Smart Agricultural Technology 9 (2024) 100676

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

# Hive behaviour assessment through vector autoregressive model by a smart apiculture system in the Mediterranean area

Filippa Bono [a], Mariangela Vallone [b,*], Maria Alleri [b], Gabriella Lo Verde [b], Santo Orlando [b], Ernesto Ragusa [b], Pietro Catania [b]

[a] University of Palermo, Department of Economic Business and Statistical Sciences, Viale delle Scienze ed. 13, 90128 Palermo, Italy
[b] University of Palermo, Department of Agricultural, Food and Forest Sciences (SAAF), Viale delle Scienze ed. 4, 90128 Palermo, Italy

ARTICLE INFO

*Keywords:*
Apiary
Precision beekeeping
Smart technologies
Vector autoregressive model

ABSTRACT

Precision beekeeping is defined as an apiary management strategy based on monitoring individual bee colonies to minimize resource consumption and maximize bee productivity. This subject has met with a growing interest from researchers in recent years because of its environmental implications. Today, the use of new monitoring technologies and management systems are facilitating the beekeeper's task by reducing operating costs and increasing animal welfare. Few studies in the literature apply forecasting models that could be useful as decision support to help beekeepers effectively monitor their hives. The Vector Autoregressive Regression (VAR) models are widely used in economics, but little applications have been performed in precision beekeeping data. The aim of this study was to apply a Vector Autoregressive Model to study the interrelations among internal factors (weight, internal temperature, internal relative humidity, sound pressure level) and between internal and external environmental parameters (external temperature and relative humidity, rain, wind speed, UV index) of some hives located in three different sites in Sicily (south Italy), monitored by a proper designed smart system. Time series were studied over the period April - August 2023. The significance recorded in the relationships between weight of the hive and its internal temperature and weight of the hive and its internal relative humidity, and the good predictive capacity of the models with respect to internal temperature and internal relative humidity, allowed to build a predictive model to understand when possibly intervene on the hives. Effect and duration of a system shock on the variables of interest were effectively monitored by the impulse response function in order to understand the level of the system response.

## 1. Introduction

Technological innovations and research allowed significant improvements in terms of both productivity and intra-company management in numerous agricultural sectors. For this purpose, a key role is played by Precision Agriculture (PA).

Beekeeping has also benefited from the introduction of Precision Beekeeping (PB) technologies [1], which focuses on the apiary management strategy through the individual monitoring of bee colonies using smart hives. PB sector is expanding to minimize resources and maximize bee productivity through smart hives [2–7]. Indeed, the use of new monitoring technologies and management systems gradually tested and introduced are facilitating the beekeeper's task by reducing operating costs and increasing animal welfare [8].

Beekeepers are often forced to move their hives between fields to provide pollination services. This hive movement is very stressful for the bees and can negatively affect colony strength, i.e., the number of bees in the hive [3]. Although there are sufficient technical means and industrial products for the practical execution of PB, the process is slow due to the differing states of development of three implementation phases: data collection, data analysis and application [9].

Furthermore, thanks to the use of technologies and the application of statistical methods, beekeepers can monitor the hive remotely without openings and, above all, without disturbing the colony.

Several studies apply this technology, but most of them are limited to a descriptive analysis of the recorded information, used as a tool for the beekeeper to know the state of the hive at a specific time [10]. Research has focused more on the acquisition of sensor data and its real-time analysis than on the application of robust statistical methods for predictive purposes [9]. Data collected from the hives are mainly used for various statistical applications and essentially for the evaluation of

* Corresponding author.
*E-mail address:* mariangela.vallone@unipa.it (M. Vallone).

https://doi.org/10.1016/j.atech.2024.100676
Received 23 July 2024; Received in revised form 23 October 2024; Accepted 28 November 2024
Available online 29 November 2024

**Nomenclature**

| | |
|---|---|
| Tint | Internal temperature |
| RHint | Internal relative humidity |
| SPL | Sound pressure level |
| Text | External temperature |
| RHext | External relative humidity |
| WS | Wind speed |
| UVI | UV index |
| p50 | Median |
| iqr | Interquartile range |
| sd | Standard deviation |
| min | Minimum value |
| max | Maximum value |
| cv | Coefficient of variation |
| $R^2$ | Coefficient of determination |

colony is facing a problem. The results showed that the proposed model could predict temperature values 24 h in advance with a Root Mean Squared Error (RMSE) of only 0.5 %.

The Vector Autoregressive Regression (VAR) model, widely used in economics [13], finds little application in the beekeeping sector research. One of the prerequisites in the application of VAR is to test the relationships between the variables using the sims causality test (1986). In short, the Granger causality test [14] demonstrates whether there is a relationship between the variables of the model [15]. Ziegler et al. (2022) [16] investigated whether the VAR model can help understand the interrelation between climate variables and the weight of *Apis mellifera* hives. The authors demonstrate that it is possible to apply econometric statistical models to apiary data by relating them to climate data and thus contribute significantly to applied statistics in beekeeping. Interesting is the research by Robustillo et al. (2022) [8] who compare various predictive models, namely, vector autoregressive models, VAR model with dynamic coefficients or time-varying VAR model (tvVAR), this being a modification which assumes that the coefficients involved in the response generating process are dynamic, Dynamic Linear Model

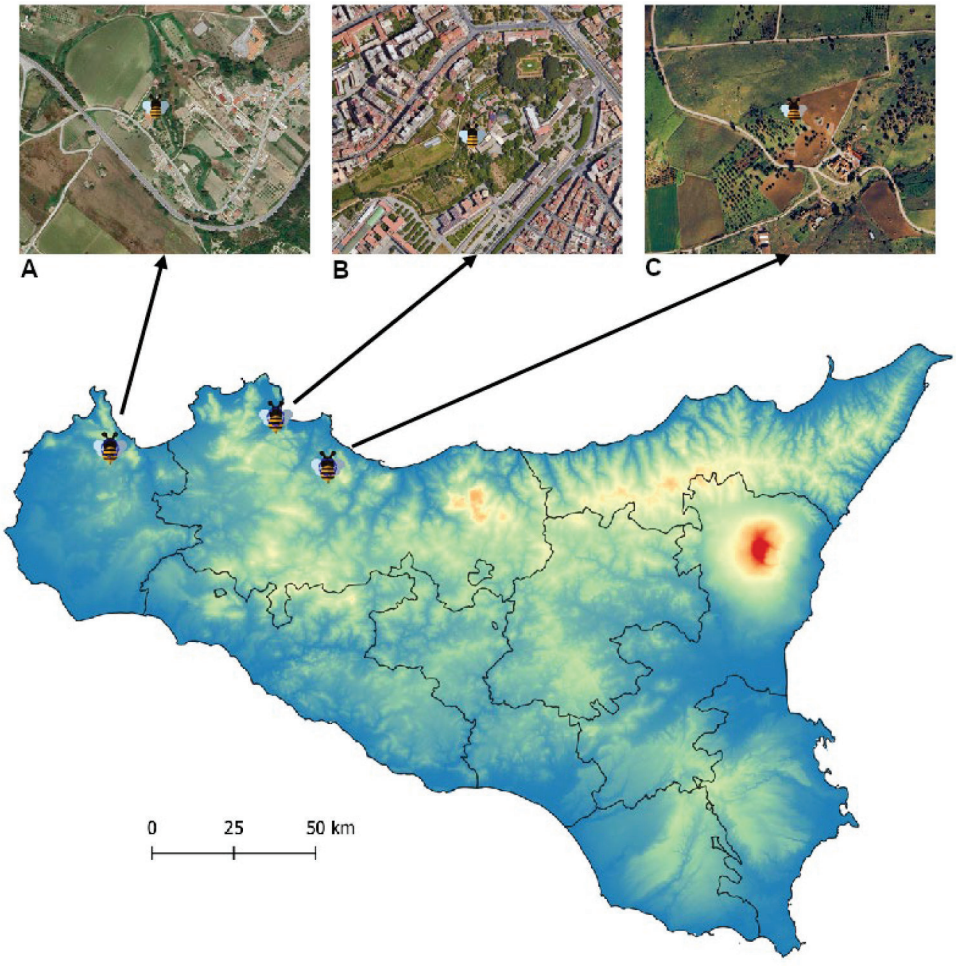


**Fig. 1.** Study areas with indication of the three sites named A (38.0240°N, 12.7855°E), B (38.1076°N, 13.3506°E) and C (37.9736°N, 13.5276°E).

variables that interfere with the swarm behavior [11].

Few studies in the literature apply forecasting models that could be useful as decision support to help beekeepers effectively monitor their hives. Braga et al. (2021) [12] predict sudden drops in temperature inside the hive by using a Long Short-Term Memory (LSTM) algorithm to predict the internal temperature of the hives. This parameter is vital for the health of bees; a decrease in temperature may indicate that the

(DLM) and Generalised Additive Model (GAM). The DLM is a particular case of dynamic regression models, which allows the regression coefficients to vary over time. Instead, GAM is a variation of generalised linear models in which the response variable is given by a sum of smooth functions of at least some (and possibly all) covariates. To compare the different models, the authors used data from three sensors installed in three hives, to predict some internal parameters of the hive such as

**Fig. 2.** Dadant Blatt type hives during positioning in site C.

temperature, relative humidity and weight. The tvVAR and VAR models provided accurate predictions of large weight losses, significant temperature drops, or important changes in relative humidity that could endanger the hive health [8].

The aim of this study is to monitor hives located in three different and distant apiaries in Sicily (south Italy), in order to investigate the interrelations among internal factors and between internal and external environmental parameters using a proper designed smart system, applying a Vector Autoregressive Model. The ability of the apiary to rebalance itself after the variation of the internal parameters of the system is also assessed.

## 2. Materials and methods

This section describes the characteristics of the experimental sites under study, the remote smart monitoring system developed by the authors, the statistical methods applied.

### 2.1. Study sites and apiary description

The study was carried out in Sicily (Italy) by identifying three different study areas (Fig. 1) characterized by semi-arid climate with mild winters and long, dry summers.

- Site A, located at the Stabile farm in the province of Trapani (municipality of Castellamare del Golfo), at an altitude of 197 m a.s.l.; it

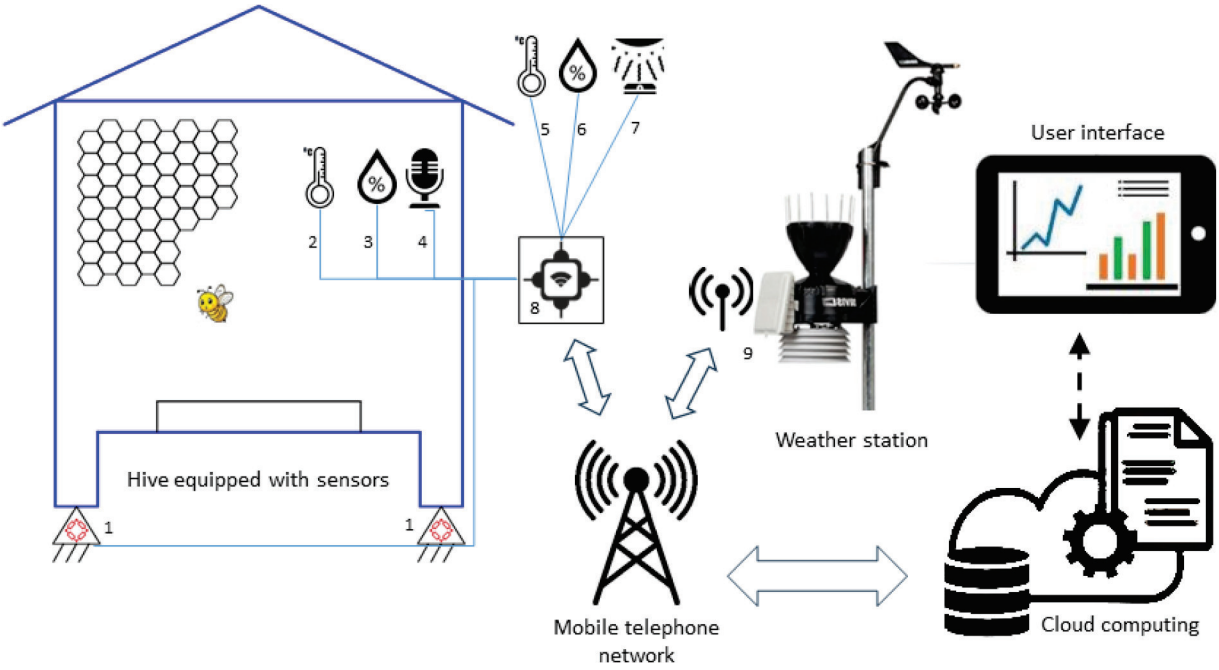


**Fig. 3.** Scheme of the system. 1 load cell; 2 internal thermometer, 3 internal hygrometer, 4 microphone, 5 external thermometer, 6 external hygrometer, 7 luminosity sensor, 8 microcontroller, 9 weather station.

