CORRELAID

GOOD CAUSES. BETTER EFFECTS.

Bee Observer

Diren Senger, Alexander Goncharskiy, Alexandros Melemenidis

Diren Senger
UNIVERSITY OF BREMEN

Alexander Goncharskiy
COMMERZBANK AG

Alexandros Melemenidis
EUROPEAN CENTRAL BANK

Introduction - Bee Observer

Introduction - The honey bee

The European Honeybee (*apis mellifera*)

©Guy Pracros / Adobe Stock

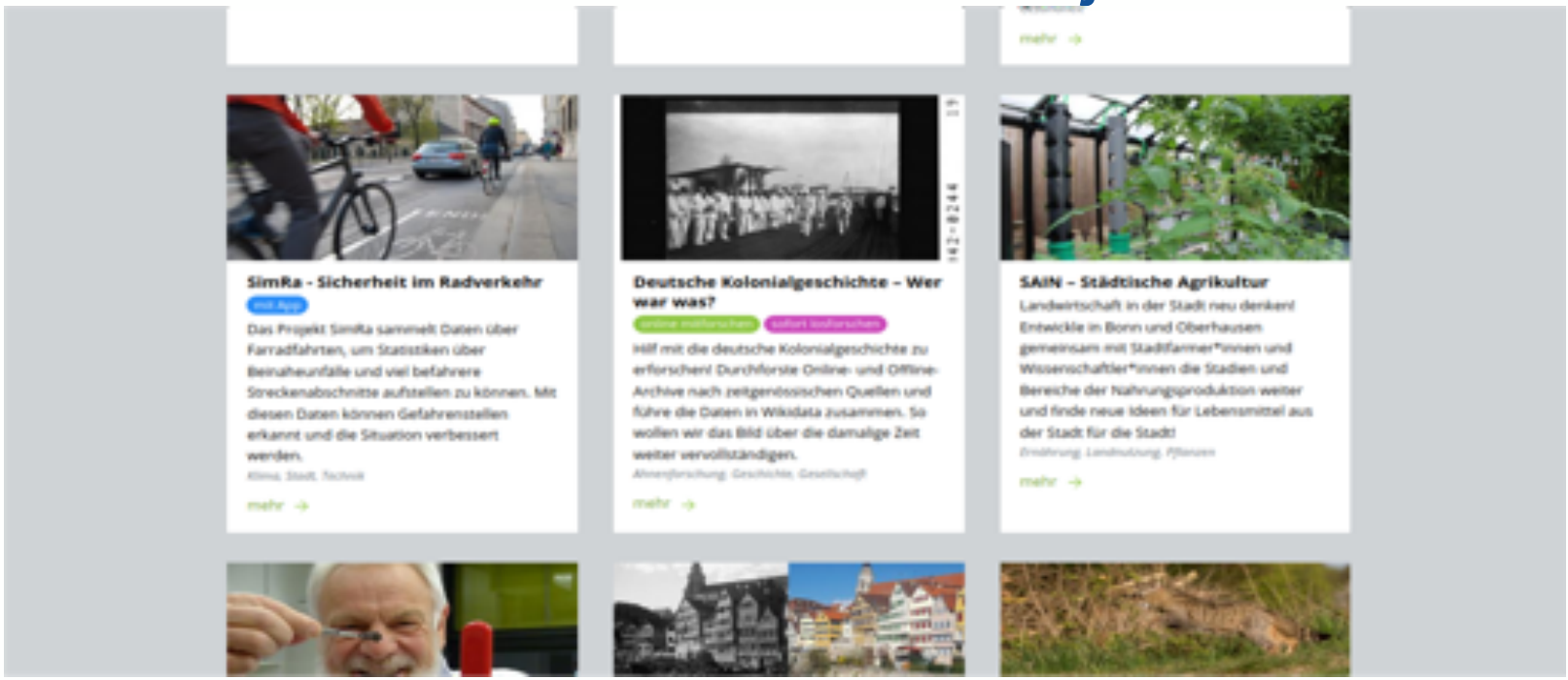

