IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Journal Européen des Systèmes Automatisés
Vol. 59, No. 10, October, 2026, pp. 2819-2834
Journal homepage: http://iieta.org/journals/jesa

# Machine Learning for Humidity Forecasting in Oyster Mushroom Cultivation: Sensor Validity, Feature Representation, and the Limits of Model Capacity

Ahmad Hashim Hussein Aal-Yhia

[1] Department of Big Data, College of Artificial Intelligence, University of Baghdad, Baghdad 10001, Iraq
[2] Department of Computer Science, College of Sciences, University of Baghdad, Baghdad 10001, Iraq

Corresponding Author Email: ahmedhashem@pgiafs.uobaghdad.edu.iq

Copyright: ©2026 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/jesa.591006

**Received:** 17 August 2026
**Revised:** 27 September 2026
**Accepted:** 22 October 2026
**Available online:** 31 October 2026

***Keywords:***
*gradient boosting, humidity forecasting, Internet of Things, machine learning, mushroom cultivation, neural networks, sensor validation, vapour pressure deficit*

**ABSTRACT**

Mushroom cultivation depends on humidity control, yet humidity is the variable that published greenhouse models forecast least well. This study asks a narrow question: does supplying a model with psychrometric quantities derived from temperature and humidity, rather than the raw pair, improve forecast accuracy? We answer it using an open dataset of two oyster mushroom cultivation cycles logged at 15 s resolution, one of which carries a Sensirion Sensor Humidity Temperature (SHT)-series alongside two Digital Humidity Temperature (DHT21) units, permitting assessment against a better-specified reference. The derived quantities gave no reliable improvement at any horizon, nor did four independent channels, including substrate moisture and carbon dioxide, for linear, tree-ensemble and neural models alike. Both DHT21 units read above the reference and spent a quarter of the record at their ceiling, while the reference never reached its own. Condensation risk, computed from surface temperature and dew point, occupied 1.13% of the record but fell almost entirely on two consecutive days. We argue that derived humidity features repay their cost only where a moisture balance is actively driven, and that low-cost humidity sensors require validation against a reference before their readings are treated as ground truth.

## 1. INTRODUCTION

Cultivated edible fungi occupy a distinctive position in food systems. Mushrooms are heterotrophic: they require neither sunlight nor arable topsoil, and growers can produce them in vertically stacked, climate-controlled structures on land unsuitable for conventional agriculture. Substrates consist mostly of lignocellulosic residues that would otherwise be burned or landfilled, and spent substrate retains value as a soil conditioner. Production has expanded rapidly, with Pleurotus species accounting for a substantial share of global output [1, 2]. Cultivation is nonetheless technically demanding. Development proceeds through phases with distinct and partly conflicting environmental optima, and relative humidity must be sustained above 85-90% without producing free water on the cap surface, which is the precondition for bacterial blotch caused by Pseudomonas tolaasii [1]. The narrow tolerance and severe consequences of excursions have made mushroom houses an attractive target for Internet of Things instrumentation and machine learning since 2020 [2-5]. A consistent asymmetry appears within this literature. Deep learning models achieve strong accuracy on temperature and carbon dioxide but perform worse on relative humidity, with reported coefficients of determination of 0.974 for temperature, 0.993 for $CO_2$ and 0.804 for humidity [6]. The asymmetry has a physical basis. Temperature and $CO_2$ are bulk air properties with well-mixed dynamics and clearly identified sources and sinks. Humidity is governed by phase-change processes at surfaces, namely evaporation from substrate and crop and condensation onto cooler surfaces, which ambient sensors observe only partially [7]. The crop itself is a large, wet, transpiring surface whose behaviour changes as it develops.

### 1.1 The representation hypothesis

We set out to test whether part of this shortfall is a representation problem admitting a cheap remedy. Relative humidity (RH) is a ratio rather than a quantity of water, so its relationship to evaporative load at the crop surface depends on temperature [7]. Air vapour pressure deficit (VPD) at 90%RH and 22 ℃ exceeds that at 90%RH and 15 ℃ by a factor of 1.55. A model supplied with raw relative humidity must infer this nonlinear relationship from data; a model supplied with VPD receives it directly. Absolute humidity (AH) contributes independently, being conserved under sensible heating and cooling, which linearises the effect of thermal actuation on the moisture balance. These quantities are standard in plant science and have been used in horticultural modelling for decades [7, 8]. The proposition under test is not their discovery but whether the prevailing convention of supplying raw relative humidity to forecasting models leaves recoverable accuracy unexploited.

### 1.2 Contribution

This paper reports a secondary analysis of the open dataset published by Duman et al. [9], covering two oyster mushroom cultivation cycles. The contribution in this study is based on empirical analysis, sensor evaluation, and the investigation of negative results. We make five contributions. First, we validate two low-cost Digital Humidity Temperature (DHT21) humidity sensors against a co-located Sensirion Sensor Humidity Temperature (SHT) reference across 63,809 matched minutes, using independent temperature probes on all nodes to establish that the nodes are effectively co-located and that observed discrepancies are instrumental rather than spatial. Second, we test the representation hypothesis under a pre-specified analysis plan incorporating an outcome-neutral quality gate, a matched-dimensionality control, and hierarchical fixed-sequence testing. Third, we evaluate whether exogenous measurements, namely substrate temperature, substrate moisture, and $CO_2$, improve humidity forecasting, a question prior reasoning suggests should be answered affirmatively. Fourth, we compare three machine learning model classes, regularised linear regression [10], gradient-boosted trees [11] and a feedforward neural network, to establish whether the feature results depend on model capacity. Fifth, we quantify condensation risk using a margin computed from substrate surface temperature and air dew point, and characterise its temporal distribution. We report a substantially negative result on the representation hypothesis and offer an account of the conditions under which it may nonetheless hold.

## 2. RELATED WORK

### 2.1 Environmental monitoring in mushroom cultivation

Internet of Things (IoT) deployments in mushroom houses have been reported consistently since 2020, typically built on low-cost microcontrollers with relay-based actuation and remote interfaces [3, 4]. Reported implementations cover home-scale cultivation chambers [3], community greenhouses, and commercially scalable edge-computing architectures [4], generally motivated by reducing the labour burden of manual environmental monitoring. Control has advanced incrementally from fixed thresholds to fuzzy logic [5], offering graded rather than binary actuation without requiring a learned model. Two limitations characterise this body of work. Control remains predominantly reactive, acting on present rather than predicted state, and validation is typically demonstrative, resting on a working prototype over a single cultivation cycle rather than statistically powered comparison against a control treatment [2].

### 2.2 Environmental forecasting

The principal methodological advance of the period has been the shift from monitoring toward forecasting, motivated by the observation that the physical lag of climate control equipment cannot be overcome by faster sensing alone [4, 6]. Architectures combining attention mechanisms with convolutional and recurrent layers have been applied to mushroom greenhouse data, with reported prediction errors of 0.17 °C for temperature, 8.367 ppm for $CO_2$ and 2.06%RH for relative humidity [6].

Forecasting architectures for environmental series divide into three families, and the mushroom literature has drawn on them unevenly.

Recurrent models came first. Long short-term memory networks [12] address the vanishing-gradient problem that defeated earlier recurrent units on long sequences, and they remain the default in greenhouse forecasting. Their sequential structure is also their cost: training does not parallelise across time steps, and dependencies beyond a few hundred steps remain difficult to capture. Convolutional models answer that cost. Temporal convolutional networks [13] stack dilated causal convolutions to reach a receptive field that grows exponentially with depth, train in parallel across the sequence, and have matched or exceeded recurrent baselines on several sequence-modelling benchmarks. Attention-based models take a different route: the Transformer [14] discards recurrence entirely in favour of self-attention, and the Temporal Fusion Transformer [15] adapts that idea to multi-horizon forecasting with static covariates and interpretable variable selection, both of which suit greenhouse data where actuator states and crop stage act as slowly varying inputs. Hybrid designs dominate current practice, and the mushroom-specific work sits here. The attention-augmented convolutional-recurrent model proposed in the study [6] combines all three ideas: convolution for local feature extraction, recurrence for temporal dependence, attention for weighting. Reviews of the wider field report similar hybrids achieving precision above 90% across monitoring and forecasting tasks [1, 2]. Two observations about this literature bear on the present study. The first is that architectural comparison in it is almost always between deep models, rarely against a regularised linear alternative, so the marginal value of the added complexity is seldom quantified. The second is the absence of naive baselines. Environmental series sampled at minute resolution are strongly autocorrelated, and a persistence forecast, predicting that the next value equals the current one, is a demanding competitor at short horizons [16]. Where no such baseline is reported, a high coefficient of determination cannot be distinguished from a strongly autocorrelated target. We report both a persistence baseline and a linear alternative throughout, and Section 5.2.4 shows why that matters here.

### 2.3 Sensor selection

The DHT family of capacitive humidity sensors appears widely in this literature [3, 5], with manufacturer-stated accuracy of ±2–5%RH. Mushroom cultivation operates in the 85–98%RH band, near the upper limit of these devices' useful range, yet published deployments rarely report validation against a reference instrument. Section 5.1 addresses this directly.

## 3. DATA

### 3.1 Source

Analysis used the open oyster mushroom environmental dataset published by Duman et al. [9] under a Creative Commons Attribution licence. The dataset comprises two complete Pleurotus ostreatus cultivation cycles conducted in Mersin/Akdeniz, Turkey, instrumented with ESP-family microcontroller nodes logging to the Arduino IoT Cloud. We

did not design the instrumentation, control the growing environment, or determine the sampling regime. All characterisations below are inferred from the released data and its accompanying notes.

### 3.2 Cycle composition

Cycle I ran from 26 December 2022 to 2 March 2023 (65 days) with two nodes. Each carried a DHT21 temperature and humidity sensor, a Dallas Semiconductor (DS18B20) substrate temperature probe, and a moisture sensor. No humidity reference was present. The accompanying notes file documents four node outages with restoration times, and records the final day of the cultivation cycle as 2 March 2023. Cycle II ran from 11 March to 1 May 2023 (50 days) with three nodes. Two carried DHT21 sensors; the third carried a Sensirion SHT temperature and humidity sensor. DS18B20 probes were present on all three nodes. The cycle additionally recorded capacitive and resistive substrate moisture, and $CO_2$ and total volatile organic compound estimates from two independent sensor families. Both cycles logged at approximately 15-second intervals.

