Computers and Electronics in Agriculture 246 (2026) 111593

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

A multivariate autoregressive multilayer perceptron model for predicting internal beehive conditions from sensor data

M. Carmen Robustillo [a],*, Diren Senger [b], M. Isabel Parra [c], Carlos J. Pérez [d]

[a] Departamento de Análisis Matemático, Estadística e Investigación Operativa, y Matemática Aplicada, Universidad de Málaga, Campus de Teatinos s/n, Málaga, 29071, Spain
[b] AG Cognitive Neuroinformatics, University of Bremen, Enrique-Schmidt-Str. 5, 28359 Bremen, Germany
[c] Departamento de Matemáticas, Universidad de Extremadura, Avenida de Elvas s/n, Badajoz, 06006, Spain
[d] Departamento de Matemáticas, Universidad de Extremadura, Campus Universitario s/n, Cáceres, 10003, Spain

ARTICLE INFO

Keywords:
Autoregressive models
Forecasting
Neural networks
Precision beekeeping
Time series

ABSTRACT

The global decline in bee populations poses a critical threat to biodiversity and ecosystem stability, motivating the adoption of precision beekeeping strategies that combine sensor networks with data-driven models to optimise hive management and reduce colony losses. This study introduces a multivariate autoregressive multilayer perceptron (AMLP) model that integrates historical internal hive variables (temperatures, weight, humidity, and pressure) with external climatological data to forecast future states of these endogenous variables. Data were collected from 13 sensor-equipped hives of the BeeObserver project. The AMLP was evaluated against a standard multilayer perceptron (MLP) and a vector autoregressive (VAR) model using 10-fold rolling-window cross-validation. Forecast performance was assessed using two different error metrics for 1- and 3-day horizons. Across all hives, the AMLP reduced the mean percentage error by approximately 6%–7% relative to the MLP and up to 1.3% relative to the VAR, achieving superior predictive accuracy, with statistically significant improvements for most internal variables. By combining autoregressive lags with neural network flexibility, the AMLP captures both temporal dependencies and specific patterns while supporting incremental retraining as new data arrive. This approach provides scalable, adaptive, and real-time prediction of hive dynamics, offering a robust tool for proactive decision-making in precision beekeeping. The results demonstrate that integrating temporal and environmental information through AMLP models enhances predictive accuracy and supports timely interventions, ultimately improving colony health and resilience. These findings highlight the potential of advanced data-driven forecasting models to strengthen sustainable apiculture practices and contribute to the conservation of bee populations.

1. Introduction

In recent decades, a marked decline in the diversity and abundance of bee species has been documented worldwide, raising serious concerns about pollinator sustainability and ecosystem health (Zattara and Aizen, 2021). The loss of bee colonies has direct implications for global food security, as it leads to a decrease in food availability and an increase in agricultural costs (Richardson et al., 2023). Estimates from recent studies suggest that the current honeybee population would need to increase approximately three to five times to meet the growing demand for crop pollination driven by rising global food needs (Nath et al., 2023).

This threat to bee populations has spurred the development of precision apiculture (or smart beekeeping). This approach integrates advanced technologies such as embedded sensors, wireless Internet of Things (IoT) networks, and machine learning tools to collect and exploit data on the behaviour and health of bee colonies. By employing precision beekeeping tools, beekeepers can optimise their management routines by reducing unnecessary visits to apiaries, thereby decreasing colony stress and enhancing operational efficiency (Stojanova et al., 2025).

The correct implementation of precision beekeeping systems not only aims to support beekeepers in their management tasks, but also addresses broader issues related to environmental sustainability, food security, and the preservation of bee colonies (Urban and Chlebo, 2024). Precision beekeeping employs in-hive and on-hive sensors, such as weight scales, temperature and humidity probes, and occasionally acoustic or video systems connected through low-power IoT networks for continuous non-invasive monitoring (Alleri et al., 2023). These

* Corresponding author.
E-mail addresses: mcrobustillo@uma.es (M.C. Robustillo), diren@uni-bremen.de (D. Senger), mipa@unex.es (M.I. Parra), carper@unex.es (C.J. Pérez).

https://doi.org/10.1016/j.compag.2026.111593
Received 3 November 2025; Received in revised form 12 January 2026; Accepted 20 February 2026
Available online 25 February 2026
0168-1699/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

devices transform colony dynamics (thermoregulation, foraging, brood development, …) into high-resolution data streams used to detect anomalies, make predictions, and guide management decisions. Recent studies demonstrate their practical value for real-time hive assessment (Danieli et al., 2024) and short-term forecasting (Kulyukin et al., 2024).

Despite the significant progress that precision beekeeping represents for the apiculture sector, beekeepers have not yet adopted it widely in their daily practice (Verbeke et al., 2024). This is primarily due to the limited network connectivity in remote apiary locations and the relatively high cost of advanced sensing systems. Furthermore, ongoing developments should transition from retrospective analysis to predictive health monitoring, with the goal of providing beekeepers with early warnings that enable preventive intervention (Sattler et al., 2025). This would address one of the concerns of beekeepers regarding the economic cost of these systems and their long return on investment periods (Vardakas et al., 2025).

Currently, the most widespread application of precision beekeeping systems is real-time monitoring of hive conditions, providing beekeepers with up-to-date information on the status of their colonies (Urban and Chlebo, 2024). As a result, most research articles in this field focus on the development and improvement of classification models aimed at detecting potential adverse events, such as the detection of hive collapse (Colin et al., 2025), varroa mite detection (Ghezal and König, 2025; Scutaru et al., 2025), queenless hive detection (Otesbelgue et al., 2025; Sad et al., 2025), identification of changes in weight patterns that may indicate honey theft (Liu and Lin, 2025) or potential health issues within the colony (Degenfellner and Templ, 2024), or the identification of optimal locations for hive to maximise honey production (Sari and Sarvia, 2025; Kotovs et al., 2025), among others. These advances in the development and implementation of classification models represent a significant step forward for the field of precision beekeeping.

In spite of notable scientific progress, there remains a considerable lack of development and application of models dedicated to predicting internal hive conditions (Šabić et al., 2025), likely due to the challenging nature of this kind of data. Most existing studies have focused on the univariate prediction of individual hive variables, employing a wide variety of modelling techniques, including classical time-series models (Kulyukin et al., 2024; Ilyés-Vincze et al., 2025), random forests (Anwar et al., 2022), regression models (Rigakis et al., 2023), and neural network architectures (Braga et al., 2021; Anwar et al., 2023; Kulyukin et al., 2024). In contrast, multivariate prediction frameworks that jointly account for the interdependence among internal hive parameters remain scarcely explored.

Among the few studies addressing multivariate forecasting, most have relied on classical multivariate time-series models, such as vector autoregressive (VAR) (Robustillo et al., 2022; Bono et al., 2024) and vector error correction models (VEC) (Robustillo et al., 2024). However, to the best of the authors' knowledge, no previous studies have developed multivariate neural network models to assess the ability of these statistical techniques to handle large volumes of sensor data, capture complex patterns, and perform simultaneous multi-output prediction of internal hive conditions in honeybee colonies.

Considering the predictive power of neural networks, evaluating their applicability to precision apiculture datasets constitutes a relevant scientific endeavour. Equally important is ensuring that model calibration and forecasting can be performed within a practical time frame to support real-time decision-making. In this context, the multilayer perceptron (MLP) architecture was chosen over others such as Long Short Term Memory (LSTM) or Recurrent Neural Network (RNN) models due to its generally faster convergence and lower computational cost, making it more suitable for near-real-time systems. This speed advantage is supported by numerous studies in the literature (e.g. Andrade et al. (2023), Li et al. (2025)). However, the MLP model has a limitation in time-series applications due to its lack of temporal memory. To overcome this, a novel autoregressive multilayer perceptron (AMLP) model addressing lags is designed, implemented, and evaluated to capture temporal dependencies on multi-sensor data collected from beehive colonies. It is aimed to evaluate its predictive capacity within a multi-input multi-output framework, where multiple sensor variables jointly describe the internal state of the hive.

The proposed approach is conceptually inspired by the proven strengths of VAR models in handling interdependent time-series data, which leads us to incorporate lags into the MLP architecture. In this way, the conventional multilayer perceptron model is modified by incorporating time-lagged inputs, enabling it to capture both temporal dependencies as well as linear and non-linear relationships across correlated variables. In essence, the AMLP operates as a feed-forward neural network that learns from the recent history of hive parameters (temperature, humidity, weight...) to forecast their future evolution over different time horizons. The model combines the interpretability and structure of classical time-series modelling with the flexibility and power of neural networks. This hybrid formulation allows the AMLP to efficiently learn complex, non-linear interactions while maintaining computational feasibility for real-time applications.

To validate the model performance, the AMLP is compared with the standard MLP and VAR models on 13 hives monitored within the BeeObserver project (Senger et al., 2023). The used dataset includes internal measurements of temperature at various points, relative humidity, weight, and internal pressure, complemented by meteorological data that capture the external environmental conditions surrounding each hive. This creates a rich multivariate dataset that reflects both internal colony dynamics and environmental influences. The results demonstrate that the AMLP offers improved adaptability and better results for multivariate hive data.

The remainder of this article is structured as follows. Section 2 describes the datasets used in this study. Section 3 details the multivariate neural network approach proposed for forecasting, the error metrics applied to evaluate predictive accuracy, the cross-validation procedure adopted to ensure the robustness of the results, and the statistical tests employed for comparative analysis. Section 4 reports the experimental setup and results obtained from the evaluated models. Section 5 provides a critical discussion of the results. Finally, Section 6 outlines the main conclusions derived from the study.

2. Beehives and sensor data

Data used in this study were obtained from the BeeObserver project, a citizen science initiative. The dataset is publicly available on Zenodo (Senger et al., 2023) and fully described by Senger et al. (2024). Beekeepers received comprehensive support for setting up the sensors in their hives, including an installation manual and personal guidance. Beekeepers that had their bees on multiple boxes were asked to place the sensors in the box where the brood and bees were most likely to be found. This support allowed them to install the sensors and tailor them to the unique conditions of their hives (Johannsen et al., 2020). There were no instructions or limitations regarding beekeeping practices, hive types, and hive locations, aiming to have different beekeeping practices represented in the data. The database comprises nine internal variables in total, eight recorded within the hive, and one measured outside.