Superorganism
Honey bee colony


Environmental influences
Parasites
Living conditions

Citizen Science – Lead Your Project

Science by nonprofessionals, citizens as data collectors or thinkers

Citizen Science – Find Your Project

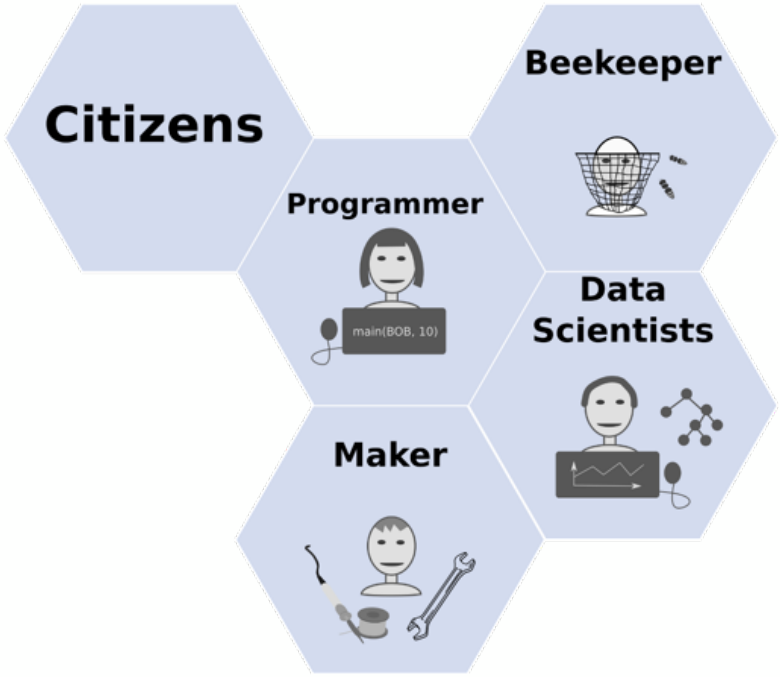


https://www.buergerschaffenwissen.de/projekte

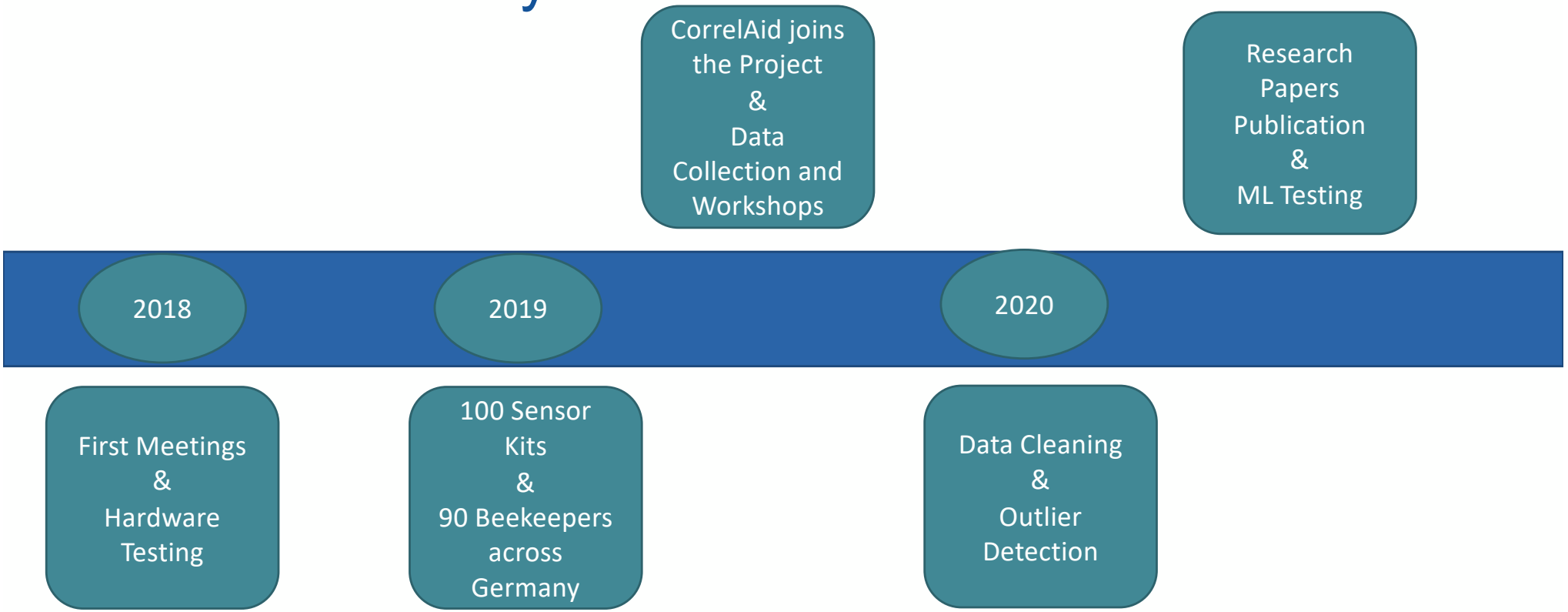
Citizen Science – Challenges and Chances

