Multivariate Time-Series Forecasting of Beehive Microclimate Parameters Using Bayesian Vector Autoregression

Punith Kumar
Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bengaluru, Karnataka, India
punithkumar87@gmail.com (corresponding author)

H. N. Champa
Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bengaluru, Karnataka, India
champahn@yahoo.co.in

Received: 31 December 2025 | Revised: 4 February 2026 | Accepted: 11 February 2026

 DOI: https://doi.org/10.48084/etasr.17270

ABSTRACT

Enhancing precision in beekeeping involves integrating technology and statistical models to assess honeybee health and mitigate the risk of colony loss. Monitoring the rate of hive population decline is a crucial tool for bee health management, providing early warnings of potential abnormalities affecting colonies. Through the we4bee project, data on humidity, temperature, and weight were collected using interior sensors placed inside the beehive. These datasets were analyzed to predict internal hive variables using Vector Autoregressive (VAR) and statistical models. This work introduces a Bayesian VAR (BVAR) model, a new framework that applies Bayesian statistics to improve the VAR model by integrating prior information, managing uncertainty, and enhancing parameter estimation. A comparative analysis of time-series data, performed with 75-fold cross-validation, showed that the new BVAR model yielded the most accurate predictions and required less computation time. Given the need for accurate predictive models, these approaches could help beekeepers prevent hive collapse and improve overall hive management.

Keywords-Bayesian inference; vector autoregression; beehive microclimate; rolling window approach

I. INTRODUCTION

Beehives provide a secure and organized environment for bees to flourish, offering many benefits, including honey, pollination, and beeswax production [1]. The benefits extend beyond the beekeeper, benefiting agriculture, the economy, and the environment by supporting biodiversity. Beekeeping plays a crucial role in nature, making beehives an important tool for maintaining a healthy ecosystem and food supply. Bees are one of the most efficient pollinators for crops, flowers, and plants. To improve crop yields, beekeepers can use hives to encourage bees to pollinate fields. Beekeepers can support the health and growth of bee colonies by providing a controlled environment. However, bee populations have recently declined due to disease, habitat loss, and pesticide use [2]. As a result, beehives provide opportunities for research and education about the biology, ecology, and behavior of bees. Scientists and beekeepers can study the behaviors, interactions, and ecological roles of bees. Pollination is essential for agriculture, food supply, and biodiversity in European countries. The former is also significant for increasing crop yield and quality in many European crops. An estimated 35% of food crops and 75% of flowering plants rely entirely on insect pollination. Therefore, pollination is crucial for both the quantity and quality of many agricultural products. The economic importance of pollination services in Europe is significant. The total annual value of pollination services is estimated at 22 billion euros [3]. This includes direct contributions from agricultural production, as well as wider effects on biodiversity, ecosystem health, and tourism. During hibernation, bees lose up to 30% of their colonies in some European countries and between 8% and 30% worldwide. Preserving pollinators is important for human survival. Authors in [4] explored how computer vision can be used to analyze colony activity through images. In the pursuit of effective management of beekeeping colonies, the concept of precision beekeeping has emerged [4]. Decision-making frameworks based on fuzzy logic have been employed for control purposes. This study employs various predictive models to analyze beekeeping sensor data to forecast hive temperature, humidity, and weight.

Authors in [5] examined long-term sensor data from native Iberian beehives to investigate internal microclimate control under different climatic conditions. Using CATPCA and

