Empirical Dynamic Modeling of Beehive Coordination Dynamics

Djourdan Gomes-Johnson
TiU CSAI / snr: 2033032
d.a.e.d.s.g.johnson@tilburguniversity.edu

Abstract

This study investigates the internal dynamics of a sensor-instrumented beehive by applying empirical dynamic modeling (EDM) to high-resolution time series data. Focusing on three key signals: hive power, humidity, and temperature, this analysis assesses intrinsic predictability, nonlinearity, and potential causal relationships. Using a combination of Simplex projection, sequential locally weighted global linear maps (S-Map), Convergent Cross Mapping (CCM), and multiview embedding, the study reveals distinct dynamic roles for each signal. Hive power exhibited strong internal regularity (ρ = 0.993), consistent with its role as a proxy for colony-level metabolic activity. Humidity, by contrast, demonstrated lower self-predictability (ρ = 0.63) and showed only a modest, variable dependence on hive power (ρ = 0.66), suggesting that it is more susceptible to external environmental influences and less governed by internal colony dynamics. Temperature dynamics were internally predictable (p = 0.991), but contributed minimally to predicting hive power. Multiview embedding consistently improved forecasting skill across variables. These results highlight the utility of EDM in uncovering asymmetrical coupling, attractor behavior, and internal regulation within biological systems, providing new insights into the coordination strategies of beehives.

Keywords: EDM, Honeybee Sensor Data, Nonlinear Time Series Analysis

1 Introduction

Bee colonies represent paradigmatic examples of complex biological systems, where environmental regulation and collective behavior emerge through decentralized coordination (Camazine et al., 2001). Within this dynamic framework, hive power serves as a macroscopic proxy for colony-level metabolic output, encompassing both thermoregulatory and communicative processes (Bencsik et al., 2011). In parallel, environmental conditions such as humidity and micro-vibrational patterns provide windows into both external perturbations and internal response dynamics (Meikle et al., 2015), (Ramsey et al., 2020).

Recent advances in sensor technology have enabled continuous monitoring of beehive environments, offering high-resolution time series data of mechanical and environmental variables (Wario et al., 2015). While traditional time series methods often assume stationarity or linear dependence, such assumptions fall short in capturing the intrinsic nonlinearity and time-varying nature of biological coordination (Sugihara et al., 2012). Addressing this gap, this study applies tools from empirical dynamic modeling (EDM)—a framework rooted in state space reconstruction and attractor-based analysis—to investigate the coupling and causal structure between hive power, humidity, and vibrational signals.

Building on previous studies that utilized phase-based analyses (e.g., Hilbert transform and continuous relative phase) to characterize coordination in environmental signals (Lamb and Stöckl, 2014), we adopt an alternative yet complementary approach that reconstructs system behavior directly from observed time series data. The primary goals are twofold: (1) to assess the intrinsic predictability and complexity of humidity and vibrational signals, and (2) to evaluate the dynamic influence of hive power on these variables. In doing so, this paper aims to uncover signal interdependencies that reveal the structure and coordination mechanisms underpinning beehive dynamics.

2 Methods

2.1 Dataset and Variables

This study utilized the `D1_sensor_data` dataset, comprising approximately 960,809 time-stamped observations collected from sensor-instrumented beehives. The dataset includes more than 30 variables representing environmental conditions and internal hive metrics. The analysis focused on four key variables:

`hive_power` serves as a proxy for colony-level metabolic activity and thermoregulatory output, while `humidity` reflects the internal microclimate critical for brood development and colony homeostasis. `temperature` captures external or ambient thermal influence, contributing to assessments of regulatory feedback mechanisms.

2.2 Preprocessing

To ensure comparability across variables and prepare the data for empirical dynamic modeling, all time series were normalized using z-score transformation:

$$\text{Normalized Value} = \frac{\text{Value} - \mu}{\sigma}$$

where μ and σ represent the mean and standard deviation of each variable. The following normalized signals were used throughout the analysis: `hive_power_norm`, `humidity_norm`, and `temperature_norm`. A contiguous 1000-observation subset (rows 10,000 to 11,000) was extracted from the full dataset to ensure computational feasibility and facilitate detailed state-space exploration. The selected window contained no missing values.

2.3 Empirical Dynamic Modeling (EDM)

Empirical dynamic modeling (EDM) techniques, implemented via the `rEDM` package in `R`, were used to reconstruct underlying system dynamics and evaluate inter-variable dependencies. EDM reconstructs attractor manifolds using time-delay embeddings and does not rely on parametric assumptions, making it suitable for nonlinear ecological systems (**?**).