### 3.3 Preprocessing

We removed manufacturer error codes prior to analysis: −127 ℃ (DS18B20 no-reading), −50 ℃ and 149.9 ℃ (DHT out-of-range), and humidity values of 0. We resampled each series to one-minute means. We then segmented records into contiguous runs of at least 301 minutes, this being the minimum required for a four-hour input window plus a one-hour forecast horizon. Windows never cross a segment boundary, so no forecast draws on inputs separated from the target by a data gap. For Cycle II, segmentation on temperature and humidity yielded 63 segments totalling 51,412 minutes. Requiring substrate moisture and $CO_2$ additionally yielded 66 segments totalling 39,358 minutes; the segment count rises because additional gaps subdivide longer runs. The distribution of segment lengths is shown in Section 5.2.

## 4. METHODS

The analysis pipeline is summarised in Figure 1.

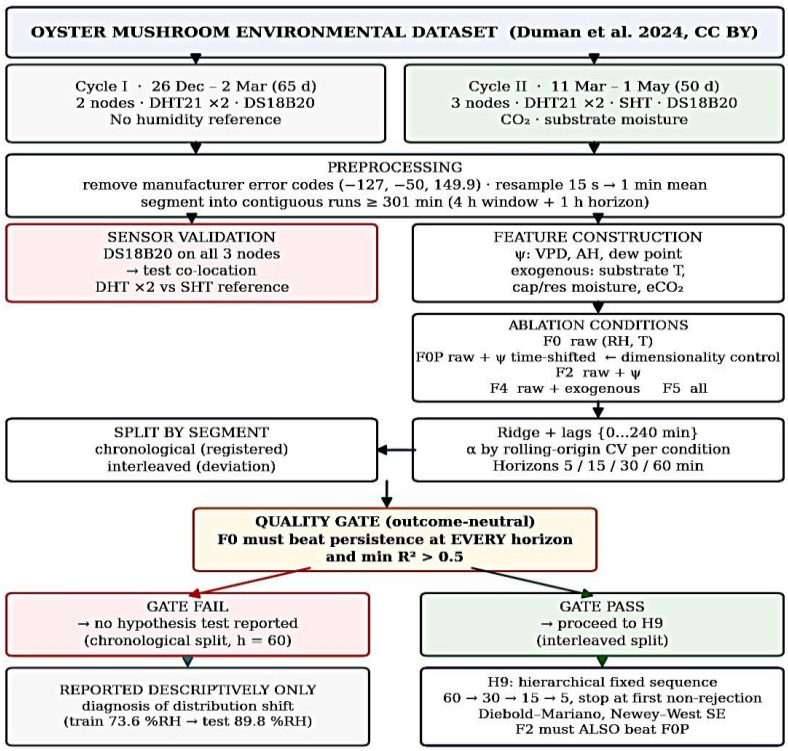


**Figure 1.** Analysis pipeline

Note: Data from two cultivation cycles undergo error-code removal, resampling and segmentation into contiguous runs. Cycle II supports both sensor validation against the Sensor Humidity Temperature (SHT) reference and the feature ablation. Five feature conditions are evaluated under two partitioning schemes; the outcome-neutral quality gate determines whether hypothesis testing proceeds.

### 4.1 Derived quantities

Psychrometric quantities were computed from temperature and relative humidity using standard relations [7, 8]. Saturation vapour pressure follows Tetens [8], with temperature T in degrees Celsius and the result in kilopascals, shown by Eq. (1):

$$e_s(T) = 0.6108\exp\left(\frac{17.27\,T}{T + 237.3}\right) \quad (1)$$

Actual vapour pressure follows from Eq. (1) and measured relative humidity RH in percent, as shown in Eq. (2):

$$e_a = e_s(T)\,\frac{RH}{100} \quad (2)$$

Air VPD, the driving force for evaporation, is the difference between Eqs. (1) and (2), as shown in Eq. (3):

$$VPD = e_s(T) - e_a \quad (3)$$

AH, expressed in grams per cubic metre with $e_a$ in kilopascals, is given by Eq. (4):

$$AH = \frac{2167\, e_a}{T + 273.15} \tag{4}$$

Dew point temperature ($T_{dew}$) is obtained by inverting the Magnus relation, using the intermediate term defined in Eq. (5) and the result in Eq. (6):

$$\gamma = ln\left(\frac{RH}{100}\right) + \frac{17.27\, T}{237.3 + T} \tag{5}$$

$$T_{dew} = \frac{237.3\, \gamma}{17.27 - \gamma} \tag{6}$$

The condensation margin is defined as the substrate surface temperature minus the air dew point in Eq. (6), as shown in Eq. (7):

$$M_c = T_{surface} - T_{dew} \tag{7}$$

A negative value of Eq. (7) implies condensation onto the surface. All of Eqs. (1) to (7) were verified numerically against independent reference values prior to use, including saturation vapour pressure at 0, 20 and 30 °C, AH of saturated air, and exactness of the inverse transformations.

**4.2 Sensor validation**

4.2.1 Status of the reference instrument

The SHT device on node ESP3 serves as the reference throughout, so its standing requires justification rather than assumption. Three arguments support it, and one caveat qualifies it. First, the Sensirion SHT3x family to which it belongs carries a manufacturer-stated tolerance of plus or minus 1.5 to 2%RH across the 10-90%RH band, against 2 to 5%RH for the DHT21, with the DHT specification degrading further above 90%RH. The reference is the better-specified instrument by roughly a factor of two before any measurement is taken. Second, and independently of the datasheet, the device behaves as a sound humidity sensor should in this environment. Across 63,809 matched minutes it never reached its measurement ceiling, attaining a maximum of 99.02%RH (Section 5.1.3), while both DHT units saturated for approximately a quarter of the record. A sensor reporting its ceiling value has stopped measuring; one that approaches without reaching it has not. Third, its temperature channel agrees with the two DHT temperature channels to within 0.033 and 0.392℃ (Table 6), and the DS18B20 probes place all three nodes within a median 0.75℃ of one another (Section 5.1). The reference is not thermally displaced from the instruments it is used to judge. The caveat is material and we state it plainly. The source dataset [9] reports neither the specific SHT variant nor any calibration certificate, and no reference-grade hygrometer was co-deployed. We therefore establish that the SHT device is the better-behaved instrument, not that it is metrologically traceable. The offsets in Section 5.1.2 should be read as departures from a superior sensor rather than as absolute errors. Resolving this requires a calibrated reference; Section 6.9 records that. Because humidity sensors of different types occupied different nodes, any observed discrepancy could in principle reflect spatial variation rather than instrument error. The DS18B20 probes, present on all three Cycle II nodes and electrically and physically independent of the humidity channels, permit this to be tested. If inter-node temperature agreement is close, spatial separation is insufficient to explain large humidity differences. Results are reported in Table 3. We computed confidence intervals (CIs) on daily means rather than on minute-level observations, to accommodate the strong serial correlation present in the underlying series.

**4.3 Forecasting task and feature conditions**

The task is to forecast SHT relative humidity at horizons of 5, 15, 30, and 60 minutes, given a four-hour input history at five-minute resolution with lags at {0, 5, 10, 15, 30, 45, 60, 90, 120, 180, 240} minutes. All conditions predict the same target on the same scale, relative humidity in percentage points. Derived quantities from Eqs. (1) to (7) are used as model inputs only. The constraint matters: AH as defined in Eq. (4) is a smoother signal than relative humidity, so comparing error across different target variables would be uninterpretable. The six conditions are defined in Table 1.

The design is set out schematically in Figure 2 and enumerated in Table 1.

Condition F0P in Table 1 requires explanation. It carries the same column count as F2, but each psychrometric feature is circularly shifted in time by a large random offset within its segment. This preserves marginal distribution, variance and autocorrelation exactly while destroying temporal alignment with the target. It therefore isolates the cost or benefit of added dimensionality from the contribution of the features themselves. Without it, a null result for F2 would be ambiguous between the features carrying no information and the additional parameters costing more than they contribute. Note that F0P is randomised; the values reported here correspond to seed 0.

4.3.1 When a deterministic transformation can carry information

VPD, AH and dew point are exact functions of temperature and relative humidity, so they add no information in the Shannon sense. Their value, where it exists, is representational rather than informational: they change what a model of given capacity can express cheaply. Three conditions make that difference operative, and each is testable. The first is model class. A linear model in temperature and relative humidity cannot represent their product, whereas AH is approximately such a product and VPD is a saturation curve scaled by relative humidity. Supplying these terms hands a linear model an interaction it could not otherwise form. A sufficiently flexible nonlinear model recovers the same structure from the raw pair, so the benefit should shrink as capacity rises. Section 5.2.4 tests this directly. The second is the operating range. Over a narrow temperature band the saturation curve of Eq. (1) is nearly linear, and the derived quantities become near-affine in the raw inputs, carrying almost nothing new. Their advantage should therefore scale with the temperature range spanned by the data. Cycle II spans 15.19 to 17.99℃ in daily means, a range over which Eq. (1) departs only modestly from linearity. The third, and the one we judge decisive here, is whether the moisture balance is actively driven. AH is conserved under sensible heating and cooling. In a house where heaters, chillers and humidifiers act on the air, that conservation separates thermal actuation from moisture actuation, which the raw pair conflates. Where no such actuation occurs and indoor

humidity follows outdoor weather, there is no mechanism to separate and the transformation is redundant. These conditions predict where the effect should and should not appear, and Section 6.1 evaluates that prediction against the contrast between the passively managed cycles analysed here and the actively controlled greenhouse of study [6].

