

OP-157 [Bee Health]

Will my colony survive this season? A benchmark analysis of statistical and computational models to predict honey bee mortality

Maxime Fraser Franco, Nico Coallier

Nectar Technologies Inc., 6250 Rue Hutchison #301, Montréal, QC, H2V 4C5, Canada

For the past decade, managed honey bee populations have experienced substantial yearly mortality worldwide (~40%). While scientists have identified a series of potential causes underlying these declines, predicting how and when bee mortality should occur remains a challenge. This is in part due to the difficulty of collecting data over a colony's lifetime, the complex nature of longitudinal datasets collected in the field, and a lack of standardized protocols and guidelines for analyzing such data. As a result, beekeepers face considerable uncertainty, as they cannot reliably anticipate when and how many of their colonies will perish. To address these problems, we built a longitudinal dataset using the Nectar application, a data platform working with beekeeping operations across Canada and the U.S. Nectar allows us to monitor a colony's lifespan from its creation to its potential death, and assess how beekeeping practices and environmental features influence survival. Our objectives were 1) to test the performance of models tailored for survival analysis on a field honey bee dataset, 2) to provide a clear protocol for the analysis of honey bee colony survival, and 3) to provide guidelines for beekeepers to maximize colony survival. Preliminary results indicate that computational models outperform traditional statistical models (Cox proportional hazards) at predicting hive survival. Models using extreme gradient boosting (XGBoost) displayed the best performance, achieving high predictive accuracy (~0.90 Concordance Index) over independent test data. On average, the chances of survival declined to 50% when colonies reached approximately 325 days of age. The most important features for predicting colony survival were air quality, precipitation, and the rate of kilometers travelled for crop pollination. This suggests that beekeepers should closely monitor colonies older than one year and aim to mitigate the impacts of practices that may compromise honey bee health. Hence, our analyses reveal that monitoring hive level data is essential to obtain accurate predictions of honey bee survival.

nectar



Will my colony survive this season?

A benchmark analysis of statistical
and computational models to predict
honey bee mortality

September 2025

Honey bees face elevated mortality worldwide

60%
U.S.

30%
Lat.Am.

16%
Eur.

Project Apis m. 
HONEY BEE HEALTH RESEARCH
(2024-2025)

scientific reports
(Requier et al. 2024)
(2016-2018)



(2019-2020)

Honey bees face elevated mortality worldwide

60%
U.S.

30%
Lat.Am.

16%
Eur.

Numerous stressors



Honey bees face elevated mortality worldwide

**60%
U.S.**

**30%
Lat.Am.**

**16%
Eur.**

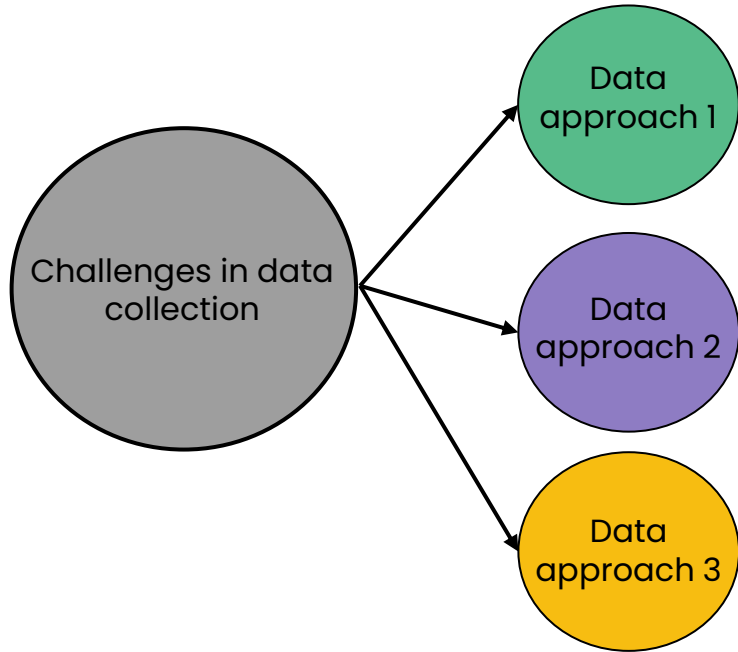
**Difficult to predict when
mortality will occur and
how important it will be!**