CATREG models, they demonstrated that brood-nest temperature stays consistently regulated around 34 °C, whereas internal humidity and hive weight are more affected by external climate, emphasizing the need for dependable predictive models for beehive microclimate parameters. Authors in [6, 7] examined how hive temperature and humidity affect Varroa destructor dynamics through ecological and statistical analyses. Their findings demonstrated that specific microclimate ranges significantly influence parasite growth and resistance patterns, suggesting that accurate prediction of internal hive conditions can facilitate early identification of environments prone to disease. Authors in [8] studied how microclimate, habitat features, and beekeeping methods affect honey yield across different areas, showing that higher temperatures and invasive plants decreased production, while traditional practices and more mother colonies greatly increased it. Authors in [9, 10] developed IoT-enabled monitoring systems that continuously tracked temperature, humidity, and hive weight using sensor data. Model evaluation using Mean Absolute Error (MAE) and accuracy metrics demonstrated enhanced reliability and quicker detection of adverse hive conditions compared to manual monitoring methods. Authors in [11] used deep learning models, including Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures, to predict microclimatic variables in beehives. Using MAE and Root Mean Square Error (RMSE) for evaluation, their results showed higher prediction accuracy than traditional statistical models, especially under non-linear and seasonal conditions, but with increased computational complexity and less interpretability.

In this study, a Bayesian VAR model is presented that uses Bayesian priors, such as normal and Wishart priors, which shrink coefficients toward zero, thereby reducing overfitting in regular VAR models. To address unstable estimates arising from the lack of a structured approach to parameter estimation in the VAR model, Bayesian inference is employed to explicitly quantify uncertainty by modeling the posterior distribution of the parameters.

II. METHODOLOGY

Bio-Inspired Algorithms, such as the Artificial Bee Colony (ABC) Algorithm, have been widely used for optimization problems due to their efficiency and adaptability, especially in complex, non-linear search spaces [12].

A. Preprocessing

Before being input into the predictive models, we4bee Researcher App dataset [13], which consists of temperature, humidity, and weight of beehives in two different locations, underwent preprocessing. Invalid values resulting from sensor malfunctions or any data collection errors are excluded. These are identified as outside the plausible range (e.g., negative weights or unrealistic humidity or temperature values). Missing data, on the other hand, are imputed using the MICE-Ranger method, which employs an iterative predictive modeling process to estimate the missing values.

B. Models

Let T denote the time at which n variables are observed, and $y_{t,} x_t$, and c_t represent the response, predictor, and covariate variable vectors, respectively. If a vector is univariate, it is written without boldface notation, as $y_{t,}$ $x_{t,}$ and c_t.

1) Vector Autoregressive Model

This is a robust statistical technique for identifying linear relationships among multiple time series [13]. The Vector Autoregressive (VAR) model of order p is defined as:

$$y_t = \nu + A_1 y_{\{t-1\}} + \cdots + A_p y_{\{t-p\}} + C\, c_t + e_{t,} \quad t = p+1, \ldots, \mathrm{T} \tag{1}$$

where $A_1 \ldots A_p$ are the fixed coefficient matrices, ν represents a vector of intercept terms. The covariates are associated with the coefficient matrix C. The error term e_t is a vector with an expected value of zero, i.e. $E(e_t) = 0$, and a positive definite covariance matrix given by $E(e_t e_t') = \sigma$ indicates a white noise process.

2) Bayesian VAR Model

The Bayesian VAR (BVAR) model framework, illustrated diagrammatically in Figure 1, is created by applying Bayesian statistics to VAR models to enhance prediction accuracy. BVAR incorporates prior knowledge or beliefs about the parameters into the model, helping to regularize estimates and reduce over-fitting. In a Bayesian framework, the parameters of the VAR model are considered random variables with prior distributions [14]. The BVAR model uses Normal-Wishart priors, in which the regression coefficients have a Gaussian prior, and the precision matrix follows a Wishart distribution. Shrinkage is applied via a diagonal prior covariance matrix to stabilize multivariate coefficient estimates when working with limited rolling-window samples. The lag order is chosen within each training window using the Bayesian Information Criterion (BIC) and remains fixed for the corresponding forecast horizon. The goal is to derive the posterior distribution of the parameters based on the data.

$$P(\theta|data) \propto P(data|\theta)P(\theta) \tag{2}$$

where $P(\theta|data)$ is the posterior distribution of parameters, $P(data|\theta)$ is the likelihood of the data, derived from the VAR model, and $P(\theta)$ is the prior distribution, encapsulating prior beliefs about the parameters. It assumes normal priors for coefficients and a Wishart prior for the error covariance matrix. Variational Inference approximates the posterior with a simpler distribution, enabling faster computation [15]. Forecasts are generated by sampling future values from the posterior predictive distribution:

$$P(y_{T+h}|y_1, \ldots, y_T) = \int P(y_{T+h}|\theta)P(\theta|data)d\theta \tag{3}$$