**Table 1**
Site A. Descriptive statistics of internal and external hive factors before preprocessing.

| Hive 1-A | | | | | Hive 2-A | | | | Hive 3-A | | | | External factors | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| stats | Weight [kg] | Tint [°C] | RHint [%] | SPL [dB] | Weight [kg] | Tint [°C] | RHint [%] | SPL [dB] | Weight [kg] | Tint [°C] | RHint [%] | SPL [dB] | Text [°C] | RHext [%] | Rain [mm] | WS [m/s] | UVI |
| missing | 0 | 0 | 0 | 0 | 60 | 60 | 60 | 60 | 139 | 140 | 140 | 500 | 0 | 0 | 0 | 0 | 0 |
| N | 1836 | 1836 | 1836 | 1836 | 1776 | 1776 | 1776 | 1776 | 1697 | 1696 | 1696 | 1336 | 1836 | 1836 | 1836 | 1836 | 1836 |
| mean | 52.6 | 33.0 | 83.9 | 41.1 | 54.0 | 31.4 | 77.9 | 35.5 | 53.7 | 32.9 | 67.3 | 38.0 | 21.5 | 67.3 | 0.1 | 0.5 | 1.75 |
| p50 | 55.8 | 34.8 | 86.7 | 43.6 | 62.5 | 35.1 | 78.7 | 35.3 | 60.1 | 35.0 | 65.6 | 37.0 | 21.3 | 70.0 | 0.0 | 0.5 | 0.00 |
| iqr | 21.7 | 0.9 | 11.8 | 15.2 | 24.8 | 6.7 | 13.9 | 16.8 | 26.3 | 0.8 | 10.7 | 15.1 | 9.3 | 26.0 | 0.0 | 0.6 | 3.00 |
| sd | 11.7 | 4.4 | 10.3 | 9.9 | 12.1 | 6.8 | 11.9 | 10.7 | 13.9 | 5.1 | 7.2 | 9.3 | 6.4 | 18.3 | 0.8 | 0.4 | 2.45 |
| min | 19.5 | 12.4 | 50.5 | 21.6 | 34.7 | 7.1 | 37.1 | 12.0 | 20.6 | 10.1 | 44.5 | 10.7 | 5.0 | 11.0 | 0.0 | 0.0 | 0.00 |
| max | 70.4 | 36.5 | 100.0 | 61.7 | 67.3 | 37.2 | 100.0 | 64.0 | 74.6 | 37.3 | 97.3 | 60.4 | 41.9 | 97.0 | 28.7 | 2.7 | 8.00 |
| cv | 0.2 | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 | 0.2 | 0.3 | 0.3 | 0.2 | 0.1 | 0.2 | 0.3 | 0.3 | 12.7 | 0.8 | 1.40 |

**Table 2**
Site B. Descriptive statistics of internal and external hive factors before preprocessing.

| Hive 1-B | | | | | Hive 2-B | | | | Hive 3-B | | | | External factors | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| stats | Weight [kg] | Tint [°C] | RHint [%] | SPL [dB] | Weight [kg] | Tint [°C] | RHint [%] | SPL [dB] | Weight [kg] | Tint [°C] | RHint [%] | SPL [dB] | Text [°C] | RHext [%] | Rain [mm] | WS [m/s] | UVI |
| missing | 0 | 0 | 0 | 360 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 372 |
| N | 1836 | 1836 | 1836 | 1476 | 1836 | 1836 | 1836 | 1836 | 1836 | 1836 | 1836 | 1836 | 1836 | 1836 | 1836 | 1836 | 1464 |
| mean | 38.87 | 34.86 | 64.30 | 46.51 | 36.02 | 34.02 | 61.90 | 46.82 | 38.99 | 34.57 | 67.19 | 31.60 | 23.06 | 63.39 | 1.71 | 0.14 | 2.22 |
| p50 | 40.45 | 35.05 | 66.44 | 49.18 | 35.82 | 34.84 | 61.91 | 48.67 | 39.08 | 35.12 | 67.85 | 28.37 | 23.00 | 65.00 | 0.00 | 0.11 | 0.00 |
| iqr | 9.99 | 0.73 | 7.24 | 14.56 | 7.51 | 1.89 | 9.40 | 14.25 | 4.77 | 0.72 | 10.19 | 9.79 | 8.00 | 18.00 | 0.00 | 0.17 | 4.00 |
| sd | 5.07 | 0.99 | 8.44 | 8.33 | 4.27 | 1.99 | 6.42 | 10.73 | 3.73 | 1.68 | 7.09 | 10.28 | 5.79 | 13.52 | 8.75 | 0.11 | 3.07 |
| min | 29.47 | 25.64 | 35.54 | 22.90 | 26.70 | 26.49 | 42.94 | 14.67 | 30.70 | 24.96 | 45.27 | 15.12 | 9.00 | 10.00 | 0.00 | 0.00 | 0.00 |
| max | 46.93 | 37.54 | 84.64 | 64.93 | 45.49 | 37.44 | 80.18 | 67.75 | 49.01 | 38.44 | 93.23 | 62.87 | 45.00 | 94.00 | 101.60 | 0.67 | 10.00 |
| cv | 0.13 | 0.03 | 0.13 | 0.18 | 0.12 | 0.06 | 0.10 | 0.23 | 0.10 | 0.05 | 0.11 | 0.33 | 0.25 | 0.21 | 5.13 | 0.77 | 1.38 |

**Table 3**
Site C. Descriptive statistics of internal and external hive factors before preprocessing.

| Hive 1-C | | | | | Hive 2-C | | | | Hive 3-C | | | | External factors | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| stats | Weight [kg] | Tint [°C] | RHint [%] | SPL [dB] | Weight [kg] | Tint [°C] | RHint [%] | SPL [dB] | Weight [kg] | Tint [°C] | RHint [%] | SPL [dB] | Text [°C] | RHext [%] | Rain [mm] | WS [m/s] | UVI |
| missing | 372 | 375 | 372 | 372 | 373 | 373 | 373 | 373 | 0 | 0 | 0 | 0 | 419 | 419 | 373 | 419 | 419 |
| N | 1464 | 1461 | 1464 | 1464 | 1463 | 1463 | 1463 | 1463 | 1836 | 1836 | 1836 | 1836 | 1417 | 1417 | 1463 | 1417 | 1417 |
| mean | 45.49 | 33.94 | 65.25 | 42.65 | 49.94 | 35.04 | 66.48 | 38.90 | 40.13 | 34.31 | 57.00 | 35.42 | 20.34 | 63.79 | 1.20 | 0.16 | 2.00 |
| p50 | 43.45 | 34.74 | 65.44 | 46.86 | 50.46 | 35.08 | 67.44 | 42.70 | 42.68 | 34.62 | 60.56 | 34.90 | 19.30 | 67.00 | 0.00 | 0.11 | 0.00 |
| iqr | 6.38 | 0.56 | 4.02 | 16.89 | 5.00 | 0.46 | 5.62 | 24.34 | 9.92 | 1.50 | 15.62 | 19.60 | 10.00 | 30.00 | 0.00 | 0.18 | 3.00 |
| sd | 7.77 | 2.69 | 3.59 | 11.03 | 3.52 | 0.50 | 6.00 | 12.85 | 6.64 | 3.16 | 13.04 | 11.14 | 7.21 | 20.96 | 5.54 | 0.16 | 3.02 |
| min | 36.98 | 22.17 | 48.72 | 12.62 | 38.35 | 32.60 | 29.25 | 13.18 | 14.33 | 20.47 | 21.62 | 10.68 | 6.00 | 10.00 | 0.00 | 0.00 | 0.00 |
| max | 63.99 | 37.56 | 78.33 | 72.02 | 54.90 | 38.43 | 78.82 | 70.77 | 50.99 | 45.76 | 79.02 | 58.13 | 42.40 | 99.00 | 56.10 | 1.63 | 12.00 |
| cv | 0.17 | 0.08 | 0.06 | 0.26 | 0.07 | 0.01 | 0.09 | 0.33 | 0.17 | 0.09 | 0.23 | 0.32 | 0.35 | 0.33 | 4.61 | 1.02 | 1.51 |

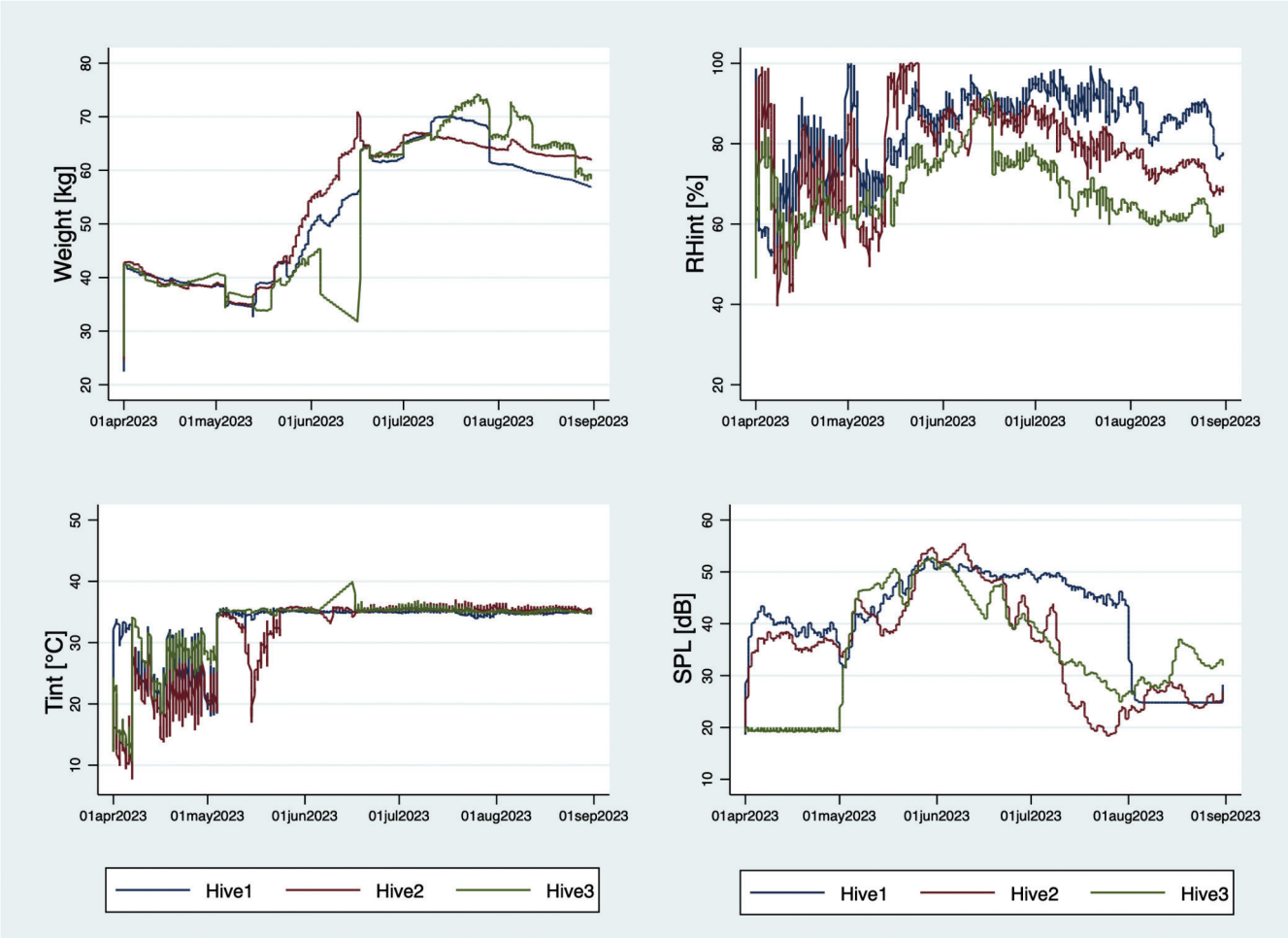


**Fig. 4.** Site A. Time plot of smoothed series by hives.

is positioned in a hilly agricultural area where olive groves, vineyards and little arable land predominate.

- Site B is located at the Department of Agricultural, Food and Forest Sciences of the University of Palermo at 160 m a.s.l., within the urban perimeter of the city of Palermo and includes the experimental fields of the Department, with herbaceous and aromatic plant species, and cultivated and ornamental trees, among them *Citrus spp., Washingtonia filifera, Ailanthus altissima*.
- Site C, located at the Basile farm in the province of Palermo (municipality of Ventimiglia di Sicilia), at an altitude of 550 m a.s.l., it is characterized by steeply sloping land towards the East and is furrowed by numerous valleys and watersheds, which sometimes deeply cut into the land. The area is largely occupied by uncultivated land, while the less steep areas are covered by arable land.

In both sites A and C, on the roadside, *Eucalyptus* spp. and *Ailanthus altissima* are present, too. The analysis of the vegetation involves both the area where the hives are positioned and the neighbouring area, so that sufficient nectariferous resources are available for the development and growth of the colonies.

The entrance to the hives is exposed to the sun from the early hours of the morning, favouring the start of the flights on the blooms and the activity inside the hives.

The study was carried out on nine hives, three per site, respectively named Hive1, Hive2 and Hive3 in each site. The positioning of the intelligent hives, which are described below, was carried out in November 2022 in order to set the system before the production season and to allow the bee communities to adapt. Monitoring began immediately, while the data reported in this paper range from April 1$^{st}$ to August 31$^{st}$, 2023.

In the hives placed in site A, the colony was replaced at the end of April, due to the heavy decrease in the number of bees recorded at the end of the winter. In the hives placed in sites A and C the honey super was placed before April.

### 2.2. Hive monitoring system

The hives used are of the Dadant Blatt type (Fig. 2) with small porch, equipped with 10 frames, the most used in Italy; the raised bees are local crossbreeds of the *Apis mellifera sp*.