Challenges in data
collection



Non-standardized data collection

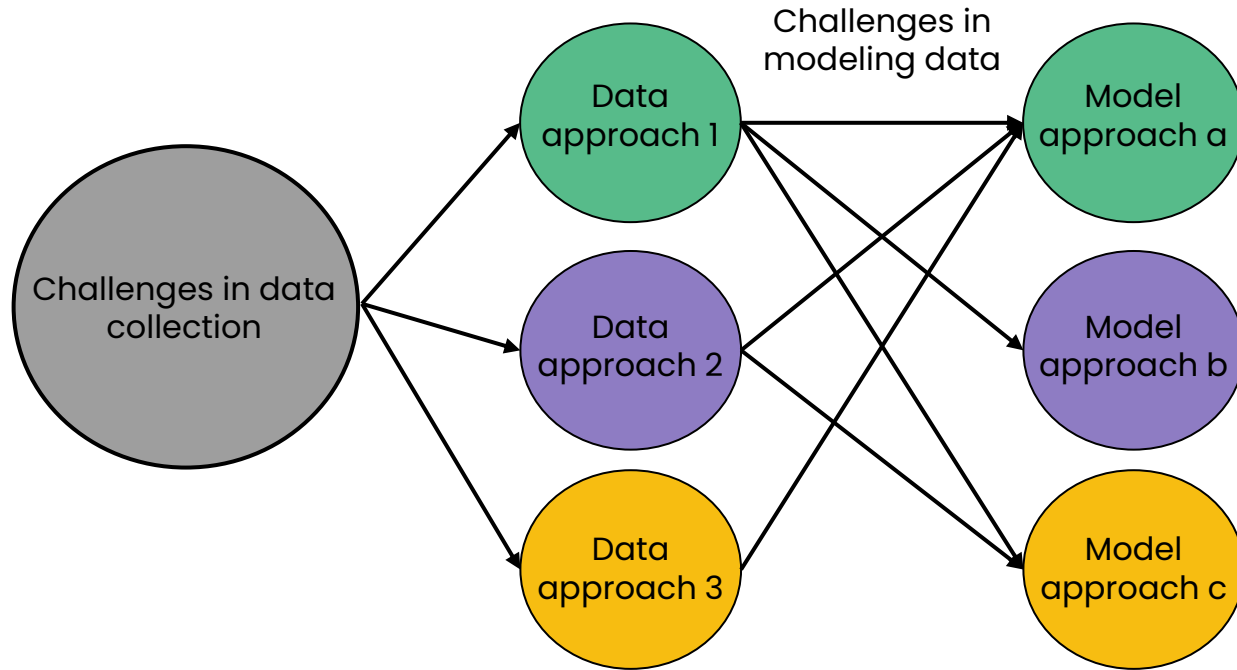




**Non-standardized
data collection**



**Non-unified
modeling practices**

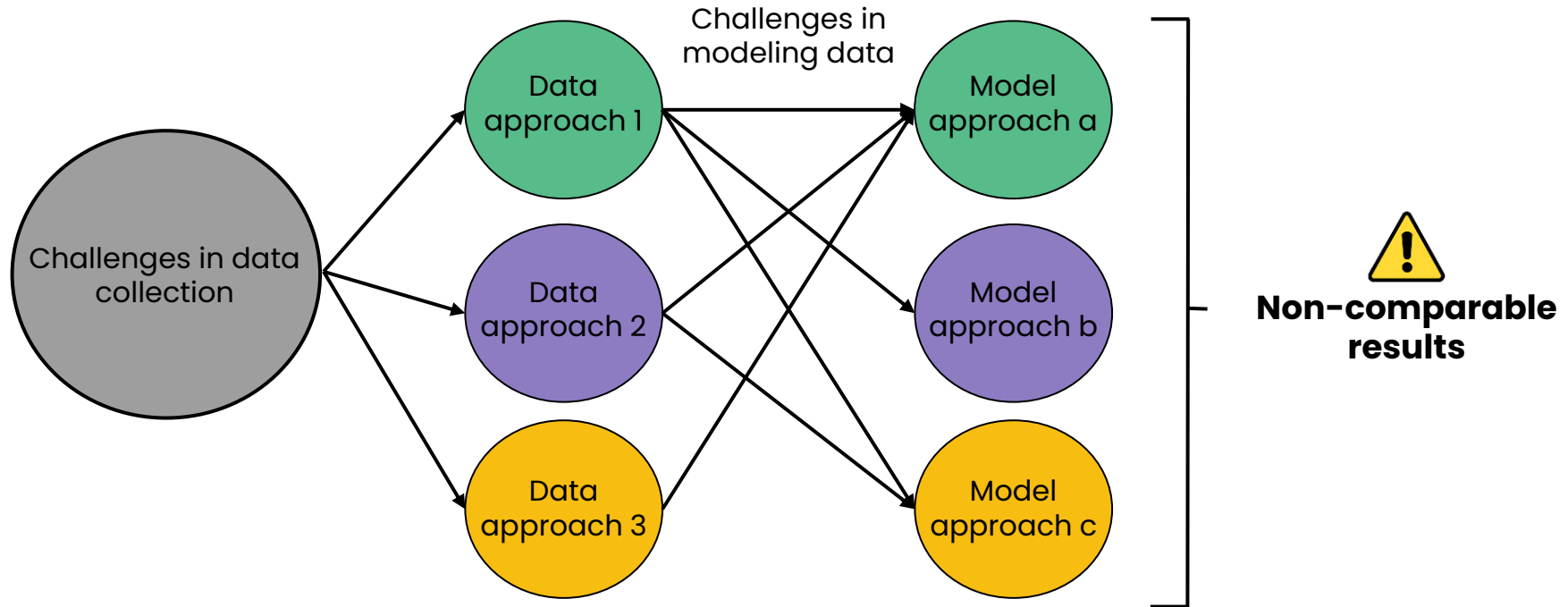




**Non-standardized
data collection**



**Non-unified
modeling practices**



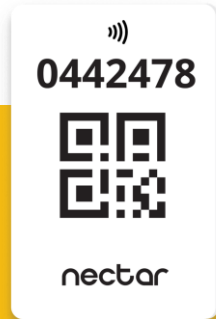
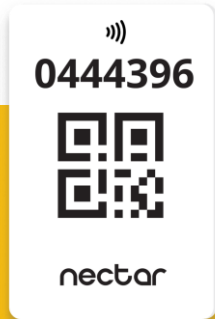
nectar

**How do we tackle
these challenges?**

nectar

300k colonies tracked across 70 operations

Canada, U.S., and New Zealand!



Objectives

- 1) Build a series of models to predict colony mortality
- 2) Compare the models' predictive accuracy
- 3) Identify the most important features in predicting mortality

Sample dataset

6

Beekeeping operations

- Variable size
- Different business models
- Canada and the U.S.

