CORRELAID
GOOD CAUSES. BETTER EFFECTS.

Bee Observer

Using machine learning to detect anomalies in beehives

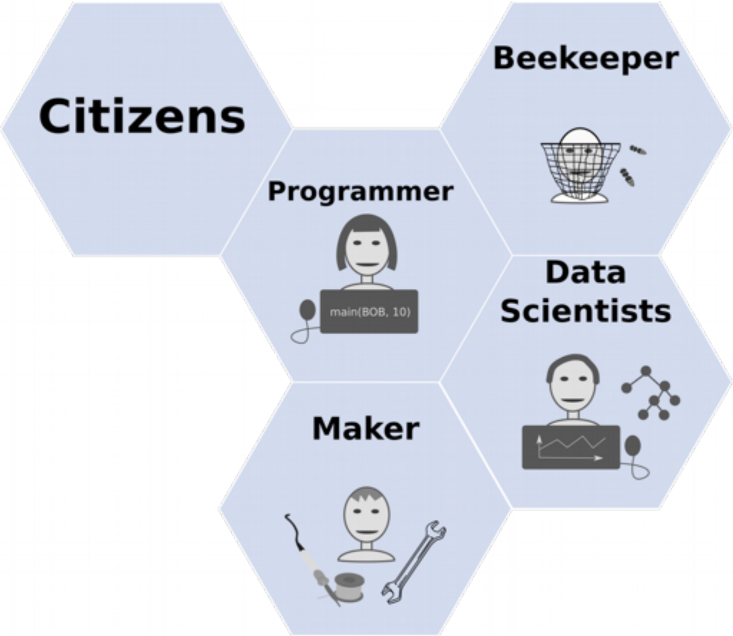
Diren Senger, Alexander Goncharskiy

Diren Senger
UNIVERSITY OF BREMEN

Alexander Goncharskiy
COMMERZBANK AG

CORRELAID
GOOD CAUSES. BETTER EFFECTS.

Beeobserver – Project Structure



Bee Observer – Project Structure

Bee Observer

CORRELAID
RHEIN-MAIN

Bundesministerium
für Bildung
und Forschung

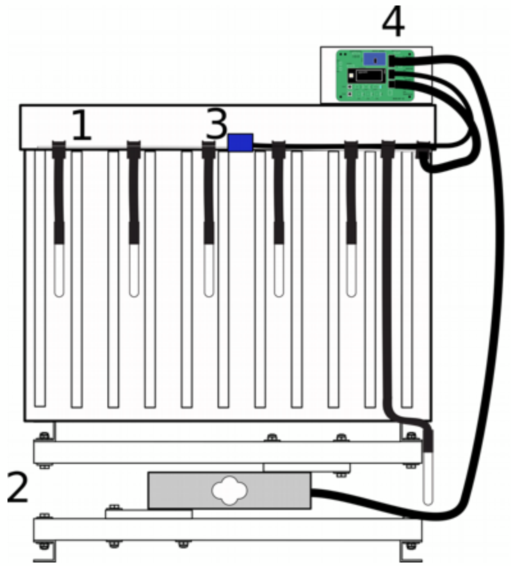
Universität Bremen

Bee Observer - Sensor Data

Data collection since February 2019

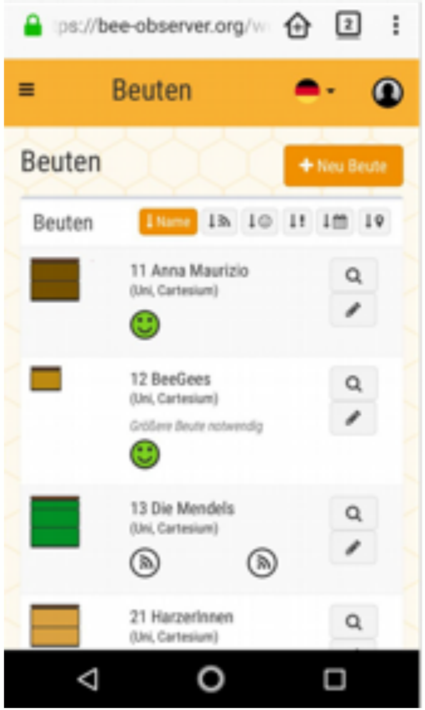
Sensor-kits:
1. six thermometers (5 in, 1 out)
2. load-cell (scale)
3. humidity sensor
4. processing unit