The developed system consisted of a series of sensors connected to a microcontroller, which acquires the data and periodically sends them to a server via modem. In particular, the sensors applied to the hive were:

- four load cells, arranged at the base of the hive in correspondence with the vertices, measuring the weight (Fig. 3 [1]). Load cell made of stainless steel, hermetically welded IP68, combined error ± 0.017 % of capacity, low sensitivity to temperature variation, accuracy class C3, nominal range of excitation voltage 1 -15 V, rated output 2 mV/V ±10 %. The maximum capacity of each load cell was 75 kg, therefore the overall system was capable of weighing up to 300 kg for each hive;
- internal (Fig. 3 [2 and 3]) and environmental (Fig. 3 [5 and 6]) thermometers and hygrometers based on Sensirion AG's (Switzerland) SHT21 sensor, offering high performance, long-term stability, and ease of integration through the I2C interface for

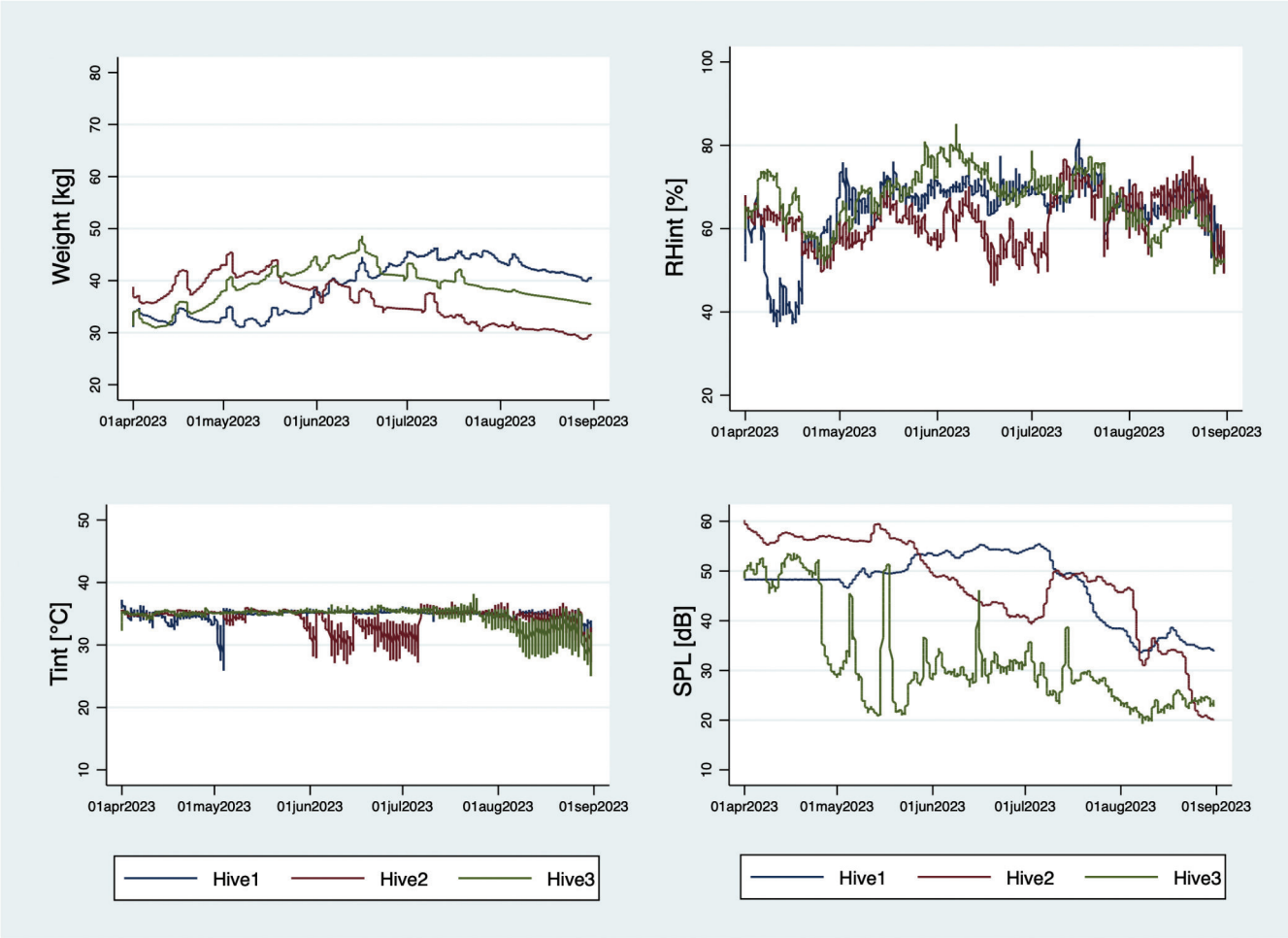


**Fig. 5.** Site B. Time plot of smoothed series by hives.

communication with the microcontroller. Temperature measurement range from -40 °C to 125 °C, relative humidity from 0 % to 100 %. It has a measurement accuracy of ± 0.3 °C for temperature and ±2 % for relative humidity. It consumes ~0.15 μA in sleep mode and ~400 μA during measurement. The response time for temperature measurement varies from 5 to 30 sec depending on the medium, while it is only 8 sec for humidity measurement;

- a digital microphone (Fig. 3 [4]) based on MEMS (Micro--Electro-Mechanical Systems) technology with an I2S interface, consisting of the GY-SPH0645 I2S sensor. The main characteristics of this sensor are frequency range 20 Hz–20 kHz, SNR (Signal-to-Noise Ratio) 65 dB, representing a good signal-to-noise ratio, sensitivity 26 dB. Power supply voltage is 1.8 V–3.6 V and consumption is < 1 mA. It is an omnidirectional microphone capable of capturing sounds from all directions, useful for ambient sound capture applications;
- the digital light sensor TSL2561 measuring ambient light. Its main features are: I2C communication interface with selectable addresses, measurement range from 0.1 to 40,000 lux, spectral response 400–700 nm, 16-bit accuracy and resolution, automatic temperature compensation, automatic sensitivity calibration, selectable low and high gain, selectable integration time: 13 ms, 101 ms, 402 ms, power consumption 0.24 mA during measurement and 0.005 mA in sleep mode, supply voltage: 2.7 V–3.6 V;
- the STM32WB microcontroller (Fig. 3 [8]) belongs to a family of microcontrollers specifically designed for wireless applications by STMicroelectronics, combining high performance and low-power connectivity. Based on the ARM Cortex architecture, the STM32WB integrates both a general-purpose processing core and a dedicated core for wireless communication management, making it ideal for IoT applications and smart devices. It features a Dual-core architecture, including an ARM Cortex-M4 core for application processing and an ARM Cortex-M0+ core dedicated to wireless connectivity management, supporting the Thread protocol for IoT mesh network applications. Flash Memory: Up to 1 MB, RAM: Up to 256 KB of SRAM, EEPROM: integrated for non-volatile data storage. Peripherals and I/O: GPIO, 12-bit ADC for analog signal acquisition, DAC for analog signal generation, advanced timers for PWM, event capture, and timing functions. Communication Interfaces: UART, SPI, I2C, CAN, USB, and others. Power Saving Modes: Various low-power modes, including sleep, stop, and standby. Optimized for battery-powered applications, with very low operating current;
- weather station (Fig. 3 [9]). The meteorological station EcoWitt WH6006 recorded the main weather-climatic parameters: temperature (°C), relative humidity (%), rainfall (mm), solar radiation (lux, fc or w / $m^2$), wind direction and speed (mph, km / h, m / s). The weather station was located in each study site. It consists of an external sensor body and a receiving unit with an external solar panel. The sensor body reads the values detected by the sensors and sends them to the receiving unit. It includes a battery of integrated sensors: thermo-hygrometer / rain gauge / anemometer / wind direction sensor, light and UV sensor, integrated solar panel, alkaline batteries, RF module (433 MHz) data transmission. The receiving unit acquires the data sent by the external sensor body via the radio frequency module. Data is saved as a CSV file on the SD card and transmitted to the public weather server: www.wunderground.com via 2G/3G WCDMA/GSM network. The weather conditions can be

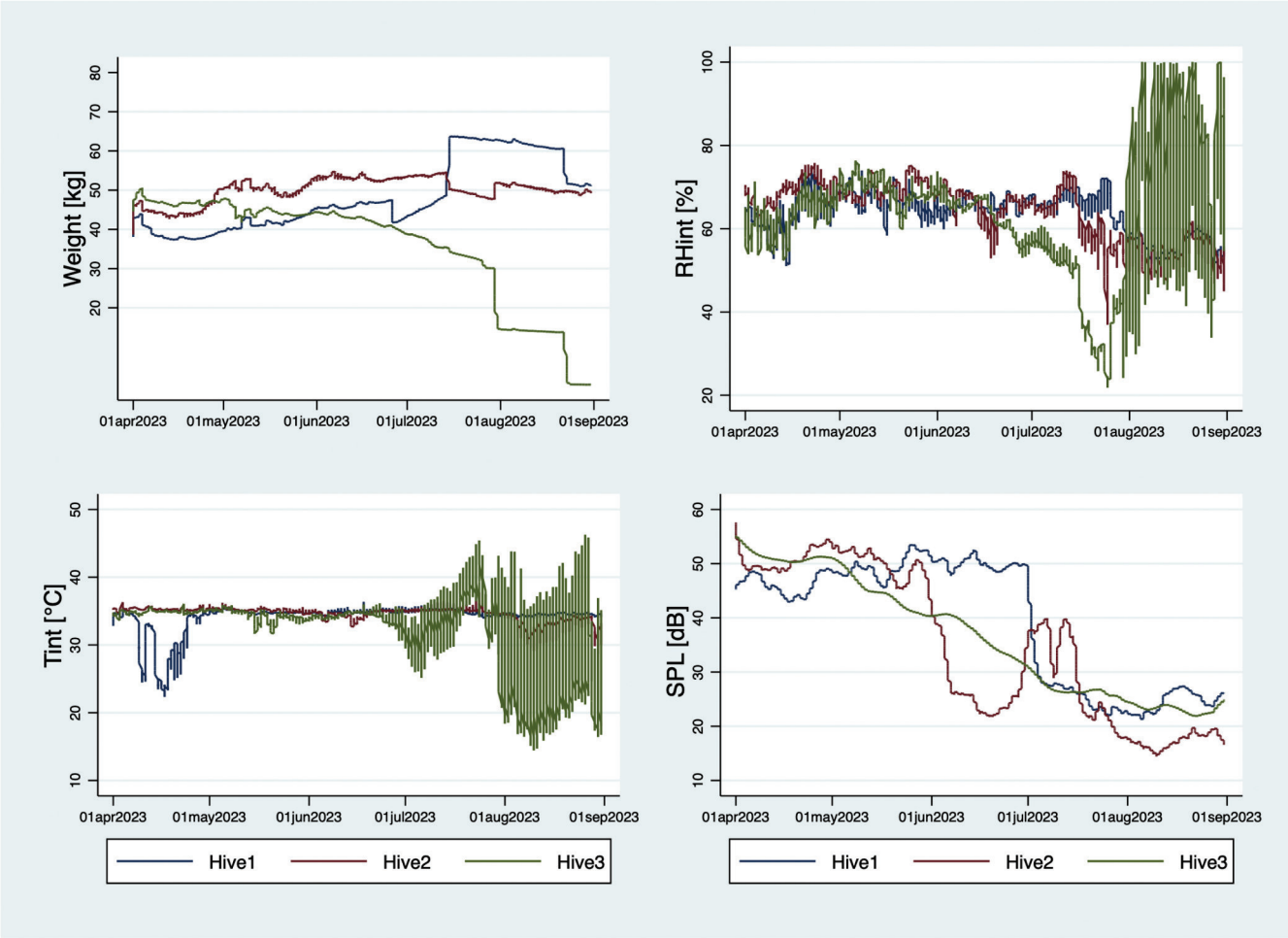


**Fig. 6.** Site C. Time plot of smoothed series by hives.

remotely monitored and data downloaded in .xlsx format. This unit consists of RF module (433 MHz) for data reception, support for data storage on SD memory card, 2G/3G GSM module with SIM card for sending data to the server, rechargeable lithium battery.

The microcontroller collected the acquired data, sending them to a cloud every 15 min via a multiband GSM modem (2G, 3G and 4G); in addition to archiving the data, it allowed for initial processing and its graphical representation. The graphs can be viewed by accessing a specifically created web interface where the trend of temperatures (internal and external), relative humidity, weight and noise level is visible [7]. The system was powered by solar energy.

### 2.3. Statistical analysis

Time series are observations that extend over time. The natural tendency of many phenomena to evolve in a more or less regular way leads to consider that the data detected at a given instant t is more similar to that detected at instant t−1 rather than at distant times; it is said that the time series has a "memory of itself". This characteristic is indicated as persistence, and it is the characteristic that differentiates time series samples from cross-section ones. In some situations, as in the case of this study, the presence of multiple time series requires tools to study the interrelationships and/or detect the cause-effect results between the detected series. When literature does not provide with a behavioural model for the analysis of a multivariate phenomenon and there are factors that can be both cause and effect of other variables observed over time, the VAR can help to understand the interactions and cause-effect relationships existing between time series.

The statistical methods used in this paper are distinguished by pre-processing phase and data analysis phase. The whole data set was obtained by downloading data from the above described hive monitoring system, i.e. weight, internal and external temperature (Tint and Text), internal and external relative humidity (RHint and RHext), Sound Pressure Level (SPL), rain, Wind Speed (WS) and UV Index (UVI).

#### 2.3.1. Data pre-processing

As descriptive statistics mean, median, minimum, maximum, standard deviation, coefficient of variation, interquartile range were calculated.

The second order exponential smoothing model [17] was considered to reduce hourly fluctuation and to identify long-term trends underlying structure of the time series and filling missing data. A second order exponential smoothing method is suitable when data have no clear trend and seasonal pattern. Based on this method, the optimal smoothing parameter was obtained by minimizing the in-sample sum-of-squared forecast errors. Missing values were filled in using the one-step-ahead predictions from the previous period. Missing values at the beginning and ending of series were not considered in the smoothing and were excluded from the sample.