3

Beekeeping seasons

- 2022-2023
- 2023-2024
- 2024-2025

1.4K

Unique yards

- Varying size
- Varying landscape
- Varying weather

Sample dataset

23.2K

Unique colonies

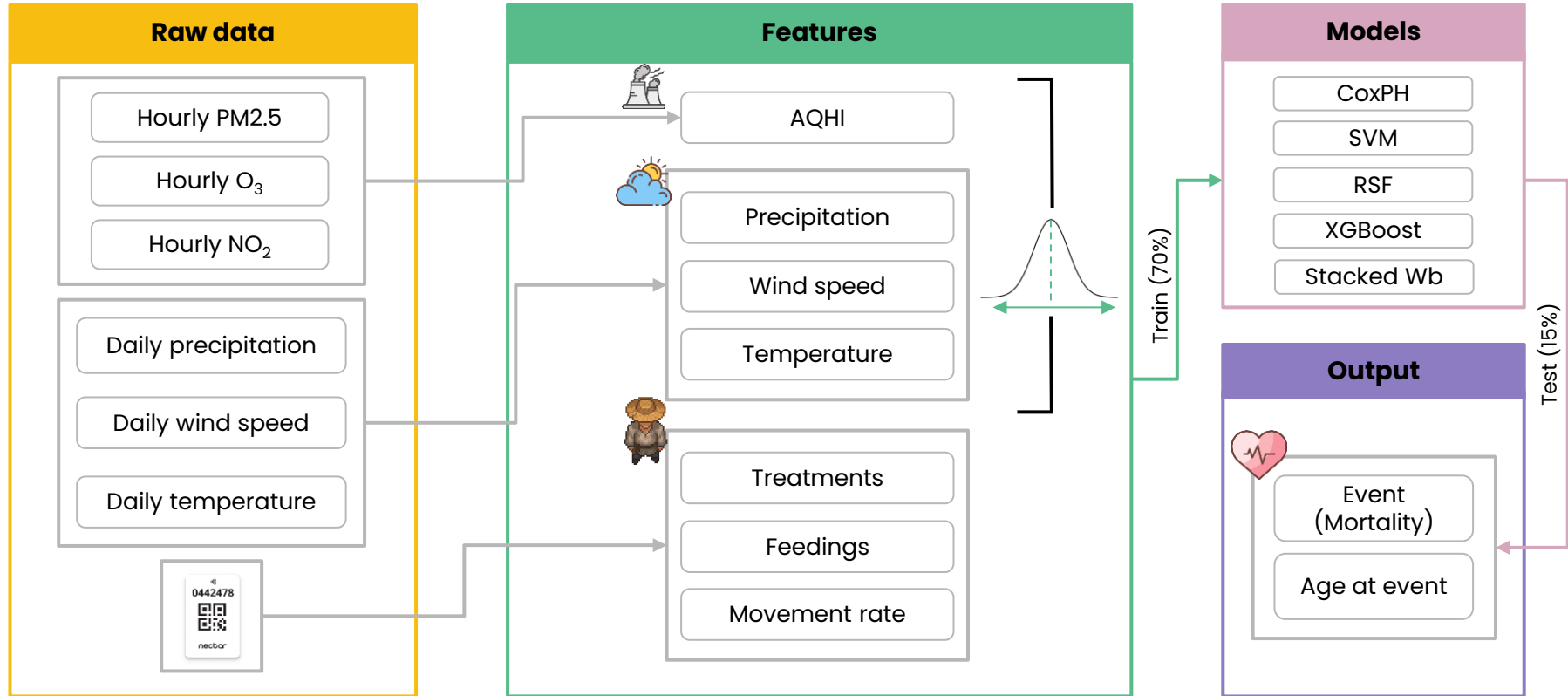
- Variable age
- Unique history
 - Treatments
 - Feeding
 - Movements

34%

Average mortality

- 15 303 colonies alive
- 7 967 colonies dead

Analysis workflow



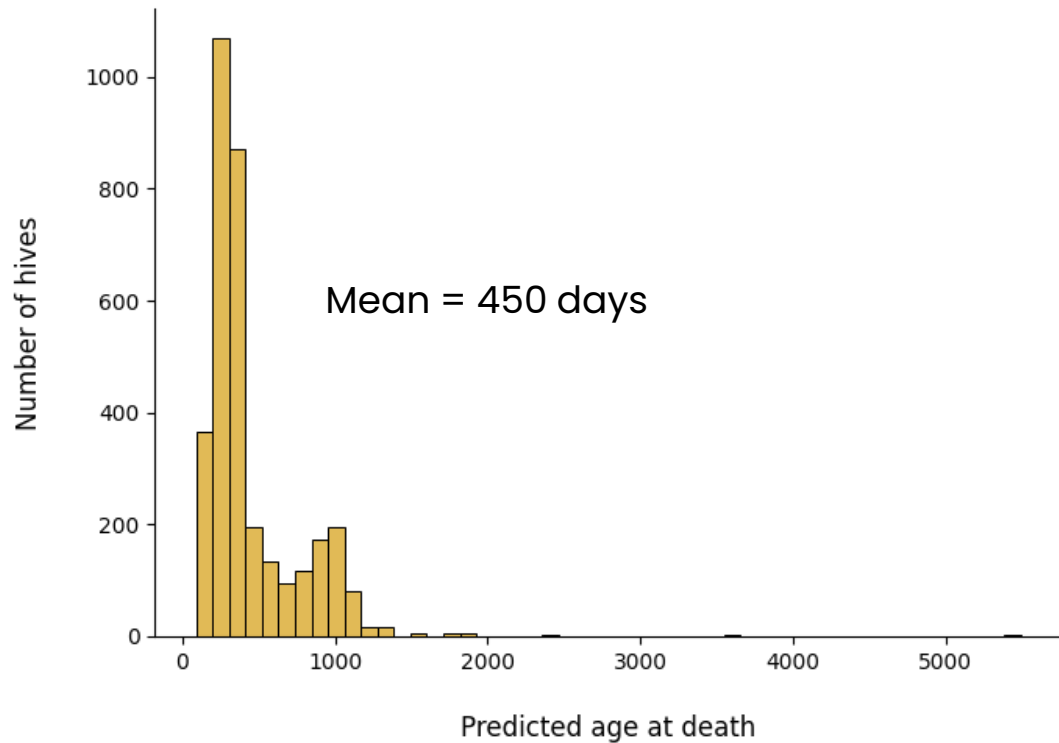
Model performance

Model	Mean Concordance Index IPCW
XGBoost 	0.933
Random Survival Forests	0.901
XGBSE – Stacked Weibull	0.875
Cox Proportional Hazards	0.746
Survival Support Vector Machine	0.474

