



CORRELAID
GOOD CAUSES. BETTER EFFECTS.

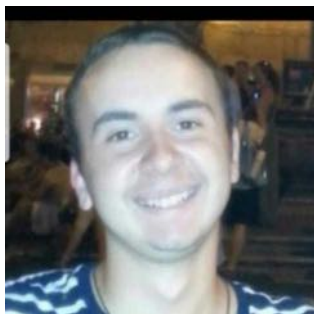
Bee Observer

*Using machine learning to detect
anomalies in beehives*

Diren Senger, Alexander Goncharskiy



Diren Senger
UNIVERSITY OF BREMEN

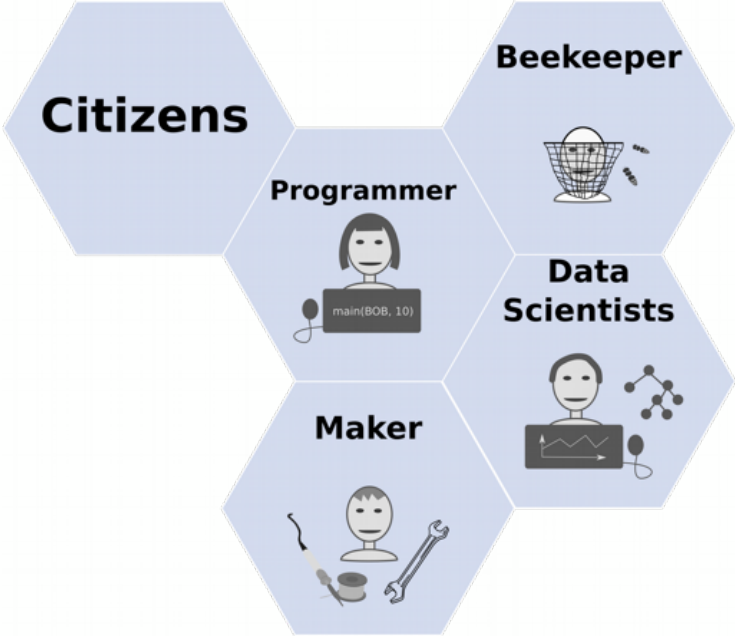


Alexander Goncharskiy
COMMERZBANK AG





Beeobserver – Project Structure



Bee Observer – Project Structure



Bee Observer



CORRELAID

RHEIN-MAIN



Bundesministerium
für Bildung
und Forschung



Universität Bremen

Bee Observer - Sensor Data

Data collection since February 2019

Sensor-kits:

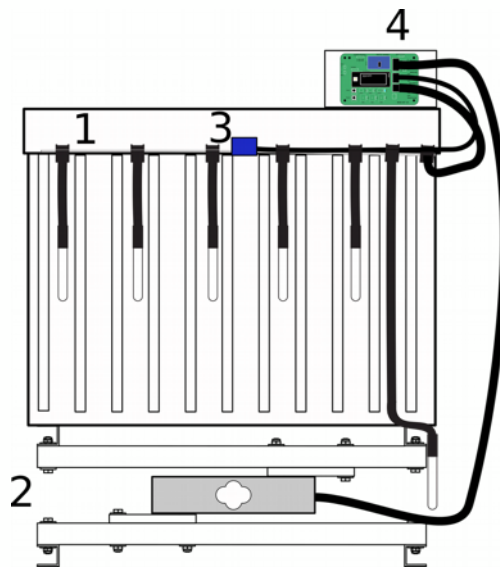
1. six thermometers (5 in, 1 out)

2. load-cell (scale)

3. humidity sensor

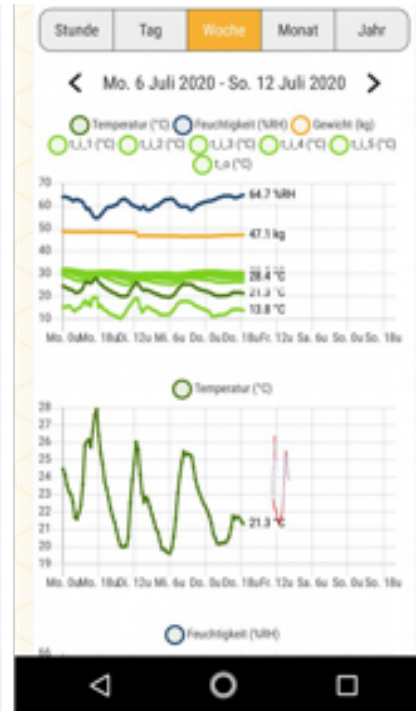
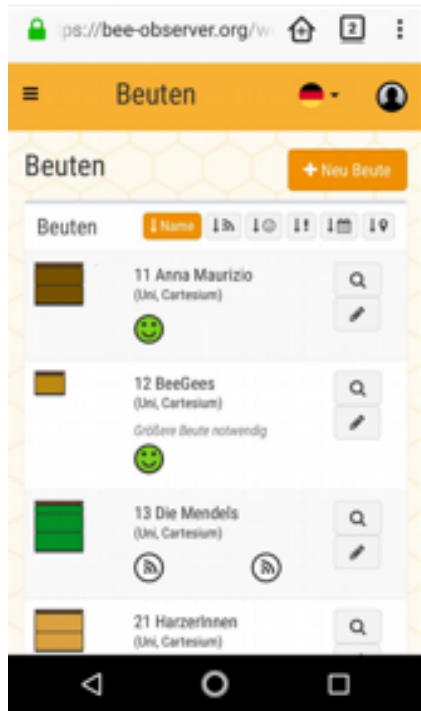
4. processing unit