Fig. 1 presents a schematic representation of the sensor distribution used in the BeeObserver project monitoring system. Five DS18B20 temperature ($°C$) sensors (1) are installed in every second inter-frame space, with t_i_3 positioned at the center, between t_i_2 and t_i_4 and t_i_1 and t_i_5 located at the margins of the hive. An additional DS18B20 sensor, hanging outside at the hive, monitors the ambient temperature (t_o, in $°C$). Load cells Bosche H30 or H40 (2) connected with a HX711 amplifier and installed beneath the hive, continuously monitor the total hive weight (w, in kg). A BME280 sensor (3), situated between two frames, simultaneously records relative humidity (h, in %), internal temperature (t, in $°C$), and internal air pressure (p, in hPa). All sensor data are processed by a microcontroller unit (4) and transmitted, along

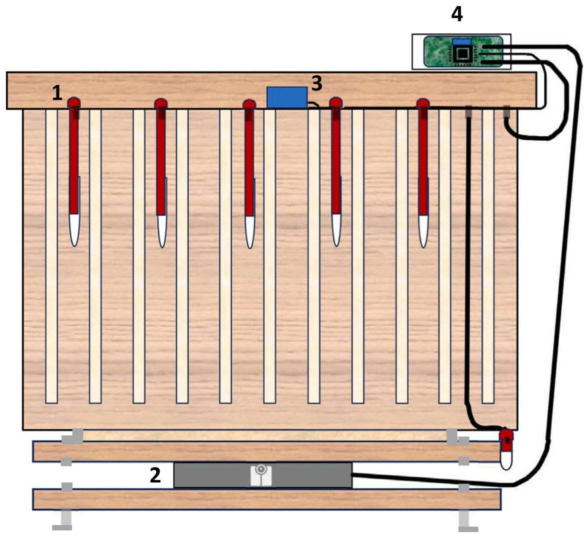


Fig. 1. Sensor setup example proposed by BeeObserver project: DS18B20 temperature sensors (1), Bosche H30/H40 load cell for weight measurements (2), BME280 sensor for humidity, internal air pressure, and temperature (3), and ESP32 micro-controller (4). Own image inspired by the original sensor setup map from the BeeObserver project, first published in Johannsen et al. (2020).

the timestamp corresponding to each recorded data point, and the hive ID, to a centralised database for further analysis using wireless LAN.

The procedure followed to select the beehives included in this study is detailed next. Initially, raw data from all available beehives were downloaded from the database. For each beehive, the data were then consolidated into a single file, as the original datasets were separated by year. Once the data were unified, hourly records were extracted for each beehive, and missing time points filled using NA values. Subsequently, the percentage of missing data and the total number of recorded measurements were calculated for each beehive. Based on this information, the final subset of selected beehives included only those that met the following criteria: (i) availability of geographical location coordinates (latitude and longitude); (ii) a minimum of six consecutive months of recorded data; and (iii) a proportion of missing values below 20% for all internal variables.

After applying these criteria, a total of 13 beehives were selected for analysis. Throughout this article, the beehives are referred to using the same identifiers (IDs) as those in the reference publication and in the corresponding Zenodo dataset. Table 1 summarises, for the 13 selected beehives, the number of months of data collected, the overall percentage of missing internal data, and the percentage of missing data for each variable.

As can be observed, most hives exhibit a similar percentage of missing data across all variables. This occurs because, at certain times, data recording was suspended for all variables simultaneously. Furthermore, since there is a sensor that simultaneously measures three variables (h, t, and p), these three variables consistently display the same percentage of missing values. When this sensor fails, none of these variables were measured. In addition, it can be observed that the sensor *t_i_5* occasionally shows a considerably higher amount of missing data than the other variables, and in hive 86, no data were recorded from this sensor at all.

As the BeeObserver database includes only one external variable (the temperature, *t_o*), it was supplemented with meteorological data from nearby German weather stations. Specifically, data were collected on external temperature, humidity, precipitation, wind, and pressure. The historical time series can be found on the official website www.dwd.de. In particular, hourly data are available for download at https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/hourly/.

Fig. 2 shows the approximate locations of the 13 beehives on a map. Some locations overlap due to hives being situated at the same or adjacent sites. Additionally, to the right of the map, a table details the coordinates of each hive and the nearby meteorological station from which the external variables were sourced. The dataset did not provide the exact locations of the apiaries for confidentiality reasons but included coordinates rounded to the first digits with an approximate spatial precision of 5–7 km on the longitude scale for central Europe.

3. Methodology

This section describes the methodology applied to preprocessed hive datasets, with particular emphasis on the design and implementation of a multivariate neural network model for forecasting the internal conditions of the beehive. Then, it introduces the evaluation metrics used to quantify predictive accuracy, followed by the cross-validation framework designed to rigorously evaluate model performance in predicting hive internal conditions.

3.1. Autoregressive multilayer perceptron for time series

This study presents a hybrid model that integrates the predictive strength of multilayer perceptron (MLP) neural networks (Garcia Cabello, 2022) with the approach of classical autoregressive time series models, with the objective of enhancing the predictive accuracy of MLP models in time series forecasting. The proposed model is an autoregressive multilayer perceptron neural network model of order p (AMLP$_p$), which is characterised by $n = n' + p \cdot m$ input neurones, H hidden layers and m output neurones, where n', m and p, are the numbers of exogenous variables, endogenous variables, and lags, respectively. This network is denoted as AMLP$_p(h_1, \dots, h_H)$, where h_i denotes the number of neurones in the ith hidden layer. The AMLP$_p$ model corresponds to a function:

$$\begin{aligned} F &: \mathbb{R}^n \longrightarrow \mathbb{R}^m \\ \mathbf{z_t}, \mathbf{y_{t-1}}, \dots, \mathbf{y_{t-p}} &\longmapsto F(\mathbf{z_t}, \mathbf{y_{t-1}}, \dots, \mathbf{y_{t-p}}) \\ &= \psi\left(f_H\left(\dots\psi\left(f_2\left(\psi\left(f_1(\mathbf{z_t}, \mathbf{y_{t-1}}, \dots, \mathbf{y_{t-p}})\right)\right)\right)\right)\right) \end{aligned}$$

where $\mathbf{z_t}$ are the n' exogenous variables $(z_{1,t}, \dots, z_{n',t})$, $\mathbf{y_{t-i}}$ denotes the ith lags of the m response endogenous variables, meaning $(y_{1,t-i}, \dots, y_{m,t-i})$, f_i is a transfer function, and ψ the activation function.

In this model, $f_i : \mathbb{R}^{h_{i-1}} \longrightarrow \mathbb{R}^{h_i}$ is an affine function of the form $f_i(\mathbf{x}) = \mathbf{w}^i\mathbf{x} + \mathbf{b}^i$, where $\mathbf{x}$ denotes an h_{i-1}−dimensional vector whose elements correspond to the outputs produced by the neurones in the preceding layer. Besides, $\mathbf{x_t}$ denotes the n input variables $(\mathbf{z_t}, \mathbf{y_{t-1}}, \dots, \mathbf{y_{t-p}})$ so that, the linear multilayer perceptron neural network model is:

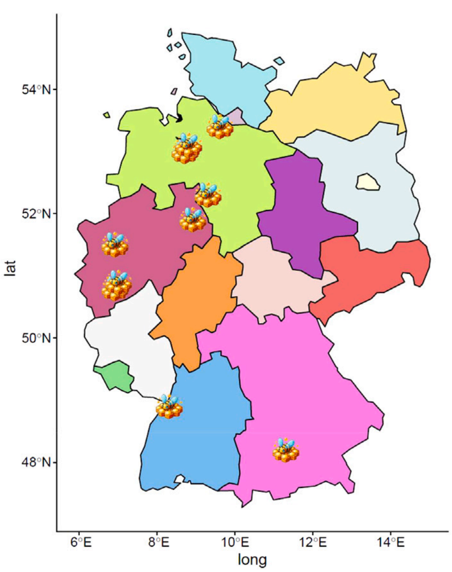
$$\mathbf{y_t} = F(\mathbf{x_t}) = \psi\left(\mathbf{w^H}(\dots\psi(\mathbf{w^2}\psi(\mathbf{w^1}(\mathbf{x_t} + \mathbf{b^1}) + \mathbf{b^2})\dots)) + \mathbf{b^H}\right)$$

The functional formulation above expresses the AMLP$_p$ model as a composition of transformations mapping an input $\mathbf{x_t} \in \mathbb{R}^n$ to an output $\mathbf{y_t} \in \mathbb{R}^m$. This perspective highlights the AMLP$_p$ as a function F constructed through successive applications of layer-specific mappings.

Another, but equivalent, way of defining an AMLP$_p$ is in terms of graph theory. In this view, the model is represented as a directed graph whose nodes correspond to neurones arranged in successive layers. Edges connect neurones in adjacent layers and are associated with weights, which encode the strength of the connection from the ith neurone in one layer to the jth neurone in the next. In this formulation, the functional mapping F described above arises naturally from the propagation of information through the graph, where each layer applies an affine transformation f, followed by the application of an activation function, whose result constitutes the new input for neurones in the subsequent layer. The activation function ψ can take multiple forms, the most common and well-known being tanh, sigmoid, ReLU, linear,

Table 1
Information about studied beehives including the data collection period (in months), the total percentage of missing values in the internal variables, and the percentage of missing data for each individual internal variable. Mean and standard deviation (Sd) are presented in the last two rows.