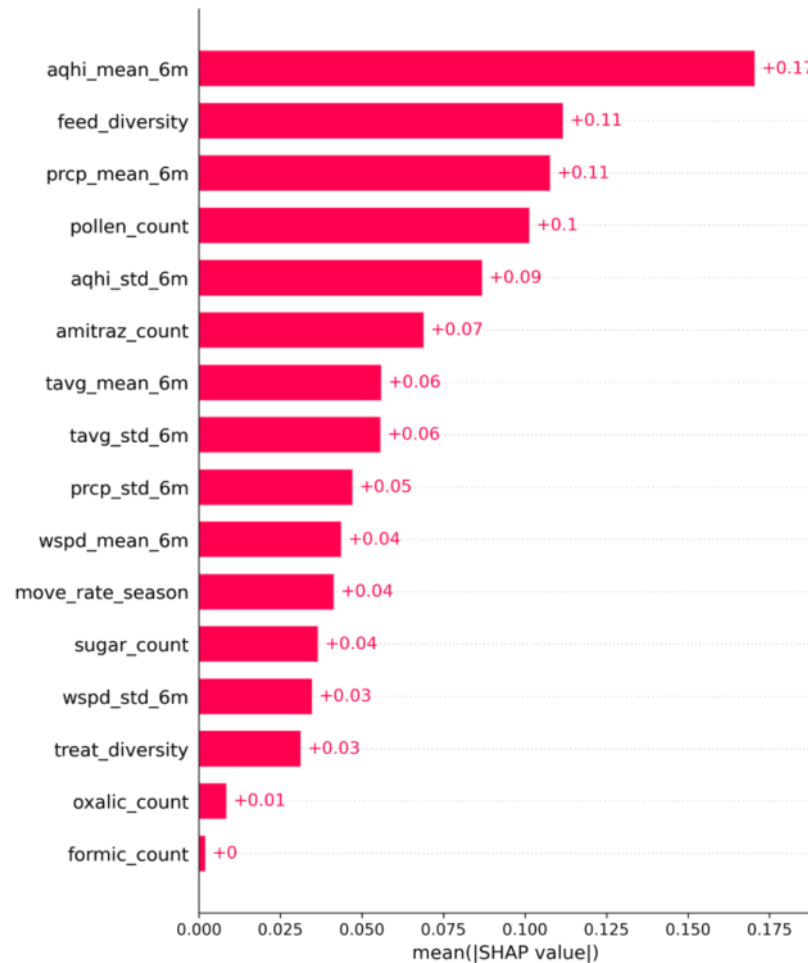
Model performance

Model		Mean Concordance Index IPCW
XGBoost 		0.933
Random Survival Forests	+19% from state-of-the-art	0.901
XGBSE – Stacked Weibull		0.875
Cox Proportional Hazards		0.746
Survival Support Vector Machine		0.474

Predictions



Feature importance



Feature importance

Mean air quality and **feeding diversity** are the **most important** predictors of hive mortality.

Oxalic and **formic** treatments **contribute the least** in predicting hive mortality.