Data stored in *influxdb*

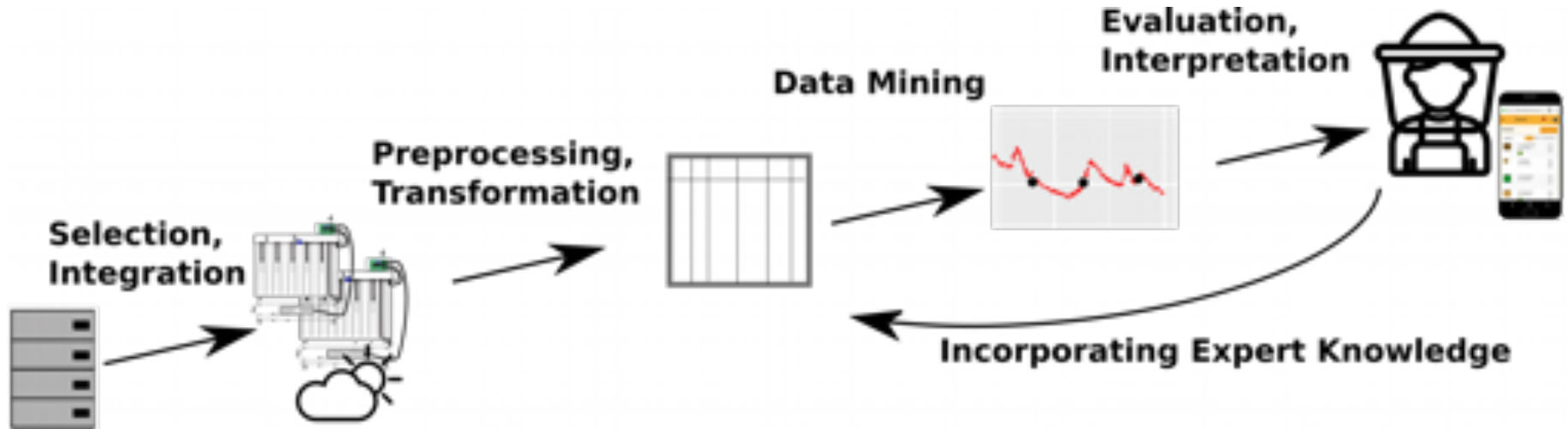


Source: Senger et al. (2020)