2.3.1 Simplex Projection

Simplex projection was used to evaluate the intrinsic predictability of each individual time series. Using lagged coordinates, embedding dimensions E were varied from 2 to 8. Each configuration was assessed by computing the Pearson correlation coefficient (ρ), mean absolute error (MAE), and root mean square error (RMSE) between predicted and observed values. The dataset was divided into a training library (observations 1–500) and a prediction interval (observations 501–1000). The optimal embedding dimension was determined by the E value that maximized ρ.

2.3.2 S-Map Analysis

The S-Map (sequential locally weighted global linear map) algorithm was applied to examine system nonlinearity and estimate time-varying interaction strengths. S-Map fits locally linear approximations across the reconstructed manifold, modulated by the nonlinearity parameter θ. Values of θ were tested from 0.0 to 2.0 in increments of 0.1.

Two modeling configurations were implemented: (1) univariate S-Map, where predictions for each variable were based solely on its own lagged history; and (2) multivariate S-Map, which incorporated `hive_power_norm` as an additional input to assess its causal influence. Coefficients derived from S-Map regressions were interpreted as time-varying partial derivatives, offering insight into the dynamic influence of hive power on target variables such as humidity and vibration.

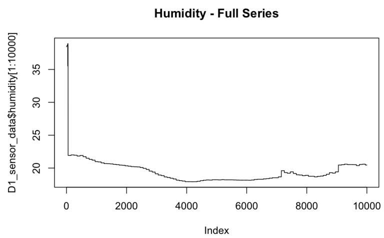
2.4 Model Evaluation and Software/Tools

Model performance was evaluated using standard forecast diagnostics (p, MAE, RMSE). Visualization, conducted using the `ggplot2` package, included time series plots comparing observed and predicted values, p curves across embedding dimensions (Simplex), forecast accuracy across nonlinearity values θ (S-Map), and temporal trajectories of S-Map coefficients reflecting interaction strength and variability.

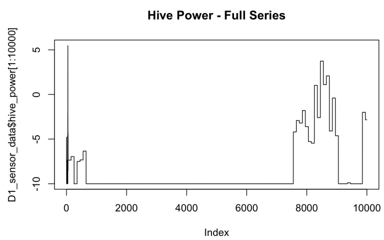
All computations were performed in `R` using the following libraries: `rEDM`, `tidyverse`, `ggplot2`, and `Kendall`.

3 Results

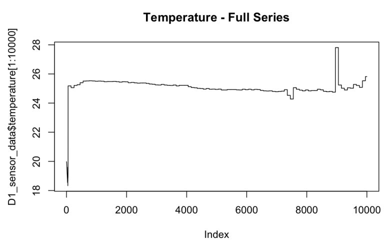
3.1 Dataset Overview



(a) Humidity (normalized)



(b) Hive Power (normalized)



(c) Temperature (normalized)

Figure 1: Normalized time series data for key internal and environmental signals.

The `D1_sensor_data` dataset consists of 960,809 rows of time-stamped measurements from instrumented beehives. Recorded metrics include hive activity (e.g., `hive_power`), and environmental parameters (e.g., `temperature`, `humidity`). A subset of 1,000 consecutive, complete observations was selected for analysis following normalization.

3.2 Univariate Predictability: Humidity

Intrinsic predictability of the `humidity_norm` signal was evaluated using Simplex projection. Forecasting accuracy peaked at embedding dimension $E = 6$ with a Pearson correlation coefficient of $p = 0.63$, mean absolute error (MAE) of 0.148, and root mean square error (RMSE) of 0.183. These values indicate moderate self-predictability, likely reflecting both internal regulation and influence from unmeasured environmental drivers.

Further evaluation with univariate S-Map revealed only a slight gain in forecasting skill across nonlinearity parameters, with a maximum $p = 0.64$ near $\theta = 0.5$. This weak response to varying θ suggests limited nonlinear structure and supports the interpretation of externally modulated or smooth, near-linear dynamics.

3.3 Univariate Predictability: Temperature

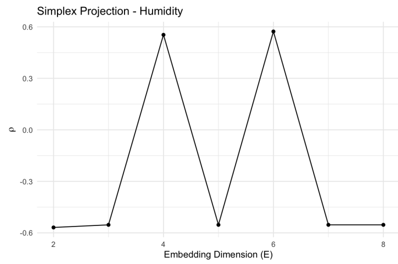
The `temperature_norm` signal exhibited strong deterministic structure under Simplex projection. The optimal embedding dimension was $E = 6$, producing $p = 0.991$, MAE of 0.035, and RMSE of 0.046. These results indicate highly regular behavior, consistent with gradual environmental variation or internally buffered thermoregulation.