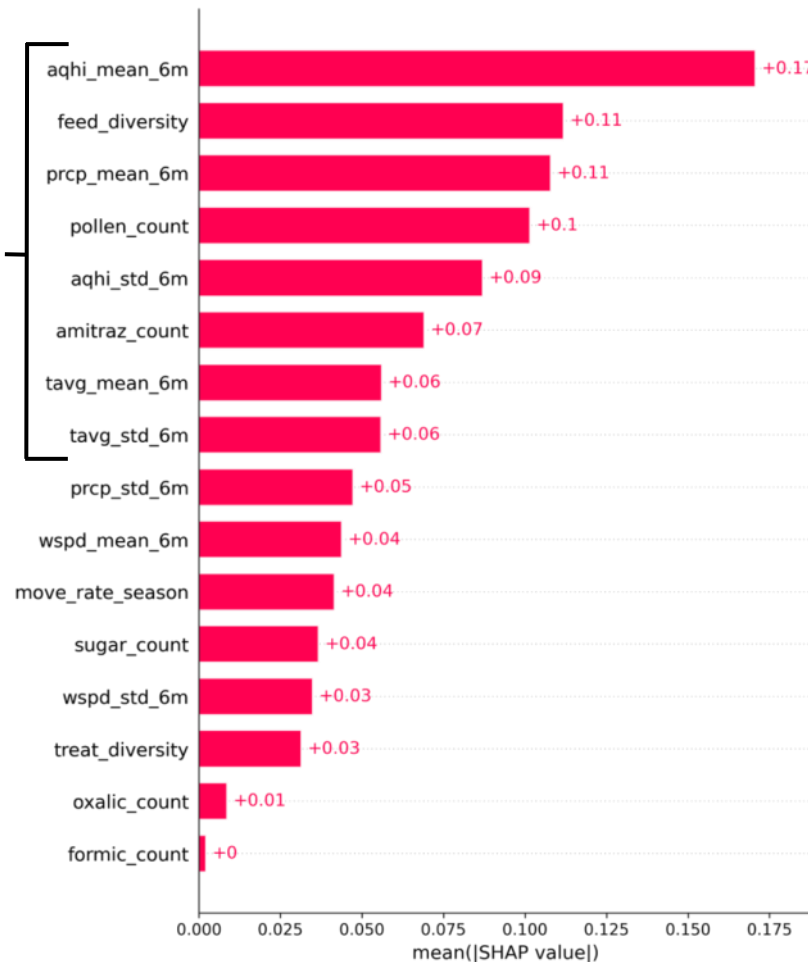
2x



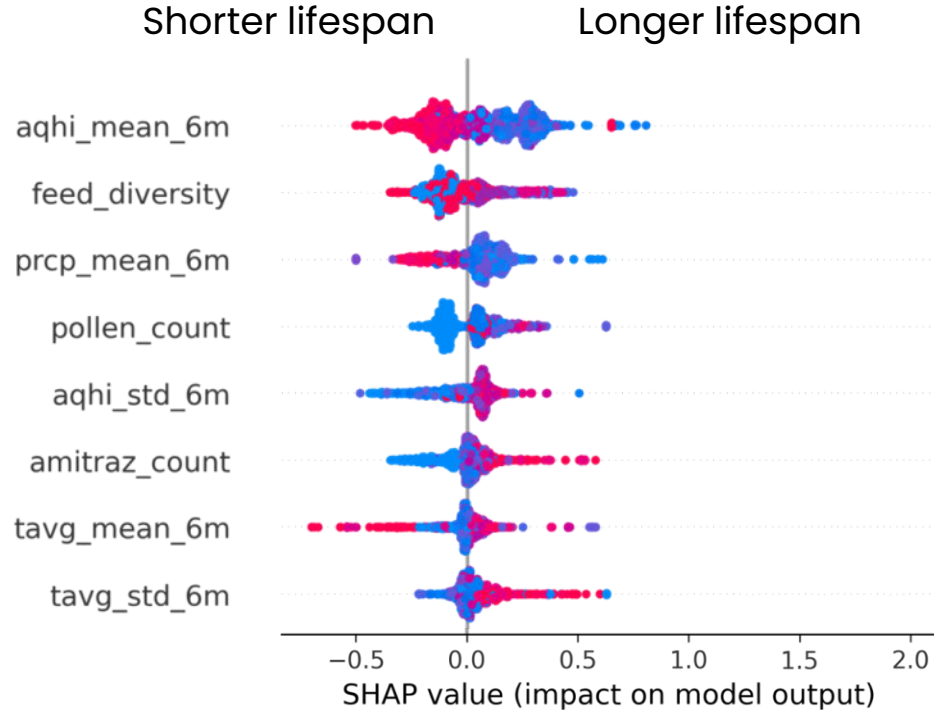
3x



3x

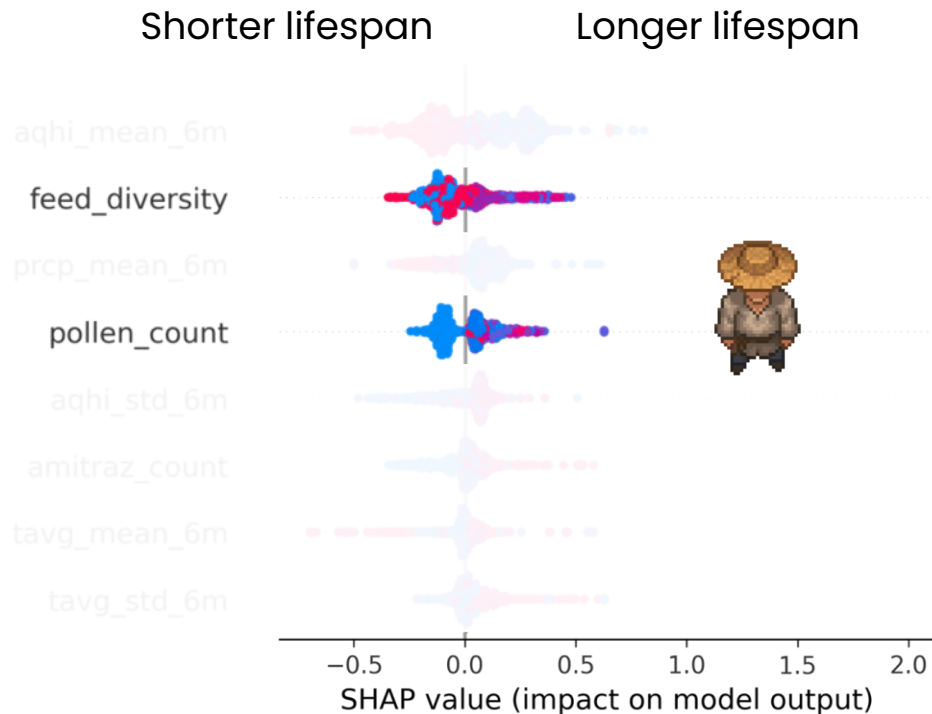