After smoothing, missing values at the beginning or at the ending of time series were estimated using the predictive mean matching imputation methods (PMM). The PMM [18–20] combines the standard linear regression and the nearest-neighbour imputation approaches and is preferable to linear regression when the normality of the underlying model is suspect. It is a partially parametric method that matches the

**Table 4**
VAR coefficient estimation.

| | Site A | | | Site B | | | Site C | | |
|---|---|---|---|---|---|---|---|---|---|
| dlnWeight | H1 $R^2 = 0.77$ | H2 $R^2 = 0.85$ | H3 $R^2 = 0.43$ | H1 $R^2 = 0.79$ | H2 $R^2 = 0.77$ | H3 $R^2 = 0.80$ | H1 $R^2 = 0.63$ | H2 $R^2 = 0.53$ | H3 $R^2 = 0.87$ |
| $\theta_{11}$ | **0.7383** | **1.0306** | **0.6658** | **1.0066** | **0.9819** | **1.0133** | **0.9153** | **0.7830** | **1.1390** |
| $\theta_{12}$ | -0.0034 | **-0.2563** | **-0.0539** | **-0.1781** | **-0.1372** | **-0.1542** | **-0.2011** | **-0.2288** | **-0.2306** |
| $\theta_{13}$ | -0.0001 | **-0.0001** | -0.0002 | **-0.0008** | -0.0001 | 0.0001 | -0.0004 | 0.0004 | **-0.0002** |
| $\theta_{14}$ | 0.0000 | 0.0001 | 0.0002 | 0.0000 | 0.0000 | -0.0001 | 0.0003 | -0.0005 | 0.0000 |
| $\theta_{15}$ | -0.0001 | 0.0000 | 0.0000 | **0.0001** | **0.0001** | 0.0001 | 0.0001 | -0.0002 | 0.0000 |
| $\theta_{16}$ | **0.0002** | 0.0000 | 0.0001 | 0.0001 | 0.0000 | -0.0001 | 0.0001 | 0.0001 | 0.0000 |
| $\theta_{17}$ | 0.0008 | 0.0014 | -0.0001 | -0.0010 | 0.0030 | 0.0002 | -0.0006 | 0.0010 | 0.0271 |
| $\theta_{18}$ | -0.0006 | -0.0010 | **0.0000** | 0.0018 | -0.0027 | -0.0001 | 0.0005 | -0.0009 | -0.0248 |
| $\gamma_{11}$ | 0.0000 | 0.0000 | 0.0000 | **0.0001** | 0.0000 | 0.0000 | 0.0000 | -0.0001 | 0.0000 |
| $\gamma_{12}$ | **0.0000** | 0.0000 | 0.0001 | 0.0000 | **0.0000** | 0.0000 | **0.0001** | **0.0002** | -0.0003 |
| $\gamma_{13}$ | 0.0067 | **0.0112** | 0.0002 | **-0.0021** | -0.0007 | -0.0013 | 0.0004 | **0.0157** | -0.0001 |
| A | 0.0001 | 0.0001 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0000 |
| **dTint** | $R^2 = 0.51$ | $R^2 = 0.55$ | $R^2 = 0.56$ | $R^2 = 0.48$ | $R^2 = 0.49$ | $R^2 = \mathbf{0.24}$ | $R^2 = 0.54$ | $R^2 = \mathbf{0.26}$ | $R^2 = \mathbf{0.64}$ |
| $\theta_{11}$ | 4.9223 | -9.2953 | **-4.0144** | **-33.3783** | **-25.9667** | 1.2561 | -2.9814 | 1.5322 | -2.3728 |
| $\theta_{12}$ | -2.9163 | 4.3462 | **4.0411** | **27.3805** | **20.1303** | 0.9382 | 2.6290 | -0.6589 | 11.0708 |
| $\theta_{13}$ | **0.7515** | **0.6854** | **0.8571** | **0.6862** | **0.5385** | **0.2611** | **0.8273** | **0.3800** | **0.9097** |
| $\theta_{14}$ | **-0.2004** | **-0.0768** | **-0.2543** | **-0.2006** | -0.0091 | **-0.0776** | **-0.2191** | **-0.1128** | **-0.2863** |
| $\theta_{15}$ | -0.0304 | -0.0014 | 0.0291 | **0.0200** | **0.0305** | -0.0258 | 0.0101 | **0.0320** | 0.0123 |
| $\theta_{16}$ | **0.0686** | **0.0623** | 0.0128 | **-0.0226** | **0.0392** | 0.0166 | **0.0207** | **-0.0293** | **0.0311** |
| $\theta_{17}$ | **0.8573** | **1.0218** | **0.1128** | -0.6472 | **1.0167** | 0.0384 | -0.0462 | **0.2095** | **-29.5414** |
| $\theta_{18}$ | **-0.8206** | **-1.2235** | **-0.1372** | 0.5763 | **-1.3144** | -0.0452 | 0.0660 | **-0.2416** | **28.7137** |
| $\gamma_{11}$ | **0.0815** | **0.1199** | **0.0537** | **0.0415** | **0.0792** | **0.0096** | **0.0274** | **0.0219** | **-0.0118** |
| $\gamma_{12}$ | -0.0057 | **-0.0159** | -0.0036 | **-0.0067** | **-0.0078** | **0.0450** | -0.0019 | **-0.0023** | -0.3549 |
| $\gamma_{13}$ | -0.0309 | -0.4300 | 0.2342 | **0.2791** | **1.3090** | **1.5781** | 0.7012 | 0.4319 | **0.1001** |
| A | 0.0035 | 0.0072 | 0.0049 | 0.0001 | -0.0040 | 0.0011 | 0.0006 | -0.0013 | -0.0129 |
| **dRHint** | $R^2 = 0.60$ | $R^2 = 0.64$ | $R^2 = 0.77$ | $R^2 = 0.48$ | $R^2 = 0.39$ | $R^2 = 0.63$ | $R^2 = 0.48$ | $R^2 = 0.51$ | $R^2 = 0.32$ |
| $\theta_{11}$ | 1.1481 | **73.8487** | -1.8483 | -23.0716 | **42.5569** | -4.4167 | -3.6250 | **11.0514** | **69.6513** |
| $\theta_{12}$ | -6.6139 | **-58.8643** | 0.0466 | 21.8207 | **-40.2443** | 6.8005 | -5.0545 | **-11.3902** | **-79.0441** |
| $\theta_{13}$ | **-0.3001** | **-0.4467** | -0.0083 | -0.0354 | **-0.1660** | **-0.1840** | 0.0896 | **0.4670** | **-0.2650** |
| $\theta_{14}$ | **-0.1069** | -0.0803 | 0.0203 | **0.3454** | 0.0653 | **0.1021** | **-0.1806** | **0.5164** | 0.0726 |
| $\theta_{15}$ | **0.7763** | **0.6769** | **1.2092** | **0.8330** | **0.6455** | **0.9948** | **0.7994** | **0.7328** | **0.5280** |
| $\theta_{16}$ | **-0.2639** | **-0.2232** | **-0.5159** | **-0.3213** | **-0.2260** | **-0.3955** | **-0.2385** | **-0.2314** | **-0.1862** |
| $\theta_{17}$ | 0.3509 | **1.1580** | -0.0540 | **-3.4604** | **2.9940** | -0.0588 | 0.9395 | 0.1899 | **89.7098** |
| $\theta_{18}$ | **-0.9437** | **-1.1305** | 0.0252 | **3.5814** | **-3.0486** | 0.0564 | -0.9769 | -0.2009 | **-89.5655** |
| $\gamma_{11}$ | **0.1250** | **0.0826** | **0.0209** | 0.0159 | **0.0504** | **0.0188** | **0.0659** | **0.0858** | **0.0378** |
| $\gamma_{12}$ | **0.0419** | **0.0888** | **0.0159** | **0.0306** | **0.0610** | **0.0231** | 0.0049 | **0.0302** | **-5.8329** |
| $\gamma_{13}$ | 0.2859 | -0.0550 | -1.8776 | -0.1748 | 0.8244 | **-0.8053** | **3.8350** | 1.1167 | **0.1295** |
| $\alpha$ | 0.0200 | 0.0031 | -0.0007 | 0.0035 | -0.0005 | 0.0006 | 0.0008 | -0.0044 | -0.0024 |
| dSPL | $R^2 = 0.83$ | $R^2 = 0.92$ | $R^2 = 0.07$ | $R^2 = 0.91$ | $R^2 = \mathbf{0.89}$ | $R^2 = 0.80$ | $R^2 = 0.95$ | $R^2 = 0.93$ | $R^2 = 0.99$ |
| $\theta_{11}$ | -0.6528 | **-2.7359** | 0.3430 | -0.6187 | -0.7502 | -4.4120 | 0.2174 | -0.4412 | 0.0080 |
| $\theta_{12}$ | **2.4336** | **2.8545** | 0.4438 | 0.4285 | **0.9268** | 7.3961 | -0.2965 | -0.5184 | -0.0243 |
| $\theta_{13}$ | **0.0127** | **0.0103** | -0.0036 | **0.0044** | **0.0069** | -0.0127 | -0.0021 | **0.0329** | **0.0003** |
| $\theta_{14}$ | **0.0079** | -0.0008 | 0.0145 | **0.0037** | **0.0033** | 0.0361 | 0.0046 | **0.0320** | **0.0004** |
| $\theta_{15}$ | **0.0043** | -0.0017 | 0.0110 | 0.0005 | 0.0009 | -0.0253 | 0.0004 | 0.0035 | **0.0001** |
| $\theta_{16}$ | 0.0034 | 0.0015 | -0.0208 | **-0.0012** | -0.0013 | **0.0282** | -0.0009 | **-0.0037** | -0.0001 |
| $\theta_{17}$ | **1.0139** | **1.1178** | **-0.0958** | **0.6443** | **1.0363** | **1.0910** | **1.1358** | **1.0530** | **1.1906** |
| $\theta_{18}$ | **-0.1575** | **-0.1859** | **0.2231** | **0.3197** | **-0.1092** | **-0.2336** | **-0.1686** | **-0.0968** | **-0.1959** |
| $\gamma_{11}$ | -0.0018 | 0.0027 | -0.0114 | -0.0001 | **0.0010** | -0.0010 | -0.0004 | 0.0018 | **0.0002** |
| $\gamma_{12}$ | **-0.0014** | **-0.0014** | -0.0028 | **0.0003** | -0.0003 | 0.0032 | **-0.0005** | -0.0005 | 0.0000 |
| $\gamma_{13}$ | 0.1721 | -0.1714 | -0.4119 | **0.0217** | **0.0635** | **0.4181** | 0.0283 | -0.1627 | 0.0065 |
| $\alpha$ | 0.0009 | -0.0005 | 0.0039 | -0.0002 | -0.0006 | -0.0029 | -0.0005 | -0.0009 | -0.0001 |

missing value to the observed value with the closest predicted mean (or linear prediction).

For each missing value in each variable in the general vector **X**, 5 imputations were taken into consideration. A set of neighbours (possible donors) were considered to calculate the linear prediction based on nearest-neighbour imputation approaches. The number of nearest neighbours regulates the trade-off between the bias and the variance of the point estimators in repeated sampling. In this paper 6 nearest neighbours seemed to be appropriate (12 h).

After data preprocessing, the Vector Autoregressive models (VAR) were considered.

### 2.3.2. Processing of data VAR model

The VAR models are multiequation models based on time series data, containing an equation for each variable, and each variable depends on past values of itself and past values of the other variables, this means that all the variables are endogenous [21].

VAR is a multivariate extension of AR (AutoRegressive) models but capturing the historical patterns of each variable and its relationship to the others.

The general VAR model with K variables, can be written as linear function of p of their own lags, p lags of the other K-1 endogenous variables and f lags of the additional M exogenous variables $\mathbf{x}_t$. In general, a p-order VAR model, VAR(p), can be written as:

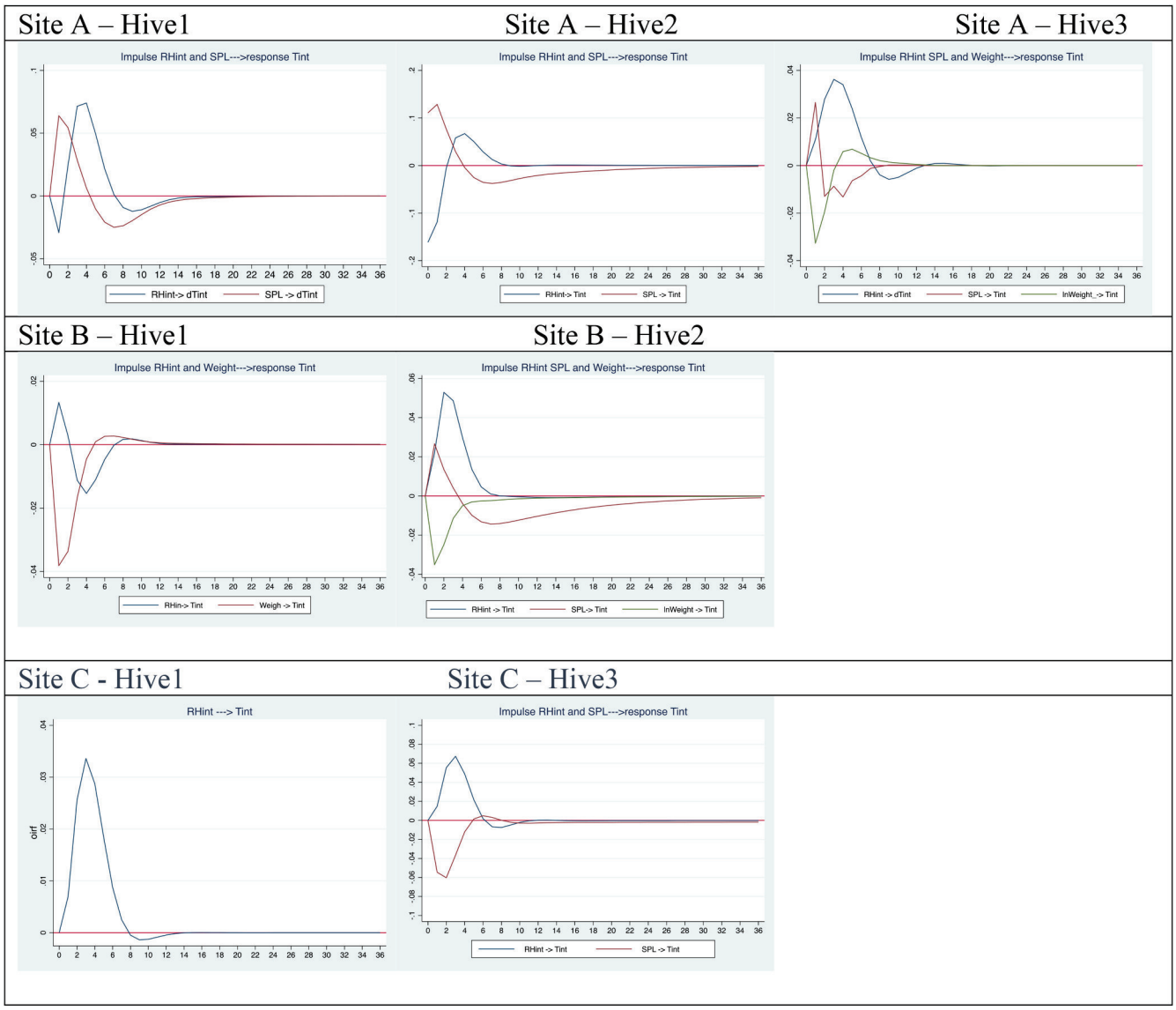


**Fig. 7.** Sites A, B, C. Percentage of variation on Impulse function for Tint (Y axis); X-axis represents 2 h period.

$$\mathbf{y}_t = \boldsymbol{\nu} + \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{B}_0\mathbf{x}_t + \mathbf{B}_1\mathbf{x}_{t-1} + \ldots + \mathbf{B}_f\mathbf{x}_{t-f} + \mathbf{u}_t\boldsymbol{t} \in \{-\infty,\ \infty\} \quad (1)$$