Hive	Months	Total % internal	NAs % for each internal variable								
ID		NAs	*t_i_1*	*t_i_2*	*t_i_3*	*t_i_4*	*t_i_5*	*w*	*h*	*t*	*p*
0	17	17.43	17.92	17.90	17.91	17.91	17.90	16.83	16.83	16.83	16.83
26	40	16.81	10.77	8.24	10.81	9.23	84.40	6.35	7.15	7.15	7.15
48	6	5.42	5.42	5.42	5.42	5.42	5.42	5.42	5.42	5.42	5.42
49	23	19.02	15.25	15.24	15.27	18.45	22.41	15.12	23.15	23.15	23.15
60	8	3.94	3.94	3.94	3.94	3.94	3.94	3.94	3.94	3.94	3.94
79	24	14.17	19.61	19.55	19.55	19.62	19.64	7.40	7.40	7.40	7.40
84	8	18.59	1.81	1.81	1.81	1.81	23.46	1.81	44.94	44.94	44.94
86	16	19.86	9.84	9.84	9.84	9.84	100	9.84	9.84	9.84	9.84
96	28	14.58	30.18	9.22	9.22	9.22	26.37	9.22	12.61	12.61	12.61
100	22	11.51	11.55	11.53	11.54	11.55	11.56	11.34	11.50	11.50	11.50
111	12	6.82	6.66	6.66	6.66	6.66	6.66	6.66	7.13	7.13	7.13
123	30	13.02	8.21	8.21	8.21	8.21	8.21	8.21	22.64	22.64	22.64
135	18	8.66	7.86	5.92	5.92	5.92	5.92	5.92	13.49	13.49	13.49
Mean	19.38	13.06	11.46	9.50	9.70	9.83	25.84	8.31	14.31	14.31	14.31
Sd	9.85	5.41	7.69	5.32	5.32	5.67	30.58	4.22	11.03	11.03	11.03



	Hive Location		DWD Station	
Hive ID	lat	long	lat	long
0	51.9	8.9	51.87	9.27
26	50.8	6.9	50.86	7.16
48	53.1	8.8	53.05	8.80
49	50.9	7.0	50.86	7.16
60	52.3	9.3	52.32	9.43
79	53.1	8.8	53.05	8.80
84	53.0	8.7	53.05	8.80
86	53.1	8.8	53.05	8.80
96	53.4	9.6	53.99	9.57
100	48.2	11.3	48.21	11.20
111	51.5	6.9	51.40	6.97
123	48.9	8.3	48.97	8.33
135	51.9	8.9	51.87	9.27

Fig. 2. Map of Germany with the location of the 13 beehives, some of them overlap (left), and a table on the right providing the exact locations of both the analysed hives and the nearby meteorological stations used to obtain external variables (right).

ramp, or step, all of which, along with others, can be found defined in Apicella et al. (2021).

Fig. 3 illustrates an example of an $\text{AMLP}_2(5, 8, 5)$ when considering the neurones in the hidden layers, with a total of 4 regressor variables ($\mathbf{x}$), 3 response variables ($\mathbf{y}$), and 2-lags of response variables.

In order to provide predictions, note that the AMLP_p model is driven by past values of the response variables. Therefore, with the exception of the first forecast in the time horizon, subsequent predictions must be based on prior forecasted values.

3.2. Evaluation metrics

To evaluate the accuracy of the models in fitting and predicting the internal variables of the hives, two error metrics are employed. The first is the mean absolute error (MAE), which quantifies the average absolute difference between predicted and observed values of each response variable. This metric provides an intuitive measure of the typical magnitude of prediction errors, offering a straightforward interpretation of model performance for each variable of interest. Its formula is

$$MAE_l = \frac{1}{\eta}\sum_{k=1}^{\eta} |\hat{y}_{l,k} - y_{l,k}|,$$

where η is the length of the forecast horizon, $\hat{y}_{l,k}$ is the prediction obtained by the model for the lth response variable at the prediction horizon k, and $y_{l,k}$ is the real value for the lth response variable at the prediction horizon k.

Additionally, to assess the overall predictive performance of the models across all variables collectively, a new error metric is considered, i.e. the corrected mean absolute percentage error (MAPEc). This metric is a modification of the conventional mean absolute percentage error (MAPE), defined as

$$MAPE_l = \frac{1}{\eta}\sum_{k=1}^{\eta} \left|\frac{\hat{y}_{l,k} - y_{l,k}}{y_{l,k}}\right|.$$

The aim of this correction is to mitigate the inflation of MAPE when the actual values approach zero. This adjustment preserves the dimensionless nature of the metric, as both numerator and denominator remain expressed in the same units. The corrected metric, MAPEc, is

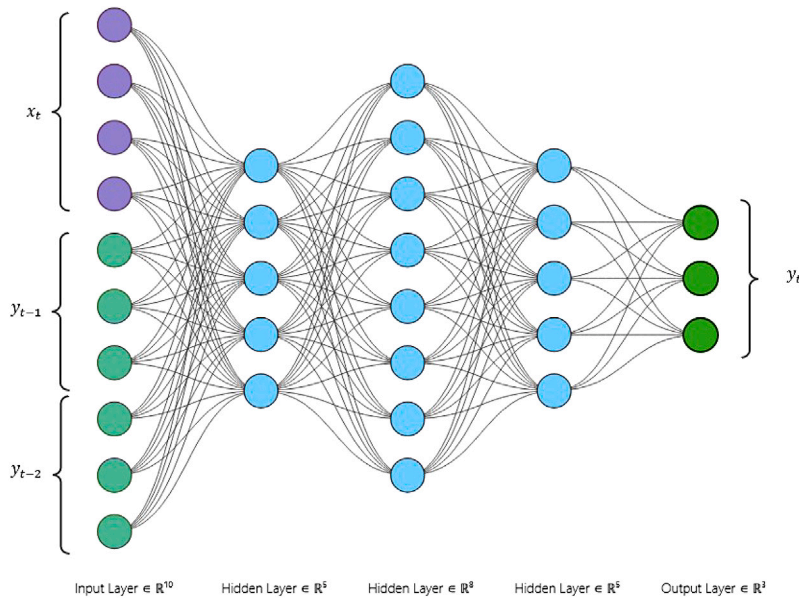


Fig. 3. An example of the graph representation of an AMLP$_2$(5,8,5), involving 10 input variables and 3 output variables.

therefore defined as

$$MAPEc_l = \frac{1}{\eta}\sum_{k=1}^{\eta} r_{l,k},$$

where

$$r_{l,k} = \begin{cases} \left|\dfrac{\hat{y}_{l,k} - y_{l,k}}{y_{l,k}}\right| & if \quad |y_{l,k}| \geq 1 \\ |\hat{y}_{l,k} - y_{l,k}| & if \quad |y_{l,k}| < 1 \end{cases}$$

On the other hand, the convergence and quality of neural networks are evaluated through the loss function. The loss function is a mathematical mapping that transforms the prediction errors produced by the model into non-negative values, thereby defining a gradient-based objective function that guides the learning process.

The parameter vector of the neural network model, $\boldsymbol{\theta}$, is optimised by minimising the distance between the actual and predicted values produced by the network. This distance is quantified using the loss function as follows:

$$\boldsymbol{\theta} = \frac{1}{T}\sum_{t=1}^{T} L(\mathbf{y_t}, f(\mathbf{x_t};\boldsymbol{\theta}))$$

where t is the time taken for each of the elements in the training set of length T.

3.3. Cross-validation

To ensure the robustness of the results, a cross-validation framework adapted to the temporal nature of the dataset is applied. The approach relies on the rolling window technique (Zivot and Wang, 2006), which adapts the principles of k-fold cross-validation to time series by explicitly accounting for temporal ordering when defining training and testing datasets. The procedure begins by dividing the time series into a training set and a test set. The model is then estimated on the training data, and η-step-ahead forecasts are produced for the test set, from which prediction errors are computed. Subsequently, the training window is advanced by a fixed increment δ, and the estimation and prediction steps are repeated using the updated data. This rolling process continues for k iterations (k-fold rolling window cross-validation) or until η-step forecasting is no longer feasible. The prediction errors obtained in each iteration are summarised using the mean absolute error and/or the corrected mean absolute percentage error, and the final results are reported as the mean and standard deviation of the metrics across all iterations.

3.4. Statistical tests

In order to evaluate whether the differences in MAE between the AMLP$_p$ and MLP models were statistically significant, paired t-tests were used when the normality assumption was met. When normality condition was violated, Wilcoxon signed-rank tests were used instead. Normality was assessed using the Shapiro–Wilk test. Statistical significance was set at p-value < 0.05.

4. Results

4.1. Experimental settings

This subsection provides detailed descriptions of the data preprocessing procedures, the fitting of statistical models, and the cross-validation method, ensuring that the entire experimental study can be replicated.

4.1.1. Data preprocessing

After obtaining the complete dataset, as described in Section 2, a data preprocessing phase is required to remove sensor-related errors and impute missing values, ensuring that the data are suitable for the subsequent training of the prediction algorithms.

Measurements clearly affected by sensor errors in the BeeObserver project were initially removed and considered as missing data. They were detected by applying plausible range limits. Specifically, internal temperatures below $0°C$ or above $60°C$, relative humidity values outside the 0–100% interval, hive weights less than 0 kg or over 100 kg, internal pressure over 1100 hPa, and external temperatures lower than $-20°C$.

Erroneous entries from external variables sourced from the German Meteorological Service website (www.dwd.de) were identified and removed based on the error codes associated with each variable, being subsequently treated as missing data.

After identifying all missing data, the imputation process was performed. First, missing external temperature values (t_o) in the BeeObserver dataset were imputed using temperature records from the nearest

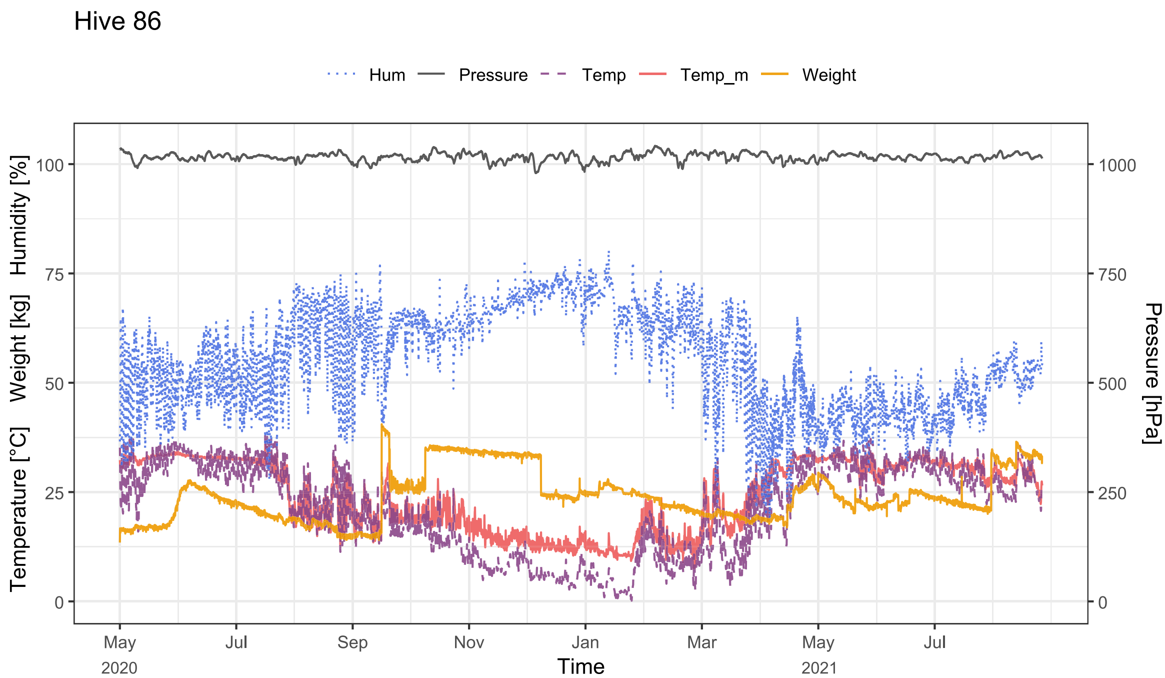


Fig. 4. Behaviour of the internal variables analysed for Hive 86. Humidity is shown as a blue dashed line, internal temperature (t) is represented by a purple dashed line, and the average temperature from sensors t_i_m is depicted as a solid pink line. Weight is shown with a solid orange line, while pressure is represented by a solid grey line. Weight, temperature, and humidity are plotted on the primary y-axis, whereas pressure is plotted on the secondary.

available meteorological stations to maximise data recovery. Subsequently, the remaining missing values for external temperature, as well as for other internal and external variables, were imputed using the Multiple Imputation by Chained Equations (MICE) method with a Random Forest algorithm (Jing et al., 2022).