- Find more relevant research questions
- Make best use of available knowledge and experience
- Collect more data from different locations
- x Mostly no experimental setup
- x How to pay citizens for their work?

Citizen Science – Beeobserver



Beeobserver Story



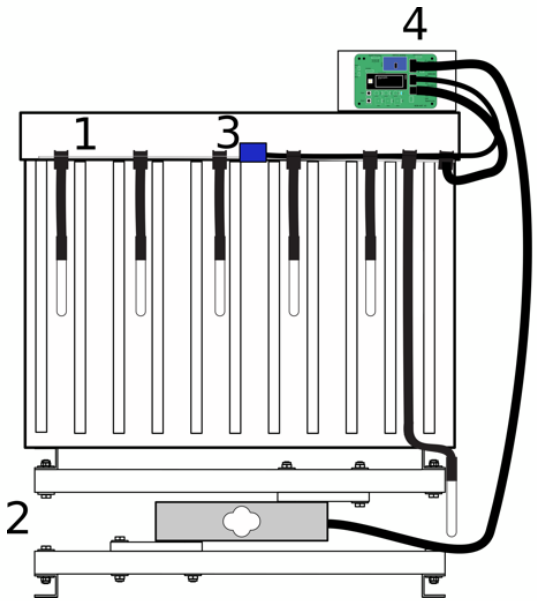
CORRELAID
GOOD CAUSES. BETTER EFFECTS.

Bee Observer - Sensor Data

Data collection since February 2019

Sensor-kits:
1. six thermometers (5 in, 1 out)
2. load-cell (scale)
3. humidity sensor
4. processing unit

Data stored in *influxdb*



Source: Senger et al. (2020)

Bee Observer - Inspections App

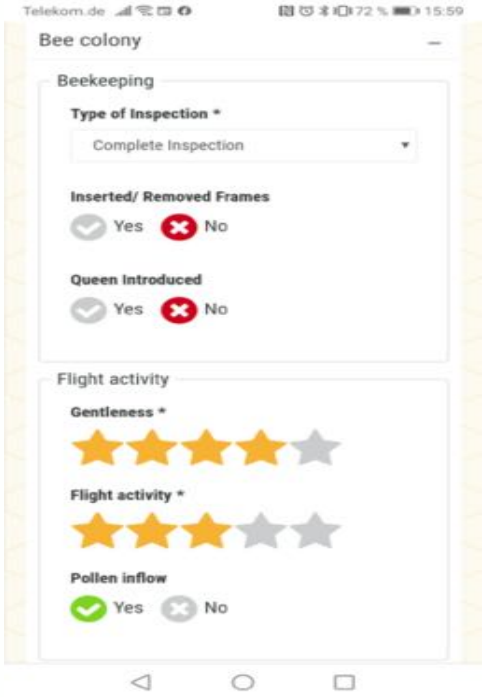
Sensor data is complemented by

Beekeeper observations

- mix of Boolean, scores, free text, categorical and numeric data
- most values voluntary