3) Other Models

Several statistical and machine learning models are used for time series prediction to enhance accuracy and uncover underlying temporal patterns. Among these, Dynamic Linear Models (DLM), Generalized Additive Models (GAM), and Time-Varying Vector Autoregression (tvVAR) are well known for their flexibility and effectiveness in modeling complex dependencies in sequential data. Missing values are imputed using the MICE-Ranger procedure (multiple imputation via chained random forests), consistent with the preprocessing stage.

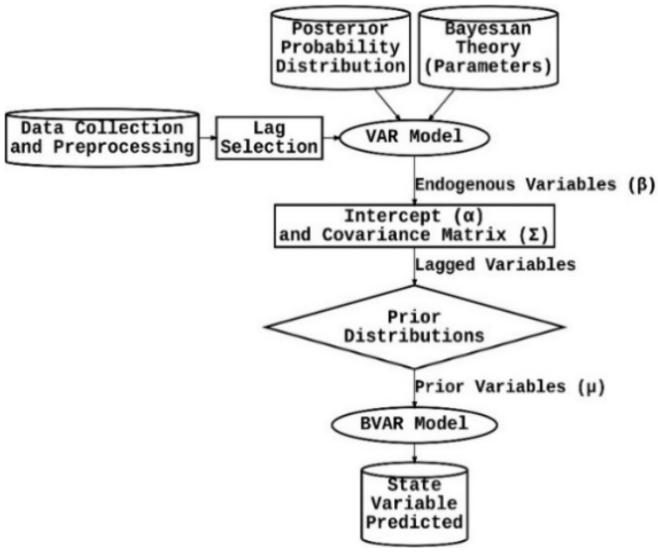

Fig. 1. Block diagram of the BVAR model.

4) Cross-Validation

The first step in the cross-validation process is to divide the time series into training and test sets [16]. The training set is used to fit the model, and γ-step-ahead predictions are generated for the test set. Errors are then calculated. The estimation and prediction process is repeated on the updated dataset after the training set has been advanced by a predetermined increment δ. This iterative process continues until further γ-step predictions are no longer feasible or until a set number of k iterations (also known as k-fold cross-validation) is completed.

III. RESULTS AND DISCUSSION

A. Experimental Setup

The we4bee project data consist of temperature, humidity, and weight measurements collected over 18 months from beehives at Schulgarten and Gymnasium LNU in Germany. A total of 2615 values for each variable were considered. The 75-fold rotation estimation was performed using a rolling window with a step size of δ=5, along with 25, 50, and 75 γ-steps for the 5-, 10-, and 15-day forecasts, respectively. Two sensors were placed inside the hives to collect temperature and humidity data. Similarly, weight data were collected using a single sensor placed below the hive. Data for both Schulgarten and Gymnasium LNU locations were collected from June 2023 to November 2024, covering all four seasons, followed by a subsequent cycle that included an additional season, as shown in Figure 2.