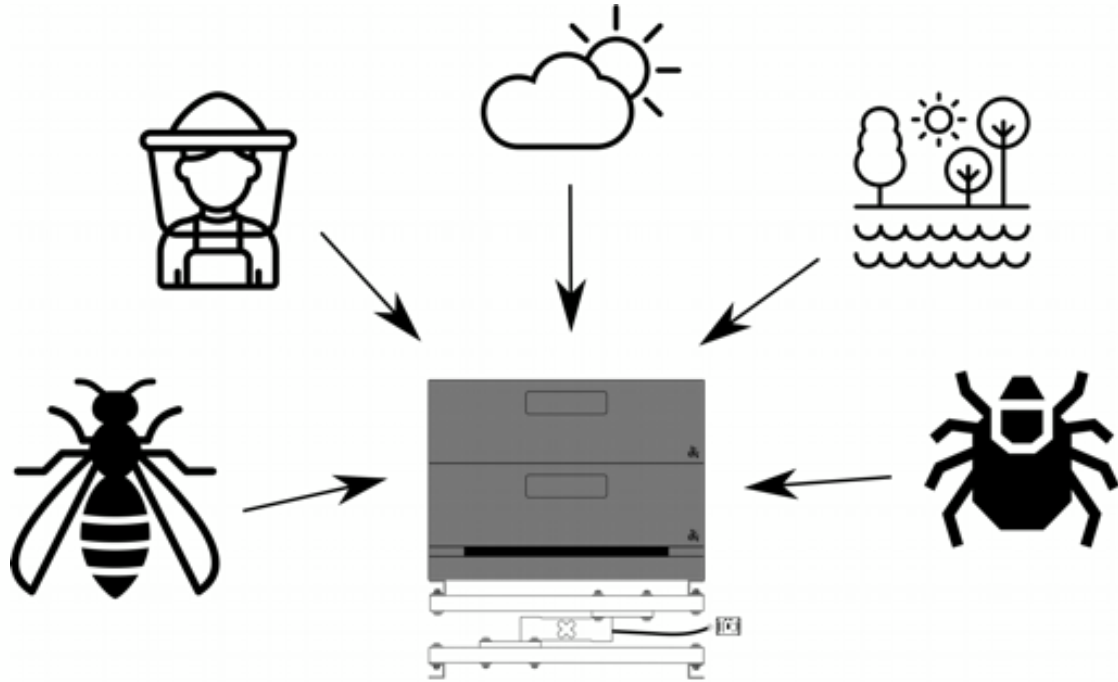
Bee Observer Application



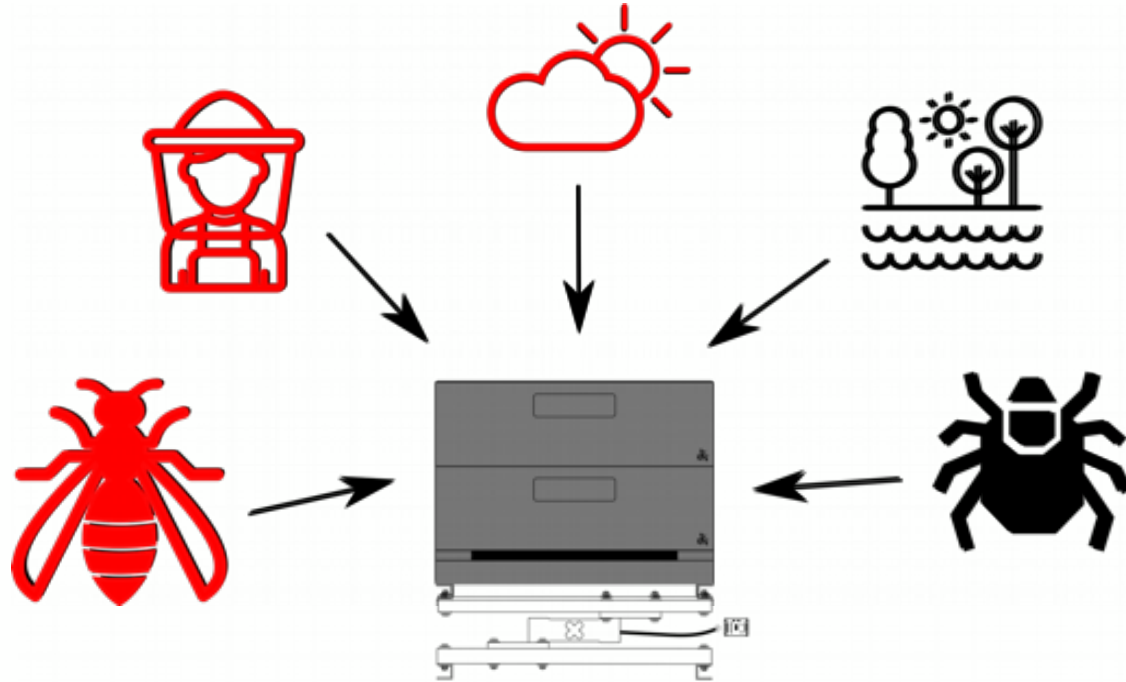
Knowledge Discovery in Datamining



Causes of changes in micro-climate



Causes of local outliers



Outlier Detection

- 1) Modelling and predicting
- 2) Generalised Extreme Student Deviation (GESD) Test

Outlier Detection - Modelling

- Auto-Regressive Integrated Moving Average (ARIMA)
 - Seasonal
 - Multivariate
 - External predictors
- Moving Median
- Seasonal Hybrid ESD

Example: ARIMA (1,1,1)

Difference to last measurement

$$\Delta y_t = \mu + \phi_1 \Delta y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$$

parameter of moving average part

error term

constant

parameter of autoregressive part

ARIMA – full definition

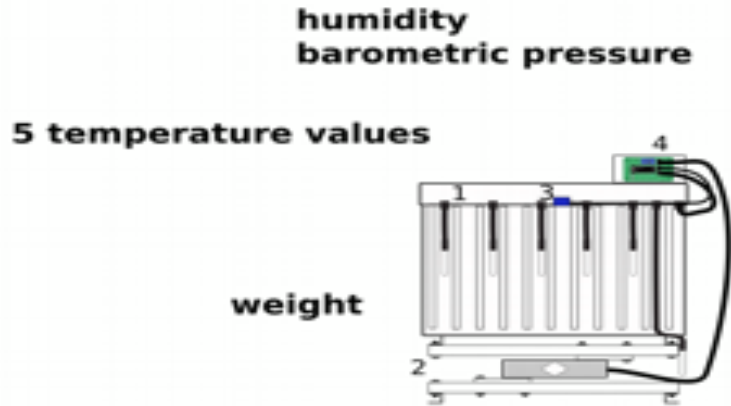
integrated measurement (d) number of lags for autoregressive part parameter of moving average part

$$y_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t$$

constant parameter of autoregressive part error term number of lags for moving average part