**Table 1.** Feature conditions evaluated in the ablation

| Condition | Inputs | Purpose |
|---|---|---|
| Persistence | — | Naive baseline; forecast equals current value |
| F0 | Relative humidity, temperature | Baseline representing prevailing convention |
| F0P | F0 plus psychrometric features, circularly time-shifted | Matched-dimensionality control |
| F2 | F0 plus apour pressure deficit (VPD), Absolute humidity (AH), dew point — Eqs. (3), (4), (6) | Test condition |
| F4 | F0 plus substrate temperature, capacitive and resistive moisture, $CO_2$ | Exogenous information |
| F5 | F0 plus psychrometric and exogenous | Combined |

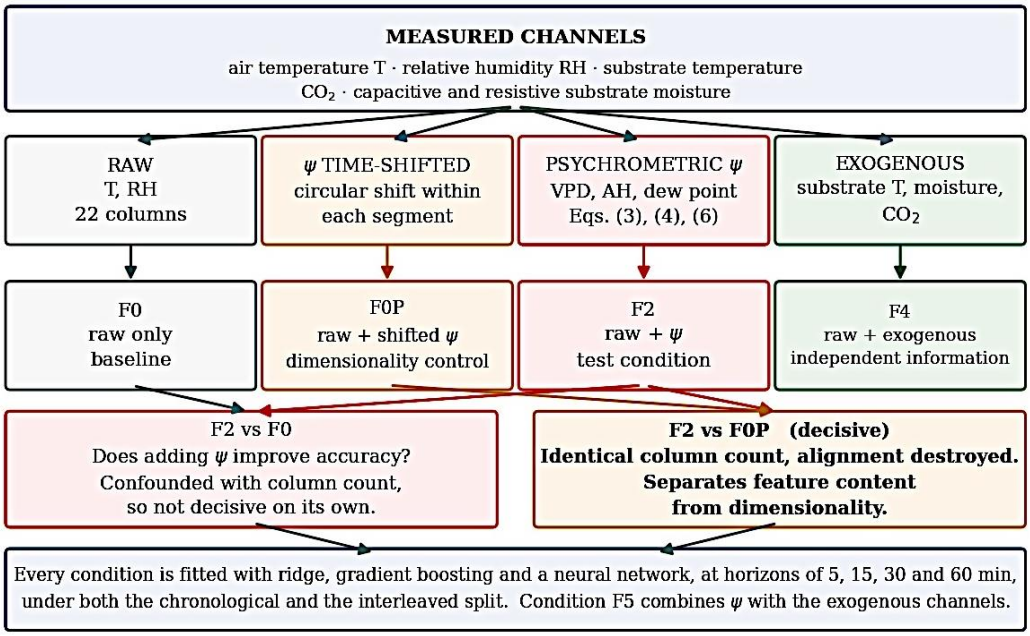


**Figure 2.** Feature ablation workflow

Note: Measured channels feed four feature groups, each defining one experimental condition. The comparison F2 against F0 asks whether psychrometric terms improve accuracy but is confounded with column count; the comparison F2 against F0P holds column count fixed while destroying temporal alignment.

### 4.4 Estimation

We used ridge regression [10] for the primary ablation, with the regularisation parameter selected independently for each condition and horizon from {0.1, 1, 10, 100, 1000} by rolling-origin expanding-window cross-validation within the training partition [17]. Per-condition tuning ensures the ablation compares feature content rather than differential tuning effort, since a condition with more columns would otherwise be penalised by a regularisation strength appropriate to a smaller design matrix. We held capacity constant by using a single model class across all feature conditions, so that any difference between conditions is attributable to features rather than to architecture, initialisation or training variance.

#### 4.4.1 Model class comparison

Holding capacity constant is necessary for the ablation but raises a separate question: whether a null result reflects the features carrying no information or the model class being unable to exploit it. This distinction matters differently for the two feature sets under test. The psychrometric quantities of Eqs. (3), (4) and (6) are exact deterministic functions of temperature and relative humidity, so they supply no information to any model class. The exogenous channels are independent measurements, and a linear model can incorporate them only additively; a nonlinear model might exploit interactions it cannot.

We therefore repeated the ablation at the 15- and 60-minute horizons across three model classes:

(1) Ridge regression [10]: regsularised linear least squares, as above.

(2) Histogram gradient boosting [11]: an ensemble of regression trees, well suited to detecting interactions in tabular data.

(3) Multilayer perceptron: a feedforward neural network with two hidden layers (128 and 64 units), Rectified Linear Unit (ReLU) activation, Adam optimisation, L2 penalty $10^{-3}$, and early stopping on a 15% validation split.

All three received identical feature matrices, identical partitions and identical targets, and all were implemented in scikit-learn [18]. Table 2 records every hyperparameter required to reproduce the fits. Values not listed were left at the scikit-learn 1.8 default. All fits used a fixed random seed of 0; the ridge results are deterministic given the seed, while the boosting and network results depend on it through subsampling and weight initialisation, respectively.

**Table 2.** Model configurations

| Parameter | Histogram Gradient Boosting | Multilayer Perceptron |
|---|---|---|
| Estimators/iterations | 400 maximum | 300 epochs maximum |
| Learning rate | 0.06 | 0.001 (Adam, initial) |
| Tree depth | unlimited | not applicable |
| Leaves per tree | 31 (default) | not applicable |
| Minimum samples per leaf | 20 | not applicable |
| Binning | 255 bins | not applicable |
| Hidden layers | not applicable | 128 and 64 units |
| Activation | not applicable | ReLU |
| Batch size | not applicable | 256 |
| L2 penalty | 0 | 0.001 |
| Early stopping | enabled | enabled |
| Validation fraction | 0.15 | 0.15 |
| Patience | 10 iterations | 15 epochs |
| Random seed | 0 | 0 |

### 4.5 Partitioning

The two partitioning schemes answer different questions, and conflating them would misread both. The chronological scheme trains on the first 70% of segments and tests on the remainder, preserving temporal order. It answers a deployment question: given a model fitted on the early part of a cultivation cycle, how well does it forecast the later part? That is the situation a grower faces, and it is the pre-registered primary analysis. The interleaved scheme assigns segments alternately to the two partitions. It answers a modelling question: holding the environmental regime constant across partitions, does a given feature set carry predictive information? A cultivation cycle progresses through phases with materially different humidity regimes, so a chronological split places systematically different conditions in the two partitions and confounds feature content with cycle phase. The interleaved scheme removes that confound at the cost of no longer simulating deployment. Neither scheme is correct in general. The chronological split is the honest test of deployment performance and the conservative one, since it also evaluates extrapolation. The interleaved split isolates the feature question this study was designed to answer. We report both, label the interleaved analysis a declared deviation from the registered plan, and treat the chronological result as primary. The magnitude of the confound is quantified in Section 5.2.1. Splitting is by segment in both cases, so no forecast window spans the partition boundary. Random splitting of auto-correlated series was avoided throughout, as it leaks information across the partition boundary and yields optimistic accuracy estimates [17]. Treating individual observations within a segment as independent units would also constitute a form of pseudoreplication [19], since observations drawn from the same contiguous run share the environmental state that generated them.

### 4.6 Pre-specified analysis plan

Quality gate. Before any hypothesis test, two outcome-neutral conditions must hold at every horizon: condition F0 must achieve lower error than the persistence baseline, and every fitted condition must attain a coefficient of determination above 0.5. If either fails, no hypothesis test is reported and the failure is diagnosed and disclosed.

Hypothesis. The representation hypothesis states that psychrometric features improve humidity forecast accuracy at equal model capacity. Testing is one-tailed, the direction being specified in advance on the grounds that condition F2 in Table 1 is computed deterministically from F0 and a sufficiently flexible model cannot be harmed by the addition.

Multiplicity. Horizons are tested in a fixed sequence, 60 → 30 → 15 → 5 minutes, stopping at the first non-rejection. Family-wise error is controlled at $\alpha = 0.05$ without correction, since each test is conducted only if all earlier tests rejected. The ordering reflects the a priori expectation that the effect increases with horizon: at short horizons the target is dominated by its own autocorrelation, leaving little structure for derived features to exploit.

Dimensionality control. Condition F2 must outperform F0P for any advantage to be attributed to feature content. Where it does not, the hypothesis is reported as not supported regardless of the primary comparison.

Test statistic. The Diebold–Mariano test [20] on the squared-loss differential, with the small-sample correction of Harvey et al. [21] and Newey–West standard errors [22] to accommodate serial correlation in the forecast error series.

## 5. RESULTS

### 5.1 Sensor validation

5.1.1 Nodes are effectively co-located

Inter-node agreement of the DS18B20 probes is reported in Table 3.

**Table 3.** Inter-node temperature agreement, Dallas Semiconductor (DS18B20) digital temperature probes, Cycle II (n = 63,809 minutes)

| Node Pair | Mean Difference (°C) | Correlation |
|---|---|---|
| ESP1 − ESP2 | +1.153 | 0.955 |
| ESP1 − ESP3 | +1.016 | 0.971 |
| ESP2 − ESP3 | −0.137 | 0.985 |

The median pairwise spread across the record was 0.75 °C. As Table 3 shows, temperature was near-uniform across the three positions, and spatial separation is therefore insufficient to account for humidity discrepancies of the magnitude reported below.

5.1.2 Both DHT units read above the reference

Humidity sensor performance against the SHT reference is summarised in Table 4.

**Table 4.** Humidity sensor comparison, Cycle II (n = 63,809 minutes over 49 days). RH denotes relative humidity; CI denotes confidence interval; SHT denotes the Sensirion Sensor Humidity Temperature SHT-series reference sensor

| Sensor | Mean (%RH) | Median | Difference from SHT | 95% CI | sd | r with SHT |
|---|---|---|---|---|---|---|
| DHT21 (ESP1) | 87.38 | 92.40 | +6.33 | [+5.80, +7.55] | 2.84 | 0.979 |
| DHT21 (ESP2) | 83.09 | 88.20 | +2.04 | [+0.07, +3.89] | 7.01 | 0.940 |
| SHT (ESP3) | 81.05 | 85.22 | reference | — | — | — |

Note: Relative Humidity (RH); confidence interval (CI); Sensor Humidity Temperature (SHT).

Both DHT units read above the reference by different amounts. Their mutual difference was +4.29%RH (sd 5.46, r = 0.965). Because the reference lies below both, neither can be regarded as correct and the error is not common-mode. Dispersion differs as much as bias does. As Table 4 shows, the ESP2 unit's offset varied with standard deviation 7.01 against 2.84 for ESP1, approximately 2.5 times more variable. Inter-unit discrepancy is therefore not a fixed calibration constant correctable by a single offset.

5.1.3 Ceiling saturation

Across the same 63,809 minutes under identical conditions, the SHT sensor never reached its ceiling while both DHT units spent approximately a quarter of the record saturated. Figure 3(a) shows the two DHT units tracking the reference in dynamics while offset in level, with repeated ceiling contact; Figure 3(b) shows the resulting accumulation of probability mass at 100%RH for both DHT units and its absence for the reference. Figure 3(c) confirms that both units lie predominantly above the 1:1 line, and Figure 3(d) presents the ceiling proportions of Table 5 graphically.

Ceiling behaviour is quantified in Table 5 and illustrated in Figure 3.