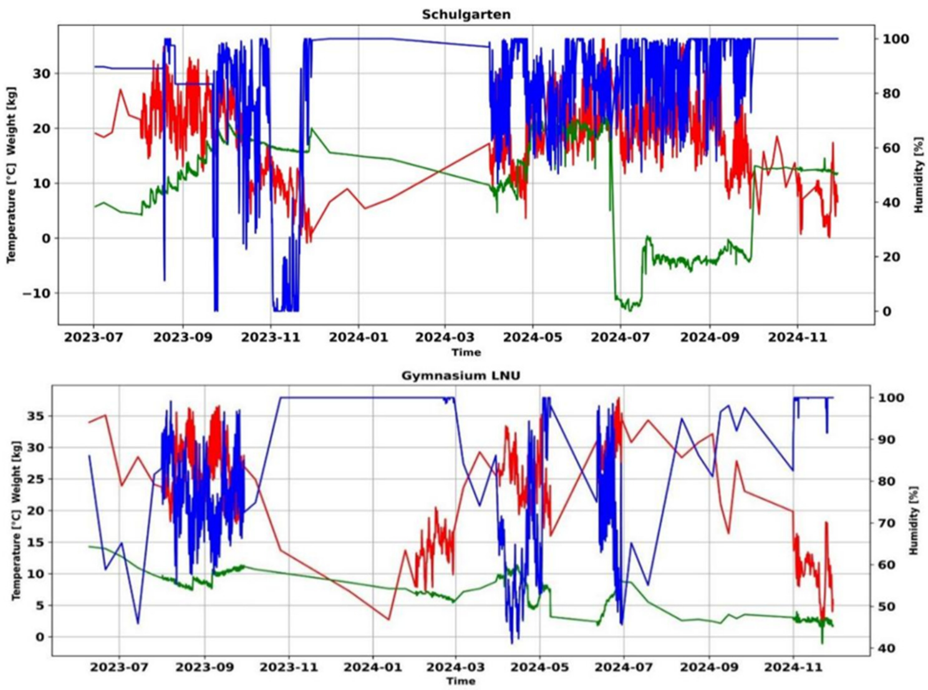

Fig. 2. Time series of humidity (blue), temperature (red), and weight (green) for two colonies, recorded from June 2023 to November 2024.

B. Experimental Results

The MAE of different models is depicted in Table I, expressed as mean ± standard deviation. The errors are fairly similar within each model. Table II displays the computation time and the ratio of computation time for each model. The BVAR model is more computationally efficient than the other five models in fitting the variables of Schulgarten and Gymnasium LNU hives. BVAR enhances humidity and temperature forecasting because these variables show strong contemporaneous and lagged cross-dependencies, which are effectively captured through multivariate modeling and shrinkage regularization.

TABLE I. TABLE I. COMPARISON OF MAE ACROSS MODELS FOR DIFFERENT VARIABLES

Hive	Variables	VAR	TvVAR	DLM	GAM	BVAR
Schulgarten	Humid	2.577 ± 1.656	2.229 ± 1.849	2.529 ± 1.895	2.443 ± 1.803	2.200 ± 1.800
	Temp	1.080 ± 0.688	0.307 ± 0.343	0.480 ± 0.386	0.508 ± 0.409	0.300 ± 0.334
	Weight	0.353 ± 0.282	0.210 ± 0.248	0.334 ± 0.377	0.412 ± 0.412	0.250 ± 0.220
Gymnasium LNU	Humid	0.683 ± 1.691	0.319 ± 1.067	0.250 ± 0.853	0.456 ± 1.097	0.229 ± 0.787
	Temp	3.188 ± 2.577	2.508 ± 2.311	0.945 ± 1.105	1.152 ± 1.210	0.922 ± 1.010
	Weight	0.522 ± 0.662	0.432 ± 0.541	0.435 ± 0.336	0.459 ± 0.344	0.500 ± 0.500

TABLE II. COMPARISON OF TIME (IN MIN) AND RATIO ACROSS MODELS FOR SCHULGARTEN AND GYMNASIUM LNU

Models	Schulgarten		Gymnasium LNU	
	Time (min)	Ratio	Time (min)	Ratio
VAR	2.7134	259.578	2.4715	102.340
tvVAR	19.4480	1860.438	19.9472	825.991
DLM	0.5118	48.962	0.7608	31.505
GAM	0.2187	20.926	0.3864	16.003
BVAR	0.0104	1.000	0.0241	1.000