The imputation process incorporated temporal information such as year, month, day, and hour of data collection, as their inclusion was observed to considerably improve the accuracy of the imputed values. The global imputation of the entire dataset was implemented using the `miceRanger()` function from the R package `miceRanger` (Wilson, 2021), applying the parameters `m` = 1, `maxiter` = 10, `valueSelector` = εvalueε, and `num.trees` = 200. Performing a single global imputation ensured that the model benefited from all available data while maintaining the dataset as realistic as possible for subsequent predictive modelling.

Once the imputation process was completed, an artificial variable was incorporated to detect abrupt weight changes likely due to beekeeper interventions (such as honey collection or feeding). This variable highlights weight differences that fall below the 2.5th percentile or above the 97.5th percentile, capturing the 5% most extreme observations.

After imputation, the five internal temperature variables (t_i_1 to t_i_5) were averaged to form a single variable, t_i_m. In the particular case of Hive 86, only four temperatures were considered for averaging due to the total absence of data from sensor t_i_5. Consequently, the dataset analysed included a total of five variables for prediction: two temperatures (t_i_m and t), weight (w), humidity (h), and internal pressure (p). This leads to preprocessed data that are used to model fitting and prediction purposes.

Fig. 4 illustrates the behaviour of the internal variables in hive 86 after data imputation. This hive exhibits the maximum rate of missing values (19.86%) in its internal variables.

It can be observed that the internal hive variables exhibit relevant irregularities, fluctuations, and abrupt changes. This presents a considerable challenge when modelling and forecasting this type of real-world time series data.

4.1.2. Statistical modelling

The number of variables to be predicted is $m = 5$. In the MLP model, the input vector $\mathbf{x_t}$ includes five external climatic conditions, one external variable representing extreme weight variations due to manual interventions on the hive, 5 temporal variables related to measurement time, and a constant variable of 1's, summing up to $n' = 12$. As a result, the AMLP model will consist of a total of $n = 12 + 5p$ input variables, where p is the number of lags, i.e., 6 or 12. In this case, each time point corresponds to one hour, so these lags represent one to six (or one to twelve) hours prior to the current prediction time.

After extensive testing, the neural network configuration that achieved the best performance for both the MLP and AMLP models comprised three hidden layers with 10, 10, and 5 neurones. To identify the most effective configuration, multiple combinations of activation functions in the hidden layers were evaluated. The architecture supports non-linear activation functions, such as the hyperbolic tangent, ReLU, and sigmoid, as well as their combinations. The models with the highest predictive accuracy were those using either the hyperbolic tangent or linear activation functions across all hidden layers. However, due to the lower computational cost – around 48 s of training for the hyperbolic tangent versus approximately 17 s for the linear model – the linear model was ultimately chosen. Consequently, the MLP(10,10,5) and AMLP$_p$(10,10,5) where finally implemented. MAE was used as the loss function evaluated over 1500 epoch, the linear function as the activation function ψ, and the Adaptive Moment Estimation (Adam) algorithm as the optimiser, as it adaptively adjusts the learning rate individually for each parameter, resulting in faster and more efficient convergence in complex problems compared to other optimisers such as Root Mean Square Propagation (RMSprop) and Stochastic Gradient Descent (SGD) (Desai, 2020).

Source code in R was developed to implement the models. `ANN2` library was used for neural networks, whereas the VAR model was implemented using the `vars` package in R (Pfaff, 2008). Source code is available on request to the corresponding author.

4.1.3. Cross-validation framework

For each of the analysed hives, a three-month training series was used, containing a total of 2160 values per variable after preprocessing.

Table 2
Mean ± standard deviation of the MAE produced by the different neural network models when performing one-day-ahead predictions for the five variables analysed across the 13 beehives studied.

Hive	Model	Temp_m	Weight	Humidity	Temp	Pressure
0	$AMLP_{12}$	0.331 ± 0.248	0.522 ± 0.366	3.634 ± 2.130	0.332 ± 0.263	0.621 ± 0.369
	$AMLP_6$	0.326 ± 0.238	0.574 ± 0.442	3.450 ± 2.059	0.327 ± 0.254	0.664 ± 0.408
	MLP	0.318 ± 0.241	2.119 ± 0.820	5.882 ± 2.756	0.347 ± 0.263	0.560 ± 0.300
26	$AMLP_{12}$	1.291 ± 0.938	0.087 ± 0.053	3.793 ± 2.458	1.666 ± 1.040	2.796 ± 1.767
	$AMLP_6$	1.494 ± 0.943	0.093 ± 0.061	3.810 ± 2.471	2.119 ± 1.293	3.523 ± 2.157
	MLP	5.421 ± 1.536	0.427 ± 0.115	8.419 ± 3.624	3.609 ± 1.311	127.26 ± 13.06
48	$AMLP_{12}$	1.468 ± 1.062	0.496 ± 0.244	2.945 ± 2.095	1.655 ± 1.013	0.392 ± 0.284
	$AMLP_6$	1.629 ± 1.316	0.363 ± 0.203	3.331 ± 2.227	1.587 ± 1.118	0.397 ± 0.276
	MLP	5.988 ± 3.426	4.150 ± 0.338	6.829 ± 3.953	4.943 ± 2.938	0.240 ± 0.176
49	$AMLP_{12}$	2.811 ± 2.048	0.862 ± 0.468	1.486 ± 0.822	4.016 ± 2.590	161.61 ± 104.24
	$AMLP_6$	2.854 ± 2.056	0.653 ± 0.352	1.147 ± 0.781	4.251 ± 2.819	158.91 ± 96.121
	MLP	2.462 ± 1.819	28.115 ± 1.521	4.816 ± 0.900	2.729 ± 1.522	322.22 ± 143.30
60	$AMLP_{12}$	0.946 ± 0.580	0.474 ± 0.384	1.863 ± 1.060	0.258 ± 0.140	0.340 ± 0.216
	$AMLP_6$	1.118 ± 0.609	0.477 ± 0.379	1.862 ± 1.028	0.256 ± 0.141	0.328 ± 0.230
	MLP	2.208 ± 0.716	0.887 ± 0.468	1.524 ± 0.892	0.465 ± 0.331	0.254 ± 0.161
79	$AMLP_{12}$	0.466 ± 0.340	0.402 ± 0.237	2.394 ± 1.598	1.069 ± 0.791	0.562 ± 0.408
	$AMLP_6$	0.463 ± 0.343	0.451 ± 0.238	2.651 ± 1.754	1.150 ± 0.825	0.518 ± 0.402
	MLP	3.446 ± 1.696	18.223 ± 0.292	3.028 ± 2.229	3.304 ± 1.876	0.331 ± 0.300
84	$AMLP_{12}$	2.965 ± 1.334	0.293 ± 0.204	2.432 ± 1.427	2.841 ± 1.231	2.078 ± 1.005
	$AMLP_6$	2.933 ± 1.306	0.311 ± 0.191	2.956 ± 1.650	3.074 ± 1.280	2.686 ± 1.386
	MLP	6.270 ± 0.981	0.348 ± 0.237	2.143 ± 1.385	6.492 ± 0.879	1.495 ± 0.257
86	$AMLP_{12}$	0.880 ± 0.463	0.409 ± 0.281	1.901 ± 1.292	1.406 ± 0.768	0.308 ± 0.203
	$AMLP_6$	0.792 ± 0.438	0.363 ± 0.222	2.060 ± 1.347	1.307 ± 0.681	0.290 ± 0.203
	MLP	1.365 ± 0.905	4.288 ± 0.488	2.317 ± 1.417	3.364 ± 1.271	0.250 ± 0.159
96	$AMLP_{12}$	0.384 ± 0.301	0.675 ± 0.450	0.835 ± 0.555	0.314 ± 0.217	4.559 ± 2.666
	$AMLP_6$	0.389 ± 0.308	0.669 ± 0.421	0.768 ± 0.555	0.289 ± 0.201	4.328 ± 2.716
	MLP	0.250 ± 0.118	0.306 ± 0.426	0.954 ± 0.551	0.183 ± 0.095	5.242 ± 2.733
100	$AMLP_{12}$	0.649 ± 0.705	0.703 ± 0.337	2.786 ± 1.974	1.266 ± 0.791	1.412 ± 0.691
	$AMLP_6$	0.657 ± 0.700	0.535 ± 0.245	2.721 ± 1.917	1.159 ± 0.773	1.459 ± 0.780
	MLP	0.951 ± 0.794	5.033 ± 1.222	2.624 ± 2.118	0.925 ± 0.587	1.135 ± 0.669
111	$AMLP_{12}$	1.857 ± 1.498	0.727 ± 0.580	4.596 ± 2.885	2.354 ± 1.491	1.252 ± 0.933
	$AMLP_6$	2.200 ± 1.541	0.648 ± 0.433	4.140 ± 2.582	2.078 ± 1.331	2.405 ± 1.290
	MLP	2.741 ± 1.671	1.287 ± 0.943	5.423 ± 2.777	4.640 ± 2.475	0.379 ± 0.324
123	$AMLP_{12}$	0.678 ± 0.417	0.399 ± 0.318	1.707 ± 0.902	0.658 ± 0.402	3.588 ± 1.795
	$AMLP_6$	0.807 ± 0.432	0.407 ± 0.290	1.798 ± 0.915	0.703 ± 0.392	3.540 ± 1.719
	MLP	1.123 ± 0.663	2.563 ± 0.425	2.464 ± 0.744	1.329 ± 0.771	4.071 ± 1.494
135	$AMLP_{12}$	1.175 ± 0.680	0.413 ± 0.325	1.634 ± 1.068	0.983 ± 0.544	0.713 ± 0.510
	$AMLP_6$	1.191 ± 0.614	0.452 ± 0.332	1.887 ± 1.083	0.938 ± 0.483	0.681 ± 0.516
	MLP	1.574 ± 0.678	0.309 ± 0.291	2.788 ± 1.147	1.597 ± 0.747	0.594 ± 0.383

A 10-fold rolling window cross-validation was conducted, with observation increments to generate 24- and 72-step ahead predictions, which correspond to 1- and 3-day forecasts, respectively.

