System Complexity and Information Theory (Entropy)
Entry 8 - Module 12

Group 6: Djourdan Johnson, Jacuqot Qiu, Lotte Michels, Nawat Nawati Azhati, Nuo Xu, Xuelin Wei

1 Introduction

Honey bee colonies (*Apis mellifera*) are essential for global food production, yet face threats from climate change and parasitic mites [1]. Non-invasive, sensor-based monitoring provides a promising way to assess hive health dynamically [2]. In this study, we analyze Hive 202204 using data from the MSPB dataset [3], which includes temperature, humidity, and audio-derived hive power measures. We focus on three entropy metrics—Shannon Entropy, Approximate Entropy (ApEn), and Sample Entropy (SampEn)—to quantify the complexity and predictability of hive dynamics. Our goal is to identify seasonal trends from April to November 2020, with particular attention to hive power as an indicator of bee activity. We also compare our findings to previously reported seasonal behavior patterns in the full MSPB cohort. By combining entropy metrics with time series and transfer entropy analyses, we aim to evaluate the sensitivity of each method in capturing dynamic shifts in colony behavior.

The code used for the computations and visualizations in this entry can be found here.

2 Data Preprocessing

To conduct entropy and information flow analysis, we preprocessed the D1_sensor.csv file from the MSPB dataset. This file contains approximately 960,809 records, spanning from April 16 to November 5, 2020. We first filtered the data to include only records from hive 202204 and retained key columns (published_at, temperature, humidity, hive_power, and tag_number). To ensure temporal uniqueness of the time series, we removed duplicate timestamps. The time series was then aggregated to hourly averages, sorted chronologically, and grouped by month to enable monthly analyses.

Building on this, we further cleaned the data to meet the requirements for transfer entropy analysis. Specifically, we removed any records that contained missing values in the three main variables: temperature, humidity, or hive_power. This resulted in a clean, hourly-sampled dataset consisting of 3,209 records covering the main summer and early autumn months. We visualized missingness using a missingno matrix to confirm that the working dataset was complete. Monthly counts confirmed sufficient data coverage from June to October 2020, while May and November were excluded from the *transfer entropy* and *characteristic time scale* analyses due to an insufficient number of hourly records (fewer than 30).

3 3. Shannon Entropy ApEn and SampEn Analysis

Entropy measures were computed for temperature, humidity, and hive power using R, with parameters m = 2, r = 0.2 * SD for ApEn and SampEn, and 10 bins for Shannon Entropy [4]. Results were aggregated monthly and visualized as line plots.