In Tables III and IV, the MAE for 5-, 10-, and 15-day forecasts is provided for each model at two locations, respectively, based on 75-fold cross-validation. The rolling window size is set to 5. Figure 3 shows the distribution of the MAE outlined in Tables III and IV. The tvVAR model is shown to be the top performer in weight prediction across both locations. Figure 4 presents the forecasted values over a 15-day span, with actual data points marked as dots and model predictions depicted as lines. Compared with previous sensor-driven studies that analyze hive microclimate relationships but do not focus on multivariate probabilistic forecasting [5-8], the proposed BVAR explicitly models the interdependencies among temperature, humidity, and weight, providing uncertainty-aware forecasts. Relative to deep recurrent models (e.g., LSTM/GRU), which can achieve strong accuracy but often require higher compute and offer limited interpretability [11], BVAR provides an interpretable multivariate structure and extremely low training time (Table II) while achieving competitive or superior MAE for humidity/temperature across horizons (Tables I, III, IV). Variational inference is used instead of Markov Chain Monte Carlo because it requires repeated model training across 75 rolling-window folds and multiple forecast horizons. Full MCMC sampling becomes too computationally expensive in this context.

TABLE III. MAE OF 5, 10, 15 DAY FORECASTS FOR SCHULGARTEN

Schulgarten	Days	VAR	TvVAR	DLM	GAM	BVAR
Humidity	5	2.530 ± 0.537	2.469 ± 1.526	3.000 ± 1.809	3.214 ± 1.718	2.428 ± 0.588
	10	2.933 ± 0.838	2.779 ± 1.501	3.677 ± 1.269	3.881 ± 1.621	2.851 ± 0.809
	15	3.247 ± 0.893	2.990 ± 1.988	3.739 ± 1.659	3.968 ± 1.383	2.449 ± 0.949
Temp	5	0.712 ± 0.532	0.644 ± 0.520	0.883 ± 0.466	0.916 ± 0.599	0.455 ± 0.502
	10	0.965 ± 0.438	0.879 ± 0.988	1.157 ± 0.445	1.380 ± 0.706	0.877 ± 0.402
	15	1.261 ± 0.423	1.175 ± 0.842	1.380 ± 0.466	1.469 ± 0.664	1.101 ± 0.136
Weight	5	2.250 ± 1.156	2.007 ± 1.401	3.126 ± 2.170	7.855 ± 0.359	2.630 ± 1.767
	10	2.417 ± 1.133	2.170 ± 1.360	3.083 ± 2.131	8.042 ± 0.320	2.989 ± 2.897
	15	3.949 ± 1.141	3.514 ± 1.220	4.205 ± 2.244	8.139 ± 0.369	3.778 ± 2.833

TABLE IV. MAE OF 5, 10, 15 DAY FORECASTS FOR GYMNASIUM LNU

Gymnasium LNU	Days	VAR	TvVAR	DLM	GAM	BVAR
Humidity	5	1.026 ± 0.009	0.723 ± 1.185	1.632 ± 2.652	1.369 ± 0.710	0.694 ± 0.923
	10	1.125 ± 1.198	1.048 ± 1.206	2.566 ± 3.668	2.082 ± 2.437	1.043 ± 1.070
	15	2.656 ± 2.596	2.182 ± 2.472	3.968 ± 3.223	3.921 ± 2.331	2.096 ± 2.437
Temp	5	2.997 ± 2.628	2.794 ± 1.978	3.089 ± 2.742	4.981 ± 2.390	2.603 ± 1.120
	10	5.805 ± 3.393	3.688 ± 2.116	4.300 ± 2.145	5.753 ± 2.097	3.387 ± 2.240
	15	6.388 ± 2.895	4.552 ± 2.200	6.720 ± 2.942	7.364 ± 2.823	4.514 ± 2.109
Weight	5	0.315 ± 0.073	0.177 ± 0.099	0.193 ± 0.087	0.211 ± 0.055	0.238 ± 0.313
	10	0.209 ± 0.136	0.159 ± 0.087	0.216 ± 0.109	0.251 ± 0.087	0.290 ± 0.313
	15	0.255 ± 0.148	0.152 ± 0.091	0.190 ± 0.104	0.237 ± 0.081	0.211 ± 0.333