ARIMA modelling

Univariate or Multivariate Modelling



Arima with external predictors

outside temperature

precipitation



cloud amount and visibility

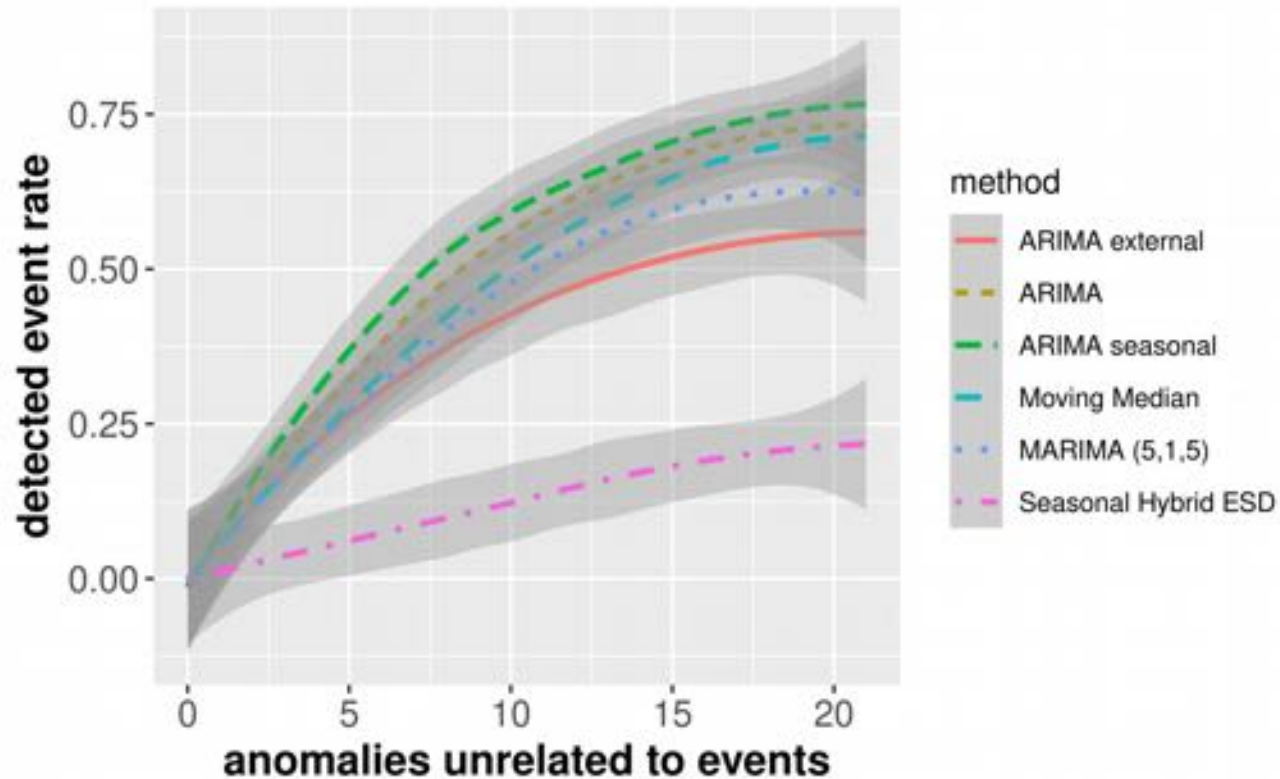
barometric pressure

humidity

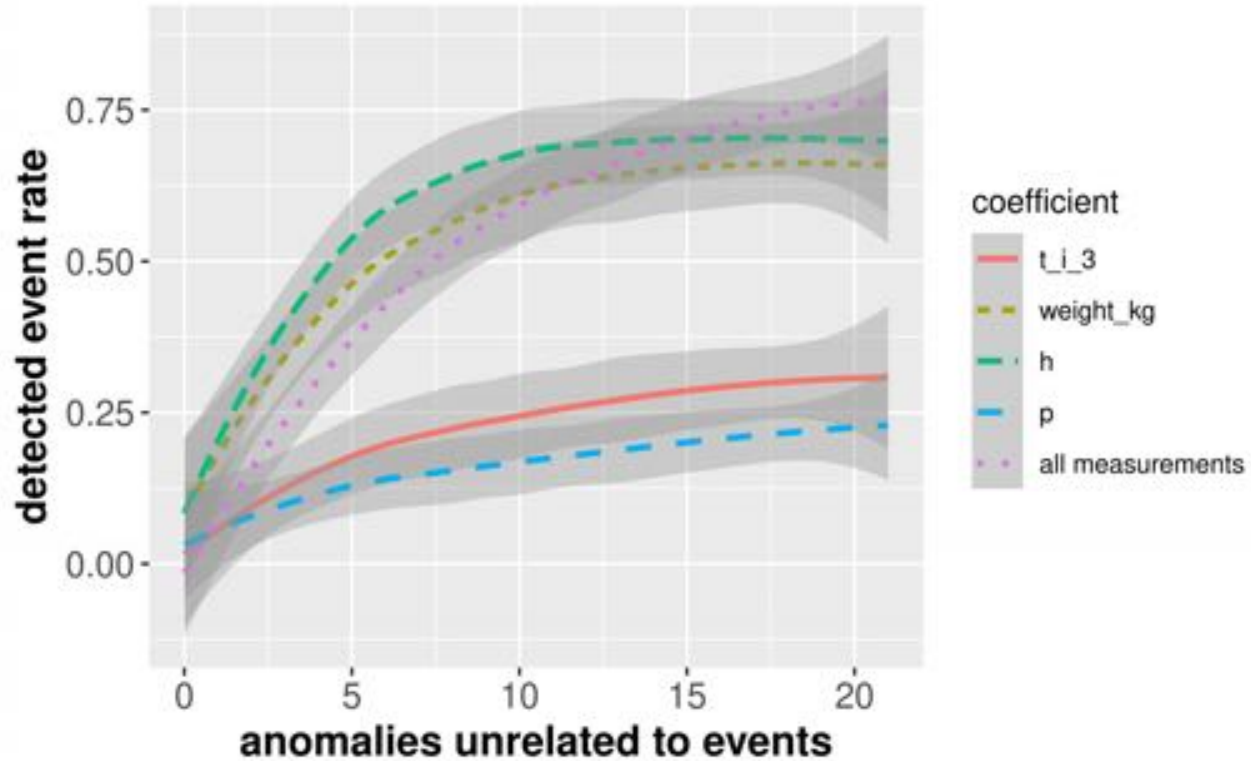


Seasonal Arima

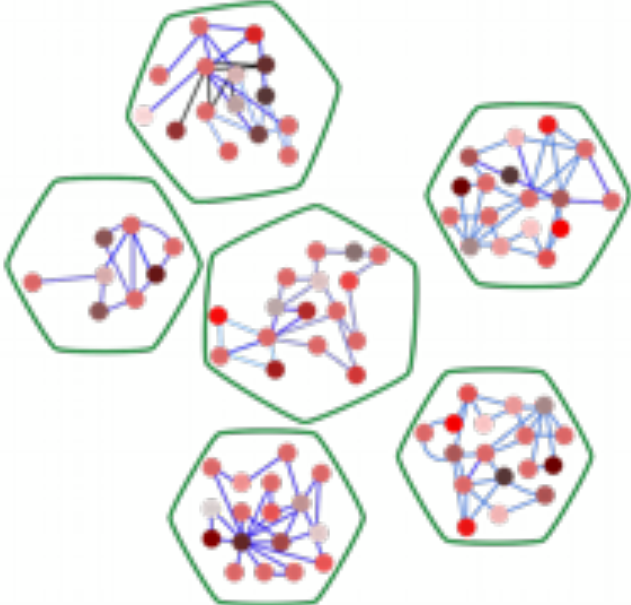
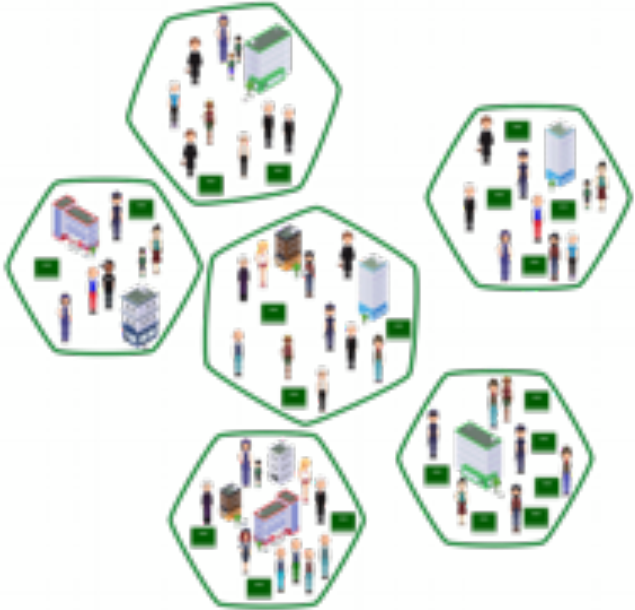
Detected events using all measurements