Where $\mathbf{y}_t = \left(y_{1t,\ldots..}, y_{Kt}\right)'$ is $K \times 1$ random vector of variables of the system;

$\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_p$, are $K \times K$ matrices of parameters;

$\mathbf{x}_t = \left(x_{1t,\ldots..}, x_{Mt}\right)'$ is a $M \times 1$ vector of M exogenous variable at time t;

$\mathbf{B}_0$, $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_f$ are the $K \times M$ matrices of coefficients;

$\boldsymbol{\nu}$ is $K \times 1$ vector of intercepts and $\mathbf{u}_t$ is the white noise disturbance with: expected mean $E(\mathbf{u}_t) = 0$; covariance matrix $E(\mathbf{u}_t\mathbf{u}'_s) = \Sigma$ if t=s, and $E(\mathbf{u}_t\mathbf{u}'_s) = 0$ if $t \neq s$. This means that $\mathbf{u}_t$ and $\mathbf{u}_s$ are uncorrelated that it is the same to say: as $\mathbf{y}_t$ is an autoregressive (AR) process it is assumed that the forecast errors for different periods are uncorrelated. This last assumption indicates that all useful information in the past $\mathbf{y}_t$'s is used in the forecasts so that there are no systematic forecast errors.

The number of equations depends on the number of the endogenous variables k, whereas p and f in (1) represent the lags of the endogenous and exogenous variables of the model. In our model the exogenous variables are observed at time t and no lagged effects have been considered, therefore s=0.

The parameters matrix $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_p$ serves to study the dependence between endogenous variables and measure the dynamic effect between the K variables and the variable itself, $\mathbf{B}_0$ and the variance matrix $\Sigma$, measures the contemporary effects.

Estimating the parameters in a VAR requires that the variables in $y_t$ and $y_{t-p}$ are covariance stationary, meaning that their first two moments (Cross-Correlation Functions ($CCF_{ij}(p)$) and Auto-Correlation Functions ($ACF_{ii}(p)$), Section 2.3.3) exist and are time invariant.

In general VAR(p) is stationary if all the eigenvalues of the coefficient matrix of the estimated VAR(p) have modulus less than 1. The Dickey–Fuller unit-root test (1979) [22] was considered to test the stability condition of the time series. The test statistic is rejected if $\alpha \leq 5$ %.

If the series is stationary, the times up to the second order of the process are t-independent.

### 2.3.3. *Covariance, variance, cross-correlation and autocorrelation function*

The cross-covariance function of lag p for the generic time series $y_{i,t}$, $y_{j,t-p}$, can be written as

$$R_{ij}(p) = Cov\left(y_{i,t},\ y_{j,t-p}\right) \quad (2)$$

varying $i$ and $j$ described as a $K \times K$ matrix of lag p.

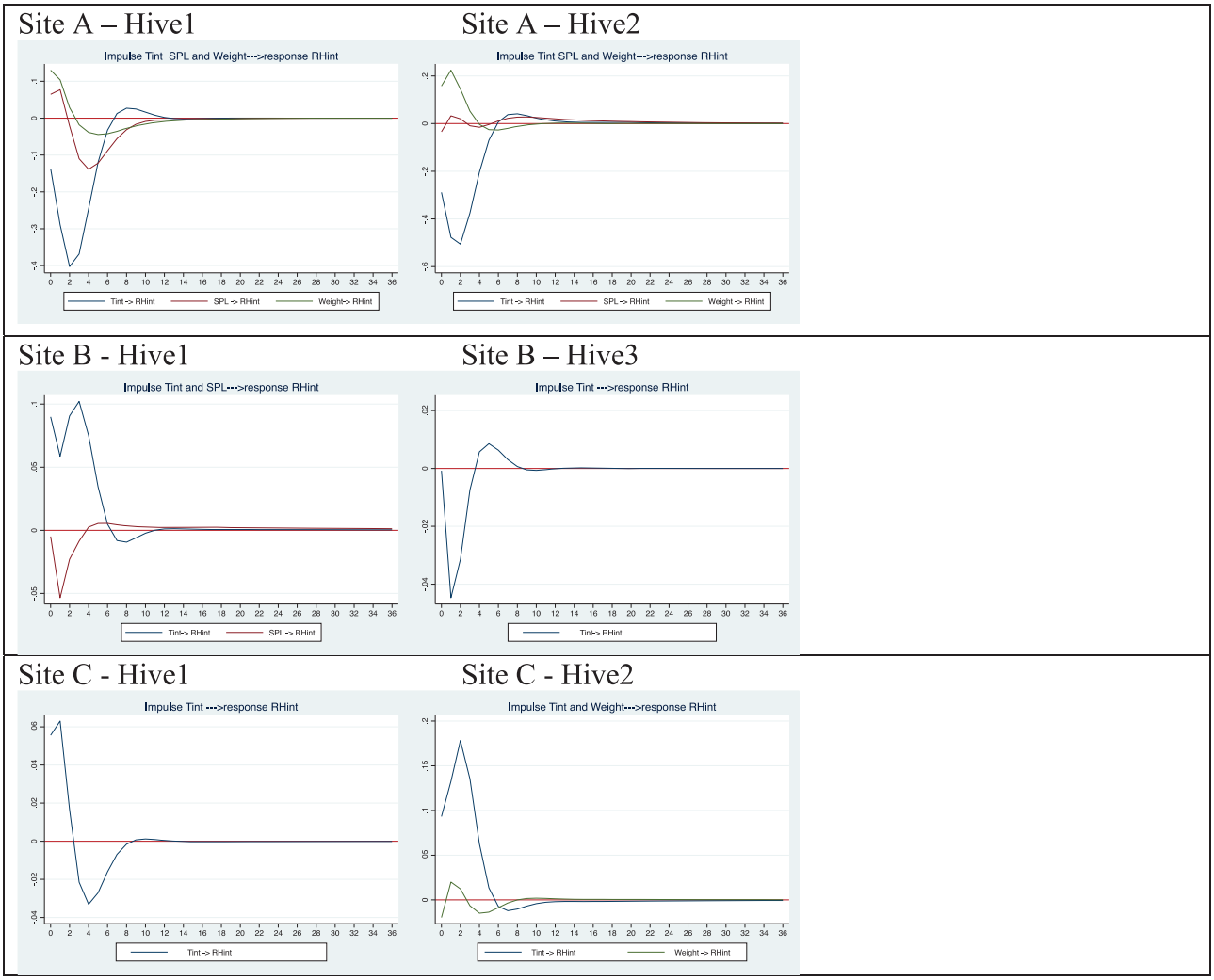


**Fig. 8.** Sites A, B and C. Percentage of variation on Impulse function for RHint (Y axis); X-axis represents 2 h period.

For i=j, $R_{ij}(p)$ is the autocovariance function of $y_{i,t}$; for i≠ *j*, $R_{ij}(p)$ is the cross-covariance function between $y_{i,t}$ and $y_{j,t}$.

If $R_{ij}(p) = 0$ for all p, the series are stationary and the function does not depend on t.

For p=0, $R_{ij}(0)$ represents the matrix of contemporary covariances between $y_{i,t}$ and $y_{j,t}$.

And, $R_{ii}(0)$ is the variance of $y_{it}$

The cross-correlation function $CCF_{ij}(p)$) can be defined as the correlation between one series at time t ($y_{i,t}$) and another series *j* at time t-p $\left(y_{j,t-p}\right)$ and it is a function of the time t and lag p:

$$CCF_{ij}(p) = \frac{R_{ij}(p)}{\left[R_{ii}(0)R_{jj}(0)\right]^{\frac{1}{2}}} \tag{3}$$

So, if i=j, $CCF_{ij}(p)$ became the $ACF_{ii}(p)$ is the autocorrelation function of $y_{i,t}$; for i ≠ *j*, $CCF_{ij}(p)$ is the cross-correlation function between $y_{i,t}$ and $y_{j,t}$. And in the same way, $CCF_{ji}(p)$ is the cross-correlation function between $y_{j,t}$ and $y_{i,t}$.

If $CCF_{ij}(p) = 0$ for all p, the series are stationary, and the function does not depend on t.

To choose the order p of the VAR model different approaches can be used. In this work the Akaike Information Criterion (AIC) and the Lagrange-multiplier test [23] of the differenced time series were considered. To check the adequacy of the estimated VAR model we check the stationarity of the pairwise Cross-Correlation Functions (3) of the time series.

### 2.3.4. *Granger-causality*

To test whether a time series offers a useful information in forecasting another time series, the Granger causality test is used. When a VAR model is identified because stationary, the Granger causality test helps us to identify if one variable helps to predict each other's. Granger causality test does not provide insight about the true causal relationship between two variables but about the forecasting ability.

A variable X is said to Granger-Cause (GC) a variable Y, if given the past values of Y, past values of X are useful for predicting Y.

The null hypothesis for Granger-causality test is:

$H_0 : X\ GC\ Y = 0$ against

$H_1 : X\ GC\ Y > 0.$

If X GC Y and Y GC X, the process $(X^{'},\ Y^{'})$ is called a feedback system or instantaneous causality [21].

### 2.3.5. *Impulse response function*

Significantly, Granger-cause variables were analyzed through the Impulse Response Function (IRF) that allows us to trace out the time path of variables in our models to a one unit increase of another in the system.

Let $Y_t$ the MA (Moving Average) representation of the VAR model:

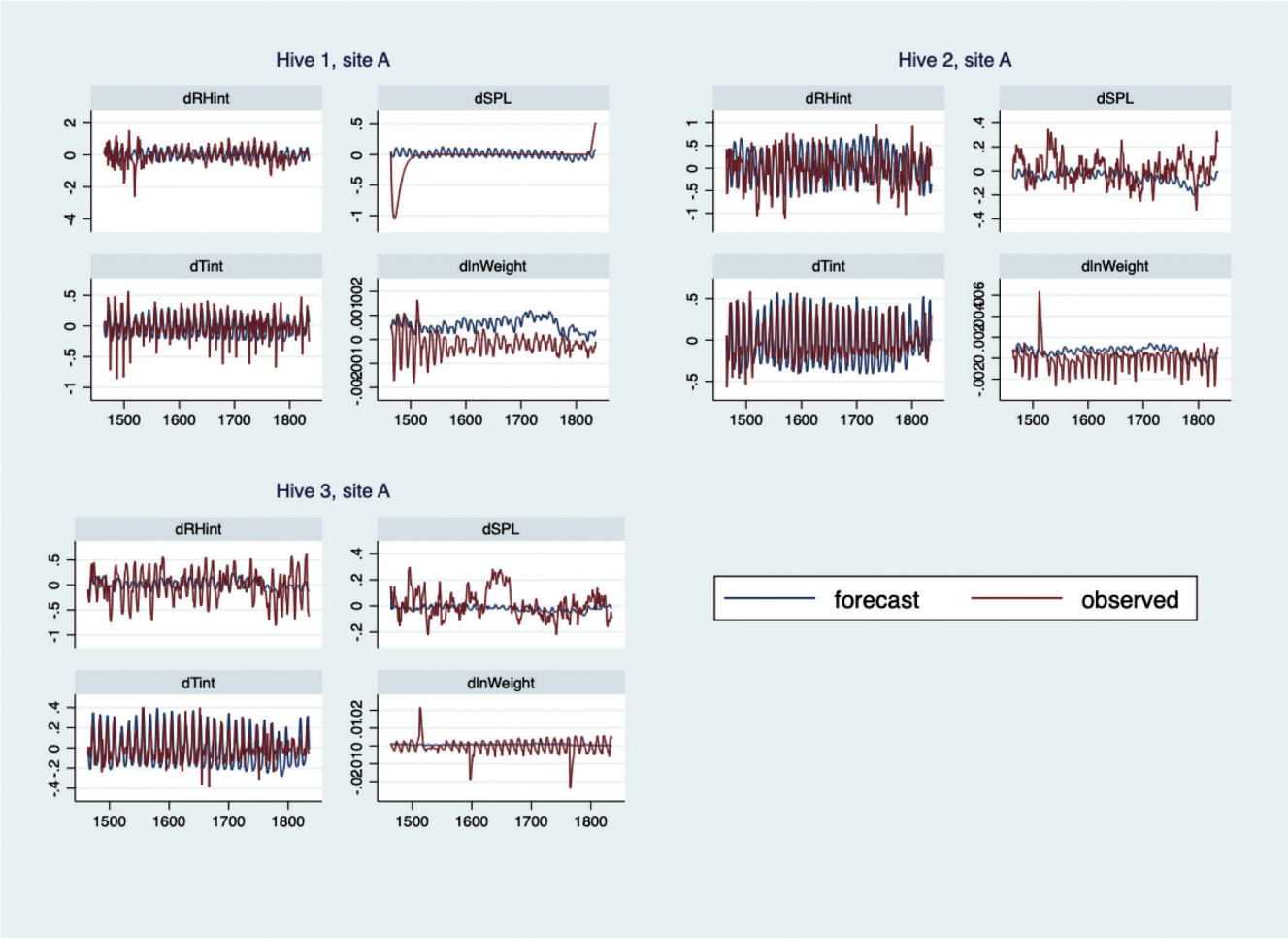


**Fig. 9.** Forecasts VAR models, Hive1, Hive2 and Hive3, site A.

$$Y_t = \mu + \sum_{i=1}^{\infty} \theta_i \varepsilon_{t-i} \tag{4}$$

For a stationary process

$$A(L)Y_t = \varepsilon_t \tag{5}$$

Let's define the impulse function:

$$h(i,j,n) = \frac{\partial y_{it}}{\partial \epsilon_{jt-n}} \tag{6}$$

It can be interpreted as the response of the i-th variable to the j-th shock after n periods.