**Table 5.** Proportion of readings at or above 99.9% relative humidity (RH), Cycle II

| Sensor | Readings ≥ 99.9%RH | Maximum Observed |
|---|---|---|
| DHT21 (ESP1) | 23.3% | 100.00 |
| DHT21 (ESP2) | 26.1% | 99.90 |
| SHT (ESP3) | 0.0% | 99.02 |

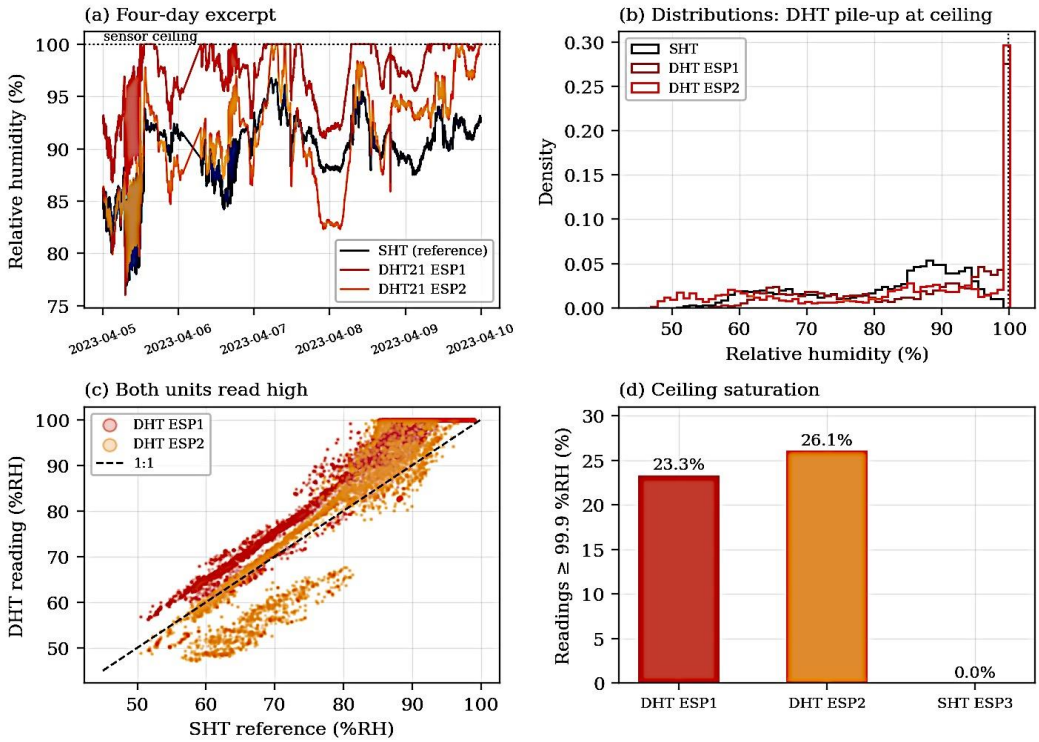


**Figure 3.** Humidity sensor comparison, Cycle II. Digital Humidity Temperature (DHT) denotes the DHT21 sensors and Sensor Humidity Temperature (SHT) the Sensirion SHT-series reference; RH denotes relative humidity. (a) Four-day excerpt showing both DHT units tracking the SHT reference in dynamics while offset in level, with repeated ceiling contact. (b) Marginal distributions; note the mass accumulated at 100%RH for both DHT units and its absence for SHT. (c) DHT readings against the SHT reference; both units lie predominantly above the 1:1 line. (d) Proportion of readings at or above 99.9%RH

5.1.4 The deficiency is specific to the humidity channel

Temperature agreement for the same devices is reported in Table 6.

**Table 6.** Temperature agreement with the Sensirion Sensor Humidity Temperature (SHT) reference sensor, Cycle II

| Sensor | Mean Difference (°C) | sd | r |
|---|---|---|---|
| DHT21 T (ESP1) | −0.033 | 0.507 | 0.988 |
| DHT21 T (ESP2) | −0.392 | 1.325 | 0.937 |

As Table 6 shows, DHT temperature measurement was accurate to within a few hundredths of a degree on one unit and 0.4 °C on the other. Comparison of Table 6 with Table 4 localises the deficiency to the humidity channel: it cannot be attributed to node placement, wiring, power supply or firmware, all of which are shared with the temperature channel.

5.1.5 Cycle I corroboration

Cycle I carried two DHT21 units and no reference. Across 38,872 matched minutes over 33 days, the uncensored inter-unit offset was +10.45%RH (95% CI [+9.89, +11.22], sd 2.46, r = 0.895), with no drift across thirteen three-day blocks (Spearman ρ = −0.077, p = 0.803). One unit spent 46.2% of the record at its ceiling; the other, 0.0%. The offset persisted across four documented power cycles of the affected node, at +11.35, +10.49, +9.49, +10.35 and +9.76%RH in successive intervals, eliminating warm-up drift, accumulated condensation and firmware state as explanations. Contrary to the common expectation that capacitive humidity sensors degrade specifically at high humidity, the offset declined as

humidity rose (Spearman $\rho = -0.546$, $p < 0.001$), from +11.24%RH at 70–80%RH to +7.02%RH at 90–95%RH. Part of this compression is mechanical, arising as the affected unit approaches its ceiling, but no evidence of range-dependent degradation was found. Cycle I's inter-unit offset (+10.45%RH, stable) and Cycle II's (+4.29%RH, sd 5.46, from Table 4) differ substantially in magnitude and stability, indicating that inter-unit discrepancy is unit-specific rather than characteristic of the sensor family.

### 5.2 Humidity forecasting

5.2.1 Pre-registered analysis: chronological split

Results under the pre-specified chronological partition are shown in Table 7, reported as root mean squared error (RMSE).

As Table 7 shows, condition F0 outperformed persistence at 5, 15 and 30 minutes but failed at 60 minutes (3.062 against 2.845). The quality gate therefore failed and no hypothesis test is reported for this split. The failure is attributable to distribution shift between partitions, illustrated in Figure 4. Figure 4(a) shows the test window occupying a distinctly wetter regime than the training window, and Figure 4(b) quantifies the resulting difference in marginal distributions: the test partition averaged 89.75%RH against 73.55%RH in training, with 14.9% of readings above 95%RH compared with 0.2%. The model was therefore extrapolating, and the penalty scaled with horizon: F0 achieved lower squared error than persistence on 50.5% of test points at 30 minutes but only 37.9% at 60 minutes.

**Table 7.** Forecast root mean squared error (RMSE), in percentage points of relative humidity (%RH), chronological split, Cycle II

| Horizon | Persistence | F0 | F0P | F2 | F4 | F5 |
|---|---|---|---|---|---|---|
| 5 min | 2.691 | 1.946 | 1.981 | 2.043 | 2.036 | 2.106 |
| 15 min | 2.507 | 2.357 | 2.283 | 2.522 | 2.435 | 2.577 |
| 30 min | 3.337 | 2.550 | 2.547 | 2.611 | 2.591 | 2.649 |
| 60 min | 2.845 | 3.062 | 2.980 | 3.123 | 3.207 | 3.331 |

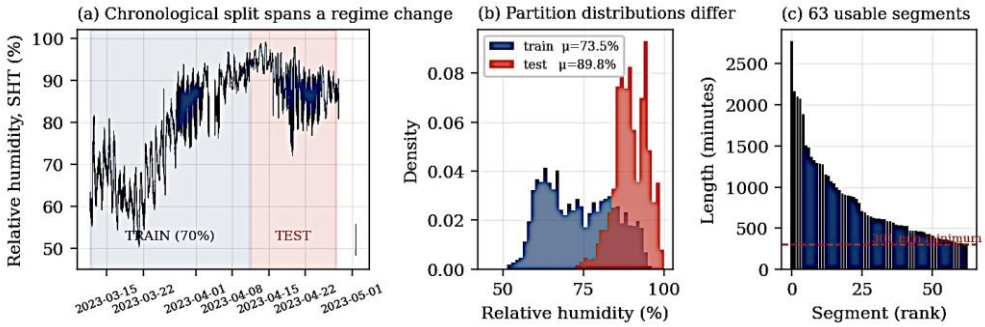


**Figure 4.** Partition structure and the chronological confound. (a) Sensor Humidity Temperature (SHT) humidity across Cycle II with training and test windows shaded; the test period occupies a distinctly wetter regime. (b) Humidity distributions by partition. (c) Length distribution of the 63 usable contiguous segments against the 301-minute minimum

5.2.2 Declared deviation: interleaved split

To separate the feature question from the phase confound evident in Figure 4, we repeated the analysis with segments assigned alternately to partitions. This is a deviation from the pre-specified plan and is not treated as the primary analysis. Results are shown in Table 8.

The gate passed at all four horizons in Table 8, permitting testing to proceed. At the 60-minute horizon, condition F2 improved on F0 by 0.026%RH with a Diebold–Mariano statistic yielding $p = 0.055$, failing to reject at $\alpha = 0.05$.

**Table 8.** Forecast root mean squared error (RMSE), in percentage points of relative humidity (%RH), interleaved split, Cycle II

| Horizon | Persistence | F0 | F0P | F2 | F4 | F5 |
|---|---|---|---|---|---|---|
| 5 min | 2.156 | 1.555 | 1.571 | 1.545 | 1.625 | 1.606 |
| 15 min | 2.154 | 1.883 | 1.942 | 1.904 | 1.929 | 1.928 |
| 30 min | 2.616 | 2.160 | 2.054 | 2.095 | 2.178 | 2.118 |
| 60 min | 2.547 | 2.376 | 2.353 | 2.350 | 2.467 | 2.425 |

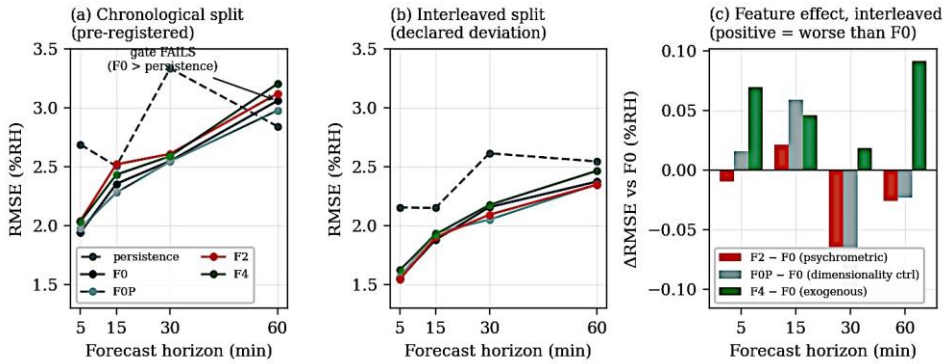


**Figure 5.** Feature ablation results. (a) Chronological split, where the baseline condition F0 fails to beat persistence at the 60-minute horizon and the quality gate halts testing. (b) Interleaved split, where the gate passes at all horizons. (c) Difference in root mean squared error (RMSE) relative to F0 under the interleaved split; positive values indicate worse performance

Under fixed-sequence testing the procedure stops at the first non-rejection; horizons of 30, 15 and 5 minutes were therefore not tested. Independently, F2 did not outperform the matched-dimensionality control F0P defined in Table 1 ($p = 0.472$). The pre-specified plan states that where F2 fails to beat F0P, any apparent advantage is attributed to dimensionality rather than feature content. The representation hypothesis is not supported. The pattern across both partitions is summarised in Figure 5. Figure 5(a) shows the gate failure at the 60-minute horizon under the chronological split, and Figure 5(b) the passing gate under the interleaved split. Figure 5(c) plots the differences relative to the baseline. The dimensionality control F0P tracks F0 closely, confirming that added columns are not themselves costly, while F2 and F4 degrade accuracy at most horizons.