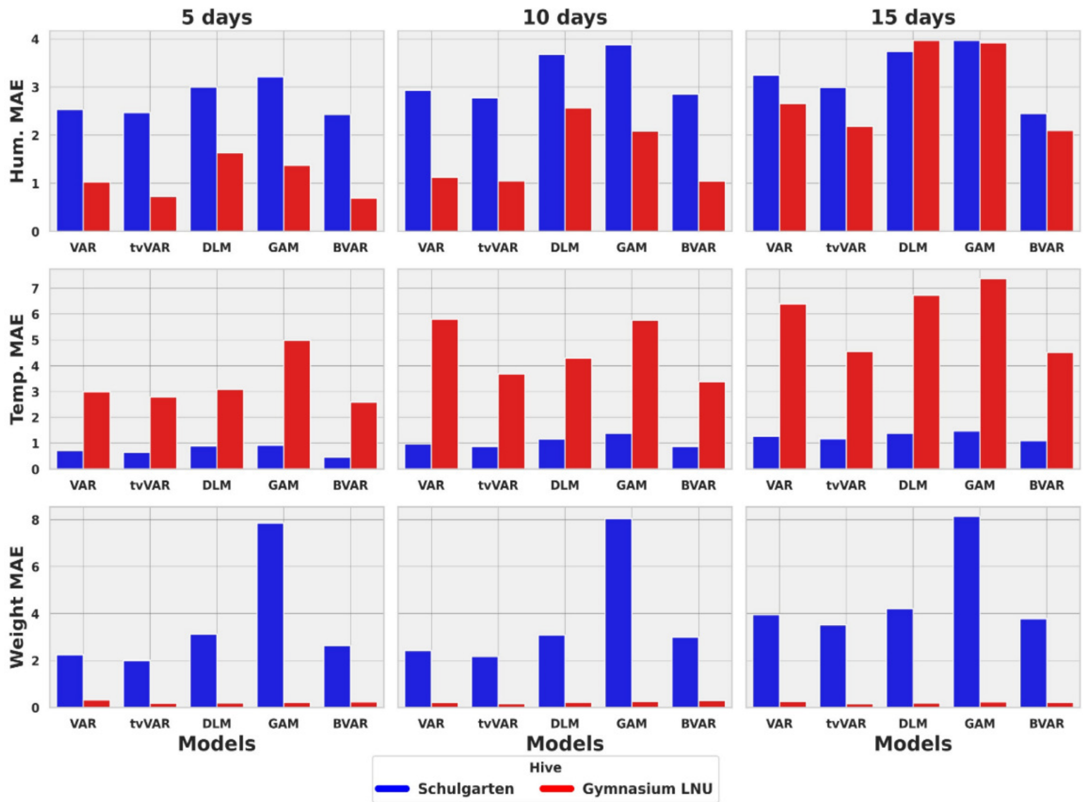

Fig. 3. Distribution of MAE values for 5, 10, and 15 day forecasts of three variables at two locations.