4.2. Experimental results

This subsection presents the predictive performance of the analysed models for the 5 internal variables across the 13 beehives. First, Table 2 reports the mean and standard deviation of the mean absolute errors obtained over 10 cross-validation iterations when applying the proposed autoregressive multilayer perceptron models of order 6 and 12, as well as the standard multilayer perceptron model.

Given the large number of variables and beehives analysed with the three models, interpreting the results requires careful consideration. To facilitate this, Tables 3 and 4 summarise the mean differences in MAE values between MLP and $AMLP_6$ (Table 3), and between MLP and $AMLP_{12}$ (Table 4). These tables also report the results of statistical significance testing, using the Student's t-test when the normality assumption could be assumed, and the Wilcoxon test otherwise. This analysis allows identification of both the direction and statistical significance of the differences. A positive value indicates that the MLP model exhibits a higher average error than the corresponding AMLP model, suggesting superior predictive accuracy of the latter. Negative values indicate better performance by the MLP model. Significance levels are denoted using the following asterisk code: * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001. For easier visual interpretation, statistically significant differences are highlighted in green for positive values (AMLP outperforms the MLP model) and in pink for negative values (MLP outperforms the AMLP model).

Overall, a total of 45 positive differences were observed, 71.11% of which were statistically significant. In contrast, only 20 negative differences were found, 55% of which were statistically significant.

The $AMLP_6$ model demonstrates a notably superior predictive performance compared to the MLP model for the primary variables (temperature, weight, and humidity) used to monitor hive health. Although its performance is weaker for internal pressure, where the MLP model often achieves slightly better results, the negative differences that were found to be statistically significant correspond to very small variations (less than $2.1hPa$, when the pressure fluctuates around $1000hPa$). Conversely, when $AMLP_6$ significantly outperforms MLP for this variable, the positive improvement exceeds $100hPa$, representing a substantial and meaningful difference. The relatively stable nature of pressure over time, with less pronounced temporal dependencies than the other variables, helps explain why the inclusion of time lags generally provides a smaller benefit for this variable. Improvements were observed when pressure exhibited more substantial oscillations, as observed in other beehives.

Table 3
Mean absolute error differences between MLP and $AMLP_6$ in one-day-ahead forecasts. Positive values denote superior performance of $AMLP_6$. Significance levels are indicated using the following asterisk code: * p-value < 0.05, ** p-value < 0.01, and *** p-value < 0.001. Statistically significant differences are colour-coded: green indicates positive outcomes ($AMLP_6$ better) and pink negative ones (MLP better).

Hive	Temp_m	Weight	Hum	Temp	Pressure
0	-0.009	1.546**	2.433*	0.020	-0.104
26	3.927***	0.333***	4.609**	1.490**	123.736***
48	4.359***	3.787**	3.498***	3.357***	-0.157**
49	-0.392	27.461***	3.669**	-1.522**	163.304**
60	1.070*	0.410**	-0.338	0.209**	-0.074**
79	2.984***	17.772**	0.377	2.153***	-0.188**
84	3.336**	0.036	-0.813	3.418**	-1.191**
86	0.573*	3.925***	0.258	2.058**	-0.040
96	-0.139**	-0.363*	0.186	-0.106**	0.914
100	0.294	4.498*	-0.096	-0.233**	-0.324
111	0.541	0.639	1.283	2.562**	-2.026**
123	0.316*	2.156**	0.666	0.626***	0.531
135	0.383	-0.143*	0.901**	0.659*	-0.087

Table 4
Mean absolute error differences between MLP and $AMLP_{12}$ in one-day-ahead forecasts. Positive values denote superior performance of $AMLP_{12}$. Significance levels are indicated using the following asterisk code: * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001. Statistically significant differences are colour-coded: green indicates positive outcomes ($AMLP_6$ better) and pink negative ones (MLP better).

Hive	Temp_m	Weight	Hum	Temp	Pressure
0	-0.013	1.598***	2.248*	0.015	-0.061
26	4.130***	0.340***	4.626**	1.943***	124.463**
48	4.520***	3.654**	3.883***	3.289***	-0.152*
49	-0.349	27.253***	3.330**	-1.288*	160.602**
60	1.262*	0.413*	-0.338	0.208**	-0.086
79	2.980***	17.820***	0.634	2.234***	-0.232**
84	3.305**	0.055	-0.289	3.651**	-0.583
86	0.484**	3.879***	0.416	1.958**	-0.058*
96	-0.134*	-0.369**	0.119	-0.131**	0.683
100	0.302	4.330*	-0.162	-0.340***	-0.278
111	0.885	0.560	0.827	2.286**	-0.873***
123	0.445**	2.165**	0.757	0.671***	0.484
135	0.399	-0.104	1.154**	0.614*	-0.119

Subsequently, Table 4 presents the differences in mean absolute error obtained by the MLP and $AMLP_{12}$ models for one-day-ahead predictions, using the same interpretative criteria as the previous table.

Similar results were obtained when comparing $AMLP_{12}$ and MLP, providing 45 positive and 20 negative differences; 71.11% of positive differences were significant, compared to 45% of negative ones.

Overall, the results presented in Tables 3 and 4 suggest that the AMLP models provide higher predictive accuracy for internal hive conditions than the MLP model for one-day-ahead forecasts, regardless of whether 6 or 12 lags are used. Consistent results are obtained when the prediction horizon is extended to three days.

As the models are evaluated across several variables with different measurement scales, a global error metric is required to provide a consistent measure of overall predictive accuracy. For this purpose, the corrected MAPEc, described in Section 3.2, is used. Fig. 5 shows the corrected mean absolute percentage errors averaged across all variables, associated with the $AMLP_{12}$, $AMLP_6$, and MLP models when making 1-day and 3-day forecasts.

The error area associated with the MLP model is substantially higher than that of the $AMLP_6$ and $AMLP_{12}$ models, which not differ substantially from each other. For one-day-ahead forecasts, the average MAPEc difference between the MLP and $AMLP_{12}$ models – computed over the 13 hives analysed – indicates a 6.72% reduction (Fig. 5, top left), confirming the superior performance of the $AMLP_{12}$ model. Similarly, when comparing the MLP and $AMLP_6$ (Fig. 5, bottom left), the averaged MAPEc is reduced by 6.74%, reporting a comparable but slightly greater improvement.

In the case of three-day-ahead predictions, the use of 12 lags results in a 5.31% reduction in averaged MAPEc (Fig. 5, top right), whereas the use of 6 lags leads to a 5.69% reduction (Fig. 5, bottom right). These outcomes further support the conclusion that the AMLP model yields more accurate predictions than the MLP model, regardless of whether 6 or 12 lags are employed, with the configuration using 6 lags exhibiting a slightly superior performance.

Overall, the AMLP models produce lower percentage errors than the MLP model, regardless of the hive under analysis and the prediction horizon. This supports the use of the AMLP model over the standard MLP model for multivariate time series forecasting.

To further assess the performance of the AMLP model, a comparison is conducted with the VAR model—an approach that has been extensively employed for time series forecasting. The analysis included configurations with both 6 and 12 lags. The MAPEc was also used here, as it effectively summarises model performance. Fig. 6 presents the comparison, where AMLP model results are displayed in purple and VAR model results in blue.

Visually, the area associated with the MAPEc across all hives is generally greater for the VAR model (violet) than for the AMLP model (blue), indicating a higher predictive accuracy of the AMLP model in most cases.

Averaging the MAPEc over the 13 hives for one-day forecasting shows that the AMLP model slightly outperforms the VAR model by 0.77% when 12 lags are included (Fig. 6, top left), and by 1.12% when using 6 lags (Fig. 6, bottom left). A similar difference between models is observed for three-day-ahead forecasts, with an improvement of 0.71% when using 12 lags (Fig. 6, top right) and 1.38% when using 6 lags (Fig. 6, bottom left).

These final results highlight that the AMLP model produces accurate and reliable predictions of internal hive conditions, outperforming both

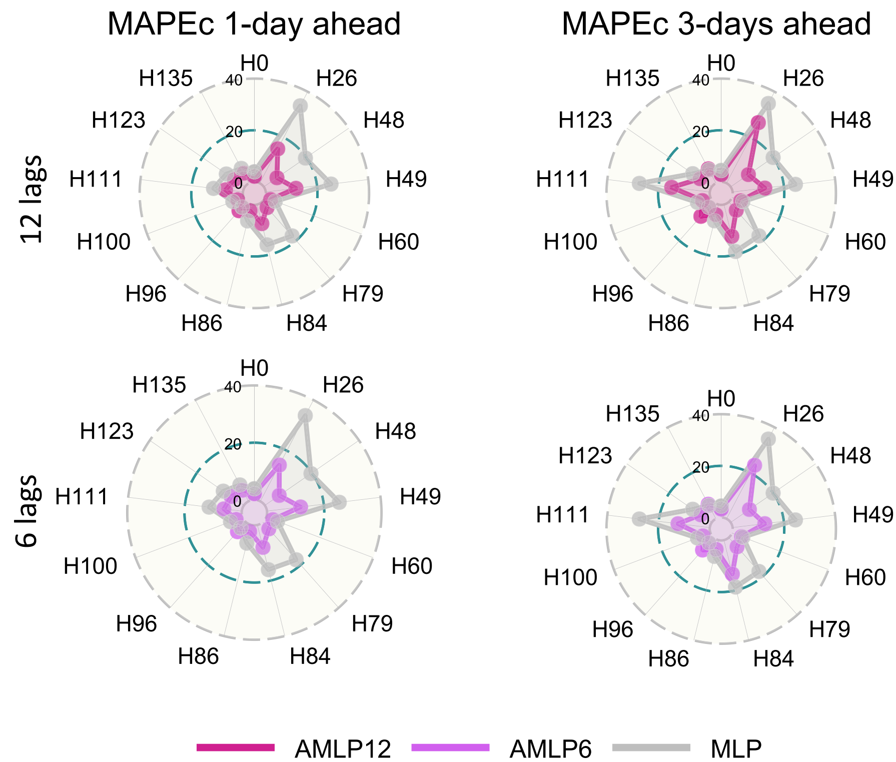


Fig. 5. The MAPEc area associated with the $AMLP_{12}$ (violet), $AMLP_6$ (pink), and MLP (grey) models for the 13 hives analysed, considering 1-day-ahead (left) and 3-day-ahead (right) predictions.