5.2.3 Exogenous channels

Condition F4 of Table 1 added substrate temperature, capacitive and resistive substrate moisture, and $CO_2$, all measurements that are not deterministic functions of temperature and relative humidity and therefore carry information unavailable to F0. Results are shown in Table 9.

**Table 9.** Condition F4 relative to F0, interleaved split

| Horizon | ΔRMSE (F4 − F0) | p |
|---|---|---|
| 5 min | +0.069 | 1.00 |
| 15 min | +0.046 | 1.00 |
| 30 min | +0.018 | 0.964 |
| 60 min | +0.091 | 1.00 |

Every addition degraded accuracy. No exogenous channel improved humidity forecasting at any horizon.

5.2.4 Model class comparison

To establish whether the nulls of Tables 8 and 9 reflect the features or the model class, we repeated the ablation across ridge regression [10], gradient boosting [11] and a neural network at the 15- and 60-minute horizons. Absolute accuracy is reported in Table 10 and differences relative to the baseline in Table 11.

The 24 pairwise comparisons that Table 11 tabulates are shown graphically in Figure 6, which is the more legible presentation and the one we recommend the reader work from; Table 11 is retained so the underlying values remain available. Table 12 is similarly compressed into Figure 7(a), and the interval and seed-spread results of Table 13 into Figure 8. Where a table and a figure present the same quantities, the figure carries the argument, and the table serves as the record.

As Table 11 shows, no model class extracted usable information from either the psychrometric derivations or the exogenous channels. The gradient boosting result is the informative one for the exogenous question, since tree ensembles detect interactions that a linear model cannot represent [11]: its best outcome was a 0.003%RH improvement at the 60-minute horizon, with $p = 0.482$. Two secondary observations follow from Table 10. Ridge regression outperformed both nonlinear alternatives on the baseline condition at the 60-minute horizon (2.376 against 2.685 and 2.958) and was within 0.014%RH of gradient boosting at 15 minutes; the linear model was therefore not the limiting factor. Section 5.2.6 reports a capacity sweep that does not support over-parameterisation. These results are shown graphically in Figure 6.

**Table 10.** Forecast root mean squared error (RMSE), in percentage points of relative humidity (%RH), by model class and feature condition, interleaved split

| Horizon | Model Class | F0 (Raw) | F2 (+Psychrometric) | F4 (+Exogenous) |
|---|---|---|---|---|
| 15 min | Ridge (linear) | 1.883 | 1.904 | 1.929 |
| 15 min | Gradient boosting | 1.869 | 1.885 | 1.904 |
| 15 min | Neural network | 2.124 | 2.986 | 2.935 |
| 60 min | Ridge (linear) | 2.376 | 2.350 | 2.467 |
| 60 min | Gradient boosting | 2.685 | 2.773 | 2.683 |
| 60 min | Neural network | 2.958 | 4.167 | 4.731 |

**Table 11.** Effect of added features relative to F0, by model class

| Horizon | Model Class | ΔRMSE (F2 − F0) | p | ΔRMSE (F4 − F0) | p |
|---|---|---|---|---|---|
| 15 min | Ridge | +0.021 | 0.958 | +0.046 | 1.00 |
| 15 min | Gradient boosting | +0.016 | 0.688 | +0.036 | 0.861 |
| 15 min | Neural network | +0.862 | 1.00 | +0.812 | 1.00 |
| 60 min | Ridge | −0.026 | 0.055 | +0.091 | 1.00 |
| 60 min | Gradient boosting | +0.087 | 0.992 | −0.003 | 0.482 |
| 60 min | Neural network | +1.209 | 1.00 | +1.773 | 1.00 |

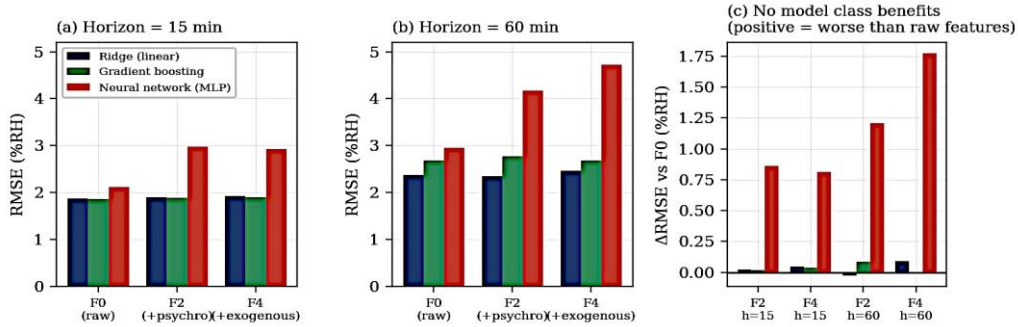


**Figure 6.** Model class comparison. (a, b) Forecast root mean squared error (RMSE) by model class and feature condition at the 15- and 60-minute horizons. (c) Change in RMSE relative to the raw-feature baseline F0; positive values indicate degradation. No model class benefits from either the psychrometric derivations or the exogenous channels, and the neural network degrades substantially as features are added

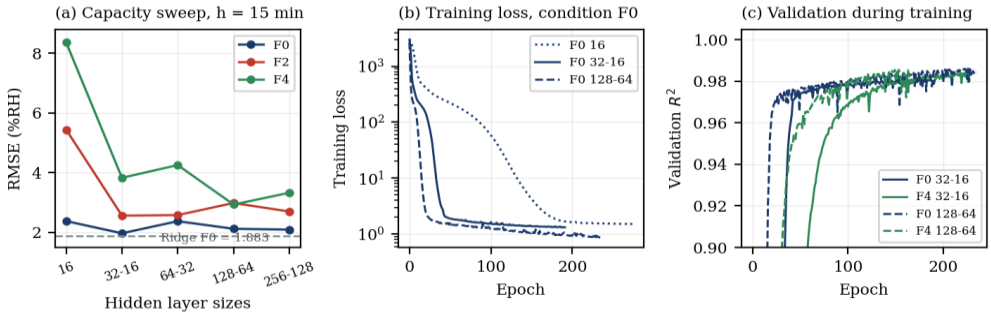


**Figure 7.** Neural network capacity sweep. RMSE denotes root mean squared error, in percentage points of relative humidity. (a) Test RMSE against hidden-layer size for the three feature conditions at the 15-minute horizon, with the ridge baseline marked. (b) Training loss for three F0 architectures; the smallest network plateaus highest, indicating underfitting. (c) Validation performance during training, showing monotone improvement without late divergence

Figure 6(a) and (b) present the absolute accuracies of Table 10 at the two horizons, and Figure 6(c) presents the differences of Table 11, making visible both the uniform absence of benefit and the neural network's disproportionate degradation.

5.2.5 Contrast with a controlled-environment dataset

For context, the same ablation was applied to an independent dataset from an actively controlled mushroom greenhouse, released alongside a published forecasting study [6]. That file comprises 4,656 records of indoor and outdoor temperature, humidity, $CO_2$ and binary actuator states, of which 2,244 are complete. In that setting, condition F2 improved on F0 by 0.503%RH (2.191 → 1.688, $p < 0.0001$) and also outperformed the matched-dimensionality control ($p < 0.0001$). Two limitations preclude treating this as confirmatory. The file contains no timestamp column, so the forecast horizon is undefined. Missing values are interleaved at single-row intervals, the longest run of complete consecutive rows being six, so retained records are not temporally contiguous, and the nominal one-step-ahead relation does not correspond to a fixed interval. The result is reported as a contrast, not a test. We note separately that on that dataset an ordinary least-squares regression on the same inputs achieved RMSE 2.191%RH ($R^2 = 0.809$) for humidity and 0.180 °C ($R^2 = 0.953$) for temperature, comparable to the published figures of 2.06%RH ($R^2 = 0.804$) and 0.17 °C ($R^2 = 0.974$) [6]. No persistence baseline is reported in that study; we computed persistence $R^2$ of 0.917 for temperature and 0.929 for $CO_2$ under the same one-step formulation.

5.2.6 Network capacity and the limits of the over-parameterisation account

We ran a capacity sweep at the 15-minute horizon across five architectures spanning 385 to 50,177 trainable parameters, holding every other setting at the values in Table 2. Results appear in Table 12 and Figure 7. The training partition contains 12,064 observations. The sweep does not support over-parameterisation. Had capacity been the binding constraint, accuracy should improve as the network is made smaller. It does the opposite. The 16-unit network, the smallest tested and the only one with fewer parameters than training observations in every condition, is the worst performer in all three feature conditions: 2.380%RH for F0, 5.440%RH for F2 and 8.356%RH for F4. Accuracy improves as capacity rises to 32-16 units and is then approximately flat to 256-128, with no systematic decline. The best network in each condition uses an intermediate architecture. Figure 7(b) and 7(c) explain why. The smallest network converges slowly and its training loss plateaus at a higher value than the larger networks reach, which is an underfitting signature rather than an overfitting one. Validation performance rises monotonically for every architecture and shows no late-training divergence from training loss; early stopping halted the fits at 97 to 300 epochs without any run exhibiting the validation turnover that would indicate memorization. Two observations survive the sweep. First, the best network obtained anywhere in it, 1.972%RH for F0 at 32-16 units, still fails to beat ridge regression at 1.883%RH. Second, adding features degrades the network at every capacity tested, so the degradation reported in Table 10 is not an artefact of the particular architecture chosen. The network underperforms a regularised linear model across the capacity range examined, and the relationship it must recover, being close to linear in the lagged inputs, is one that ridge regression represents directly while a network must approximate it. Why the network fails to recover that structure is not established by these data. Candidate explanations include the strong collinearity among lagged inputs, the limited training volume a single cultivation cycle affords, and the optimisation settings themselves; distinguishing them would require an ablation over optimiser, schedule and regularisation that we did not perform.