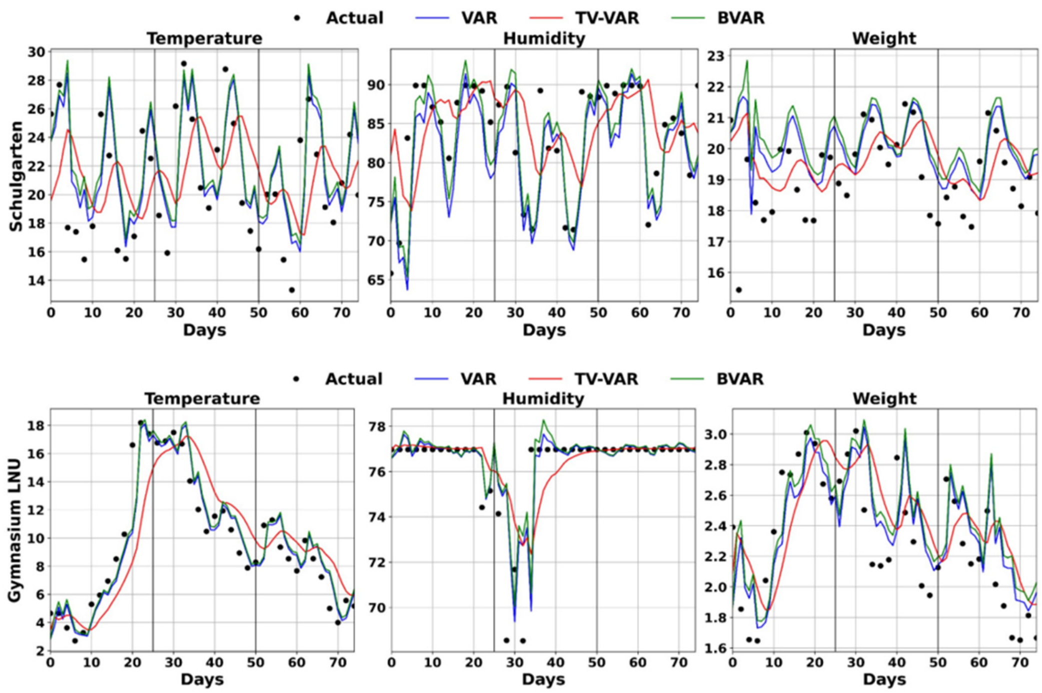

Fig. 4. Forecasts for the last 15 days at both locations. Vertical lines mark the 5, 10, and 15 day limits; the x axis shows 75 forecasts (5 per day).

IV. CONCLUSIONS

This study examines the effectiveness of different predictive models for analyzing beehive microclimates, focusing on Bayesian Vector Autoregression (BVAR) and other time-series models. The results show that BVAR offers a probabilistic framework that improves interpretability and uncertainty measurement, making it especially useful for small to medium-sized datasets. When it comes to temperature and humidity variables, BVAR performs better than the Vector Autoregression (VAR) and time-varying Vector autoregression (tvVAR) models; however, for the weight variable, tvVAR yields a lower Mean Absolute Error (MAE). Additionally, BVAR enhances hive event detection while achieving higher predictive accuracy. Models like BVAR, Generalized Additive Models (GAM), and Dynamic Linear Models (DLM) perform the best when working with small datasets where interpretability is essential. The basic VAR model took less than 1 min to fit the variables over 16 months of data. The proposed BVAR model took slightly less than 1 min (0.01 minutes) but achieved high prediction accuracy, reducing MAE to 0.30, 2.22, and 0.25 from 0.89, 4.9, and 0.44 for temperature, humidity, and weight, respectively. Overall, the obtained conclusions emphasize the importance of choosing models that fit the data volume, computational efficiency, and interpretability needs. Unlike traditional VAR and tvVAR models that offer only point forecasts, this study introduces a BVAR–based forecasting framework for beehive microclimate that combines Normal–Wishart shrinkage priors with variational Bayesian inference to facilitate rapid rolling-window training and probabilistic predictions. The proposed approach explicitly models cross-variable dependencies among temperature, humidity, and weight, and provides posterior predictive uncertainty estimates, thereby supporting risk-aware decision-making in precision beekeeping and the early detection of abnormal hive conditions.

REFERENCES

[1] A. Ilgun and T. Schmickl, "Mycelial Beehives of HIVEOPOLIS: Designing and Building Therapeutic Inner Nest Environments for Honeybees," *Biomimetics*, vol. 7, no. 2, June 2022, Art. no. 75, https://doi.org/10.3390/biomimetics7020075.

[2] G. Singh and A. Rana, "Honeybees and colony collapse disorder: understanding key drivers and economic implications," *Proceedings of the Indian National Science Academy*, vol. 91, no. 3, pp. 750–766, Sept. 2025, https://doi.org/10.1007/s43538-025-00399-x.

[3] E. Gazzea, P. Batáry, and L. Marini, "Global meta-analysis shows reduced quality of food crops under inadequate animal pollination," *Nature Communications*, vol. 14, no. 1, July 2023, Art. no. 4463, https://doi.org/10.1038/s41467-023-40231-y.