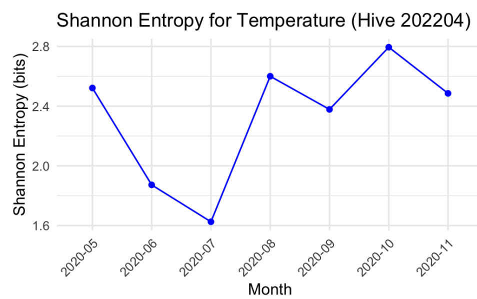

Figure 1: Shannon Entropy for Temperature

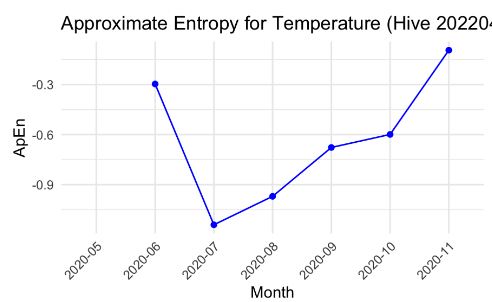

Figure 2: Approximate Entropy for Temperature

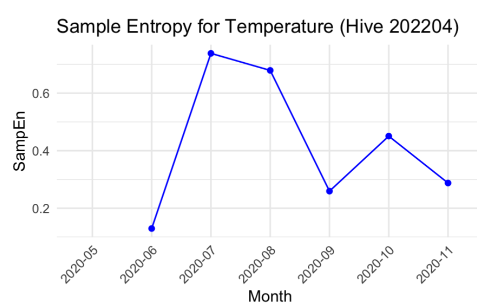

Figure 3: Sample Entropy for Temperature

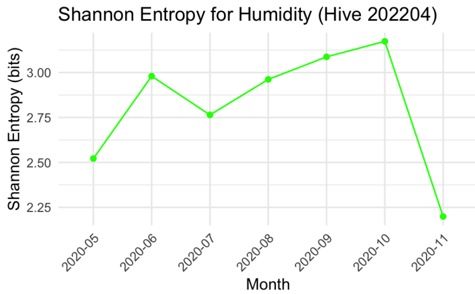

Figure 4: Shannon Entropy for Humidity

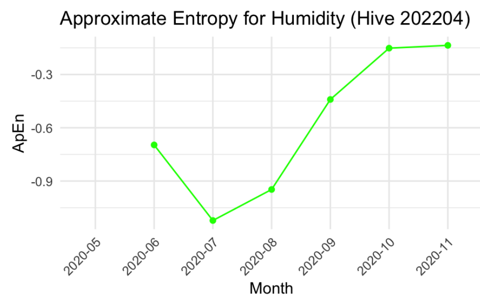

Figure 5: Approximate Entropy for Humidity

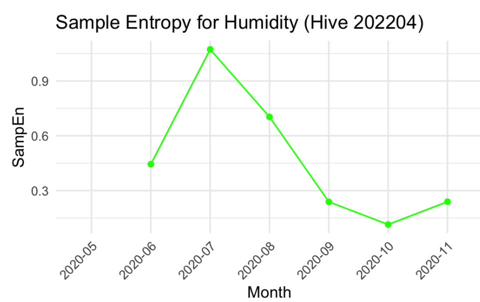

Figure 6: Sample Entropy for Humidity (Hive 202204)

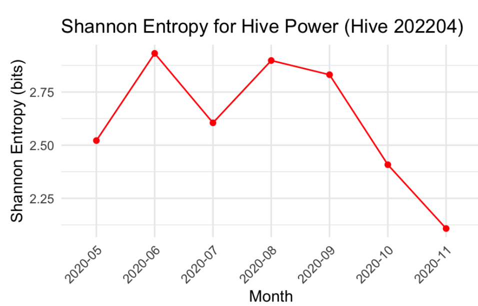

Figure 7: Shannon Entropy for Hive Power

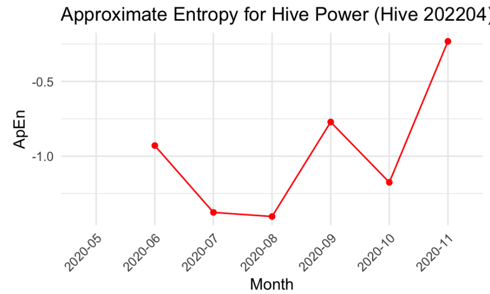

Figure 8: Approximate Entropy for Hive Power

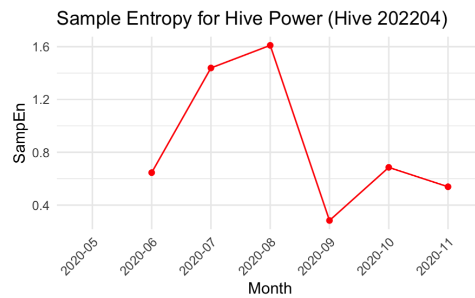

Figure 9: Sample Entropy for Hive Power

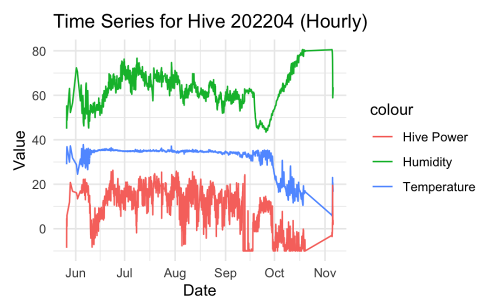

Figure 10: Time Series for Hive 202204 (Hourly)

Entropy analysis revealed distinct seasonal patterns across all three variables. For temperature, Shannon Entropy remained relatively low and stable throughout the observed period ($\tilde{2}$.0–2.5 bits), indicating effective thermoregulation by the colony. A slight rise in entropy during the summer months (June–August 2020) suggests minor external influences. Sample Entropy peaked at 0.74 in July and dropped to 0.29 in November, while Approximate Entropy values were negative, reflecting regularity but also potential calculation instability [3].

For humidity, Shannon Entropy peaked in October ($\tilde{3}$.17 bits), then dropped sharply in November, likely due to environmental stabilization inside winter chambers. Sample Entropy values followed a similar trend, with a summer peak (1.07 in July) and significant decline in fall. Approximate Entropy again showed negative values, with increasing irregularity observed later in the season [3].

Hive power, derived from acoustic features, exhibited the most dynamic changes. Shannon Entropy peaked at 2.93 bits in June and declined to 2.11 bits by November, suggesting intense summer activity followed by winter quiescence. Sample Entropy confirmed these fluctuations, with high values in August (1.61) and a marked decrease in November (0.54). These trends likely reflect biologically significant events such as swarming or brood rearing during peak activity. Approximate Entropy remained negative and erratic, indicating computational limitations for this variable [3].