Detected events using ARIMA seasonal

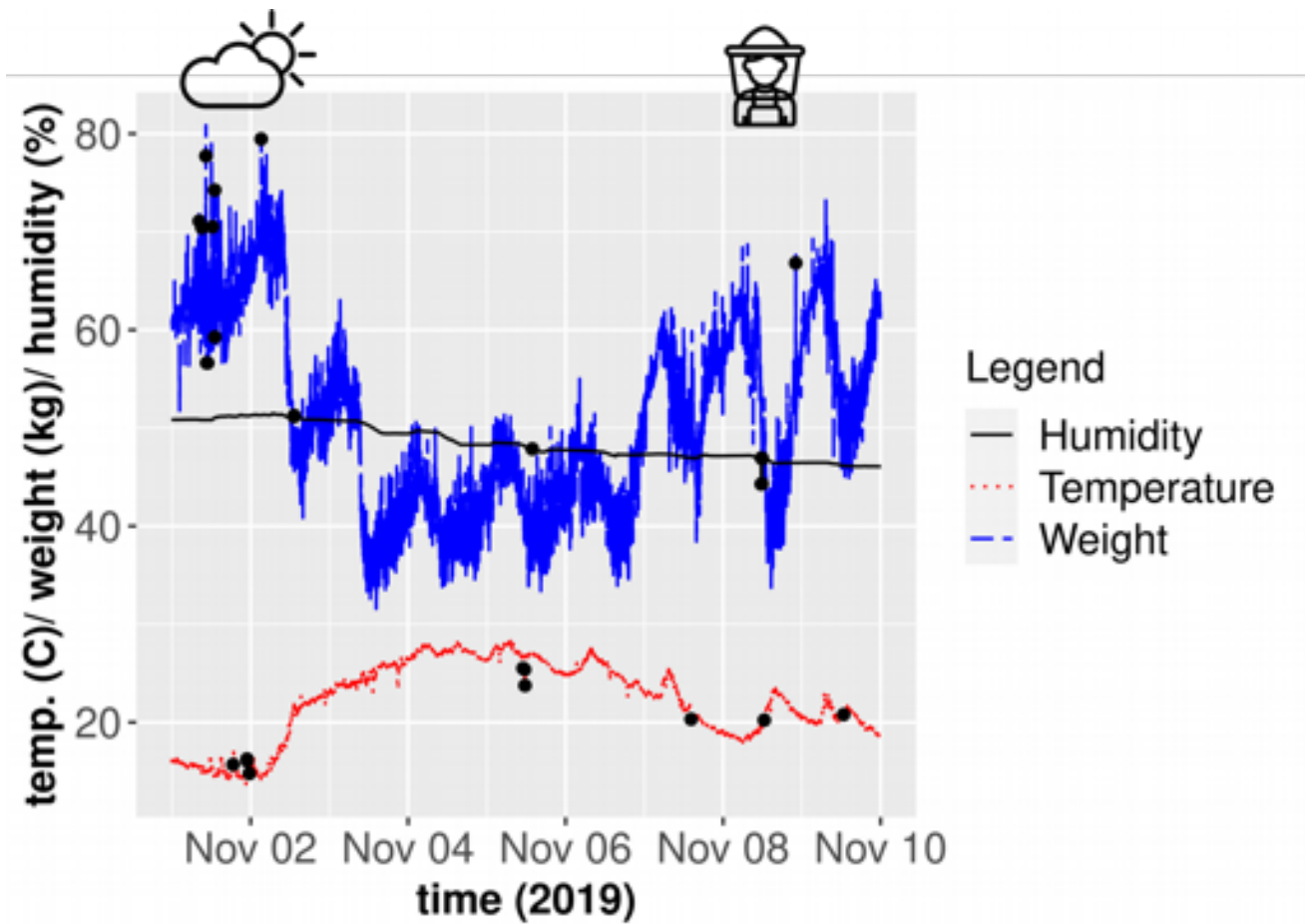


Agent based model



Example: How likely is an inspection?

- **Day and night:** inspections are least likely in the middle of the night
- **Interval:** an interval of 7 to 10 days between inspections is most likely
- **Colony health:** it is more likely that the beekeeper will inspect a hive in poor health than a fit one



Analytical Use Case for Supervised Learning - Swarm Detection

- During an swarm event half of the colony emerges from the hive with the old queen bee
- A beekeeper might loose half of their bees

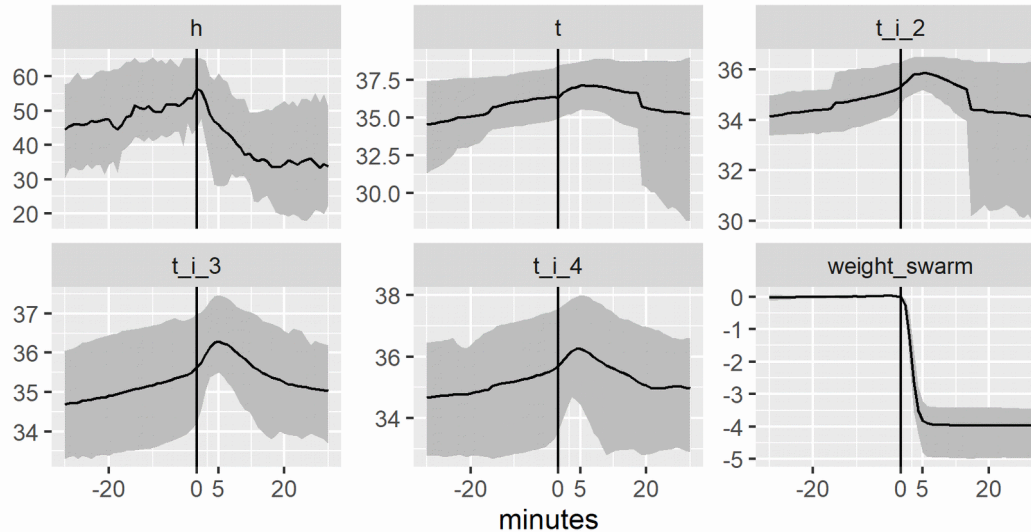


©Thomas Söllner - stock.adobe.com

What happens during a swarm?

Developments during a swarm event

line: mean developments - area: range over eight hives



Notes: The vertical line denotes the start of the colony leaving the hive.

The weight variable was standardised to zero at the beginning of the event.