**Table 12.** Neural network capacity sweep, 15-minute horizon, interleaved split. root mean squared error (RMSE) in percentage points of relative humidity (%RH)

| Hidden Units | Parameters | F0 | F2 | F4 |
|---|---|---|---|---|
| 16 | 385-1,089 | 2.380 | 5.440 | 8.356 |
| 32-16 | 1,281-2,689 | 1.972 | 2.563 | 3.829 |
| 64-32 | 3,585-6,401 | 2.376 | 2.582 | 4.255 |
| 128-64 | 11,265-16,897 | 2.124 | 2.986 | 2.935 |
| 256-128 | 38,913-50,177 | 2.095 | 2.702 | 3.327 |

Notes: 1. Parameter counts vary within a row because the input width differs by condition (22, 55 and 66 columns). 2. Bold marks the best architecture within each condition. 3. Ridge regression achieves 1.883%RH on F0 under the same partition.

5.2.7 Uncertainty and repeated-run variability

The results reported so far are point estimates. Two sources of variability qualify them and both were quantified. The first is sampling uncertainty in the test set. We resampled test segments with replacement, 600 times, preserving whole segments so that the autocorrelation within each is retained, and recomputed RMSE. Table 13 gives the resulting intervals. They are wide, and they overlap completely across the three feature conditions: at the 15-minute horizon, F0 spans 1.258

to 2.532%RH and F2 spans 1.238 to 2.558%RH. The point difference between them, 0.021%RH, is two orders of magnitude smaller than the interval width. This does not weaken the conclusion, but it does change what supports it. Overlapping marginal intervals are the wrong instrument for comparing forecasts on shared data, because they discard the pairing: both models predict the same targets from the same inputs, so their errors are strongly correlated and much of the interval width is common to both. The Diebold-Mariano test used throughout operates on the paired loss differential and is correspondingly more sensitive. The intervals in Table 13 describe how well the RMSE of any single condition is pinned down by 66 segments; they do not describe how well the conditions can be told apart. Read together, the two say that this dataset determines the absolute error level loosely and the relative ordering tightly. The second source is stochastic model fitting. Ridge regression is deterministic given the data, so its results are reproducible exactly. The other two classes are not. We refitted each across five random seeds; Table 13 reports the spread. Gradient boosting is stable, with standard deviations of 0.015 to 0.055%RH. The neural network is not: 0.078%RH for F0 at the 15-minute horizon, rising to 0.892%RH for F4 at 60 minutes. Figure 8 shows the pattern, which is that instability scales with error. The single-seed neural network figures in Table 10 should therefore be read as draws from a distribution whose width, in the worst case, approaches the magnitude of the effects under discussion. This is a further reason to treat the network results as descriptive. The dimensionality control was likewise re-evaluated across five independent shift realisations, since F0P depends on a random offset. Its RMSE is 1.861 +/- 0.007%RH at 15 minutes and 2.285 +/- 0.079%RH at 60 minutes. Both means fall below the corresponding F0 values of 1.883 and 2.376%RH. Adding columns that carry no aligned information did not degrade accuracy and marginally improved it, which is consistent with a mild regularisation effect and reinforces the conclusion that F2 gains nothing from feature content.

**Table 13.** Uncertainty in reported accuracy. Root mean squared error (RMSE) in percentage points of relative humidity (%RH); CI denotes confidence interval. Bootstrap intervals from 600 segment-level resamples of the ridge fits; across-seed spread from five refits of each stochastic model. Interleaved split

| Horizon | Cond. | Ridge RMSE | 95% Bootstrap CI | Boosting, Mean +/- sd | Network, Mean +/- sd |
|---|---|---|---|---|---|
| 15 min | F0 | 1.883 | 1.258 to 2.532 | 1.897 +/- 0.017 | 2.111 +/- 0.078 |
| 15 min | F2 | 1.904 | 1.238 to 2.558 | 1.911 +/- 0.034 | 2.659 +/- 0.236 |
| 15 min | F4 | 1.929 | 1.343 to 2.574 | 1.898 +/- 0.015 | 3.510 +/- 0.669 |
| 60 min | F0 | 2.376 | 1.711 to 3.180 | 2.687 +/- 0.020 | 2.983 +/- 0.159 |
| 60 min | F2 | 2.350 | 1.670 to 3.155 | 2.722 +/- 0.048 | 3.795 +/- 0.301 |
| 60 min | F4 | 2.467 | 1.823 to 3.231 | 2.696 +/- 0.055 | 4.943 +/- 0.892 |

Notes: 1. All values in %RH. 2. Ridge is deterministic given the data, so no seed spread is reported. 3. Bootstrap resampling is at segment level to preserve within-segment autocorrelation.

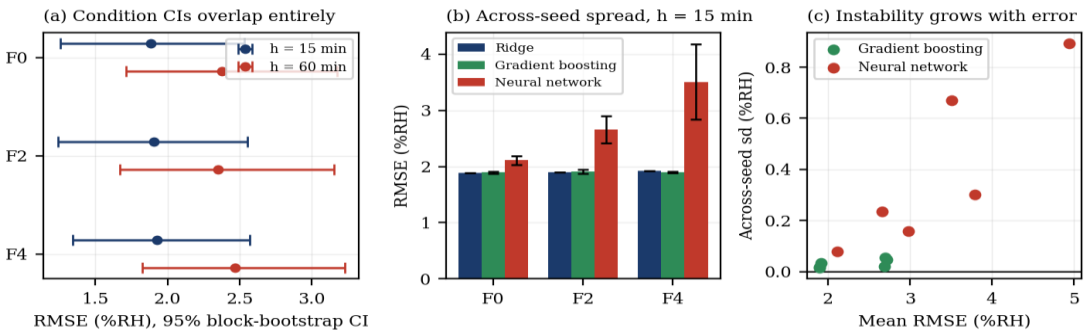


**Figure 8.** Uncertainty in the reported accuracies. (a) Bootstrap intervals for the three feature conditions overlap entirely at both horizons. (b) Across-seed spread by model class at the 15-minute horizon; error bars are one standard deviation over five refits. (c) Seed instability scales with mean error, and is confined almost entirely to the neural network

### 5.3 Condensation margin

The condensation margin of Eq. (7) was computed across 64,289 minutes of Cycle II, using the DS18B20 substrate probe as the surface temperature term. Results are shown in Table 14. Exposure was clustered in time. Figure 9(a) shows the margin across the full cycle with negative intervals shaded, confirming that excursions are rare and brief relative to the record. As Figure 9(b) shows, two consecutive days, 13 and 14 April, contributing 342 and 287 minutes, account for 87% of total negative-margin time, with only two further days contributing at all. Figure 9(c) contrasts humidity during negative-margin periods against the remainder, corresponding to the means of 98.17% and 80.87% reported in Table 14.

**Table 14.** Condensation margin, Cycle II (n = 64,289 minutes)

| Statistic | Value |
|---|---|
| Mean margin | +5.37 °C (sd 3.62) |
| Minimum | −0.90 °C |
| Minutes at negative margin | 725 (1.13% of record) |
| Total exposure | 12.1 hours over 50 days |
| Episodes | 20 (median 4 min, maximum 272 min) |
| Mean RH during negative margin | 98.17% |
| Mean RH otherwise | 80.87% |
| Correlation, margin against RH | −0.969 |

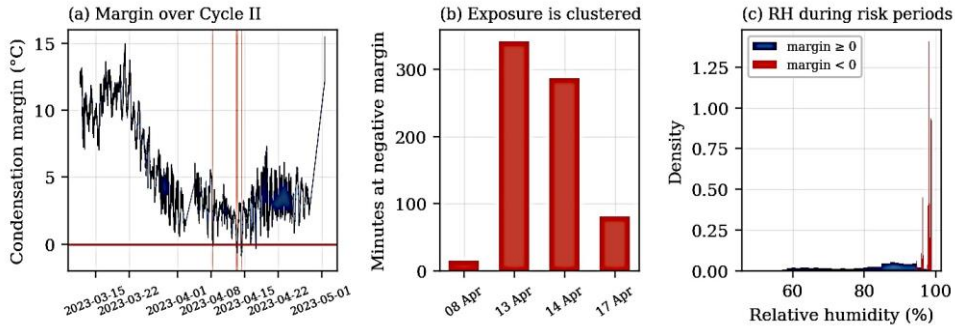


**Figure 9.** Condensation margin, Cycle II. (a) Margin across the cycle, with negative-margin intervals shaded; excursions are rare and brief. (b) Daily exposure, showing concentration on two consecutive days. (c) Humidity distributions during negative-margin periods against the remainder of the record

The DS18B20 probe measures substrate rather than cap-surface temperature, so Eq. (7) is evaluated here with a proxy. The clustering and direction are robust; absolute margins should not be interpreted as cap-surface values.

## 6. DISCUSSION

### 6.1 A conditional, not universal, feature effect

The central finding is negative and specific. Encoding psychrometric structure as model inputs did not improve humidity forecasting in either oyster cultivation cycle, as shown in Tables 8, 9 and 10, while the same manipulation produced a substantial improvement on an independent dataset from an actively controlled greenhouse (Section 5.2.5). We suggest the difference is not arbitrary. The oyster structure was passively managed: Cycle I recorded a median temperature of 11.5 ℃ with weekly medians as low as 5.8 ℃, far below any oyster cultivation setpoint [1], and no actuator states were logged in either cycle. Under such conditions indoor humidity tracks outdoor weather, and the temperature–humidity relationship encoded by Eqs. (3), (4) and (6) is already fully expressed in the raw pair. These derived quantities are exact deterministic functions of temperature and relative humidity; where they add no information beyond the raw variables, they add only parameters. Where moisture is actively controlled, the situation differs. Humidification, ventilation and heating act on the moisture balance through distinct physical pathways, and AH as defined in Eq. (4) is conserved under sensible heating and cooling, linearising the effect of thermal actuation on moisture content. In that setting, the derived quantities separate mechanisms that the raw pair conflates. The revised proposition is therefore that psychrometric encoding assists humidity forecasting where there is active moisture control to model, and not where indoor humidity essentially follows ambient conditions. This is narrower than the hypothesis originally posed, and it is a statement these data motivate rather than establish. Testing it properly requires paired controlled and passive structures with logged actuator states, a design neither available dataset provides.