Feature effects



Feature effects

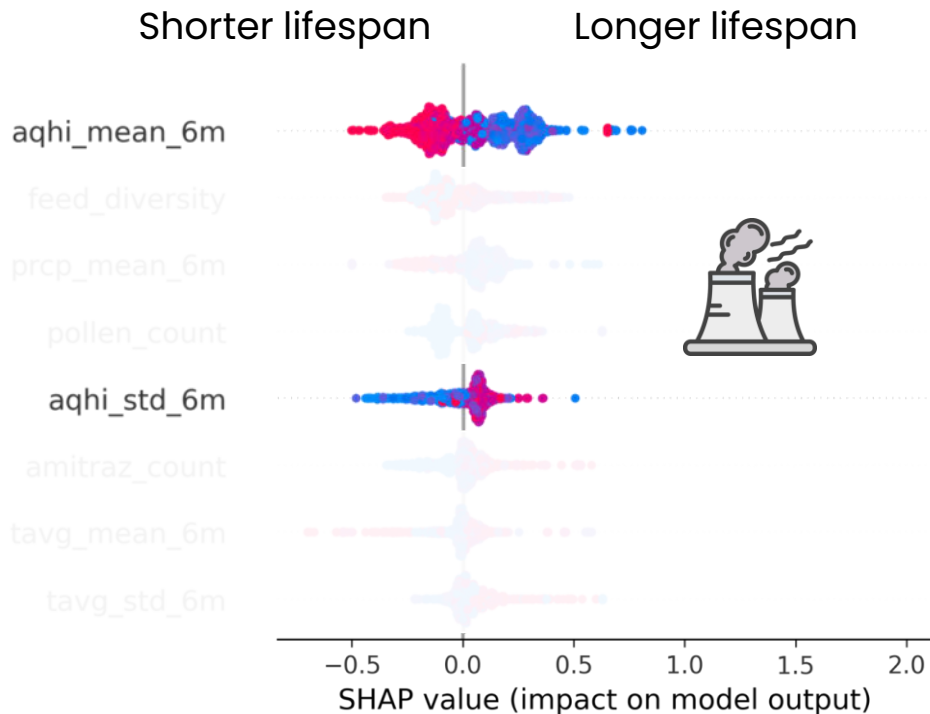
Hives with a
balanced diet have
longer lifespans