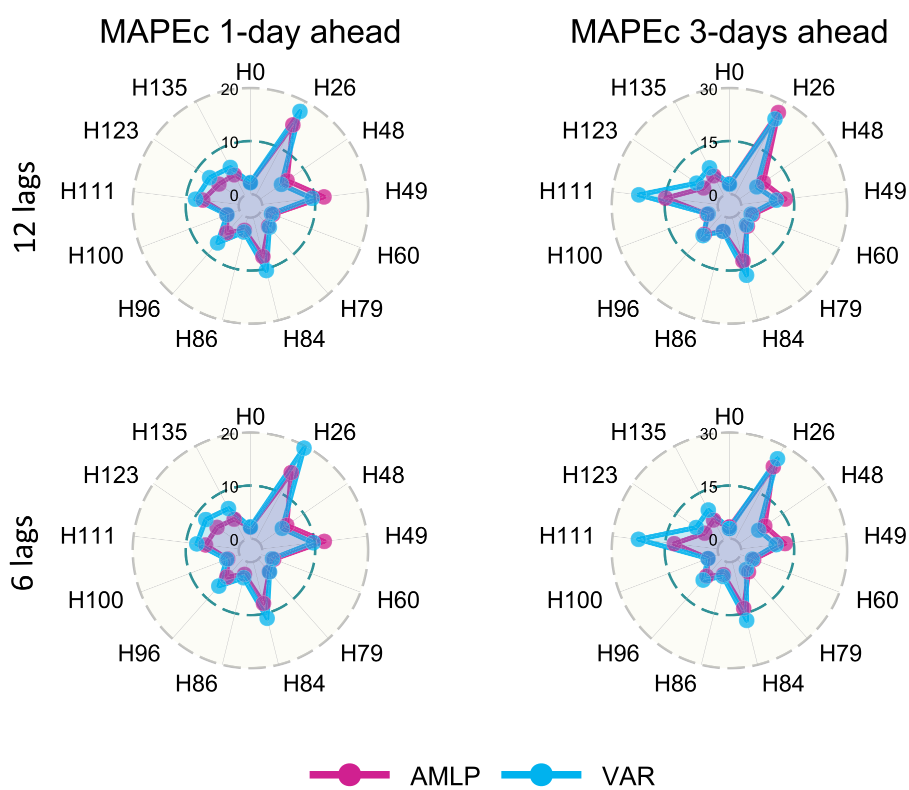


Fig. 6. The MAPEc area associated with the VAR (blue) and AMLP (violet) models for the 13 hives analysed, considering 1-day-ahead (left) and 3-day-ahead (right) predictions, with the different model including 12 lags (top) and 6 lags (bottom) respectively.

Table 5
Comparative overview of articles focused on predicting internal variables in beehives: Reference, number of hives analysed, name of the model, multi-response (MR), neural network approach (NN), temperature (T), humidity (H), weight (W), other response variables (O), weather covariates (WD), cross-validation (CV), forecast horizon (FH), and error metric (Error).

Article	Hives	Model(s)			Variable(s) response				WD	CV	FH	Error
		Name(s)	MR	NN	T	H	W	O			(days)	
Braga et al. (2021)	5	NNet	x	✓	✓	x	x	x	✓	x	1	RMSE
Robustillo et al. (2022)	3	VAR	✓	x	✓	✓	✓	x	✓	✓	1,3,7	MAE
Anwar et al. (2022)	8	Random Forest	x	x	x	x	✓	x	✓	✓	1	MAE
Rigakis et al. (2023)	1	Additive Regresion	x	x	x	✓	x	x	x	x	1	MAPE
Anwar et al. (2023)	8	Encoder	x	✓	x	x	✓	x	✓	✓	1	RMSE per frame
Robustillo et al. (2024)	4	VAR, VEC	✓	x	✓	✓	✓	x	✓	✓	1,3	MAE
Bono et al. (2024)	3	VAR	✓	x	✓	✓	✓	✓	✓	x	31	R^2
Kulyukin et al. (2024)	10	NNets, ARIMA	x	✓	✓	x	✓	✓	x	x	1,2	MSE
Ilyés-Vincze et al. (2025)	24	ARIMA, SARIMA	x	x	x	x	✓	x	✓	x	1	MAE, MAPE, RMSE
This approach	13	AMLP	✓	✓	✓	✓	✓	✓	✓	✓	1,3	MAE, MAPEc

the classic MLP model and the classical VAR time series model when the results are analysed collectively across all hives.

The models were also evaluated in terms of computational cost. The VAR(6) model was the fastest, requiring approximately 0.2 s to complete training with three months of data. The MLP and $AMLP_6$ models required around 13s and 17s, respectively. For the neural network models, 1500 epochs were used for training; however, the loss function converged prior to 500 epochs, indicating that total training time could be reduced if necessary. Regarding prediction times, the MLP model was the fastest (∼0.01 s), followed by $AMLP_6$ (∼0.25 s) and VAR (∼0.3 s). These training times are consistent with previous studies comparing VAR and MLP models (Di Mauro et al., 2024). Retraining with one additional day of data and 100 extra epochs required 0.01 s for the MLP model, 0.02 s for the AMLP model, and 0.2 s for the VAR model.

Overall, while VAR training is initially faster, the neural network models offer lower incremental retraining times, making them suitable for applications requiring frequent updates with minimal computational cost.

5. Discussion

A comparative discussion of previous studies on predicting internal hive variables is first presented to contextualise the current findings. This is followed by an examination of the main advantages and limitations of the study. Finally, the practical implications of the results are outlined, together with prospective avenues for future research.

5.1. Comparative overview of predictive beekeeping studies

Given the diversity of methodologies, variables, error metrics, number of hives analysed, and validation procedures used across studies, comparing results is inherently challenging. The use of different datasets further constrains direct comparison. To facilitate a structured discussion, a summary table has been developed to compile the key features considered in each work. Table 5 provides an overview of each reference, indicating the total number of hives analysed, the primary model(s) used, and whether the approaches are multi-response (MR) and/or based on neural network (NN) models. It also specifies the inclusion of the internal variables to be predicted – temperature (T), humidity (H), weight (W), or other response variables (O) – along with the inclusion of external meteorological covariates (WD). Furthermore, the table identifies whether cross-validation (CV) was applied to evaluate model robustness, the forecast horizon (FH) expressed in days, and the error metrics used to assess predictive performance (Error). The purpose of this table is simply to provide an overview of the key features of existing prediction studies and is not intended to establish any preference or rank studies, but rather to present them in chronological order.

As shown in Table 5, there is considerable variability in the number of hives analysed across studies. Most works rely on a limited number of hives, with one including a single hive (Rigakis et al., 2023). In contrast, Ilyés-Vincze et al. (2025) analysed 24 hives. In this context, the present study—based on 13 sensor-equipped hives largely exceeds the average sample size reported in the literature, providing a broader basis for evaluating model performance across different colonies and environmental conditions.

To contextualise the findings, it is essential to evaluate how the present study relates to other works addressing similar problems in the literature. Next, the variables and modelling strategies used in previous research on predicting internal hive conditions are compared, along with the performance metrics and cross-validation procedures employed to assess predictive capability and result generalisation. Finally, a comparative assessment of the results reported in the different studies is presented.

5.1.1. Variables and models

All existing studies aimed at forecasting internal hive conditions have focused on at least one of the key variables – weight, temperature, or humidity – highlighting their fundamental role in evaluating colony health. Furthermore, most of these studies incorporate climate-related external variables, which further reinforces the strong interdependence between environmental conditions and the internal dynamics of bee colonies. For instance, Braga et al. (2021) leveraged external weather inputs to predict in-hive temperature, illustrating how ambient climate data can enhance the accuracy of internal condition forecasts. Only Rigakis et al. (2023) and Kulyukin et al. (2024) completely omit external data, relying solely on internal sensor readings to drive their predictions.

A variety of models have been employed across studies, including classical time series models with both univariate and multivariate outputs, as well as neural network models, which, to date, have been applied exclusively in univariate forecasting settings (Braga et al., 2021; Anwar et al., 2023; Kulyukin et al., 2024). To the best of the authors' knowledge, this work represents the first documented application of a multi-output neural network model aimed at predicting the internal conditions of beehives. In general, classical time series models (e.g., ARIMA or VAR) have served as strong baselines for capturing linear temporal relationships, whereas neural networks approaches excel at modelling complex patterns in sensor data. The introduction here of a multi-response neural network architecture including lags leverages the correlations between different internal variables, potentially improving predictive power and efficiency by learning shared representations from their joint behaviour.

5.1.2. Evaluation and robustness

When considering how these models are assessed, it is important to note that several studies do not report the use of cross-validation procedures to assess the generalisation of their models (Braga et al., 2021;

Rigakis et al., 2023; Bono et al., 2024; Kulyukin et al., 2024; Ilyés-Vincze et al., 2025). This methodological limitation raises concerns, as the reported results may be strongly influenced by the particular data segment analysed, thus failing to provide a robust evaluation of model performance. As a result, the generalisability of some models remains questionable. Without techniques such as rolling-window or k-fold cross-validation, performance could be inflated by patterns specific to the training period. The remaining studies incorporate rigorous validation, increasing confidence that their reported accuracies are not driven by overfitting to a specific time period.

Most existing studies employ short-term prediction horizons, typically ranging from one to three days. An exception is the work by Bono et al. (2024), which considered a 31-day forecasting horizon. However, the primary objective of this study was not to assess predictive accuracy, but rather to explore associations between internal variables and external covariates. In fact, no error metrics were reported for the forecasts. The prevailing focus on short-term forecasts aligns with practical beekeeping needs, as accurate near-term predictions are crucial for timely interventions in hive management. Indeed, even studies that examined moderate extensions (such as a 7-day horizon in the work of Robustillo et al. (2022)) primarily emphasise the one-day-ahead results where model accuracy is the highest.