A time series plot of all three variables further supports these findings. Hive power displayed erratic variability during summer and stabilized toward late fall. Meanwhile, humidity remained relatively constant (50–70%), gradually increasing toward winter. Temperature followed a seasonal downward trend, dropping from $\tilde{3}0$°C to $\tilde{2}0$°C by November, consistent with environmental cooling [3].

Month	Temp_Sha	Humid_Sha	Power_Sha
May-20	2.52	2.52	2.52
Jun-20	1.87	2.98	2.93
Jul-20	1.63	2.76	2.61
Aug-20	2.60	2.96	2.90
Sep-20	2.38	3.09	2.83
Oct-20	2.79	3.17	2.41
Nov-20	2.49	2.20	2.11

Table 1: Shannon Entropy values for Hive 202204 (May–Nov 2020).

Month	Temp_ApE	Humid_ApE	Power_ApE
May-20	NA	NA	NA
Jun-20	-0.30	-0.70	-0.93
Jul-20	-1.14	-1.12	-1.38
Aug-20	-0.97	-0.95	-1.40
Sep-20	-0.68	-0.44	-0.77
Oct-20	-0.60	-0.15	-1.17
Nov-20	-0.09	-0.14	-0.23

Table 2: Approximate Entropy (ApEn) values for Hive 202204 (May–Nov 2020).

Month	Temp_Sam	Humid_Sam	Power_Sam
May-20	NA	NA	NA
Jun-20	0.13	0.44	0.65
Jul-20	0.74	1.07	1.44
Aug-20	0.68	0.70	1.61
Sep-20	0.26	0.24	0.28
Oct-20	0.45	0.11	0.69
Nov-20	0.29	0.24	0.54

Table 3: Sample Entropy (SampEn) values for Hive 202204.

*Note: ApEn values are omitted due to negative results, suggesting calculation issues

4 Comparison with Hive 202202 and Data Team Findings

The data team's analysis (Page 8, Figure 6) reported hive_power's seasonal variations across 53 hives:

summer peaks (June-August 2020) from super addition and swarming, fall fluctuations (September-October) from Varroa treatment, and winter stability (November onward) in indoor chambers [3]. They noted hive_power's high variability, capturing colony-specific dynamics (Page 8) [3]. For hive 202204, hive_power's Shannon Entropy (Figure 7, Page 2) peaks in June (2.93 bits), remains high in July-August (2.61-2.90), and drops to 2.11 in November, mirroring the data team's summer activity peaks and winter stabilization [3]. SampEn (Figure 9, Page 2) peaks in August (1.61), drops to 0.54 in November, reflecting complex summer patterns (e.g., swarming, evaluations) and winter regularity [3]. The time series (Figure 10, Page 3) confirms hive_power's erratic summer fluctuations and winter stability, consistent with interventions like super addition (June August) and winterization (November) [3]. ApEn (Table 1) is less reliable due to negative values, possibly from implementation issues (e.g., small r in approx_entropy). Assuming hive 202202 at the Côté apiary shares identical management (e.g., super addition, Varroa treatment), its hive_power entropy would likely follow similar trends: high summer entropy (June-August) and low winter entropy (November) [3]. Without 202202's specific data, we hypothesize consistency based on environmental and management similarities. Our entropy analysis replicates the data team's findings, with Shannon Entropy and SampEn capturing hive_power's seasonal dynamics more robustly than ApEn [3]. The monthly granularity provides finer insights compared to the data team's daily averages, highlighting entropy's sensitivity to colony activity [3, 4].

5 Transfer Entropy Analysis and Surrogate-based significance test

Month	**TE_temp→power**	**TE_power→temp**	**DirIndex**
06	-0.0696	0.0813	**-0.151**
07	-0.1193	-0.0394	**-0.080**
08	-0.1172	-0.0068	**-0.110**
09	**0.1139**	0.0383	**+0.076**
10	**0.1385**	0.0254	**+0.113**

Table 4: Transfer Entropy and Directionality Index for Temperature → Hive Power.

Month	**TE_humid→power**	**TE_power→humid**	**DirIndex**
06	0.1057	0.0811	**+0.025**
07	0.0291	-0.0562	**+0.085**
08	0.0141	0.1713	**-0.157**
09	0.0098	0.1229	**-0.113**
10	0.0292	0.0375	**-0.008**

Table 5: Transfer Entropy and Directionality Index for Humidity → Hive Power.