### 6.2 The exogenous-channel null is the more surprising result

That the psychrometric derivations failed to help is explicable by collinearity. That substrate moisture, substrate temperature and $CO_2$ also failed, as Table 9 shows, is not. These are independent measurements. Prior reasoning, including our own, identified substrate and surface moisture as the highest-value additional instrumentation for humidity modelling, on the argument that humidity in a mushroom house is governed by evaporation and condensation at surfaces only partially observable from ambient sensors [7]. The data do not support this expectation: every exogenous channel degraded forecast accuracy at every horizon. Two readings are available. The measurements may be too indirect, since the capacitive and resistive probes sit within the substrate mass whereas the evaporating boundary is the substrate and fruiting-body surface, which may be hydrologically decoupled from the interior at the timescales examined. Alternatively, in a passively managed structure the surface may simply not be the rate-limiting term, with air exchange dominating instead. The two readings imply different next instruments: a surface wetness sensor in the first case, an air-exchange measurement in the second. These data cannot distinguish them. What can be excluded is model capacity. A linear model incorporates exogenous variables only additively, so a plausible objection to the null is that interactions were present but unrepresentable. Gradient-boosted trees are constructed precisely to capture such interactions [11], and found none: as Table 11 shows, the best outcome across both horizons was a 0.003%RH improvement with $p = 0.482$. The information is not present in a form any of the three tested model classes can use.

### 6.3 Chronological splitting confounds cycle phase with partition

The pre-registered chronological split failed the quality gate at the 60-minute horizon (Table 7), with partitions differing by 16 percentage points in mean humidity as shown in Figure 4. The problem is not peculiar to this dataset. Cultivation cycles progress through phases with materially different environmental characteristics by design [1]. Any chronological partition of cycle data therefore places systematically different regimes in training and test, and evaluates extrapolation rather than generalisation. The effect scales with horizon because longer forecasts depend more on learned structure and less on autocorrelation. The implication for practice is that a chronological split remains correct for assessing deployment performance, since deployment genuinely involves forecasting a later phase from earlier ones. It is not appropriate for isolating a feature-representation question, where phase differences act as a confound. Reporting both, as in Tables 7 and 8, distinguishes the two questions.

### 6.4 Persistence baselines are essential and frequently absent

The persistence baseline was informative at every stage and decisive twice. On the controlled-greenhouse dataset, persistence achieved $R^2$ of 0.917 for temperature and 0.929 for $CO_2$ under the published one-step-ahead formulation, approaching the reported model performance of 0.974 and 0.993 [6]. On the oyster data, the gate correctly halted analysis at the 60-minute horizon in the pre-registered split, as Table 7 shows. Autocorrelated environmental series at short horizons are largely predicted by their own most recent value [16]. A coefficient of determination above 0.9 may indicate a strong model or merely a strongly autocorrelated target, and the two are indistinguishable without an explicit baseline. We recommend that forecasting results in this literature report persistence comparison as standard.

### 6.5 The linear baseline is not the limiting factor

The comparison across model classes in Table 10 carries an implication beyond the immediate hypothesis. Ridge regression matched or outperformed both nonlinear alternatives on the baseline condition, decisively so at the 60-minute horizon, and the neural network was worst throughout, as Figure 6 makes visible. This is not an argument that neural architectures have no place in cultivation forecasting, nor is it, on the evidence of Section 5.2.6, an argument about capacity. The capacity sweep reported there found the smallest network worst in every condition and no systematic penalty for the largest, which is the reverse of what over-parameterisation predicts. What the sweep does establish is narrower and more secure: no architecture in the range tested beat a regularised linear model, and adding features degraded the network at every capacity. The relationship to be recovered is close to linear in the lagged inputs, and ridge regression represents it directly where a network must approximate it. Why the approximation falls short is not settled by these data. Published work in this area frequently adopts recurrent and attention-based architectures [1, 6] without reporting comparison against regularised linear alternatives. Our results suggest such comparison is worth making: where a linear model is competitive, the additional complexity requires justification on grounds other than accuracy.

### 6.6 Instrumentation

The sensor comparison of Tables 4 and 5 provides direct evidence for a claim more often asserted from datasheets. Two DHT21 units, co-located with a reference to within 0.75 ℃ (Table 3), read 6.33 and 2.04%RH above it, disagreed with each other by 4.29%RH, and spent 23.3% and 26.1% of the record at their ceiling while the reference never reached its own (Table 5). Their temperature channels were simultaneously accurate to within 0.03 and 0.39 ℃ (Table 6). The practical consequence is not that these sensors are inaccurate in a correctable way. It is that inter-unit discrepancy is neither constant nor predictable: +10.45%RH and stable in Cycle I, +4.29%RH with standard deviation 5.46 in Cycle II. A single-point calibration would transfer neither between units nor between deployments. Ceiling censoring compounds the problem. As Figure 3(b) illustrates, a sensor pegged at 100.0%RH for a quarter of the record provides no information precisely in the range where mushroom cultivation operates and where condensation risk arises.

### 6.7 Condensation margin as a monitoring target

Table 14 carries the most actionable result in the paper. Negative margin occupied 1.13% of the record, but 87% of that exposure fell on two consecutive days, as Figure 9(b) shows. Rare, clustered risk suits monitoring and targeted intervention: a grower does not require continuous humidity suppression, only protection on identifiable days. Detection requires one surface temperature measurement and the arithmetic of Eqs. (6) and (7), applied to quantities already available. The proxy limitation is material. Substrate temperature is not cap-surface temperature, and the fruiting body, being wet, exposed and transpiring, will not track substrate temperature exactly. Confirmation requires infrared surface measurement.

### 6.8 From forecast accuracy to control action

Nothing in this study demonstrates that a better humidity forecast produces better environmental regulation, and the gap between the two deserves to be stated rather than assumed. Neither cycle logged actuator states, so no control loop can be reconstructed from these data and no closed-loop evaluation is possible. What can be done is to establish what forecast accuracy would have to be achieved before it mattered for control, and these results speak to that. A forecast enters a control decision through a threshold. A controller acting on predicted humidity fires when the forecast crosses a setpoint by some margin, and that margin must exceed the model's own error or the controller will act on noise. The forecast errors reported here therefore set a floor on the deviation a predictive controller could reliably detect. At the 15-minute horizon, the best model achieves 1.883%RH, so predicted excursions smaller than roughly 2%RH are not separable from forecast error. Against a cultivation envelope commonly specified as 85 to 95%RH, a 10-point band, that floor consumes a fifth of the tolerance before any actuator responds. Three consequences follow. First, the difference between the feature conditions examined here, at most 0.16%RH, is an order of magnitude below the threshold at which a controller would behave differently, so the null reported in Section 5.2.2 has no operational cost. Second, horizon choice is a control decision rather than a modelling one: it should match the actuator's response time, since a forecast further ahead than the plant can respond to is unusable, and one shorter than the response time arrives too late. Third, the condensation margin of Eq. (7) is closer to control-ready than the humidity forecast, because it is a present-state measurement compared against a threshold with clear physical meaning rather than a prediction carrying its own error. Establishing that a forecast improves regulation requires a paired trial: two houses, identical instrumentation, control law as the sole difference, with actuator states and energy logged and integrated setpoint deviation as the outcome. That is the natural successor to this work, and the sensor findings of Section 6.6 are a precondition for it, since a controller acting on a sensor saturated for a quarter of the record would be acting on a constant.

### 6.9 Limitations

No confirmatory test was completed. The pre-registered analysis failed its quality gate (Table 7), and the declared

deviation produced a marginal primary result that did not survive the dimensionality control (Table 8). Nothing here confirms or refutes the representation hypothesis in general; the finding is that it is not supported in these two passive cultivation cycles.

Secondary analysis. We did not design the instrumentation, control the environment, or determine the sampling regime. Node placement is not documented in the source dataset [9], so co-location is inferred from the temperature agreement of Table 3 rather than known.

No reference in Cycle I. The Cycle I inter-unit offset establishes disagreement, not which unit was correct.

Passive management. Neither cycle logged actuator states, and observed temperatures indicate little or no active climate control. Findings may not extend to controlled facilities. Indeed, the argument of Section 6.1 is that they do not.

The controlled-greenhouse comparison is not a test. Absent timestamps and with interleaved missingness, that dataset [6] supports contrast only.

Model classes tested. Three classes were compared (Tables 10 and 11) and none extracted value from the additional features. This addresses the most obvious alternative explanation for the nulls, but does not exhaust the model space: recurrent and attention-based architectures were not evaluated, and the data volume may be insufficient to train such models effectively, as the neural network result suggests.

Scope of generalisation. The evidence base is two cultivation cycles of one species at one site, and the boundaries of what it can support deserve stating individually rather than as a single caveat.

Across species, the sensor findings should transfer and the feature findings should not. A DHT21 saturating at its ceiling does so because of its own construction, irrespective of what is growing beneath it, so the instrumentation result in Section 6.6 is a property of the device. The feature result is not: it rests on the argument of Section 4.3.1 that psychrometric transformations pay off only where moisture is actively driven, and species differ in the humidity envelopes their cultivation demands. Agaricus bisporus is grown under tighter humidity control than Pleurotus ostreatus and would plausibly place a house nearer the actively controlled regime where the transformations do carry information. Across structures and climates, the constraint is sharper. Both cycles ran in a passively managed house through a Northern Hemisphere winter and spring, with median temperatures well below any oyster cultivation setpoint and no actuator states recorded. Where indoor humidity tracks outdoor weather, the raw temperature and humidity pair already expresses everything the derived quantities encode. A mechanically conditioned house, or the same house in a different season, may not resemble this one. The contrast with the actively controlled greenhouse of study [6] in Section 5.2.5 is consistent with that reading but does not establish it, since that dataset lacks timestamps. Across control strategies, we make no claim at all. Neither cycle logged actuation, so nothing here speaks to how a forecast would behave inside a control loop. Section 6.8 sets out what would be required to test that. The condensation-margin result occupies a middle position. The physics of Eq. (7) is general, but the observed exposure of 1.13% of the record concentrated on two days reflects this house in this season, and the clustering pattern should not be assumed elsewhere. We therefore present the sensor findings as likely to generalise across deployments of the same devices, the methodological findings on persistence baselines and chronological partitioning as applying to any cycle-structured environmental series, and the feature findings as bounded to passively managed cultivation until tested against a controlled house with logged actuation.

### 6.10 Recommendations

For researchers working with cultivation environmental data:

(1) Report a persistence baseline for every forecasting result. Without one, reported accuracy on autocorrelated series is uninterpretable [16].

(2) State the partitioning scheme explicitly, and prefer segment-interleaved splits when isolating modelling questions from cycle-phase effects, while retaining chronological splits for deployment assessment. Avoid random splitting of autocorrelated series [17].