Minimal gains were observed in the S-Map framework across the θ spectrum, confirming linearity and predictability of the temperature series. Lack of significant change in accuracy reinforces the interpretation of temperature as a stable and slow-moving variable.

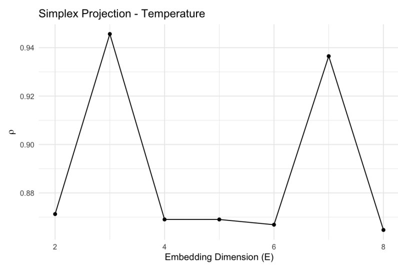
3.4 Univariate Predictability: Hive Power

The `hive_power_norm` time series demonstrated high self-predictability, with optimal performance at $E = 6$, where $p = 0.993$, MAE = 0.039, and RMSE = 0.053. These metrics reflect a deterministic and internally regulated signal, suggesting consistent colony-level activity.

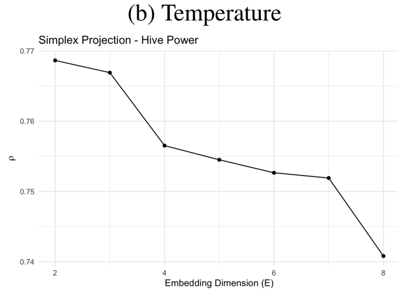
S-Map analysis showed marginal improvement with increased nonlinearity, peaking at $\theta \approx 0.9$ with $p = 0.994$. This indicates weak but present nonlinear dynamics, possibly associated with adaptive metabolic responses or behavioral variability.



(a) Humidity



(b) Temperature



(c) Hive Power

Figure 2: Simplex projection accuracy across embedding dimensions (E) for key variables.

3.5 Multivariate S-Map: Hive Power as a Driver

Multivariate S-Map was used to assess the dynamic influence of `hive_power_norm` on other signals. When hive power served as the target variable, predictive skill improved to $\rho = 0.995$ at $E = 6$ and $\theta = 0.9$, outperforming the univariate baseline. Time-varying coefficients ∂`Vibration`/∂`HivePower` were consistently positive and stable, suggesting a robust influence of hive power on vibrational output.

For `humidity`, the addition of `hive_power` as a predictor led to a marginal improvement in forecasting ($\rho = 0.66$), accompanied by fluctuating coefficient trajectories, indicative of weaker or conditional dependence.

3.6 Temperature–Hive Power Interactions

To examine whether temperature dynamically influenced hive power, a multivariate S-Map was conducted. Predictive skill increased only marginally ($\rho = 0.994$ vs. $\rho = 0.993$ in the univariate case), while estimated coefficients ∂`HivePower`/∂`Temperature` remained close to zero throughout. These findings imply minimal dynamic coupling, reinforcing the interpretation that hive power may regulate temperature, rather than being directly influenced by it.

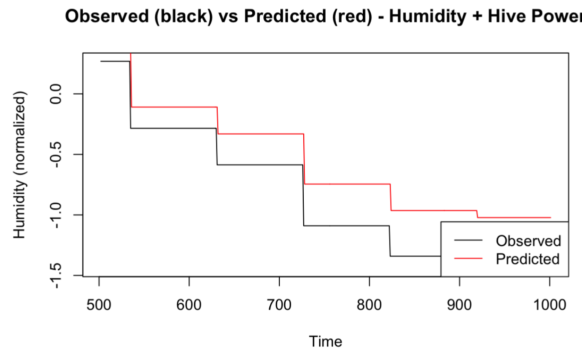


Figure 3: Observed vs. predicted humidity using multivariate S-Map with hive power as input.

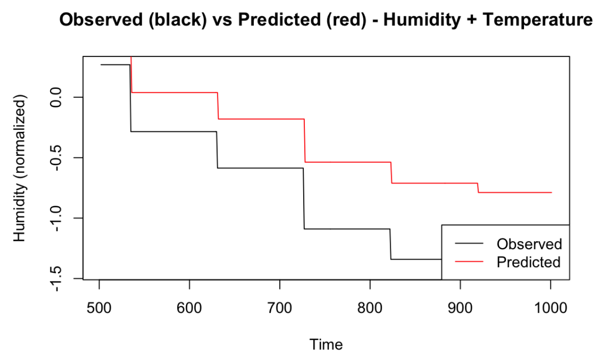


Figure 4: Observed vs. predicted humidity using temperature as input.