The heterogeneity of the error metrics used in the literature poses a major challenge to comparing results across studies. As illustrated in Table 5, the most frequently reported metrics are MAE and RMSE. Both quantify the discrepancy between observed and predicted values. However, MAE assigns equal weight to all deviations, offering a straightforward interpretation that does not require graphical support. In contrast, RMSE penalises large deviations more heavily, which can hinder interpretability in the absence of visual aids. Although MAE and RMSE are predominant in the literature, certain studies – such as (Ilyés-Vincze et al., 2025) – employ a broader set of metrics to offer a more general overview of predictive performance. One such metric is the MAPE. However, despite its intuitive appeal, MAPE is rarely reported, appearing only in Rigakis et al. (2023) and Ilyés-Vincze et al. (2025). This is primarily due to its sensitivity to zero values, which can render it undefined. This limitation has motivated the development and application of alternative metrics offering similar interpretability, such as the MAPEc used in this study. This type of metrics is especially advantageous when presenting results from a large number of hives, a task that entails considerable complexity and becomes even more challenging when multiple predictions are performed.

5.1.3. Results comparison

Although the use of distinct datasets precludes a direct comparison of results, a comparative assessment is carried out with studies that implemented cross-validation procedures, under the assumption that the reported results are more robust and generalisable than those that do not use cross-validation. First, the studies by Anwar et al. (2022, 2023) present a distinct approach to predictive modelling. Both works involve forecasting tasks, but their primary aim is to eliminate the need for physical scales in hive monitoring by estimating hive weight using data from alternative sensors. Compared to these works, which focused on single-step daily prediction, our study addresses a more complex and realistic forecasting task by predicting 24 hourly values per day. This difference must be considered when interpreting the results, as forecasting accuracy tends to degrade over extended horizons. Anwar et al. (2022) report a MAE of approximately 200 grams, whereas our model's performance ranges from 93 g to 650 g across different hives. Despite the increased prediction difficulty, $AMLP_6$ achieves a MAE below 100 g for hive number 26, demonstrating the potential of the model under stable conditions. In contrast, Anwar et al. (2023) used a Root Mean Square Error per frame, a highly specific and non-standard metric that is not generalisable and prevents direct comparison with other studies.

Robustillo et al. (2022) applied a VAR model for forecasting, with evaluations limited to five prediction points per day. In contrast, the present study extends this approach by considering 24 hourly prediction points, thereby substantially increasing the temporal resolution and the overall complexity of the forecasting task. Furthermore, this work expands the number of predicted variables from three to five, introducing a more demanding multivariate scenario. Despite these additional challenges, the proposed $AMLP_6$ model achieves superior predictive accuracy compared to the VAR model. Later, Robustillo et al. (2024) considered only five data points per day across four different hives. A straightforward way to establish a comparison is to compute the average MAE for each variable common to both studies (weight, temperature, and humidity) and compare the resulting values. For temperature, the VEC model proposed by Robustillo et al. (2024) achieves an average one-day MAE of $1.227°C$, whereas the $AMLP_6$ model reports an average MAE of $1.388°C$—both within a similar range. Regarding hive weight, the VEC model yields an average MAE of 120 g across the four hives analysed, while the $AMLP_6$ model obtains a higher average MAE of 461 g. This increase can likely be attributed to the larger number of predicted values per day in the present work, given that the average prediction error for hive weight in the present study increases to 212 g when forecasting over 15 time steps (3 days). In terms of humidity, the VEC model reports an average error of 4.86%, whereas the $AMLP_6$ model reduces this to 2.51%, despite forecasting over a longer horizon and across a greater number of hives. These results suggest that, although the $AMLP_6$ model faces a more challenging prediction scenario, it maintains competitive accuracy, particularly in the estimation of relative humidity.

5.2. Advantages and limitations

The study was conducted using data collected from a variety of hive types employed by the participating beekeepers, including Zander, Segeberger, Hohenheim, and other designs, made either from wood (8 hives) or plastic (5 hives). As this was a citizen science project, the sensor system was distributed independently of hive type and accompanied by a general installation diagram. Beekeepers were encouraged to adapt this diagram to the specific characteristics of their hives, which may have resulted in undocumented variations in sensor placement or configuration. In addition, because sensor operation and maintenance were managed by the beekeepers themselves, a degree of control over data quality was inevitably lost, potentially increasing the proportion of missing data due to handling, maintenance issues, or connectivity disruptions. This heterogeneity introduces additional variability into the data and increases modelling complexity; however, it also ensures that the proposed system is applicable across a wide range of hive designs, materials, and management practices, thereby enhancing its robustness and real-world relevance.

The imputation process involved performing a single global imputation on the entire dataset prior to training and validation. This approach allowed the model to leverage all available data and produce a data set that closely reflects real-world conditions. However, this approach risks introducing optimistic bias, since future data may inform missing values in earlier records. In contrast, performing imputation individually for each train–test subset (by fold) would avoid this potential bias, but reduce the amount of information available for imputation, likely resulting in less accurate and less realistic imputations. Thus, while the global imputation may slightly overestimate predictive performance in some folds, it exposes the model to realistic conditions overall.

Beyond performance metrics, this neural approach offers substantial operational advantages in precision apiculture, particularly in scenarios where models must be updated frequently as new data become available. Although the AMLP and MLP models initially require more computational resources than the VAR model, the cost remains low enough to be easily handled by any system, and their capacity for efficient incremental retraining reduces the impact of this difference over time. In contrast, the VAR model's training cost increases as the historical dataset grows, because it must reprocess the entire historical dataset. These characteristics suggest that the AMLP and MLP models are more suitable for near real-time decision-support systems in precision beekeeping, where daily updates would incur negligible computational costs. Therefore, an initial training time of about 15–25 s in these experiments is perfectly acceptable given the minimal time required for subsequent retraining. This capability allows continuous learning from high-frequency multi-sensor data streams, ensuring that forecasts remain up-to-date as hive and environmental conditions evolve. Such adaptive retraining – whether performed through real-time online updates or rapid batch learning cycles – allows the AMLP to incorporate new information while preserving prior knowledge, effectively mitigating data drift in non-stationary environments. This behaviour aligns with findings in IoT research, where incremental learning approaches have demonstrated the ability to update models without complete retraining (Waseem et al., 2025). Such flexibility supports adaptive decision-making, allowing the system to quickly adjust to emerging patterns and provide timely alerts for intervention, thereby promoting proactive and resilient management of colony health.

While the AMLP has shown strong predictive capabilities and supports incremental retraining, the current implementation does not inherently manage missing data or irregular sampling intervals. Developing models capable of learning directly from incomplete time series – without the need for external imputation procedures – would improve robustness and reduce preprocessing overhead.

5.3. Practical implications and future research

The improved forecasting accuracy directly supports decision-making in hive management. By anticipating internal temperature, humidity, and weight dynamics, the model allows early detection of anomalies such as brood interruption or resource depletion. The approach thus contributes to proactive and sustainable hive management, facilitating interventions based on predictive rather than reactive strategies.

Future research should focus on ensuring reliable and continuous data collection to minimise outliers, sensor failures, and connectivity-related data loss, thereby contributing to the development of more robust databases. Beyond these data-quality aspects, a key line of future work is the extension of the proposed framework to larger-scale studies involving a higher number of hives and longer monitoring periods, which would allow for a more rigorous assessment of model robustness and generalisability. This scaling-up will require the development of concise strategies to summarise, aggregate, and visualise multivariate, multi-output results across many colonies, while simultaneously enabling a systematic evaluation of input-variable relevance and redundancy.

6. Conclusion

In recent years, the field of precision beekeeping has experienced notable advances. However, further research is required to transition from systems focused primarily on monitoring to comprehensive decision-support systems. To date, most available solutions provide real-time monitoring of hive conditions. Although considerable progress has been made in detecting events detrimental to colony health, these are typically identified after they have occurred, when effective intervention is no longer possible. Future efforts should therefore aim to integrate real-time monitoring with predictive capabilities, enabling the early detection of potentially harmful conditions. This would allow beekeepers to take preventive actions and avoid significant economic losses. Consequently, the development of robust predictive models that can accurately estimate the internal dynamics of beehives represents a key step towards enhancing early intervention strategies and improving overall hive management.

The proposed approach advances in this direction, achieving accurate predictions while maintaining a low computational cost for network training and forecast generation. The consistent improvement of the proposed AMLP model over the standard MLP underscores the methodological relevance of incorporating time delays into neural network models for time-series analysis, leading to superior predictive accuracy and robustness across varying conditions. Moreover, the incremental retraining capability of the AMLP model represents a key advantage in the Big Data context, where large volumes of sensor data are continuously generated. This feature allows the model to remain up to date without requiring full retraining, thereby supporting scalable, efficient, and adaptive management strategies in precision-beekeeping applications.

CRediT authorship contribution statement

M. Carmen Robustillo: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Diren Senger:** Writing – review & editing, Writing – original draft, Resources, Data curation. **M. Isabel Parra:** Writing – review & editing, Visualization, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Carlos J. Pérez:** Writing – review & editing, Validation, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Acknowledgements

This research was funded by Ministry of Science, Innovation and Universities - Spain, and State Research Agency - Spain (Projects PID2021-122209OB-C32 and PID2024-155179NB-C21), funded by MICIU/AEI/10.13039/501100011033 and European Union (European Regional Development Fund). Funding for open access charge: Universidad de Málaga/CBUA.

Data availability

We are not the owners of the data, but it is available online (cited in the article) and the code will be available in a public repository.

References

Alleri, M., Amoroso, S., Catania, P., Lo Verde, G., Orlando, S., Ragusa, E., Sinacori, M., Vallone, M., Vella, A., 2023. Recent developments on precision beekeeping: A systematic literature review. J. Agric. Food Res. 14, 100726. http://dx.doi.org/10.1016/j.jafr.2023.100726.

Andrade, C.H.T.d., Melo, G.C.G.d., Vieira, T.F., Araújo, c.B.Q.d., Medeiros Martins, A.d., Torres, I.C., Brito, D.B., Santos, A.K.X., 2023. How does neural network model capacity affect photovoltaic power prediction? a study case. Sensors 23 (3), 1357. http://dx.doi.org/10.3390/s23031357.

Anwar, O., Keating, A., Cardell-Oliver, R., Datta, A., Putrino, G., 2022. Design and development of low-power, long-range data acquisition system for beehives - BeeDas. Comput. Electron. Agric. 201, 107281. http://dx.doi.org/10.1016/j.compag.2022.107281.