Since the vector $\epsilon$ represents the gap between $Y_t$ and its expected value conditional on the information set $I_{t-1}$, $\varepsilon_t$ is often simply referred to as the "one-step forecast error". More formally, we might think of our forecast error vector as a function of movements in behavioural relationships, which we call structural shocks, so we might write:

$$\varepsilon_t = \mathrm{B}u_t \tag{7}$$

If B was known, we could reconstruct the history of structural shocks (through $u_t = B^{-1}\varepsilon_t$), but above all we could calculate the structural impulse responses: by putting together Eqs. (5) and (7) we have:

$$A(L)Y_t = \mathrm{B}u_t \tag{8}$$

$$Y_t = [A(L)]^{-1}\mathrm{B}u_t = \mathrm{B}u_t + C_1\mathrm{B}u_{t-1} + C_2\mathrm{B}u_{t-2} + .. \tag{9}$$

So

$$IRF(i,j,n) = \frac{\partial y_{it}}{\partial u_{jt-n}} = (C_n \cdot B)_{ij} \tag{10}$$

The impulse response to the structural shock would allow us to assess how observable quantities respond over time (in our case the endogenous variables Tint, RHint, SPL) with respect to a shock that impacts on a behavioural relationship (relationship between behaviour of the bees and variables within the hives), and for this reason it is called "structural". Given two variables $Y_{1t},\ Y_{2t}$, IRF responds to the question: what is the effect of one unit shock in $Y_{1t},\ Y_{2t}$?

Moreover, the scenario analysis of the effect of a shock to the system with the impulse response functions can help gain insight into how long the effect of a shock will last.

In general, to calculate the impulse functions, the variables should be ordered from the least to the most reactive. In our work there is a certain arbitrariness in choosing the ordering of the variables because we do not know which is the most reactive in the system, so we decided to consider the following order of the variables RHint, SPL, Tint, Weight.

To ensure the stability condition, transformation in first difference was applied to all the variables considered. For the weight series, transformation in difference of logarithm of weight was applied.

Finally, prediction in the test set were considered to test the prediction capability of the model, p-value $< 0.05$ was considered significant.

All the statistical analyses were carried out with Stata 16.1 statistic software (StataCorp LLC, College Station, Texas 77845 USA).

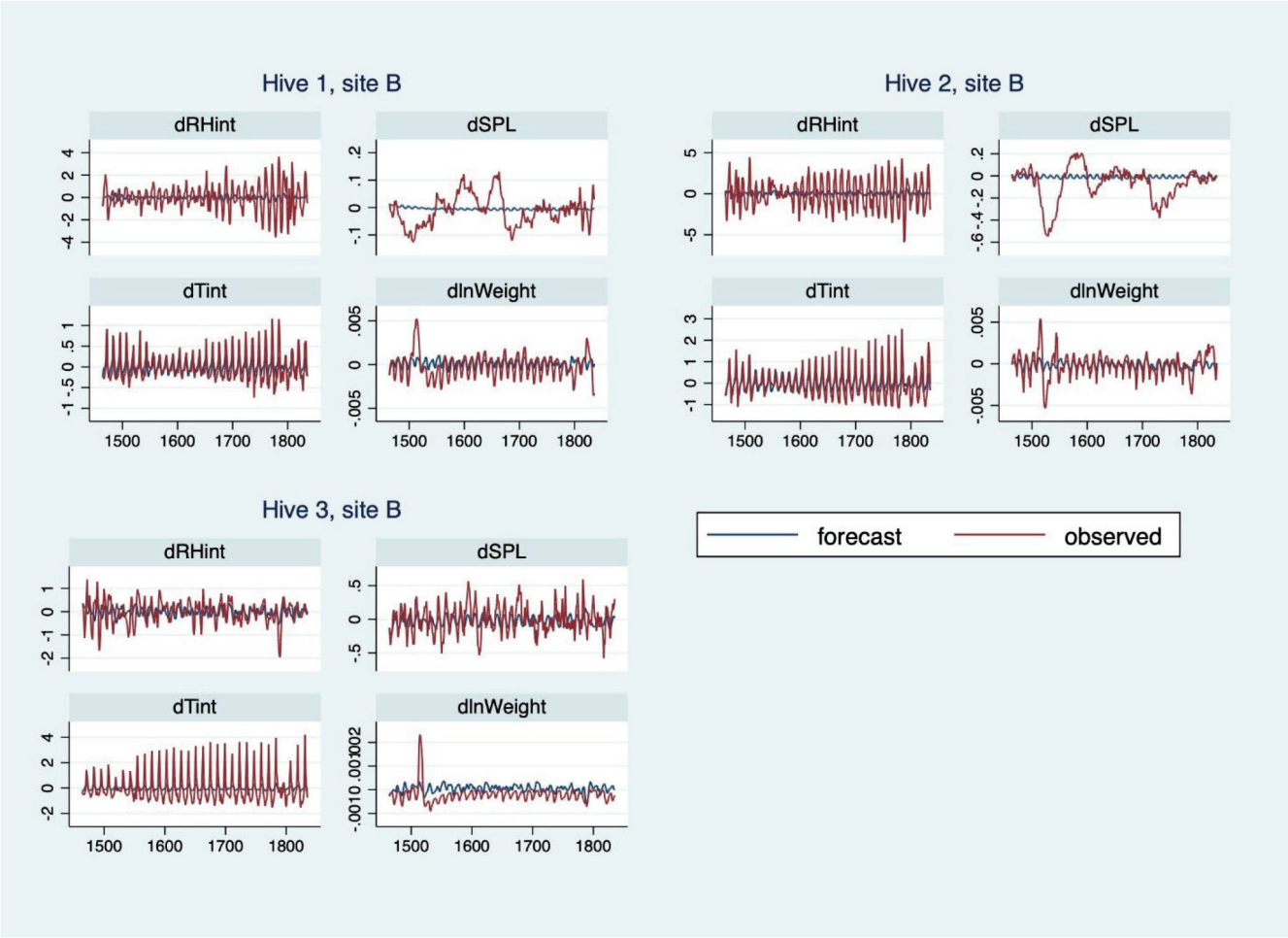


**Fig. 10.** Forecasts VAR models, Hive1, Hive2 and Hive3, site B.

## 3. Results and discussion

### 3.1. Data monitoring and post processing results

The descriptive statistics of the monitored parameters before processing, both internal and external factors, are reported in Tables 1, 2 and 3 respectively for the three sites A, B and C. The line "missing" reports the number of missing data that have not been detected due to problems of different nature independent of the human factor (sensor malfunction, interruption of data transmission, etc.).

After performing data smoothing and imputation of the missing values for all the internal parameters, the complete data sets for weight, internal temperature, internal RH, SPL were represented in the following Figs. 4 (site A), 5 (Site B) and 6 (Site C).

In site A (Fig. 4), in April the weight slightly decreased in all the three hives, probably due to a slow development of the families linked to both the colony status at the end of the winter period and to the external temperature, which ranged between 5 and 22 °C. Indeed, during the cold season bees rarely go out to forage, consuming the reserves inside the hive, and causing a consequent weight decrease of the colony by 30–80 g per day [24]. Ochoa et al. (2019) [10] argue that exposure to low temperatures during operculate reproductive states induces high mortality and shortening in worker longevity. When the weather temperature was higher than 25 °C, the temperature distribution in the beehive was relatively uniform [24]. In the following months, i.e. starting from May until July, an increase in weight was recorded in all the three hives, respectively 21 kg for Hive1, 25 kg for Hive2 and 28 kg for Hive3. Afterwards, the weight of each hive decreases until the end of August.

Furthermore, sudden changes in weight have been recorded due to operator interventions such as, for example, the removal of real cells, placing of pollen traps or the usual bee colony monitoring. Due to the same reasons, an irregular trend was recorded in April also for internal temperature and humidity. Internal temperature during April was low, due to the heavy decrease in the number of bees recorded at the end of winter, while, from May on, internal temperatures stabilize at 35 °C, remaining constant until the end of the observed period. In Hive2 an internal temperature decrease was noticed in mid-May, probably due to a technical problem in data recording.

Also internal Relative Humidity shows high variability in April (ranging from 40 % to 100 %), while it had a more homogeneous trend starting from the second half of May, although the three hives, from June on, show similar trend, with ranges of 78–100 %, 65–92 % and 58–72 % in 1, 2 and 3, respectively. A decrease in RH in all the three hives was recorded from the second half of June, probably due to the temperature and humidity regulation activity of bees inside the hive [25].

SPL shows a trend similar to weight, with lower and variable levels in April, higher and more stable levels in May and June, and a decrease in July-August, particularly evident in hives 2 and 3.

In site B (Fig. 5), in which honey super were not used, the weight of the three hives was similar in the first period (about 35 kg), although Hive1 had only 7 frames. While until the end of June all hives show an increasing weight, although the maximum weight was recorded in a different period. Hive1 started to increase in June, after the number of frames was increased to 10 (first decade of May), reaching its maximum weight in July; Hive2 had a peak in May, followed by a decrease,

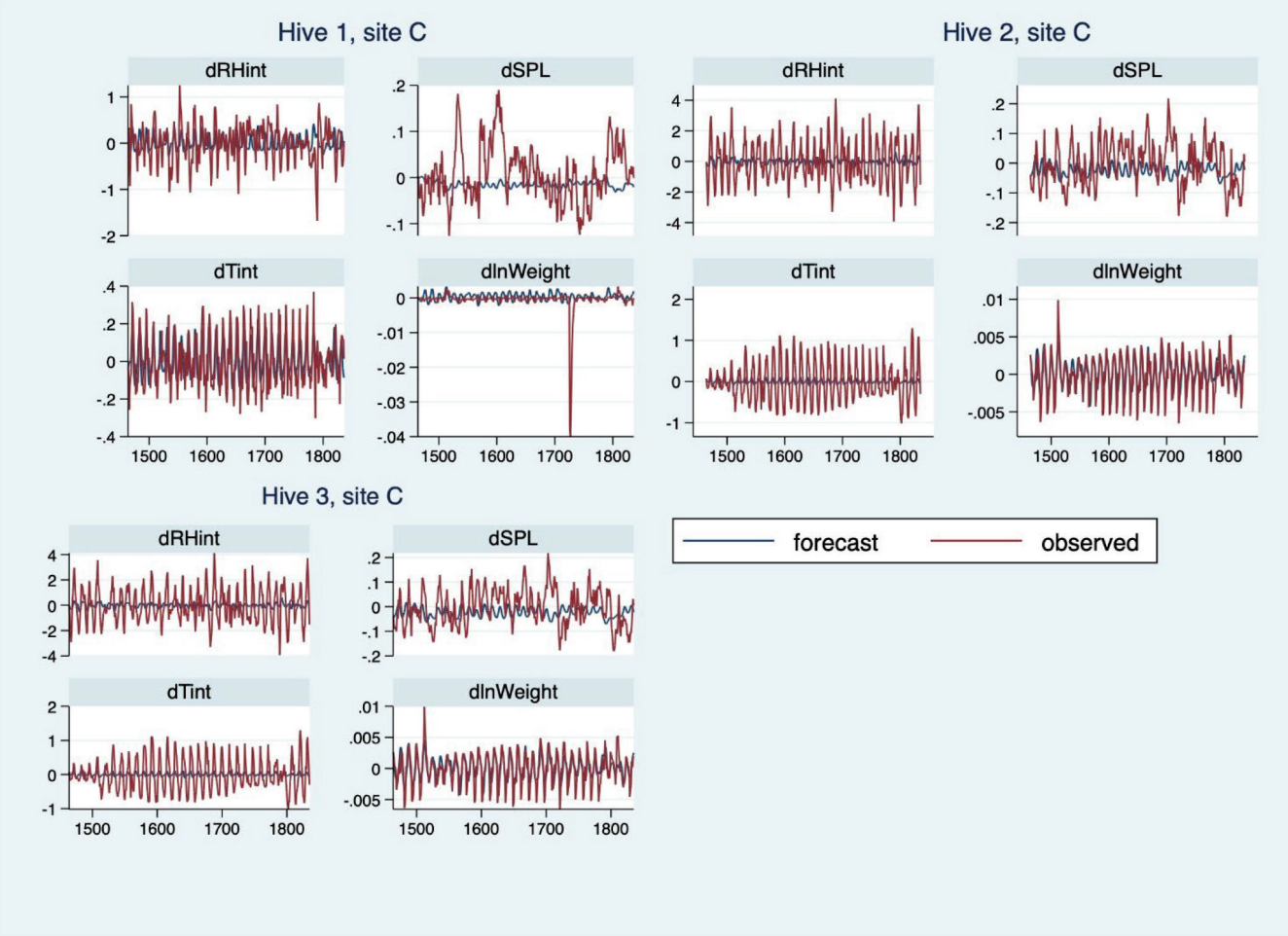


**Fig. 11.** Forecasts VAR models, Hive1, Hive2 and Hive3, site C.

probably due to a swarming. Finally, Hive3 was almost constant until June, afterwards increasing its weight during July. Starting from July, all the hives showed a decrease in weight, mainly due to intense predation by *Vespa orientalis*.