Apiary metadata

- type of hive, race of queen
- geolocation (voluntary)

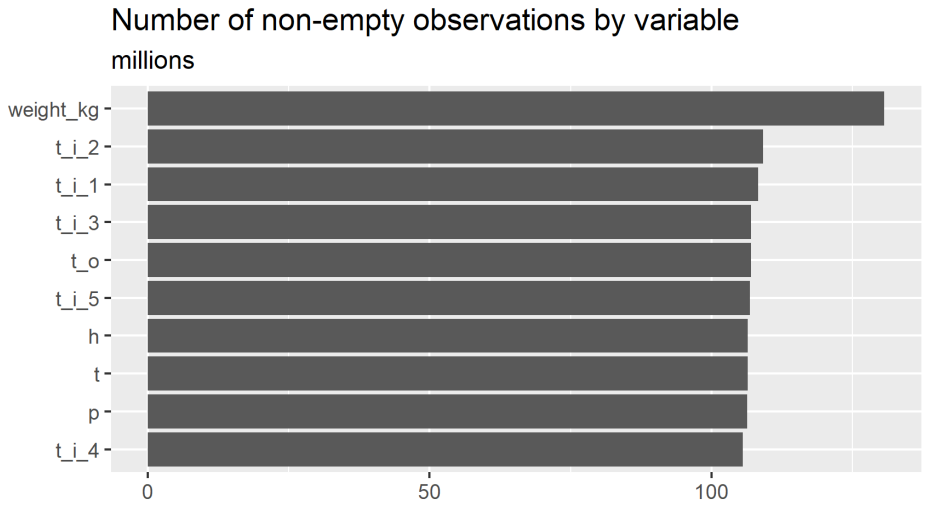


Data Profiling - Sensor Data

hundreds of millions of observations from 129 beehives

BUT

- variation across variables
- regular gaps
- some sensor kits do not have all sensors
- values outside the technical sensor range

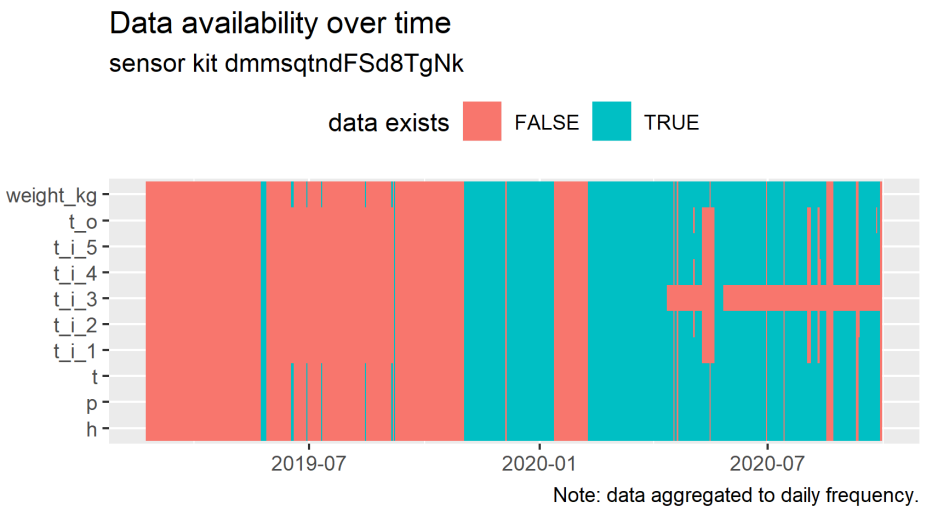


Note: Data until 30 September 2020.