Outlook

- Integrate visualisation of outliers in App
- Look at other abnormal trends
- Include more information about flowering period and habitat structure
- ...

Q&A



Bee Observer



CORRELAID

RHEIN-MAIN

In Deutschland gibt es

33.300 Insektenarten ...

... das sind ca.

**70 % aller
Tierarten!**

42 % der Insektenarten gelten als bestandsgefährdet, extrem selten oder bereits ausgestorben!

© BMU



Bei **45 %** der Insektenarten ist der Bestand rückläufig ...

... z. B. bei

96 % der Köcherfliegen,


62,5 % der Tagfalter,

60,2 % der Ameisen,

42,6 % der Großschmetterlinge

und **41,8 %** der Wildbienen.

Gesamtwerk: BMU | Ameisen: dmpark/Shutterstock.com | Bienen: Aurelija Džider/Shutterstock.com | Pflanzen/Schmetterlinge: Val_leva/Shutterstock.com | Motte, Hummel, Marienkäfer: Olga Oboia/Shutterstock.com



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



SimRa - Sicherheit im Radverkehr

[Web App](#)

Das Projekt SimRa sammelt Daten über Fahrradfahrten, um Statistiken über Beinaheunfälle und viel befahrene Streckenabschnitte aufstellen zu können. Mit diesen Daten können Gefahrenstellen erkannt und die Situation verbessert werden.

Klima, Stadt, Technik

[mehr →](#)



Deutsche Kolonialgeschichte – Wer war was?

[Online recherchieren](#)

[Lokal recherchieren](#)

Hilf mit die deutsche Kolonialgeschichte zu erforschen! Durchforste Online- und Offline-Archive nach zeitgenössischen Quellen und führe die Daten in Wikidata zusammen. So wollen wir das Bild über die damalige Zeit weiter vervollständigen.

Altenforschung, Geschichte, Gesellschaft

[mehr →](#)

[zusammen
mehr →](#)



SAIN – Städtische Agrikultur

Landwirtschaft in der Stadt neu denken
Entwickle in Bonn und Oberhausen
gemeinsam mit Stadtlerner*innen und
Wissenschaftler*innen die Stadten
und Bereiche der Nahrungsproduktion weiter
und finde neue Ideen für Lebensmittel aus
der Stadt für die Stadt

Ernährung, Landwirtschaft, Pflanzen

[mehr →](#)

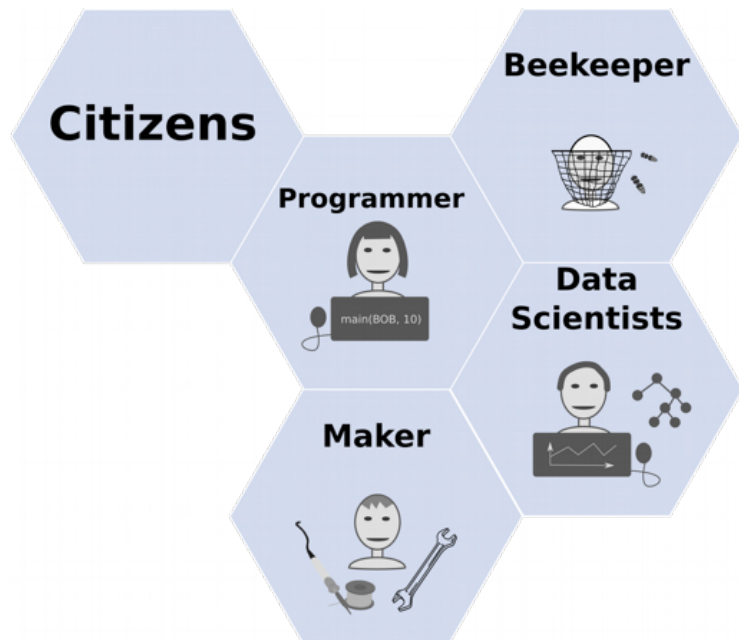


Citizen Science – Challenges and Chances

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



Citizen Science – Beeobserver



Beeobserver Story

CorrelAid joins
the Project
&
Data Collection
and
Workshops

Research
Papers
Publication
&
ML Testing

2018

2019

2020

First Meetings
&
Hardware
Testing

20 Sensor Kits
&
90 Beekeepers
across
Germany

Data Cleaning
&
Outlier
Detection



Bee Observer - Sensor Data

Data collection since February 2019

Sensor-kits:

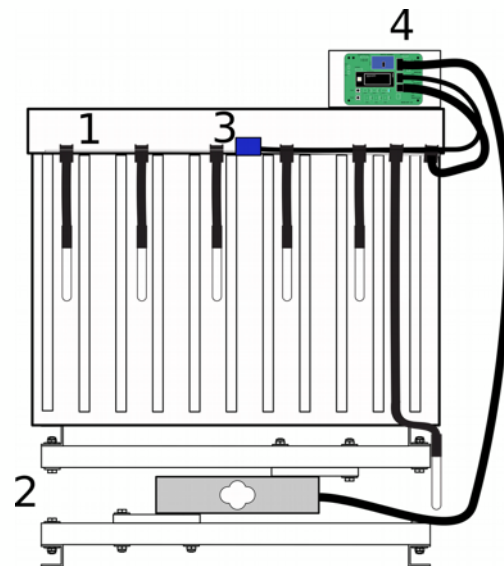
1. six thermometers (5 in, 1 out)

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)

The screenshot shows a mobile application interface for recording beekeeping observations. The title is "Bee colony". The form is divided into three sections:

- Beekeeping**:
 - Type of Inspection ***: A dropdown menu currently showing "Complete Inspection".
 - Inserted/ Removed Frames**: Two radio buttons, "Yes" (checked) and "No" (unchecked).
 - Queen Introduced**: Two radio buttons, "Yes" (checked) and "No" (unchecked).
- Flight activity**:
 - Gentleness ***: A 5-star rating system with 4 stars filled (yellow) and 1 star empty (grey).
 - Flight activity ***: A 5-star rating system with 3 stars filled (yellow) and 2 stars empty (grey).
 - Pollen inflow**: Two radio buttons, "Yes" (checked) and "No" (unchecked).