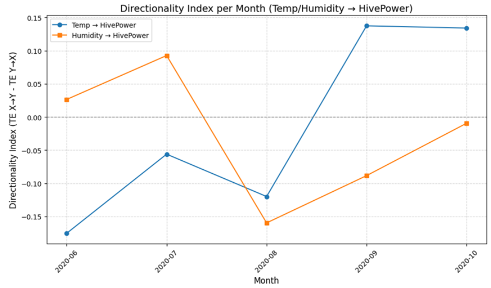

Figure 11: Directionality Index per Month(Tem/Humidity to HivePower)

Month	TE_temp→power	p-value
2020-06	-0.077698	0.88
2020-07	-0.099055	0.85
2020-08	-0.121769	0.95
2020-09	0.137959	0.10
2020-10	0.150305	0.12

Table 6: Surrogate-based significance test for Transfer Entropy (Temperature → Hive Power).

Month	TE_humid→power	p-value
2020-06	0.104871	0.09
2020-07	0.028398	0.40
2020-08	0.013782	0.51
2020-09	0.006512	0.48
2020-10	0.034606	0.40

Table 7: Surrogate-based significance test for Transfer Entropy (Humidity → Hive Power).

Although the Transfer Entropy analysis did not yield statistically significant results, visual inspection of the directionality index plot suggests possible trends in information flow between hive behavior and environmental variables. Specifically, from June to August 2020, the directionality index for temperature is negative, indicating slightly stronger information flow from hive_power to temperature. This may reflect active thermoregulatory behavior during peak colony activity. In September and October, the index reverses and turns positive, suggesting that temperature increasingly predicts hive_power, potentially signaling a shift toward passive environmental adaptation. The humidity → hive_power directionality index, on the other hand, fluctuates across months, with a pronounced dip in August, hinting at transient feedback, but lacking a consistent pattern—supporting the idea that humidity plays a relatively minor or background role in behavioral modulation.

In Transfer Entropy analysis, surrogate testing is used to determine whether the observed TE value is significantly greater than would be expected by chance. This involves randomizing the source variable (e.g., temperature or humidity) to break any temporal structure and generate a null distribution. If the actual TE exceeds most surrogate values, it suggests a meaningful directional influence rather than a spurious correlation.

As shown in the tables, none of the p-values fall below the 0.05 significance threshold. This indicates that neither temperature nor humidity exhibits statistically significant information flow to hive_power during any month. Although certain months (e.g., September and October 2020) show slightly elevated TE values, these should be interpreted cautiously as exploratory findings without statistical support.

Given the lack of statistical significance, these patterns should be interpreted as exploratory, offering a basis for future investigations using larger or higher-resolution datasets.

Although the Transfer Entropy analysis did not yield statistically significant results, the directionality indices across months revealed systematic fluctuations, particularly a reversal around the summer-to-fall transition. Since the temporal scale of information transfer may vary across conditions, relying solely on aggregate TE values could obscure meaningful dynamic patterns. Therefore, we proceeded with a Characteristic Time Scale analysis to identify the optimal lag at which temperature and humidity most strongly influence hive_power. This allows us to explore whether the timing of environmental-behavior coupling shifts across seasons—a question that remains relevant even when the overall TE magnitude lacks statistical significance.

6 Characteristic Time Scale

Month	Best Lag (hours)	Max TE
2020-06	20	0.091447
2020-07	6	0.053804
2020-08	20	0.099630
2020-09	3	0.061664

Table 8: Characteristic Time Scale for Temperature → Hive Power.

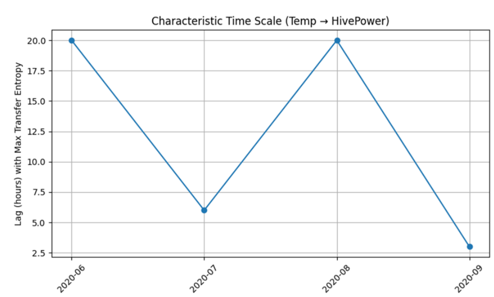

Figure 12: Characteristic Time Scale(Temp to Hivepower)

The Characteristic Time Scale analysis indicates notable month-to-month variation in the optimal lag at which temperature most strongly predicts hive_power. In June and August 2020, the highest transfer entropy occurred at a 20-hour lag, suggesting a prolonged delay in behavioral response to temperature shifts—potentially due to gradual thermoregulatory mechanisms. In contrast, the lag shortened to 6 hours in July and just 3 hours in September, implying a quicker reaction by the colony to temperature changes during these months. Notably, the sharp drop in September may signal a behavioral adjustment as the hive transitions into fall. Overall, the time scale follows a "longer in summer, shorter in fall" pattern, consistent with the hypothesis that the rhythm of environmental-behavior coupling accelerates as seasons change.