Internal relative humidity was quite different in the three hives during April, when Hive1, containing 7 frames, shows a lower RHint (about 40 %) compared to hives 2 and 3 (60–70 %). In the following months, the three hives show a similar RHint trend, except for Hive2, in which an appreciable decrease recorded in the second half of July could be linked to swarming. It has been observed that changes in the internal RH and temperature can be linked to sound and used as predictor for swarming of the bees [25]. On the other hand, the appreciable decrease in these two parameters in the Hive2 seems to be a consequence of swarming, as a lower number of bees in the colony are not able to guarantee a constant RHint and Tint.

Internal temperature is almost constant from May on, apart from Hive2 (see above) and a decrease in values recorded in Hive3 from middle August.

SPL shows a quite different trend in the three hives, with values always lower in Hive3 compared to hives 1 and 2.

In site C (Fig. 6), the hives 1 and 2 showed a similar trend for almost all parameters, while Hive3, from July on, shows anomalous values for all parameters. In the entire period considered, Hive1 records 11 kg increase in weight and Hive2 shows 10 kg. Furthermore, constant internal temperature values of 32–35 °C are observed for the entire period. A constant temperature of 35 °C inside the hive demonstrated that, during the test period, no critical issues occurred regarding the swarming or escape of the bees [26,5]. In Hive3, weight decreases from the second half of June, as RH, and large variability of the internal temperature is recorded, which can be traced back to possible problems within the hive or the recording system. SPLshows a decreasing trend in the three hives.

### 3.2. VAR model, parameter estimations and forecasts

The system takes into consideration the endogenous variables Weight/Tint/RHint/SPL. In particular, in this study the two-lag VAR model (VAR(2)), with 4 endogenous variables (Tint, Weight, RHint and SPL) and 3 exogenous variables (UVI, RHext, WS) was considered. The exogenous variables were not lagged.

To make the time series stationary, some transformations were performed on the original series, described below. Weight was considered as the difference of the logarithms in two consecutive periods t and t-1, and therefore represents approximately the percentage variation of the weight between time t and time t-1[1]. With $dlnWeight_t$ the differences of the logarithms of the weight from time t-1 to time t. are indicated.

All other endogenous (Tint, RHint, SPL) and exogenous (UVI, RHext, WS) variables were considered as first differences, thus representing an absolute change from time t-1 to time t. With dTint, dRHint, dSPL, dUVI and dWS we indicate the transformed variables and therefore the absolute variation of each variable between the two periods. The suffix *d* in front of the variables indicates that the differences of the historical series

[1] as $lnWeight_t - lnWeight_{t-1}$ is approximately equal to $(\frac{Weight_t - Weight_{t-1}}{Weight_{t-1}})$ when the variations are small.

have been considered.

The parameter dText was not considered as its high and significant correlation with RHext. Moreover, the external factor Rain did not show significance in all the considered models. Then, in accordance with the principle of parsimony, we excluded it from the estimated VAR models.

The pairwise Cross-Correlation Functions (CCF), Auto-Correlation Functions (ACF) (not shown) and [22] unit-root test were used to test the non-stationarity of the time series. The Lagrange-multiplier test [23] of the differenced time series lead us to identify the lag order of the VAR model for the relative percentage variation of Weight (dlnWeight), the absolute variation in Internal Temperature (dTint), the absolute variation in Internal Relative Humidity (dRHint), the absolute variation in Sound (dSPL).

Data was split into a training set (from April 1$^{st}$ to July 31$^{st}$) and a test set (from August 1$^{st}$ to August 31$^{st}$) to verify the forecast capacity of the model.

The following VAR model was estimated for the four endogenous variables in each site and hive:

$$\widehat{dlnWeight}_t = \alpha + \theta_{11} dlnWeight_{t-1} + \theta_{12} \mathrm{d}lnWeigh_{t-2} + \theta_{13} d\mathrm{Tint}_{t-1} + \theta_{14} d\mathrm{Tin}_{t-2} + \theta_{15} \mathrm{dRhint}_{t-1} + \theta_{16} \mathrm{dRhint}_{t-2} + \theta_{17} \mathrm{dSPL}_{t-1} + \theta_{18} \mathrm{dSPL}_{t-2} + (\gamma_{11} \mathrm{dUVI}_t + \gamma_{12} \mathrm{dRHext}_t + \gamma_{13} \mathrm{dWS}_t) \quad (11)$$

$$\widehat{d\mathrm{Tint}}_t = \alpha + \theta_{11} dlnWeight_{t-1} + \theta_{12} \mathrm{d}lnWeigh_{t-2} + \theta_{13} d\mathrm{Tint}_{t-1} + \theta_{14} d\mathrm{Tin}_{t-2} + \theta_{15} \mathrm{dRhint}_{t-1} + \theta_{16} \mathrm{dRhint}_{t-2} + \theta_{17} \mathrm{dSPL}_{t-1} + \theta_{18} \mathrm{dSPL}_{t-2} + (\gamma_{11} \mathrm{dUVI}_t + \gamma_{12} \mathrm{dRHext}_t + \gamma_{13} \mathrm{dWS}_t) \quad (12)$$

$$\widehat{\mathrm{dRhint}}_t = \alpha + \theta_{11} dlnWeight_{t-1} + \theta_{12} dlnWeight_{t-2} + \theta_{13} d\mathrm{Tint}_{t-1} + \theta_{14} d\mathrm{Tin}_{t-2} + \theta_{15} \mathrm{dRhint}_{t-1} + \theta_{16} \mathrm{dRhint}_{t-2} + \theta_{17} \mathrm{dSPL}_{t-1} + \theta_{18} \mathrm{dSPL}_{t-2} + (\gamma_{11} \mathrm{dUVI}_t + \gamma_{12} \mathrm{dRHext}_t + \gamma_{13} \mathrm{dWS}_t \quad (13)$$

$$\widehat{\mathrm{dSPL}}_t = \alpha + \theta_{11} dlnWeight_{t-1} + \theta_{12} \mathrm{d}lnWeigh_{t-2} + \theta_{13} d\mathrm{Tint}_{t-1} + \theta_{14} d\mathrm{Tin}_{t-2} + \theta_{15} \mathrm{dRhint}_{t-1} + \theta_{16} \mathrm{dRhint}_{t-2} + \theta_{17} \mathrm{dSPL}_{t-1} + \theta_{18} \mathrm{dSPL}_{t-2} + (\gamma_{11} \mathrm{dUVI}_t + \gamma_{12} \mathrm{dRHext}_t + \gamma_{13} \mathrm{dWS}_t \quad (14)$$

In (11) $\theta_{11}$ indicates the effect of the past percentage growth rate of the weight (between t-2 and t-1) on the current percentage growth rate (between t-1 and t), considering that both variables (dependent and independent) are expressed in terms of the difference of the logarithms. Similarly, $\theta_{12}$ expresses the effect on the current percentage growth rate of Weight of a two-period lagged (lag 2) percentage change in Weight. In the different models of the VAR system, $\theta_{11}$ and $\theta_{12}$ generally represent, respectively, the effect of the recent percentage growth rate of Weight and of the most distant percentage growth rate (of lag 2) on the percentage growth rate at time t (if the model is 11) or on the absolute change at time t (in models 12-14) of Tint, RHint and SPL. $\theta_{13}$ measures the percent variation that Weight undergoes (in model 11) in response to a unit change in temperature recorded in the previous period, and so on.

The estimated VAR models provided very different levels of significance in some cases (Table 4). For the response variable Weight (expressed in terms of differences of logarithms) it was observed a minimum $R^2$ value equal to 0.43 for Hive3 in site A up to a value of 0.87 for Hive3 in site C. With reference to dTint, $R^2$ goes from 0.24 in Hive3 of site B to a maximum value of 0.64 in Hive3 of site C. RHint shows $R^2$ values varying from 0.32 for Hive3 of site C to 0.77 for Hive3 in site A. Finally, dSPL shows $R^2$ values ranging from 0.07 for Hive3 in site A to 0.99 for Hive3 in site C. The high variability of $R^2$ values in the different sites and Hives suggests that uncontrollable factors have influenced the data collected, therefore it was considered to give more emphasis to the results with $R^2 \geq 0.5$. Table 4 presents in bold the parameters found to be significant and grey colored the models with $R^2 \geq 0.5$.

The VAR models estimates are linked to the information set considered which is a subset of the factors that influence the system; there would be unobservable and non-controllable factors that can influence the endogenous variables. Furthermore, relative to the same information set, there may be relationships that are established for time periods less than two hours; therefore, relationships that are found to be non-Granger-causal, would be causal for example if the observation had been made at one hour or half an hour.

The percentage growth rate weight in site A, was approximately 0.74 % in Hive1 and more than 1.03 % in Hive2, in Hive3 the estimated model for Weight shows $R^2 < 0.5$. Since the parameters $\theta_{11}$ *and* $\theta_{12}$ are the effect (11-14) of a variable expressed as difference of logarithms on a difference of logarithms, it is commented as the percentage variation produced on the response variable for a 1 % variation of the independent variable. In the model for Weight, we note that a 1 % growth rate of the weight in the previous period produces a growth of about 0.74 % in the current period (Hive1, site A). In general, all the models show that a positive growth rate in the previous period has a positive effect on the current period, while a positive growth rate in a more distant period (at the time lag of two previous periods) reduces the current growth rate. Therefore, the effects of positive variations that are more distant in time (for example $\theta_{12}$ or $\theta_{14}$) act on the system as adjustment variables. Furthermore, in the sites where there is a higher growth compared to the previous period, the adjustment or correction effect is higher (site A, Hive2, site B Hive1 and Hive2). The most important information provided in the Weight model is the percent increase rate of weight compared to previous periods, while the variations of the other explanatory variables (dTint, dRHint dSPL and dUVI) do not seem to have significant effects on the percentage change in weight; only the coefficient of WS ($\gamma_{13}$), in H2 site A and H2 site C, shows a positive and significant effect on the current weight change. Considering that coefficient $\gamma_{13}$ measures the effect of WS variation of 1 m/s on the percentage growth rate of weight, it is possible to say that a WS variation of 1 m/s produces a percent increase in weight equal to 1.1 % in Hive2 site A and 1.6 % in Hive2 site C, while it has a negligible negative effect (equal to -0.002 %) in in Hive1 site B. We recall that the effects of the exogenous variables are contemporary effects.

Continuing to observe the results of the model for Weight, it is noted that site B shows the highest percentage growth rates (respectively 1.01 % in Hive1, 0.98 % in Hive2 and 1.01 % in Hive3 for the 1 % growth recorded in the previous period). The percentage growth rates of more distant periods (t-2) also have an adjustment effect reducing the current growth by a percentage rate between approximately 0.14 % and 0.18 %.

Regarding internal temperature (Tint), in almost all the hives with $R^2 \geq 0.5$, it seems that in the 4 hours of observation the bees are able to regulate the temperature through a feedback with SPL. In other words, Tint influences SPL, which in turn influences Tint. RHint shows feedback with Tint and Weight dynamic effects for some sites and hives, but the external factors (UVI and RHext) are decisive in all the hives.

In the model for Tint the delayed effects of the variation in Tint itself ($\theta_{13}$, $\theta_{14}$) are significant at all sites and hives; the coefficient $\theta_{14}$ shows a corrective effect opposite to $\theta_{13}$. As regards the effects of the other lagged variables on dTint, it is noted that a 1 % increase in weight produces a reduction in the temperature variation of approximately 4.01 °C in Hive3 of site A, while more distant percent weight variations have a corrective effect on temperature (+4.04) adjusting the equilibrium of the system. In the hives with $R^2 \geq 0.5$, the UVI coefficient $\gamma_{11}$ was found to be significant with positive effects: a unit change in UVI causes a change in temperature ranging from 0.03 for Hive1 in site C to 0.12 for Hive2 in site A. Internal hive temperature is the primary indicator of a colony's health. The ability of a colony to thermoregulate is influenced by the subspecies and, within this, by the genetic diversity of the colony [27]. Bees maintain a temperature of 34 ± 1.5 °C near the brood [28]. The RH coefficient $\gamma_{12}$ has a negative effect on Tint (except for H3 in site B). On the other hand, the relationship between wind speed and internal temperature in site B is significantly positive. (coefficient $\gamma_{13}$).