Anwar, O., Keating, A., Cardell-Oliver, R., Datta, A., Putrino, G., 2023. Apis-prime: A deep learning model to optimize beehive monitoring system for the task of daily weight estimation. Appl. Soft Comput. 144, 110546. http://dx.doi.org/10.1016/j.asoc.2023.110546.

Apicella, A., Donnarumma, F., Isgrò, F., Prevete, R., 2021. A survey on modern trainable activation functions. Neural Netw. 138, 14–32. http://dx.doi.org/10.1016/j.neunet.2021.01.026.

Bono, F., Vallone, M., Alleri, M., Lo Verde, G., Orlando, S., Ragusa, E., Catania, P., 2024. Hive behaviour assessment through vector autoregressive model by a smart apiculture system in the Mediterranean area. Smart Agric. Technol. 9, 100676. http://dx.doi.org/10.1016/j.atech.2024.100676.

Braga, A.R., Freitas, B.M., Gomes, D.G., Bezerra, A.D., Cazier, J.A., 2021. Forecasting sudden drops of temperature in pre-overwintering honeybee colonies. Biosyst. Eng. 209, 315–321. http://dx.doi.org/10.1016/j.biosystemseng.2021.07.009.

Colin, T., Dakos, V., Barron, A.B., Meikle, W.G., Altmann, E.G., Latty, T., 2025. Temperature as an early warning signal of honeybee colony failure. Ecol. Inform. 92, 103445. http://dx.doi.org/10.1016/j.ecoinf.2025.103445.

Danieli, P.P., Addeo, N.F., Lazzari, F., Manganello, F., Bovera, F., 2024. Precision beekeeping systems: State of the art, pros and cons, and their application as tools for advancing the beekeeping sector. Animals 14, 70. http://dx.doi.org/10.3390/ani14010070.

Degenfellner, J., Templ, M., 2024. Modeling bee hive dynamics: Assessing colony health using hive weight and environmental parameters. Comput. Electron. Agric. 218, 108742. http://dx.doi.org/10.1016/j.compag.2024.108742.

Desai, C., 2020. Comparative analysis of optimizers in deep neural networks. Int. J. Innov. Sci. Res. Technol. 5, 959–962, URL: https://www.researchgate.net/publication/345381779_Comparative_Analysis_of_Optimizers_in_Deep_Neural_Networks.

Di Mauro, M., Galatro, G., Postiglione, F., Song, W., Liotta, A., 2024. Multivariate time series characterization and forecasting of voip traffic in real mobile networks. IEEE Trans. Netw. Serv. Manag. 21, 851–865. http://dx.doi.org/10.1109/TNSM.2023.3295748.

Garcia Cabello, J., 2022. Mathematical neural networks. Axioms 11, 80. http://dx.doi.org/10.3390/axioms11020080.

Ghezal, A., König, A., 2025. A comparative study of hybrid machine-learning vs. Deep-learning approaches for varroa mite detection and counting. Sensors 25, 5075. http://dx.doi.org/10.3390/s25165075.

Ilyés-Vincze, C., Leelőssy, A., Mészáros, R., 2025. ARIMAX modeling of hive weight dynamics using meteorological factors during robinia pseudoacacia blooming. Atmosphere 16, 918. http://dx.doi.org/10.3390/atmos16080918.

Jing, X., Luo, J., Wang, J., Zuo, G., Wei, N., 2022. A multi-imputation method to deal with hydro-meteorological missing values by integrating chain equations and random forest. Water Resour. Manag. 36, 1–15. http://dx.doi.org/10.1007/s11269-021-03037-5.

Johannsen, C., Senger, D., Kluss, T., 2020. A DIY sensor kit, Gaussian processes and a multi-agent system fused into a smart beekeeping assistant. In: 2020 16th International Conference on Intelligent Environments. IE, pp. 92–99. http://dx.doi.org/10.1109/IE49459.2020.9154974.

Kotovs, D., Krievina, A., Zacepins, A., 2025. Enhancing precision beekeeping by the macro-level environmental analysis of crowdsourced spatial data. ISPRS Int. J. Geo-Inf. 14, 47. http://dx.doi.org/10.3390/ijgi14020047.

Kulyukin, V.A., Coster, D., Kulyukin, A.V., Meikle, W., Weiss, M., 2024. Discrete time series forecasting of hive weight, in-hive temperature, and hive entrance traffic in non-invasive monitoring of managed honey bee colonies: Part I. Sensors 24, 6433. http://dx.doi.org/10.3390/s24196433.

Li, J., Bortnik, J., Wang, Q., Wu, Y., Lizarraga, A., Angel, M., Wang, B., Wen, Q., Jiang, J., 2025. Modeling ring current proton distribution using mlp, cnn, lstm, and transformer networks. Front. Astron. Space Sci. 12, http://dx.doi.org/10.3389/fspas.2025.1629056.

Liu, Y.L., Lin, T.T., 2025. Pattern analysis of daily beehive weight variation for colony health assessment. Comput. Electron. Agric. 239, 111016. http://dx.doi.org/10.1016/j.compag.2025.111016.

Nath, R., Singh, H., Mukherjee, S., 2023. Insect pollinators decline: an emerging concern of anthropocene epoch. J. Apicult. Res. 62, 23–38. http://dx.doi.org/10.1080/00218839.2022.2088931.

Otesbelgue, A., de Lima Rodrigues, Í., dos Santos, C.F., Gomes, D.G., Blochtein, B., 2025. The missing queen: a non-invasive method to identify queenless stingless bee hives. Apidologie 56, 28. http://dx.doi.org/10.1007/s13592-025-01148-1.

Pfaff, B., 2008. VAR, SVAR and SVEC models: Implementation within R package vars. J. Stat. Softw. 27, 1–32. http://dx.doi.org/10.18637/jss.v027.i04.

Richardson, K., et al., 2023. Beekeeping role in enhancing food security and environmental public health. Health Econ. Manag. Rev. 4, 69–79. http://dx.doi.org/10.61093/hem.2023.4-06.

Rigakis, I., Potamitis, I., Tatlas, N.A., Psirofonia, G., Tzagaraki, E., Alissandrakis, E., 2023. A low-cost, low-power, multisensory device and multivariable time series prediction for beehive health monitoring. Sensors 23, 1407. http://dx.doi.org/10.3390/s23031407.

Robustillo, M.C., Naranjo, L., Parra, M.I., Pérez, C.J., 2024. Addressing multidimensional highly correlated data for forecasting in precision beekeeping. Comput. Electron. Agric. 226, 109390. http://dx.doi.org/10.1016/j.compag.2024.109390.

Robustillo, M.C., Pérez, C.J., Parra, M.I., 2022. Predicting internal conditions of beehives using precision beekeeping. Biosyst. Eng. 221, 19–29. http://dx.doi.org/10.1016/j.biosystemseng.2022.06.006.

Šabić, J., Perković, T., Šolić, P., Šerić, L., 2025. Buzzing with intelligence: A systematic review of smart beehive technologies. Sensors 25, 5359. http://dx.doi.org/10.3390/s25175359.

Sad, C., Kampelopoulos, D., Sofianidis, I., Kanelis, D., Nikolaidis, S., Tananaki, C., Siozios, K., 2025. Deep edge IoT for acoustic detection of queenless beehives. Electronics 14, 2959. http://dx.doi.org/10.3390/electronics14152959.

Sari, F., Sarvia, F., 2025. Lavender field detection via remote sensing and machine learning for optimal hive placement to maximize lavender honey production. Earth 6, 107. http://dx.doi.org/10.3390/earth6030107.

Sattler, C., Barron, A.B., Colin, T., 2025. Beekeeping in the digital age: prospects and pitfalls of hive sensors in commercial beekeeping. J. Apicult. Res. 1–11. http://dx.doi.org/10.1080/00218839.2025.2552530.

Scutaru, D., Bergonzoli, S., Costa, C., Violino, S., Costa, C., Albertazzi, S., Capano, V., Kostić, M.M., Scarfone, A., 2025. An AI-based digital scanner for varroa destructor detection in beekeeping. Insects 16, 75. http://dx.doi.org/10.3390/insects16010075.

Senger, D., Gruber, C., Kluss, T., Johannsen, C., 2023. Weight, temperature and humidity sensor data of honey bee colonies in Germany, 2019 – 2022 [data set]. Zenodo http://dx.doi.org/10.5281/zenodo.10407693.

Senger, D., Gruber, C., Kluss, T., Johannsen, C., 2024. Weight, temperature and humidity sensor data of honey bee colonies in Germany, 2019–2022. Data Brief 52, 110015. http://dx.doi.org/10.1016/j.dib.2023.110015.

Stojanova, S., Superina, A., Verhovnik, J., Duh, E.S., 2025. Leveraging IoT technologies for smart hive management and sustainable beekeeping. Elektroteh. Vestn. 92, 68–72. http://dx.doi.org/10.5281/zenodo.15223058.

Urban, M., Chlebo, R., 2024. Current status and future outlooks of precision beekeeping systems and services. Rev. Agric. Sci. 12, 165–181. http://dx.doi.org/10.7831/ras.12.0_165.

Vardakas, P., Mainardi, G., Minaud, E., Patalano, S., Rebaudo, F., Requier, F., Steffan-Dewenter, I., Hatjina, F., 2025. Unveiling beekeepers' use and preference of precision apiculture systems. J. Apicult. Res. 64, 647–656. http://dx.doi.org/10.1080/00218839.2024.2355813.

Verbeke, W., Diallo, M.A., van Dooremalen, C., Schoonman, M., Williams, J.H., Van Espen, M., D'Haese, M., de Graaf, D.C., 2024. European beekeepers' interest in digital monitoring technology adoption for improved beehive management. Comput. Electron. Agric. 227, 109556. http://dx.doi.org/10.1016/j.compag.2024.109556.

Waseem, Q., Wan Din, W.I.S., Aamir, M., 2025. A generalized three-tier hybrid model for classifying unseen (IoT devices) in smart home environments. Sci. Rep. 15, 35388. http://dx.doi.org/10.1038/s41598-025-19303-0.

Wilson, S., 2021. Miceranger: Multiple imputation by chained equations with random forests. URL: https://CRAN.R-project.org/package=miceRanger. r package version 1.5.0.

Zattara, E.E., Aizen, M.A., 2021. Worldwide occurrence records suggest a global decline in bee species richness. One Earth 4, 114–123. http://dx.doi.org/10.1016/j.oneear.2020.12.005.

Zivot, E., Wang, J., 2006. Rolling Analysis of Time Series. Springer New York, New York, NY, pp. 313–360. http://dx.doi.org/10.1007/978-0-387-21763-5_9, (Chapter 9).