Month	Best Lag (hours)	Max TE
2020-06	10	0.087963
2020-07	1	0.059427
2020-08	22	0.122362
2020-09	2	0.066048
2020-10	20	0.021934

Table 9: Characteristic Time Scale for Humidity → Hive Power.

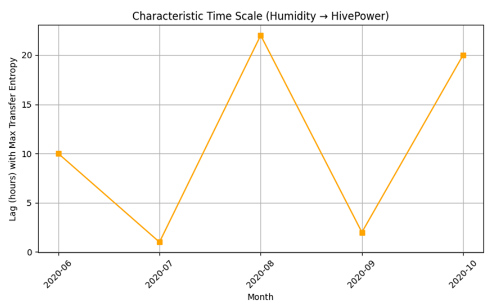

Figure 13: Characteristic Time Scale(Humidity to Hivepower)

In the analysis of characteristic time scale for Humidity → HivePower, the observed lag times varied greatly across months, suggesting an unstable or inconsistent influence of humidity on hive activity. As shown in the figure, the peak information transfer occurred at a 10-hour lag in June, dropped sharply to 1 hour in July, spiked to 22 hours in August, and again dropped to 2 hours in September before rising to 20 hours in October. These abrupt fluctuations imply that hive responses to humidity changes were neither consistent nor robust, potentially indicating that humidity is not a primary behavioral driver. Moreover, none of the transfer entropy values for these months passed the statistical significance threshold (p ¿ 0.05), which supports the view that these lags reflect exploratory temporal trends rather than reliable causal dynamics. Nevertheless, the results raise interesting possibilities for further investigation, suggesting that under specific conditions, humidity may exert delayed effects on hive activity without being a consistently dominant factor.

During our initial analysis of characteristic time scales, we observed that several months exhibited peak Transfer Entropy (TE) values near the upper bound of our predefined lag window (max_lag = 20 hours). This raised a valid concern that our lag range might be too narrow, potentially missing stronger information flow at longer delays. To address this, we extended the lag window to 24 hours and visualized TE values across this broader range for each month.

Results showed that in June and August 2020, TE peaked near or beyond 20 hours, suggesting that meaningful interactions may exist at longer delays. These late peaks may reflect lagged behavioral responses of the hive to humidity fluctuations, such as adjustments in internal microclimate through collective activity. In contrast, July and September exhibited TE peaks at shorter lags (1–6 hours), indicating high variability in the temporal structure of information flow across months. Although TE magnitudes varied across lags, the overall values remained low and failed to pass surrogate significance testing. Therefore, these time-scale trends should be interpreted cautiously as exploratory signals rather than definitive causal mechanisms.

7 Conclusion

Our analysis shows that entropy measures, particularly Shannon Entropy and SampEn, effectively capture seasonal changes in hive behavior. Hive power demonstrated the clearest variation, with entropy peaking in summer and declining in late fall, aligning with known colony events like swarming and winter preparation. While Approximate Entropy produced unstable results, the other metrics proved consistent. Transfer Entropy and directionality indices did not yield statistically significant patterns, though exploratory trends suggest temporal asymmetries in environmental-behavior coupling. The characteristic time scale analysis further revealed that hive responses to temperature occur over shorter lags in fall than summer, reflecting seasonal adaptation. Despite the limited statistical power, entropy-based methods showed sensitivity to subtle shifts in colony dynamics and offer promising tools for non-invasive behavioral monitoring. Future work should validate these findings across more hives and longer periods, potentially integrating higher-resolution data or experimental interventions.

References

[1] Klein, A.-M., et al. (2007). Importance of pollinators in changing landscapes for world crops. Proceedings of the Royal Society B: Biological Sciences, 274, 303-313.

[2] Abdollahi, M., Giovenazzo, P., & Falk, T. H. (2022). Automated beehive acoustics monitoring: A comprehensive review of the literature and recommendations for future work. Applied Sciences, 12, 3920.

[3] Zhu, Y., Abdollahi, M., Maucourt, S., Coallier, N., Guimarães, H. R., Giovenazzo, P., & Falk, T. H. (2023). MSPB: A longitudinal multi-sensor dataset with phenotypic trait measurements from honey bees (Apis mellifera L.). arXiv preprint, arXiv:2311.10876. Available at https://arxiv.org/pdf/2311.10876.

[4] Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology, 278(6), H2039-H2049