelling approach," *Geoderma*, vol. 451, p. 117049, 2024. https://doi.org/10.1016/j.geoderma.2024.117049

[16] M. M. Sazzad, T. Ahmed, G. Kibria, and I. Khan, "IoT based soil moisture measurement and type prediction using advanced regression and machine learning models," *Sci. Rep.*, vol. 15, no. 1, p. 35730, 2025. https://doi.org/10.1038/s41598-025-19444-2

[17] A. Ali, "Machine learning approaches for soil moisture prediction: Enhancing agricultural water management with integrated data," *Model. Earth Syst. Environ.*, vol. 12, no. 1, p. 16, 2026. https://doi.org/10.1007/s40808-025-02607-5

[18] S. Ye, Y. Chai, J. Li, J. Wang, X. Deng, and Q. Ran, "Explainable transfer learning for subsurface soil moisture prediction," *J. Hydrol.*, vol. 661, p. 133473, 2025. https://doi.org/10.1016/j.jhydrol.2025.133473

[19] A. Fragkos, D. Loukatos, G. Kargas, and K. G. Arvanitis, "Response of the TEROS 12 soil moisture sensor under different soils and variable electrical conductivity," *Sensors*, vol. 24, no. 7, p. 2206, 2024. https://doi.org/10.3390/s24072206

[20] P. A. Petsetidi and G. Kargas, "Evaluation of a multivariate calibration model for the WET sensor that incorporates apparent dielectric permittivity and bulk soil electrical conductivity," *Land*, vol. 13, no. 9, p. 1490, 2024. https://doi.org/10.3390/land13091490

[21] J. H. Barnard, J. A. Edeh, L. D. van Rensburg, and C. C. du Preez, "Calibration of an apparent electrical conductivity sensor using capacitance probes for determining soil water content," *Soil Sci. Soc. Am. J.*, vol. 89, no. 1, p. e20799, 2025. https://doi.org/10.1002/saj2.20799

[22] Y. Kojima, M. Matsuoka, T. Ariki, and T. Yoshioka, "Time domain transmissiometry-based sensor for simultaneously measuring soil water content, electrical conductivity, temperature, and matric potential," *Sensors*, vol. 23, no. 4, p. 2340, 2023. https://doi.org/10.3390/s23042340

[23] J. S. Witkowski and A. F. Grobelny, "Dual-frequency soil moisture meter method for simultaneous estimation of soil moisture and conductivity," *Sensors*, vol. 24, no. 10, p. 2969, 2024. https://doi.org/10.3390/s24102969

[24] A. Chandel, D. Swami, and N. Joshi, "Calibration complexities: Full-scale error impact and simultaneous variation of salinity, temperature, and moisture content on sensor performance in soil," *Environ. Dev. Sustain.*, vol. 27, no. 11, pp. 26 823–26 843, 2025. https://doi.org/10.1007/s10668-024-04812-1

[25] Y. S. Zakaria, S. Chen, T. A. Adongo, G. Kranjac-Berisavljevic, and H. Larijani, "Soil-specific calibration and integration of low-cost capacitive soil moisture sensors into a solar-powered sensor node," *Sensors*, vol. 26, no. 13, p. 3979, 2026. https://doi.org/10.3390/s26133979

[26] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," *Patterns*, vol. 4, no. 9, p. 100804, 2023. https://doi.org/10.1016/j.patter.2023.100804

[27] A. Apicella, F. Isgrò, and R. Prevete, "Don't push the button! Exploring data leakage risks in machine learning and transfer learning," *Artif. Intell. Rev.*, vol. 58, no. 11, p. 339, 2025. https://doi.org/10.1007/s10462-025-11326-3

[28] J. Leukel, L. Scheurer, and T. Zimpel, "Overinterpretation of evaluation results in machine learning studies for maize yield prediction: A systematic review," *Comput. Electron. Agric.*, vol. 230, p. 109892, 2025. https://doi.org/10.1016/j.compag.2024.109892

[29] L. A. Yates, Z. Aandahl, S. A. Richards, and B. W. Brook, "Cross validation for model selection: A review with examples from ecology," *Ecol. Monogr.*, vol. 93, no. 1, p. e1557, 2023. https://doi.org/10.1002/ecm.1557

[30] M. Lai, S. S. Wulff, Y. Cao, T. J. Robinson, and R. Rajapaksha, "Temporal cross-validation in forecasting: A case study of COVID-19 incidence using wastewater data," *Qual. Reliab. Eng. Int.*, vol. 41, no. 2, pp. 672–688, 2025. https://doi.org/10.1002/qre.3686

[31] D. R. Roberts, V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, J. J. Lahoz-Monfort, B. Schröder, W. Thuiller *et al.*, "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," *Ecography*, vol. 40, no. 8, pp. 913–929, 2017. https://doi.org/10.1111/ecog.02881