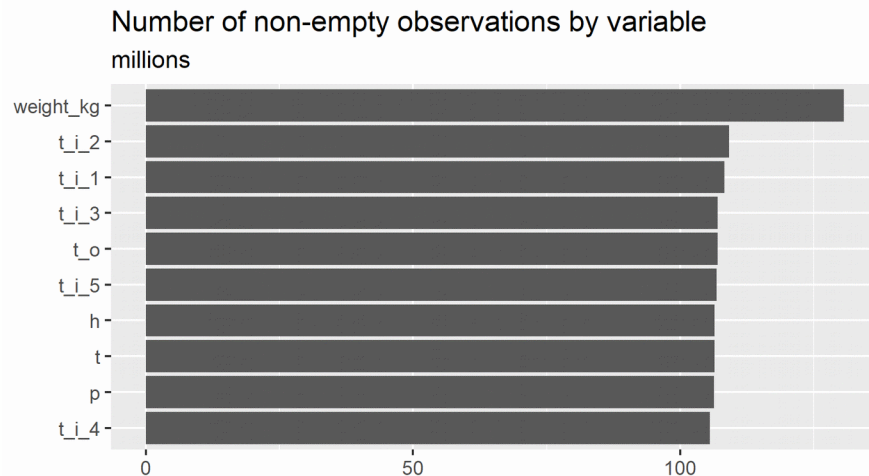
The interface includes a status bar at the top with "Telekom.de", signal strength, Wi-Fi, and battery icons, and a bottom navigation bar with back, home, and recent apps icons.

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

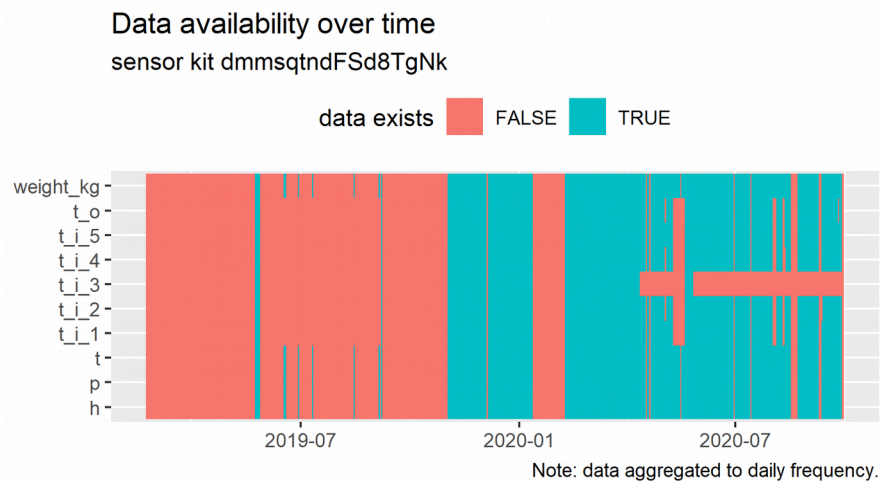


Note: Data until 30 September 2020.

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)



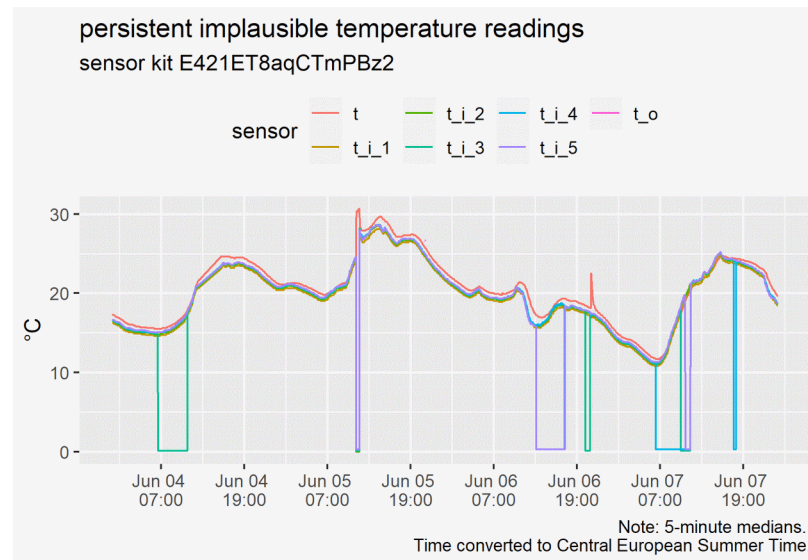
Heatmap of data availability for an example beehive

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

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!

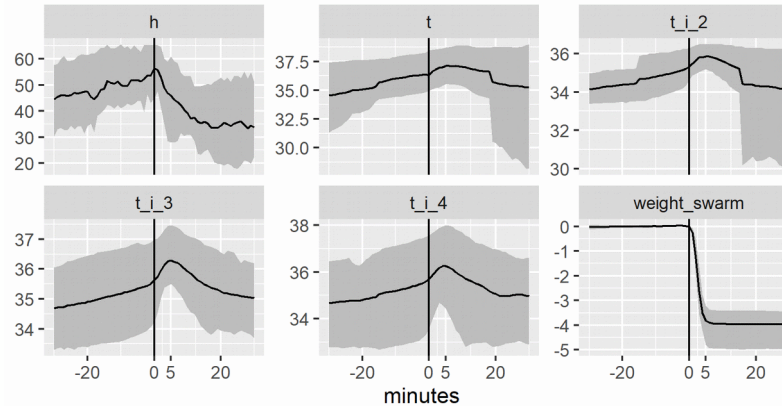


©Thomas Söllner - stock.adobe.com

What happens during a swarm?

Developments during a swarm event

line: mean developments - area: range over eight hives



Notes: The vertical line denotes the start of the colony leaving the hive.

The weight variable was standardised to zero at the beginning of the event.