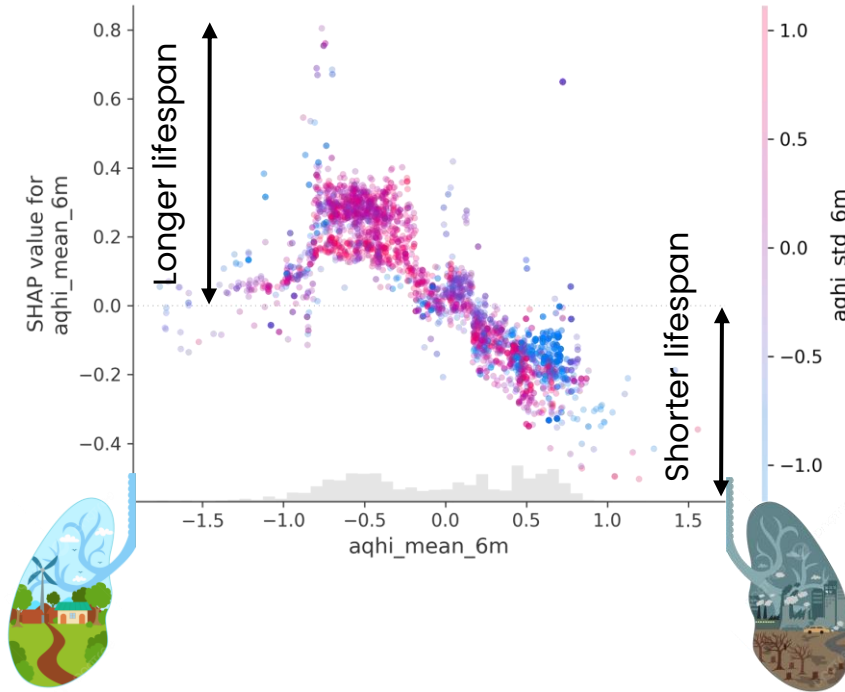
Feature effects

Hives exposed to **poor air quality** have **shorter lifespans**

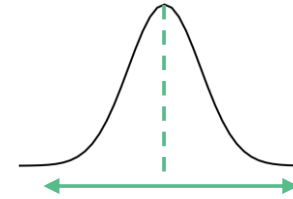
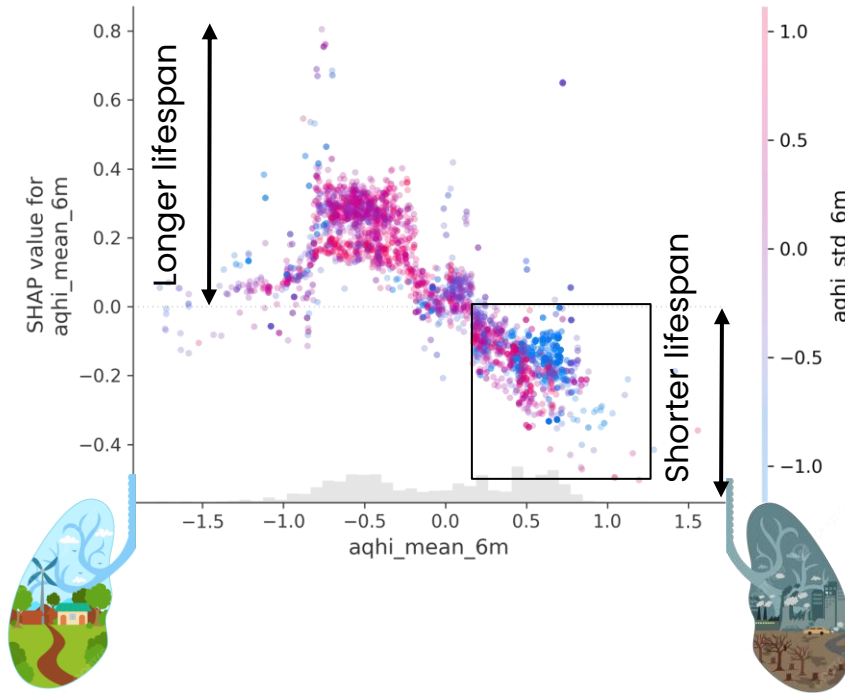
Hives **constantly** exposed to the same **air quality** have **shorter lifespans**



Feature interactions



Feature interactions



Range of exposure is key:

Hives **constantly** exposed to
poor air quality face
increased risks of mortality

Next steps

- ☐ **Integrate landscape features**
- ☐ **Fit the model to a much larger dataset**
- ☐ Collaboration with the **Computational Ecology Research Group** at Université de Montréal
- ☐ **Integrate the feature importance workflow into Nectar** as a decision-making tool available for beekeepers



contact

Thank you! 

maxime@nectar.buzz