3.7 Causal Discovery via Convergent Cross Mapping (CCM)

CCM analysis revealed the strongest causal influence from `hive_power` to `vibration`, with cross-map skill increasing steadily with library size. In contrast, the relationship between `hive_power` and `humidity` appeared bidirectional but weak, with only slight asymmetry in predictive skill. No significant causal influence was detected from temperature to hive power, as evidenced by flat CCM convergence curves.

3.8 Multiview Embedding

Application of Multiview embedding improved forecasting robustness across all variables. The method achieved $\rho = 0.994$ for `hive_power`, indicating that ensemble embedding approaches effectively synthesize multiple perspectives of the underlying dynamics.

3.9 Time-Varying Interaction Strengths

Temporal profiles of interaction strength, as inferred from S-Map coefficients, demonstrated that:

- Hive power exerted a consistently positive and stable influence on vibration.
- Hive power's influence on humidity was weak and fluctuating.
- Temperature showed minimal influence on hive power, with noisy and near-zero coefficient values.

These observations support a causal hierarchy wherein hive power acts as a central modulator of internal hive signals, while external factors such as temperature play more constrained or indirect roles.

4 Discussion and Conclusion

4.1 Summary of Findings

This study employed empirical dynamic modeling (EDM)—including Simplex projection, S-Map analysis, Convergent Cross Mapping (CCM), and Multiview embedding—to examine the internal dynamics of a sensor-instrumented beehive. The `humidity_norm` signal demonstrated moderate self-predictability, with a correlation coefficient of approximately $\rho = 0.63$. This suggests that humidity is influenced by both endogenous hive factors and exogenous environmental conditions. The S-Map analysis revealed minimal nonlinear structure in the humidity signal, indicating relatively smooth or linear dynamics. The inclusion of `hive_power` as a covariate in multivariate S-Map models improved prediction of dynamics to $\rho \approx 0.995$, with consistently positive interaction coefficients.

Temperature showed negligible dynamic influence on hive power. Both multivariate S-Map and CCM analyses yielded coefficients near zero and non-converging causal curves, reinforcing the hypothesis that hive power governs internal temperature regulation, rather than responding to it. Among all techniques applied, multiview embedding achieved the highest predictive accuracy, with $\rho > 0.995$ for `hive_power`, underscoring the utility of ensemble embeddings in noisy, nonlinear biological systems.

4.2 Interpretation

These findings reveal a hierarchical organization in beehive dynamics, where internal metabolic activity—as proxied by `hive_power`—plays a central role in driving collective vibrational behavior. Humidity and temperature exhibit weaker and less consistent associations with hive power, consistent with the biological understanding that bees actively regulate their internal environment and buffer external fluctuations. Overall, these results highlight the capacity of EDM to uncover structured, nonlinear dependencies that characterize complex adaptive systems such as honeybee colonies.

4.3 Limitations and Future Work

This analysis was constrained by a limited selection of variables and a short observation window. While the focus on hive power, humidity, and temperature provided meaningful insights, future work should incorporate additional ecological drivers. Furthermore, the analysis was conducted on a 1000-row subset of the full dataset for computational tractability. Extending these methods to longer time windows or seasonal timescales could reveal slower dynamics, including regime shifts or adaptive transitions in colony behavior.

References

M. Bencsik, M. Baxter, A. Lucian, J. Romieu, and M. Millet. 2011. Honeybee colony vibration monitoring using accelerometers. *Sensors*, 11(5):4788–4806.

Scott Camazine, Jean-Louis Deneubourg, Nigel R. Franks, James Sneyd, Guy Theraulaz, and Eric Bonabeau. 2001. *Self-Organization in Biological Systems*. Princeton University Press.

Peter F Lamb and Michael Stöckl. 2014. On the use of continuous relative phase: Review of current approaches and outline for a new standard. *Clinical Biomechanics*, 29(5):484–493.

William G Meikle, Martin Weiss, and Andrew R Stilwell. 2015. Using within-day hive weight changes to measure environmental effects on honey bee colonies. *PLOS ONE*, 10(8):e0134216.

Samuel D Ramsey, Ronald Ochoa, Gary Bauchan, Christine Gulbronson, James D Mowery, Arnold Cohen, Daniel Lim, Josh Joklik, Joseph M Cicero, and James D Ellis. 2020. Mechanical and genetic pathways of resistance to varroa destructor in honey bees. *Nature Communications*, 10(1):1–13.

George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. 2012. Detecting causality in complex ecosystems. *Science*, 338(6106):496–500.

Fidelis Wario, Bernd Wild, Raúl Rojas, and Tim Landgraf. 2015. Automatic detection of bee colony activity using video and sound data. *arXiv preprint arXiv:1504.00271*.