Data Profiling

The main issues in terms of data quality are:

- no data at all (interruption of electricity or WiFi)
- incomplete observations (single components fail)



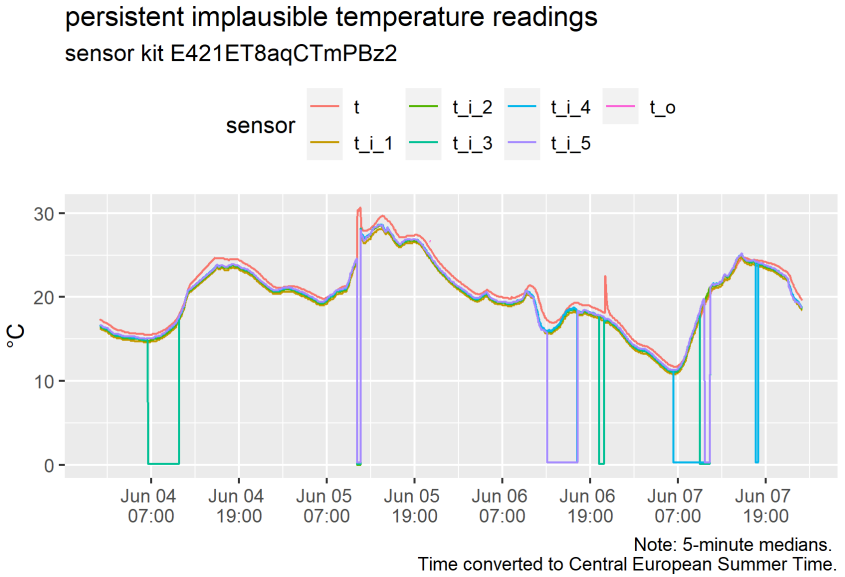
Heatmap of data availability for an example beehive

Data profiling - Data Correctness

The other problem (less frequent) is obviously incorrect data:

- values outside the technical sensor range
- values within the range, but displaying implausible fluctuation (and/or lack thereof)

Some cleaning needed!