[4] S. Haddaoui, N. Khlifa, S. Chikhi, S. Varastehpour, and F. Adjailia, "A Comprehensive Review of Beekeeping Datasets for Precision Apiculture Research," in *2024 10th International Conference on Control, Decision and Information Technologies (CoDIT)*, July 2024, pp. 2043–2048, https://doi.org/10.1109/CoDIT62066.2024.10708308.

[5] S. Gil-Lebrero, F. J. Navas González, V. Gámiz López, F. J. Quiles Latorre, and J. M. Flores Serrano, "Regulation of Microclimatic Conditions inside Native Beehives and Its Relationship with Climate in Southern Spain, " *Sustainability*, vol. 12, no. 16, Aug. 2020, Art. no. 6431, https://doi.org/10.3390/su12166431.

[6] F. Nazzi and Y. Le Conte, "Ecology of *Varroa destructor* , the Major Ectoparasite of the Western Honey Bee, *Apis mellifera*," *Annual Review of Entomology*, vol. 61, no. 1, pp. 417–432, Mar. 2016, https://doi.org/10.1146/annurev-ento-010715-023731.

[7] J. Stara, S. Pekar, M. Nesvorna, M. Kamler, I. Doskocil, and J. Hubert, "Spatio-temporal dynamics of Varroa destructor resistance to tau-fluvalinate in Czechia, associated with L925V sodium channel point mutation," *Pest Management Science*, vol. 75, no. 5, pp. 1287–1294, May 2019, https://doi.org/10.1002/ps.5242.

[8] N. A. Barahona *et al.*, "Understanding how environmental degradation, microclimate, and management shape honey production across different spatial scales," *Environmental Science and Pollution Research*, vol. 31, no. 8, pp. 12257–12270, Jan. 2024, https://doi.org/10.1007/s11356-024-31913-1.

[9] V. Gaikwad, A. Amune, V. Rajput, V. Musale, N. Rajas, and S. Kakade, "Smart Beehive Monitoring System using IoT," in *2024 Second International Conference on Advances in Information Technology (ICAIT)*, July 2024, vol. 1, pp. 1–10, https://doi.org/10.1109/ICAIT61638.2024.10690535.

[10] R. K. Singh, T. Singh, and A. Arulmurugan, "Apis Insight for Phenotype Classification and Hive Health Forecasting using IoT and Deep Learning," in *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*, Apr. 2024, vol. 1, pp. 1–9, https://doi.org/10.1109/ICKECS61492.2024.10616417.

[11] N. H. K. Anuar, M. A. M. Yunus, M. A. Baharudin, S. Ibrahim, S. Sahlan, and M. Faramarzi, "An assessment of stingless beehive climate impact using multivariate recurrent neural networks," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 2, pp. 2030–2039, Apr. 2023, https://doi.org/10.11591/ijece.v13i2.pp2030-2039.

[12] N. T. Linh and P. V. Long, "Optimizing the Location and Capacity of DGs and SOPs in Distribution Networks using an Improved Artificial Bee Colony Algorithm," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 15171–15179, Aug. 2024, https://doi.org/10.48084/etasr.7665.

[13] "We4bee." https://analytics.we4bee.org

[14] S. V. Oskin, D. A. Ovsyannikov, and I. N. Shishigin, "Modeling Beehive Microclimate at the End of Wintering," *Biophysics*, vol. 67, no. 1, pp. 85–91, Feb. 2022, https://doi.org/10.1134/S0006350922010122.

[15] C. Ziegler, R. M. Ueda, T. Sinigaglia, F. Kreimeier, and A. M. Souza, "Correlation of Climatic Factors with the Weight of an Apis mellifera Beehive," *Sustainability*, vol. 14, no. 9, Apr. 2022, Art. no. 5302, https://doi.org/10.3390/su14095302.

[16] J. Degenfellner and M. Templ, "Modeling bee hive dynamics: Assessing colony health using hive weight and environmental parameters," *Computers and Electronics in Agriculture*, vol. 218, Mar. 2024, Art. no. 108742, https://doi.org/10.1016/j.compag.2024.108742.