The model for RHint shows significant effects of the Weight coefficients $\theta_{11}$ and $\theta_{12}$, in Hive2 site A and Hive2 site C, of Tint with $\theta_{13}$ and $\theta_{14}$ in Hive1 site A, in Hive1 and Hive2 site B, and Hive2 and Hive3 site C), of RHint ($\theta_{15}$ and $\theta_{16}$) in all sites and hives, whereas the effect of SPL ($\theta_{17}$ and /or $\theta_{18}$) is significant in Hive 2 site A. In all sites and hives, the simultaneous effects of UVI and RHext were significant ($\gamma_{11}$ and $\gamma_{12}$), showing and increase of RHint variation for a unit increment in the variation of UVI and RHext. WS shows a significant and negative effect only for Hive3 site B: an increase in WS unit variation produces a reduction of 0.81 in the absolute variation of RHint. Similarly to temperature, RH is a useful parameter for predicting swarming, as it has been observed that this event is preceded by a decrease in RH due probably to ventilations, consisting in a a rapid flitting of bee wings [29].

The model for SPL shows significant effects of weight (Hive1 and Hive2 site A; Hive2 site B), of Tint (Hive1 and Hive2 site A; Hive1 and Hive2 site B; Hive2 and Hive3 site C), of the recent effects of RHint (in Hive1 site A and Hive3 site C), of the delayed ones for RHint (Hive1 and Hive3 site B; Hive2 site C). The model also shows the SPL itself lagged effects showing some stability of the current absolute variations with the lagged ones in the more recent periods, and a compensating effect of the more distant variable in time. Among the external variables, RHext has a significant effect, even in size, with the highest value in Hive3 site B where a unit variation increase of RHext produces an increase in the variation of SPL equal to 0.42.

Having considered the same model allows us to verify if the system behaves in the same way at different sites. In all the estimated VARs, the Eigenvalues of the coefficients matrix of the models are strictly less than one, so, based on [21] and [30], the model shows stability.

The Granger-causality may not tell us the complete story about the interactions between the variables of a system, therefore Impulse-Response Functions (IRFs) observation can be useful. After fitting a VAR, IRFs show how the process reacts to a single shock at time t providing important information about how the shock propagates throughout the system, and what effect it has over time. The understanding of the dynamic response of a system to a small shock is of considerable interest in the field of beekeeping.

Based on the models for which resulted $R^2 \geq 0.5$, the IRFs plots are reported below (Fig. 7).

In other words, we look at the reaction of the system to a shock of each variable and see how long it takes for the system to settle back to its original functioning, based on the estimated VAR model.

In the three hives of site A, it is noted that an increase equal to 1 % in the standard deviation of SPL leads to an increase in Tint after 2 h. After 4 h the maximum increase in temperature is reached, then it begins to reduce after approximately 10 h but the system, here represented by the bees, takes about 14 h ($t = 7$) to realize that the temperature has dropped too much under the equilibrium level and it has to increase again to return to the equilibrium state (Fig. 7). It is worth remembering that Tint and SPL were found to be simultaneous, therefore they influence each other in the same period. The same behavior is observed for Hive 2 of site B, while in reference to Hive 3 of site C an opposite behavior is observed (consider the anomaly found in hive 3 of site C).

We can also note how in site B a shock equal to one of the standard deviation of RHint produces an increase of about 1 % of Tint in Hive1 and of about 6 % in Hive2. In both hives it takes about 4 h for the system to start activating to reduce the internal temperature but, while in Hive1 after 10 h the temperature drops below the equilibrium level and requires another 10 h to bring the system to the initial equilibrium state, in Hive2 the system returns to its initial equilibrium after 18–20 h. A shock in SPL has a discordant effect in the two hives, while in Hive1 it produces a significant reduction in temperature, in Hive2 the shock in SPL produces a slight increase in internal temperature but after about 4 h the system tries to return to its initial equilibrium by reducing the internal temperature. In both hives, weight shocks produce an immediate reduction in Tint, the system restores its usual dynamics after approximately 1 day.

In site C, the two hives where $R^2$ was>0.5 show, as the other sites, shocks in RHint have an immediate effect on Tint going from over 3 % (Hive1) to over 6 % (Hive3).

Tint impulse on RHint (Fig. 8) has a different effect in site A than the others. Here a simultaneous reduction of RHint is observed in Hive1 and Hive2 (about 40 % in 4 h). The opposite trend is observed in sites B and C, where a 1 % variation of the standard deviation of Tint produces an increase of RHint that goes from about 10 % in Hive1 site B to over 15 % in Hive2 site C. SPL produces a significant effect on RHint only in Hive1 and Hive2 site A, and Hive1 site B, with non-univocal behavior.

Figs. 9,10 and 11 show the forecast in the test set for the three hives in sites A, B and C using the fitted VAR model in the test period.

Fig. 9 shows that the model predicts well all the changes in the internal factors. The graph also shows how quickly the predictions from the one-lag model of the differenced series in the test set, settle down to their mean values. We can see that forecast work good in RHint Tint, SPL and Weight. It can be noting a little overestimation for Weight (in Hive1 and Hive2) likely because in the last month, used for forecast, production was decreasing, and the training of the model was done when production had an increasing trend.

In site B (Fig. 10) we observed predictions near observed values of the variables. SPL is those with the highest error in particular in Hive1 and Hive2, it captures mean but not cycle.

In site C, the SPL had in the test set, the worst prediction in Hive3. Forecast in the differenced SPL in Hive3 are close to the true values (in red) when forecast is short term but tend to be biased the longer the forecasting period is. Although RHint and Tint have predictions that converge on average, they have a larger error between observed values and predicted values related to detection problems in the system in the last month (Fig. 6) of the difference series for Hive3.

## 4. Conclusions

In this paper we identified the interrelationships between internal and external factors of the hive in three different sites, measured by a proper designed smart system. While the relationships with external factors (UVI, RHext, WS) measure the contemporary effects on endogenous variables, the internal factors allow us to measure the dynamic relationships. The data collected by the monitoring system of internal parameters of the hives allowed us to estimate the VAR models in the field of precision beekeeping. The significance recorded in the relationships between Weight and Tint and Weight and RHint and the good predictive capacity of the models considered with respect to Tint and RHint, allow us to build a predictive model about the hive behaviour.

Not all the external parameters showed a significant effect on production. In particular, neither the parameter UVI nor the external relative humidity had significant effects on the weight of the hives, while wind speed had a positive effect in terms of weight on the hives placed in the windiest sites. UVI showed, in almost all cases, a positive correlation with internal temperature. Conversely, external relative humidity showed negative correlation with internal hive temperature. However, in one of the sites studied, the relationship between wind speed and internal temperature was significantly positive. Regarding the effect of the external parameters on relative humidity inside the hive, the study highlighted significant positive effects in almost all the hives considered.

The correlations found between internal parameters and external parameters to the hive confirm the expected signs.

The impulse plots allowed us to monitor the effect and duration of a system shock on the variables of interest (Tint and RHint) and this could help us understand the level of the system response.

Time series over longer periods would significantly improve the predictive capacity of the model, allowing effective monitoring of the health conditions of the hive. The observations collected in the next years may be of help in increasing the explanatory power of the model.

## Ethics statement

Not applicable: This manuscript does not include human or animal research.

## CRediT authorship contribution statement

**Filippa Bono:** Writing – original draft, Validation, Methodology, Formal analysis, Data curation. **Mariangela Vallone:** Writing – review & editing, Writing – original draft, Supervision, Investigation, Data curation, Conceptualization. **Maria Alleri:** Writing – original draft, Visualization, Investigation, Formal analysis, Data curation. **Gabriella Lo Verde:** Validation, Investigation, Formal analysis, Conceptualization. **Santo Orlando:** Writing – original draft, Visualization, Validation, Investigation, Data curation. **Ernesto Ragusa:** Writing – original draft, Visualization, Investigation, Data curation. **Pietro Catania:** Validation, Supervision, Resources, Project administration, 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 work was supported by the Ministero delle politiche agricole, alimentari e forestali, Italy. [Bando Miele 2021, Project title: Prototipo di arnia intelligente per migliorare la produzione di miele. CUP: J79J210131900008].

The authors are grateful to Basile farm (Ventimiglia di Sicilia, Palermo, Italy) and Stabile farm (Castellammare del Golfo, Trapani, Italy) for giving the hives and hosting the experimentation.

## Data availability

Data will be made available on request.

## References

[1] A. Zacepins, E. Stalidzans, J. Meitalovs, Application of information technologies in precision apiculture, in: *Proceedings of the 13th International Conference on Precision Agriculture* (ICPA 2012) 7, 2012.
[2] A. Brini, E. Giovannini, E. Smaniotto, A Machine Learning Approach to Forecasting Honey Production with Tree-Based Methods, arXiv preprint arXiv:2304.01215, 2023.
[3] O. Anwar, A. Keating, R. Cardell-Oliver, A. Datta, G. Putrino, We-bee: Weight estimator for beehives using deep learning, in: AAAI Conference on Artificial Intelligence2022: 1st International Workshop on Practical Deep Learning in the Wild, 2022. February).
[4] H. Hadjur, D. Ammar, L. Lefèvre, Toward an intelligent and efficient beehive: a survey of precision beekeeping systems and services, Comput. Electron. Agric. 192 (2022) 106604.
[5] P. Catania, M. Vallone, Application of a precision apiculture system to monitor honey daily production, Sensors 20 (7) (2020) 2012.
[6] M. Alleri, S. Amoroso, P. Catania, G. Lo Verde, S. Orlando, E. Ragusa, M. Sinacori, M. Vallone, A. Vella, Recent developments on precision beekeeping: a systematic literature review, J. Agric. Food Res. (2023) 100726.
[7] M. Vallone, S. Orlando, M. Alleri, M.V. Ferro, P. Catania, Honey production with remote smart monitoring system, Chem. Eng. Trans. 102 (2023) 169–174.
[8] M.C. Robustillo, C.J. Pérez, M.I. Parra, Predicting internal conditions of beehives using precision beekeeping, Biosyst. Eng. 221 (2022) 19–29, https://doi.org/10.1016/j.biosystemseng.2022.06.006.
[9] A. Zacepins, V. Brusbardis, J. Meitalovs, E. Stalidzans, Challenges in the development of precision beekeeping, Biosyst. Eng. 130 (2015) 60–71.
[10] I.Z. Ochoa, S. Gutierrez, F. Rodríguez, Internet of things: low cost monitoring beehive system using wireless sensor network, in: *2019 IEEE International Conference on Engineering Veracruz* (ICEV) 1, IEEE, 2019, pp. 1–7.
[11] A.R. Braga, D.G. Gomes, R. Rogers, E.E. Hassler, B.M. Freitas, J.A. Cazier, A method for mining combined data from in-hive sensors, weather and apiary inspections to forecast the health status of honey bee colonies, Comput. Electron. Agric. 169 (2020) 105161.
[12] A.R. Braga, B.M. Freitas, D.G. Gomes, A.D. Bezerra, J.A. Cazier, Forecasting sudden drops of temperature in preoverwintering honeybee colonies, Biosyst. Eng. 209 (2021) 315e321.
[13] C.A. Sims, Macroeconomics and reality, Econ.: J. Econ. Soc. (1980) 1–48.
[14] C.W.J. Granger, Developments in the study of cointegrated economic variables, Oxf. Bull. *Econ. Stat.* 48 (1986) 213–228.
[15] R.M. Ueda, A.M. Souza, R.M.C.P. Menezes, How macroeconomic variables affect admission and dismissal in the Brazilian electro-electronic sector: a VAR-based model and cluster analysis, Phys. *Phys. A: Stat. Mech. Appl.* 557 (2020) 124872.
[16] C. Ziegler, R.M. Ueda, T. Sinigaglia, F. Kreimeier, A.M. Souza, Correlation of climatic factors with the weight of an Apis mellifera Beehive, Sustainability 14 (9) (2022) 5302.
[17] D.C. Montgomery, L.A. Johnson, J.S. Gardiner, Forecasting and Time Series Analysis, 2nd ed., New York: McGraw–Hill, 1990.
[18] D.B. Rubin, Multiple Imputation for Nonresponse in Surveys, New York: Wiley, 1987.
[19] R.J.A. Little, Missing-data adjustments in large surveys, J. Bus. Econ. Stat. 6 (1988) 287–296.
[20] N. Schenker, J.M.G. Taylor, Partially parametric techniques for multiple imputation, Comput. Stat. Data Anal. 22 (1996) 425–446.
[21] H. Lütkepohl, New Introduction to Multiple Time Series Analysis, Springer, New York, 2005.
[22] D.A. Dickey, W.A. Fuller, Distribution of the estimators for autoregressive time series with a unit root, J. Am. Stat. Assoc. 74 (1979) 427–431.
[23] S. Johansen, Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, Oxford University Press, Oxford, 1995.
[24] L. Li, C. Lu, W. Hong, Y. Zhu, Y. Lu, Y. Wang, S. Liu, Analysis of temperature characteristics for overwintering bee colonies based on long-term monitoring data, Comput. Electron. Agric. 198 (2022) 107104.
[25] S. Ferrari, M. Silva, M. Guarino, D. Berckmans, Monitoring of swarming sounds in bee hives for early detection of the swarming period, Comput. Electron. Agric. 64 (2008) 72–77.
[26] A. Kviesis, A. Zacepins, System architectures for real-time bee colony temperature monitoring, in: *Proceedings of the 29th European Conference on Solid-State Transducers* (EUROSENSORS 2015), Freiburg, Germany, 2015, 6–9 September.
[27] W.G. Meikle, N. Holst, Application of continuous monitoring of honeybee colonies, Apidologie 46 (1) (2015) 10–22.
[28] I.Z. Ochoa, S. Gutierrez, F. Rodríguez, Internet of things: low cost monitoring beehive system using wireless sensor network, in: 2019 IEEE International Conference on Engineering Veracruz (ICEV) 1, IEEE, 2019, pp. 1–7.
[29] S. Ferrari, M. Silva, M. Guarino, D. Berckmans, Monitoring of swarming sounds in bee hives for early detection of the swarming period, Comput. Electron. Agric. 64 (1) (2008) 72–77.
[30] J.D. Hamilton, Time Series Analysis, Princeton university press, 2020.