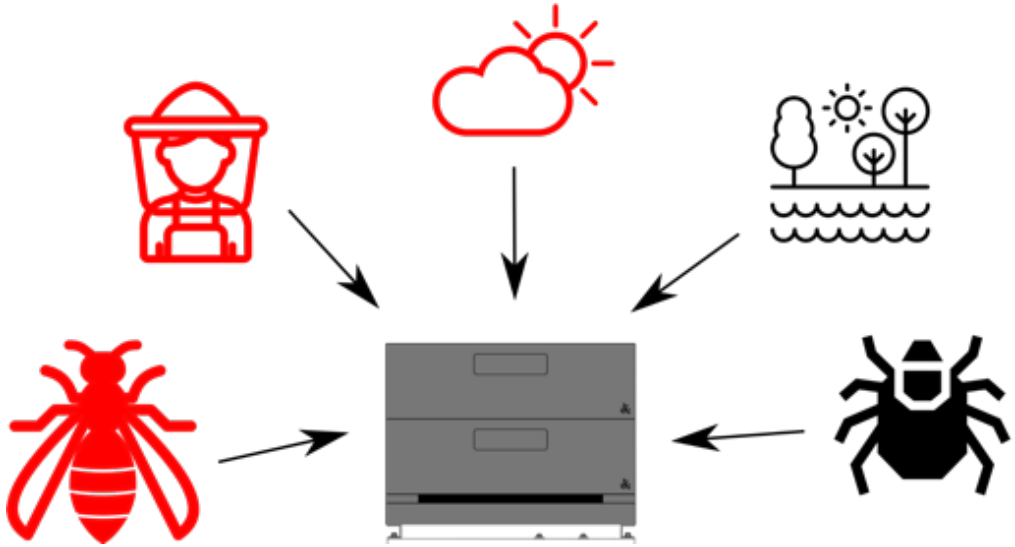


Example for apparent issues with thermometers

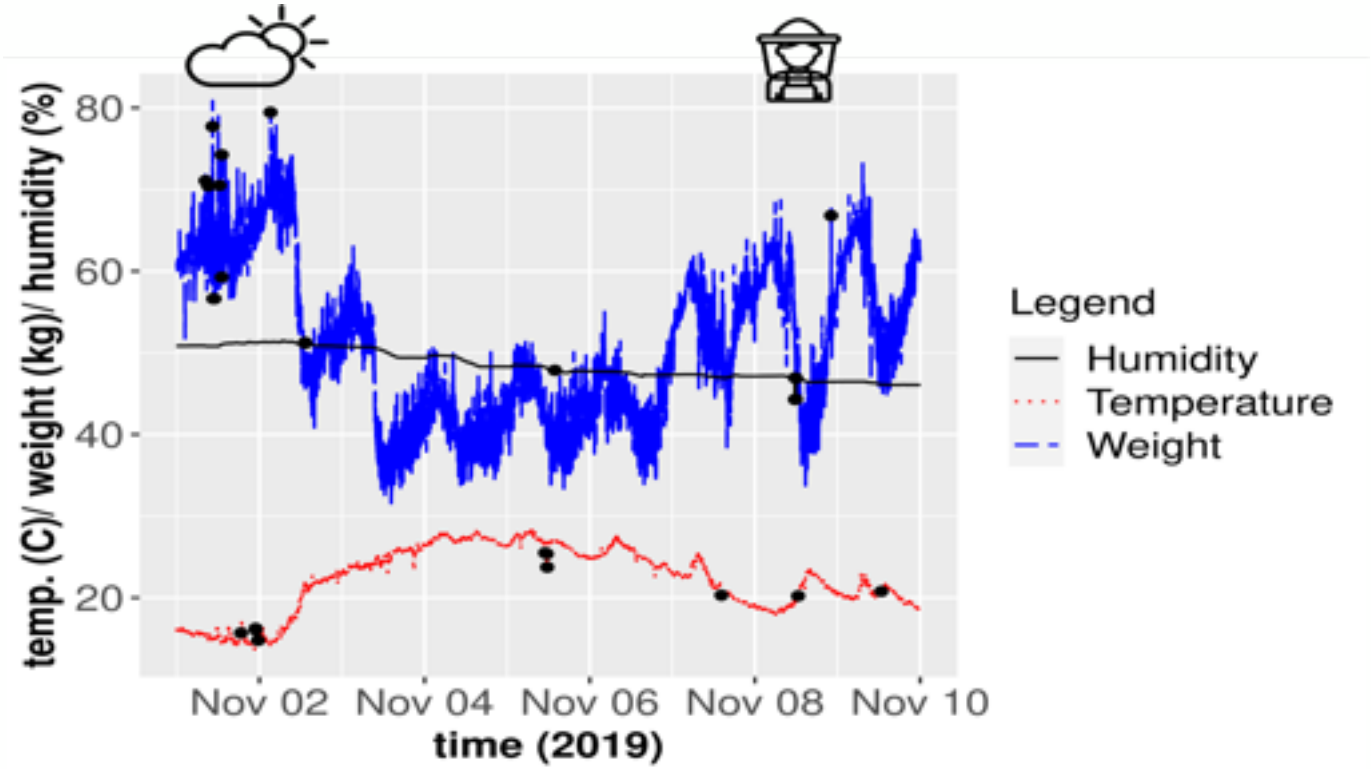
Local Outlier Detection

Data Modelling & Sensor Measurements Predictions

Difference between Prediction and the Actual Value



Local Outlier Detection



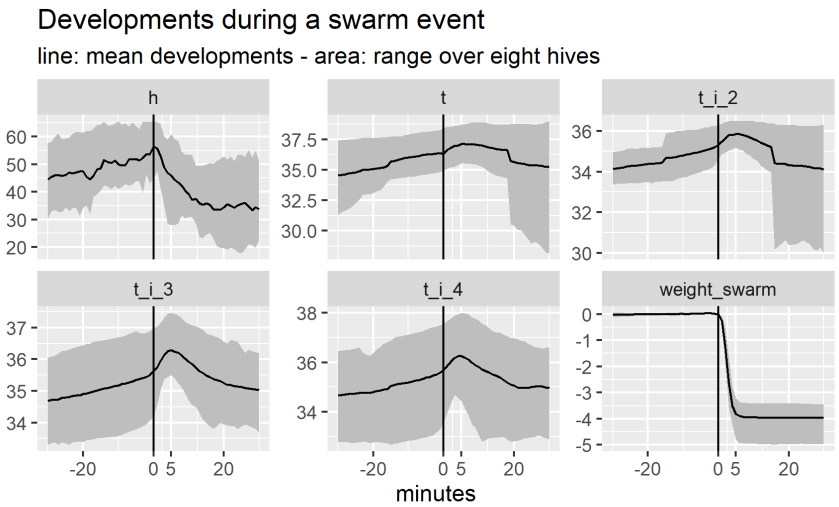
Analytical Use Case for Supervised Learning - Swarm Detection

When colonies swarm, the beekeeper has around one day to react. Otherwise the swarmed part of the colony may die or at least be lost to the beekeeper. Also the “old” part of the colony may be weakened.

Swarms can be characterised by typical developments!

©Thomas Söllner - stock.adobe.com

What happens during a swarm?



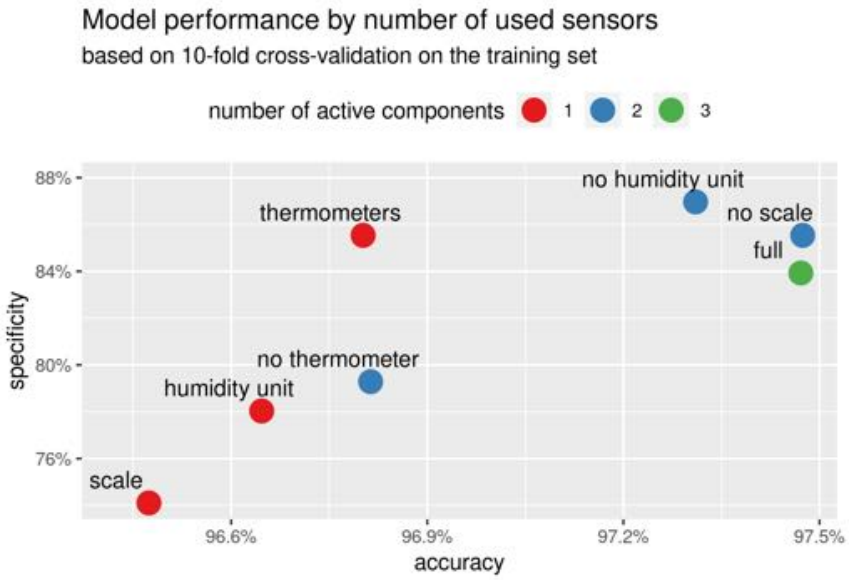
Developments in key variables during a swarm

How do we model?

Non-linear model based on key variables, their growth rates and lags, time of day.

- small sample with missing data -> dimension reduction to maximum thermometer value, but also multiple models on subsets of variables
- unbalanced sample (few observed swarms): up-sampling training set with SMOTE

Random Forest works quite well



Performance metrics of RF models using subsets of sensor components

Q&A

Introduction - Precision beekeeping

The colony is a complex super-organism

- complex organisation
- division of labour
- developments are hard to observe: inspections are stressful, in winter not possible at all

Precision beekeeping

- install **sensors within beehive** for continuous monitoring
- weight, temperature, humidity, gas concentration, video, audio
- first works using simple algorithmic rules
- later first steps using ML approaches
- most based on small samples

Introduction - The honey bee

The European Honeybee (*apis mellifera*)

- important for pollination
- producer of honey
- increasing number of colonies in DE
- but increasing cases of colony losses over the winter

©Guy Pracros / Adobe Stock

Outline

- Topical context - Bees and beekeeping
- Bee Observer as a Citizen Science project
- The data
- ML use cases