Developments in key variables during a swarm

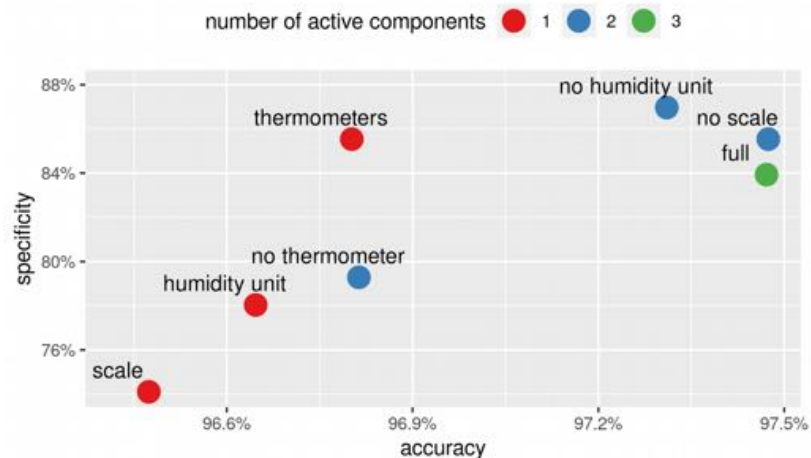
How do we model?

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

Model performance by number of used sensors
based on 10-fold cross-validation on the training set

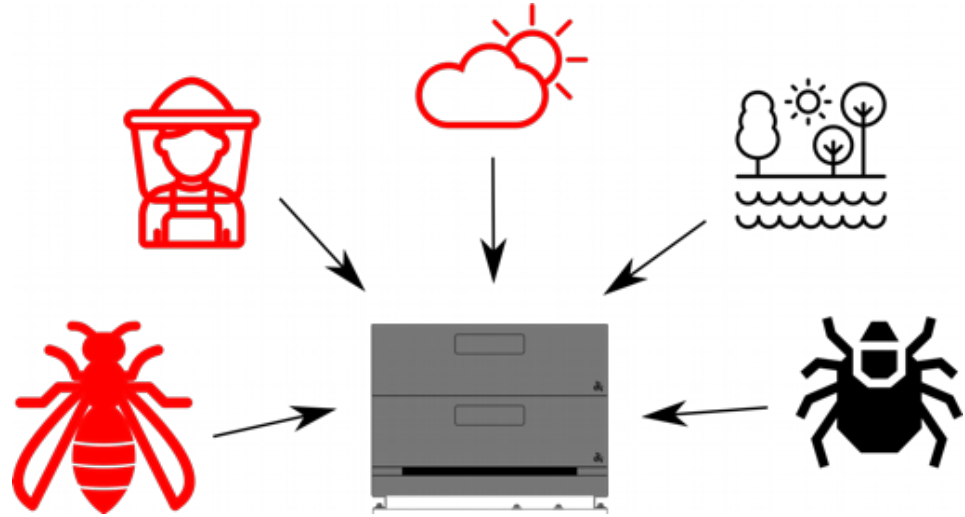


Performance metrics of RF models using subsets of sensor components

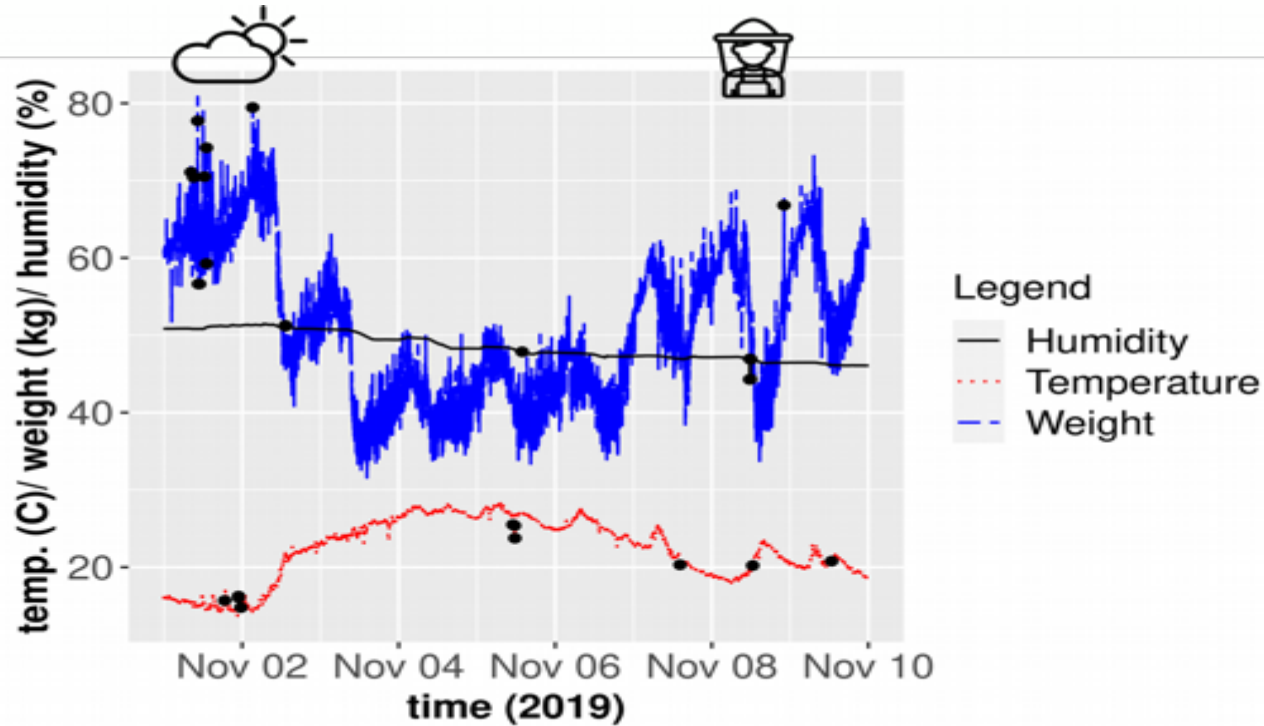
Local Outlier Detection

Data Modelling & Sensor
Measurements Predictions

Difference between Prediction
and the Actual Value



Local Outlier Detection



Q&A



Bee Observer



CORRELAID
RHEIN-MAIN

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