(3) Validate humidity instrumentation against a reference before treating measurements as ground truth, particularly above 90%RH where ceiling censoring becomes likely (Table 5).

(4) Compare against a regularised linear baseline before adopting high-capacity architectures. At single-cycle data volumes, the additional complexity may not be repaid in accuracy (Table 10).

(5) Record actuator states. Their absence prevents assessment of whether environmental variation is controlled or ambient, which Section 6.1 suggests determines whether derived features carry information.

For reusers of the source dataset [9] specifically, we note two issues. The Cycle I DHT Hum - ESP8266.xlsx series correlated at −0.128 with the corresponding ESP32-02 cloud export over their common interval, indicating divergent provenance. One Cycle I node recorded 46.2% of its humidity readings at the sensor ceiling from at least 9 January onward.

## 7. CONCLUSIONS

This secondary analysis of the open oyster mushroom cultivation dataset of Duman et al. [9] produced four findings. Psychrometric feature encoding, as defined in Eqs. (3), (4) and (6), did not improve humidity forecasting in either cultivation cycle. Under the pre-specified analysis the quality gate failed (Table 7) and no test was reported; under a declared deviation the primary comparison was marginal ($p = 0.055$) and did not survive the matched-dimensionality control ($p = 0.472$). One reading of the contrast with [6] is that the benefit reported there depends on an actively driven moisture balance, which the passively managed house studied here lacks. We put this forward as a hypothesis rather than a conclusion. It rests on a contrast between two datasets that differ in several respects besides control regime, one of which lacks timestamps and so cannot be analysed on comparable terms, and it has not been tested against a controlled house with logged actuation. The conditions under which it would be falsified are set out in Section 4.3.1; until that test is performed, the mechanism remains conjectural. Exogenous measurements, namely substrate temperature, substrate moisture, and $CO_2$, degraded forecast accuracy at every horizon (Table 9), contrary to the prior expectation that direct moisture sensing would be the highest-value addition. The result held across all three model classes tested (Table 11), including gradient-boosted trees constructed to detect interactions a linear model cannot

represent, which rules out insufficient capacity as an explanation within the range examined. It does not rule out architectures we did not test, and Section 5.2.7 shows that the neural network estimates carry across-seed spread of up to 0.892%RH, so those figures in particular should be read as indicative. Neither low-cost humidity sensor tracked the reference, and the two disagreed with each other as well (Table 4). Their temperature channels remained accurate throughout (Table 6), which places the fault in the humidity channel rather than in placement, wiring, or firmware. The offsets were of different magnitude in the two cycles, so no single calibration constant would carry across units or deployments. Condensation risk, computed from Eq. (7), occupied 1.13% of the record with 87% of exposure falling on two consecutive days (Table 14, Figure 9). If that concentration recurs, targeted intervention would be more efficient than continuous humidity suppression; whether it recurs across seasons and structures is untested, and the substrate probe used here is a proxy for the cap surface rather than a measurement of it. Regularised linear regression held its own against a gradient boosting ensemble and a feedforward neural network (Table 10, Figure 6), and the network alone worsened as inputs multiplied. At the data volumes a single cultivation cycle affords, capacity is not what limits forecast accuracy. Two methodological observations follow. Persistence baselines are necessary for interpreting forecasting results on autocorrelated environmental series and are frequently absent from this literature. Chronological partitioning of cultivation-cycle data confounds cycle phase with partition membership and evaluates extrapolation rather than generalisation. The reproducibility information is found in the Appendix section.

## ACKNOWLEDGMENT

The source dataset is published by Duman et al. [9] under a Creative Commons Attribution licence. We thank the authors of the source dataset [9] for making the data openly available.

## REFERENCES

[1] Jacob, M.S., Xu, A., Qian, K., Qi, Z., Li, X., Zhang, B. (2025). Artificial intelligence in edible mushroom cultivation, breeding, and classification: A comprehensive review. Journal of Fungi, 11(11): 758. https://doi.org/10.3390/jof11110758

[2] Yang, K., Argyropoulos, D. (2026). Towards smart mushroom farming: A comprehensive review of data, sensors, robotics and artificial intelligence. Smart Agricultural Technology, 14: 102129. https://doi.org/10.1016/j.atech.2026.102129

[3] Chong, J.L., Chew, K.W., Peter, A.P., Ting, H.Y., Show, P.L. (2023). Internet of Things (IoT)-based environmental monitoring and control system for home-based mushroom cultivation. Biosensors, 13(1): 98. https://doi.org/10.3390/bios13010098

[4] Nguyen, H.H., Shin, D.Y., Jung, W.S., Kim, T.Y., Lee, D.H. (2024). An integrated IoT sensor-camera system toward leveraging edge computing for smart greenhouse mushroom cultivation. Agriculture, 14(3): 489. https://doi.org/10.3390/agriculture14030489

[5] Irwanto, F., Hasan, U., Lays, E.S., et al. (2024). IoT and fuzzy logic integration for improved substrate environment management in mushroom cultivation. Smart Agricultural Technology, 7: 100427. https://doi.org/10.1016/j.atech.2024.100427

[6] Huang, S., Liu, Q., Wu, Y., Chen, M., Yin, H., Zhao, J. (2024). Edible mushroom greenhouse environment prediction model based on attention CNN-LSTM. Agronomy, 14(3): 473. https://doi.org/10.3390/agronomy14030473

[7] Allen, R.G., Pereira, L.S., Raes, D., Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper, 56: D05109.

[8] Tetens, O. (1930). Uber einige meteorologische. Begriffe, Zeitschrift fur Geophysik, 6: 297-309.

[9] Duman, S., Elewi, A., Hajhamed, A., Khankan, R., Souag, A., Ahmed, A. (2024). A novel dataset of annotated oyster mushroom images with environmental context for machine learning applications. Data in Brief, 57: 111074. https://doi.org/10.1016/j.dib.2024.111074

[10] Hoerl, A.E., Kennard, R.W. (1970). Ridge regression: biased estimation for nonorthogonal problems. Technometrics, 12(1): 55-67. https://doi.org/10.1080/00401706.1970.10488634

[11] Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. Annals of Statistics, 29(5): 1189-1232. https://doi.org/10.1214/aos/1013203451

[12] Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8): 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[13] Bai, S., Kolter, J.Z., Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv:1803.01271. https://doi.org/10.48550/arXiv.1803.01271

[14] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30: 5998-6008.

[15] Lim, B., Arik, S.O., Loeff, N., Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4): 1748-1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

[16] Hyndman, R.J., Koehler, A.B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4): 679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001

[17] Bergmeir, C., Benítez, J.M. (2012). On the use of cross-validation for time series predictor evaluation. Information Sciences, 191: 192-213. https://doi.org/10.1016/j.ins.2011.12.028

[18] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12: 2825-2830. https://doi.org/10.5555/1953048.2078195

[19] Hurlbert, S.H. (1984). Pseudoreplication and the design of ecological field experiments. Ecological Monographs, 54(2): 187-211. https://doi.org/10.2307/1942661

[20] Diebold, F.X., Mariano, R.S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3): 253-263. https://doi.org/10.1080/07350015.1995.10524599

[21] Harvey, D., Leybourne, S., Newbold, P. (1997). Testing the equality of prediction mean squared errors. International Journal of Forecasting, 13(2): 281-291. https://doi.org/10.1016/S0169-2070(96)00719-4

[22] Newey, W.K., West, K.D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55(3): 703-708. https://doi.org/10.2307/1913610

## NOMENCLATURE

| | |
|---|---|
| AH | absolute humidity, $g.m^{-3}$ |
| CI | confidence interval |
| DHT | digital humidity temperature |
| DS18B20 | Dallas Semiconductor |
| $e_a$ | actual vapour pressure, kPa |
| $e_s$ | saturation vapour pressure, kPa |
| IoT | Internet of Things |
| $M_c$ | condensation margin, ℃ |
| $n$ | number of observations, dimensionless |
| $p$ | probability value, dimensionless |
| $r$ | Pearson correlation coefficient, dimensionless |
| $R^2$ | coefficient of determination, dimensionless |
| ReLU | rectified linear unit |
| RH | relative humidity, % |
| RMSE | root mean squared error, %RH |
| SHT | sensor humidity temperature |
| $T$ | air temperature, ℃ |
| $T_{dew}$ | dew point temperature, ℃ |
| $T_{surface}$ | substrate surface temperature, ℃ |
| VPD | vapour pressure deficit, kPa |

### Greek symbols

| | |
|---|---|
| $\alpha$ | significance level, dimensionless |
| $\gamma$ | intermediate term in the inverse Magnus relation, dimensionless |
| $\rho$ | Spearman rank correlation coefficient, dimensionless |

### Subscripts

| | |
|---|---|
| a | actual |
| dew | dew point |
| s | saturation |
| surface | substrate surface |

## APPENDIX

Reproducibility. The analysis was performed in Python 3.12.3 using NumPy 2.4.4, pandas 3.0.2, SciPy 1.17.1, scikit-learn 1.8.0, matplotlib 3.10.8 and openpyxl for reading the source workbooks. Model hyperparameters appear in full in Table 2; values not listed there were left at the scikit-learn 1.8 defaults. Every fit used seed 0 unless a seed sweep is reported, and Section 5.2.7 quantifies the sensitivity of each model class to that choice. The processing chain runs as follows. Manufacturer error codes are removed at load time, namely -127 ℃ for the DS18B20 no-reading condition, -50 ℃ and 149.9 ℃ for DHT out-of-range, and humidity values of zero. Series are resampled to one-minute means and segmented into contiguous runs of at least 301 minutes, this being a four-hour input window plus a one-hour horizon plus one sample. Lagged design matrices are built at {0, 5, 10, 15, 30, 45, 60, 90, 120, 180, 240} minutes and never span a segment boundary. Partitioning is by whole segment under both schemes of Section 4.5. The analysis code is organised as seven modules: psychrometric feature construction implementing Eqs. (1) to (7), with a numerical verification suite checking each against published reference values; data loading; preprocessing and segmentation; sensor validation; the ablation, quality gate and hypothesis test; the statistical procedures; and figure generation. The source data remain with Duman et al. [9] under CC BY and are not redistributed. Two aspects of the pipeline are worth flagging for anyone rebuilding it. The one-minute model resolution is not a free parameter: coarsening to five minutes smooths the target and materially changes the results. And condition F0P is randomised, so a single realisation is a noisy estimate of the dimensionality penalty; Section 5.2.7 reports it averaged over five shift seeds.