[32] C. Bergmeir, R. J. Hyndman, and B. Koo, "A note on the validity of cross-validation for evaluating autoregressive time series prediction," *Comput. Stat. Data Anal.*, vol. 120, pp. 70–83, 2018. https://doi.org/10.1016/j.csda.2017.11.003

[33] X. Zhou, Q. Liu, D. Katzin, T. Qian, E. Heuvelink, and L. F. M. Marcelis, "Boosting the prediction accuracy of a process-based greenhouse climate-tomato production model by particle filtering and deep learning," *Comput. Electron. Agric.*, vol. 211, p. 107980, 2023. https://doi.org/10.1016/j.compag.2023.107980

[34] B. Morcego, W. Yin, S. Boersma, E. van Henten, V. Puig, and C. Sun, "Reinforcement learning versus model predictive control on greenhouse climate control," *Comput. Electron. Agric.*, vol. 215, p. 108372, 2023. https://doi.org/10.1016/j.compag.2023.108372

[35] S. Mallick, F. Airaldi, A. Dabiri, C. Sun, and B. De Schutter, "Reinforcement learning-based model predictive control for greenhouse climate control," *Smart Agric. Technol.*, vol. 10, p. 100751, 2025. https://doi.org/10.1016/j.atech.2024.100751

[36] B. van Laatum, S. Msaad, E. J. van Henten, R. D. Mcallister, and S. Boersma, "Stochastic model predictive control with reinforcement learning for greenhouse production systems under parametric uncertainty," *Control Eng. Pract.*, vol. 169, p. 106787, 2026. https://doi.org/10.1016/j.conengprac.2026.106787

[37] J. Hou, Y. Li, Z. Sun, H. Wang, M. Lu, J. Hu, and H. Wu, "A cooperative regulation method for greenhouse soil moisture and light using Gaussian curvature and machine learning algorithms," *Comput. Electron. Agric.*, vol. 215, p. 108452, 2023. https://doi.org/10.1016/j.compag.2023.108452

[38] E. Gerampinis, F. Kalafati, N. F. Matsatsinis, and G. Arampatzis, "Towards digital twin-enabled greenhouse management: Enhancing climate resilience and sustainability through machine learning-driven controlled environmental agriculture," *Sustain. Comput. Inform. Syst.*, vol. 50, p. 101343, 2026. https://doi.org/10.1016/j.suscom.2026.101343

[39] A. Costantino, A. L. Ghiberti, C. Bibbiani, A. Barbaresi, and E. Fabrizio, "Advancing digital twin development for controlled-environment agriculture: Comparative evaluation of two greenhouse energy models and their

readiness for digital twin implementation," *Smart Agric. Technol.*, vol. 14, p. 102009, 2026. https://doi.org/10.1016/j.atech.2026.102009

[40] R. Gund, C. M. Badgujar, S. Samiappan, and S. Jagadamma, "Application of digital twin technology in smart agriculture: A bibliometric review," *Agriculture*, vol. 15, no. 17, p. 1799, 2025. https://doi.org/10.3390/agriculture15171799

[41] J. Frontzek, Z. Wagner, and S. Streif, "Dynamic, adaptive and modular digital twin framework for resource-efficient controlled environment agriculture," *Front. Plant Sci.*, vol. 17, p. 1864757, 2026. https://doi.org/10.3389/fpls.2026.1864757

[42] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970. https://doi.org/10.1080/00401706.1970.10488634

[43] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001. https://doi.org/10.1023/a:1010933404324

[44] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, 2006. https://doi.org/10.1007/s10994-006-6226-1

[45] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001. https://doi.org/10.1214/aos/1013203451

[46] H. Drucker, "Improving regressors using boosting techniques," in *Proceedings of the 14th International Conference on Machine Learning*, San Francisco, USA, 1997, pp. 107–115. https://mlanthology.org/icml/1997/drucker1997icml-improving/

[47] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, 2004. https://doi.org/10.1023/b:stco.0000035301.49549.88

[48] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011. https://jmlr.org/papers/v12/pedregosa11a.html

[49] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, p. e623, 2021. https://doi.org/10.7717/peerj-cs.623

[50] H. R. Kunsch, "The jackknife and the bootstrap for general stationary observations," *Ann. Stat.*, vol. 17, no. 3, pp. 1217–1241, 1989. https://doi.org/10.1214/aos/1176347265

[51] P. Pernot, "Calibration in machine learning uncertainty quantification: Beyond consistency to target adaptivity," *APL Mach. Learn.*, vol. 1, no. 4, p. 046121, 2023. https://doi.org/10.1063/5.0174943

[52] H. Paasche and I. Schröter, "Quantification of data-related uncertainty of spatially dense soil moisture patterns on the small catchment scale estimated using unsupervised multiple regression," *Vadose Zone J.*, vol. 22, no. 4, p. e20258, 2023. https://doi.org/10.1002/vzj2.20258

