

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☐ ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	All colonies data was collected by commercial beekeepers using Nectar software 1.19. Weather and air quality data was obtained through OpenWeather and Airpyllution API and the coordinate of each colonies at all timestamp. NDVI was computed from satellite images obtained through Earth Engine client library.
Data analysis	Statistical analysis were computed in R version 4.4.1 with the packages with the packages lme4 (version 1-35.4), ggplot2 (version 3.5.1), and performance (version 0.12.3). Machine learning model were built in Python 3.12.0 using the lifelines (version 0.30.0) and sklearn (version 1.6.0). Map was created using ArcGIS Pro version 3.3.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Commercial and privacy restrictions impede us from making the dataset used for this study freely available. Requests for the full dataset should be addressed to Nico Coallier. The provided sample dataset contains data from Nectar research hives and can be found here: https://github.com/Nectar-Technologies/ledge-nectar/tree/main/aqi/example_data.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Reporting on race, ethnicity, or other socially relevant groupings

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

We've analyzed data from 103 477 colonies on 3 beekeeping seasons (2021-2023). We've included random slopes and intercept on the regions in which the beekeepers operates as well as on the beekeeper in statistical models. We've defined adhoc yearly experimental period were each year (or season) in order to run survival models on properly structure data. We've selected the most parsimonious statistical model using the AIC criterion. For machine learning models, we've split the dataset into a training set (85%) and a test set (15%) were each set is colonies independent, meaning that colonies are unique to their set. Afterwards, we've scaled all variable using a min-max standardization. Then we've optimized hyper-parameters of the RSF models using a grid search approach. Best models was selected using concordanc index and ROC-AUC curve. Finally, we've created the prediction map using ArcGIS and trained a ElasticNet to extrapolated risk scores and performed an inverse-distance weighted interpolation to create a continuous spatial map.

Research sample

The dataset consists of 103 477 colonies and 112 626 observations. Colonies are tracked by commercial beekeeper using Nectar software 1.19 which allow them to ensure perfect traceability of the hive and it's location during their season. The colonies are located in 7 states in the US and 3 provinces in Canada. We've included data from 2021 up to 2023 since we had the full beekeeping season logged.

Sampling strategy

One aim of this study is to provide a big data approach to beekeeping science. While we didn't try to identify an optimal sample size, this study present one of the largest dataset on honeybee findings. We validated the quality of the data for each beekeepers included in this study. The validation was done manually by blacklisting users that don't use Nectar software according to the standard.

Data collection

This data was collected from commercial beekeeping operations that employ a mobile app-based system to track the movement and management practices of their operations (Nectar Technologies Inc.). Hives included in this data set are managed colonies across the US and Canada. Data was generated through the use of cell phone QR code/ RFID readers deployed in the field and all data was

anonymized for use in this study. Data collection from the field occurred as follows. To employ data collection via the Nectar Application, apiaries affix label cards containing printed QR codes or embedded RFID chips to individual hives. To log management actions (e.g., requeening, feeding sucrose syrup, inspecting hive strength), apiary workers scan the tags with their mobile phone using the “Nectar App” and select premade actions or type in custom actions. When colonies are moved to new locations, the system captures GPS location, date, time, and user when Nectar tags are scanned. As beekeepers perform their normal management practices they utilize the same mobile app system to capture hive mortality when encountered. Hive tags are scanned as “deadout” and the system again captures GPS location, date, time, and user.

Timing and spatial scale	The data was collected by commercial beekeeper who in average inspect their colonies every two weeks during the season. We've included 3 full beekeeping season from Jan 1st 2021 to June 15th 2023. This dataset included colonies from US and Canada in locations spread in 7 states and 3 provinces.
Data exclusions	No data were excluded from the analysis other than hives that were lost. Lost hives are flagged by users meaning that they haven't track the location of the hive properly. These hives were remove during the original manual data selection process.
Reproducibility	To reproduce this study, one should have access to a large population of digitally tracked hives. Once hives are tracked, each move and location must be captured as well as the state of the colonies and age.
Randomization	No randomization was done other than when splitting training and test set for predictive models. We've used all data available during the period.
Blinding	The blinding was total since none of the users of Nectar were aware of this study at the moment of data collection.
Did the study involve field work?	☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	N/A
Wild animals	N/A
Reporting on sex	N/A
Field-collected samples	Data was collected by commercial beekeeper using Nectar software 1.19.0 on live honeybee colonies.
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A