[53] R. F. Brena, A. A. Aguileta, L. A. Trejo, E. Molino-Minero-Re, and O. Mayora, "Choosing the best sensor fusion method: A machine-learning approach," *Sensors*, vol. 20, no. 8, p. 2350, 2020. https://doi.org/10.3390/s20082350

[54] G. Nicolosi, R. Volpe, and A. Messineo, "An innovative adaptive control system to regulate microclimatic conditions in a greenhouse," *Energies*, vol. 10, no. 5, p. 722, 2017. https://doi.org/10.3390/en10050722

[55] Y. Song, J. Bi, and X. Wang, "Design and implementation of intelligent monitoring system for agricultural environment in IoT," *Internet Things*, vol. 25, p. 101029, 2024. https://doi.org/10.1016/j.iot.2023.101029

[56] M. von Bloh, D. Lobell, and S. Asseng, "Knowledge informed hybrid machine learning in agricultural yield prediction," *Comput. Electron. Agric.*, vol. 227, p. 109606, 2024. https://doi.org/10.1016/j.compag.2024.109606

[57] S. P. Friedman, "Soil properties influencing apparent electrical conductivity: A review," *Comput. Electron. Agric.*, vol. 46, no. 1–3, pp. 45–70, 2005. https://doi.org/10.1016/j.compag.2004.11.001

[58] G. Caposciutti, L. Miserocchi, F. Cela, L. Incrocci, A. Franco, and B. Tellini, "Experimental analysis of temperature distribution inside a greenhouse: Influence of sensor placement on temperature measurement and implications for LAI estimation," *Therm. Sci. Eng. Prog.*, vol. 73, p. 104629, 2026. https://doi.org/10.1016/j.tsep.2026.104629

[59] N. Kakhani, S. Alamdar, N. M. Kebonye, M. Amani, and T. Scholten, "Uncertainty quantification of soil organic carbon estimation from remote sensing data with conformal prediction," *Remote Sens.*, vol. 16, no. 3, p. 438, 2024. https://doi.org/10.3390/rs16030438

[60] A. Blažič, I. Škrjanc, and V. Logar, "Prediction interval soft sensor for dissolved oxygen content estimation in an electric arc furnace," *Appl. Soft Comput.*, vol. 167, p. 112246, 2024. https://doi.org/10.1016/j.asoc.2024.112246

Appendix

A1 Within-Campaign Permutation Diagnostics

Permutation diagnostics summarize the increase in heldout MAE after each predictor was shuffled within its own campaign. Values are means and standard deviations over 20 shuffles. Temperature importance is given in degrees Celsius and moisture importance in sensor-scale percentage points. A small or negative value means that the predictor adds little for that fitted model and test set; it is no proof that the variable is physically irrelevant.

Table A1. Within-campaign permutation increases in heldout MAE

Target	Model	Illuminance	Conductivity	Air RH	Air Temperature
Temperature	GB	0.122 ± 0.019	0.124 ± 0.019	-0.004 ± 0.004	4.603 ± 0.159
Temperature	ET	0.218 ± 0.029	0.146 ± 0.022	0.008 ± 0.013	4.670 ± 0.152
Moisture	GB	0.054 ± 0.040	2.077 ± 0.141	-0.053 ± 0.016	0.351 ± 0.046
Moisture	ET	0.039 ± 0.017	2.388 ± 0.124	-0.018 ± 0.016	0.518 ± 0.066

Note: MAE = mean absolute error; RH = relative humidity; GB = gradient boosting; ET = extremely randomized trees.

A2 Algorithmic Workload

Batch timing covers 290 records on a Linux x86-64 host with an Intel Xeon Platinum 8573C processor. There were three warm-up calls and 20 measured repetitions. The selected Extra Trees models used n_jobs = -1. Table A2 reports fitting time and median and 95th-percentile batch prediction times. These host-specific software measurements exclude sensor acquisition, communications and actuator response; complete software metadata are preserved in the reproducibility archive.

Table A2. Algorithmic workload of validation-selected estimators on the analysis host

Inputs	Target	Model	Fit Time (s)	Median Batch (ms)	95th Percentile (ms)
Four channels	Temperature	ABS	3.768	294.265	313.499
Four channels	Moisture	ET	0.177	63.016	79.311
Air channels	Temperature	ABS	2.386	242.638	308.437
Air channels	Moisture	ET	0.298	110.056	119.905

Note: ABS = AdaBoost with support vector regression (AdaBoost–SVR); ET = extremely randomized trees. Prediction batches contain 290 records. Times are host-specific software measurements and exclude acquisition, communications and actuation.

Supplementary Information

The online version contains supplementary material available at: https://doi.org/10.56578/jii040303.