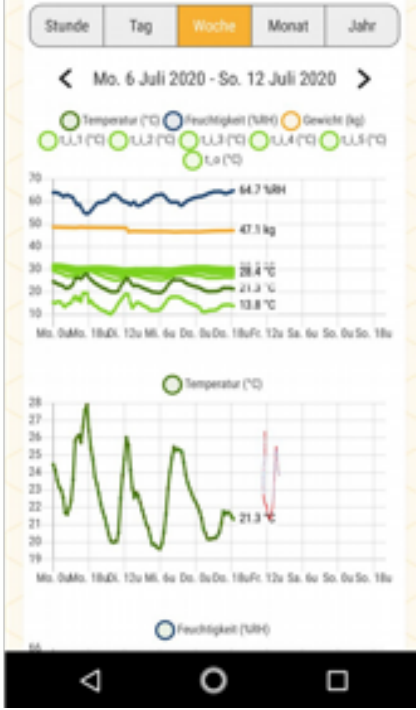
Data stored in *influxdb*



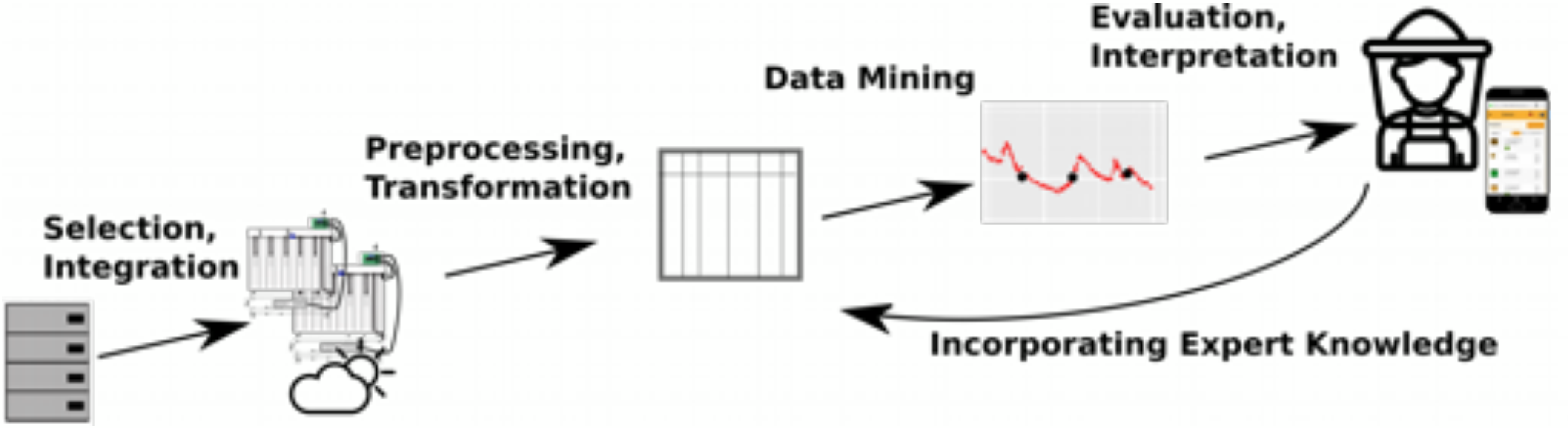
Source: Senger et al. (2020)

Bee Observer Application

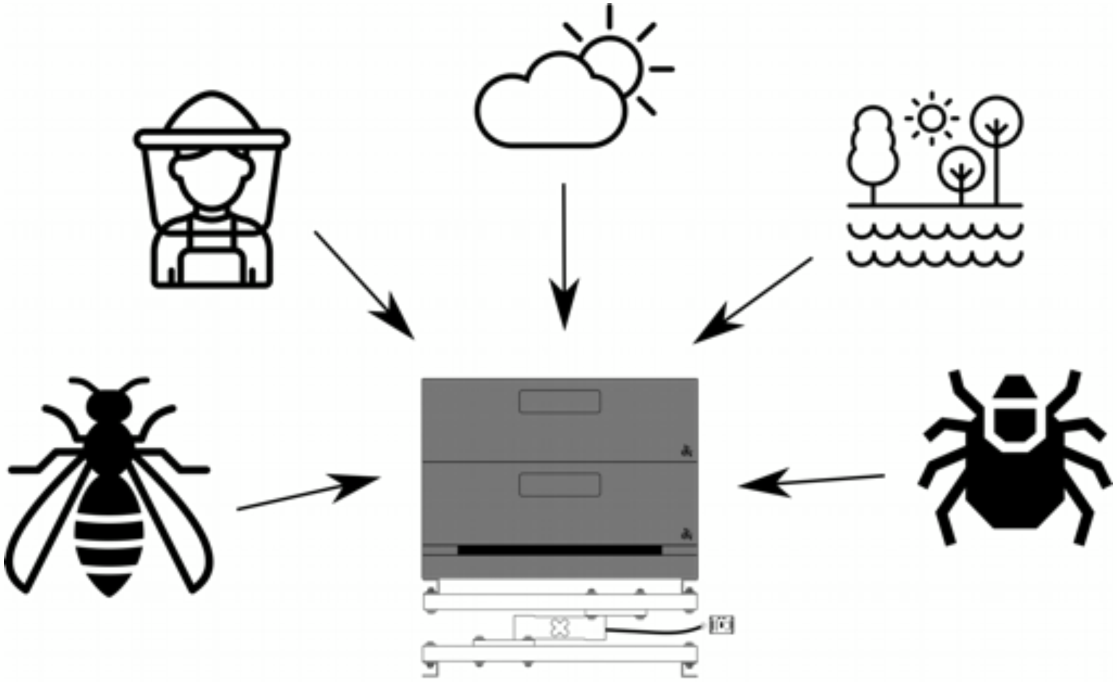




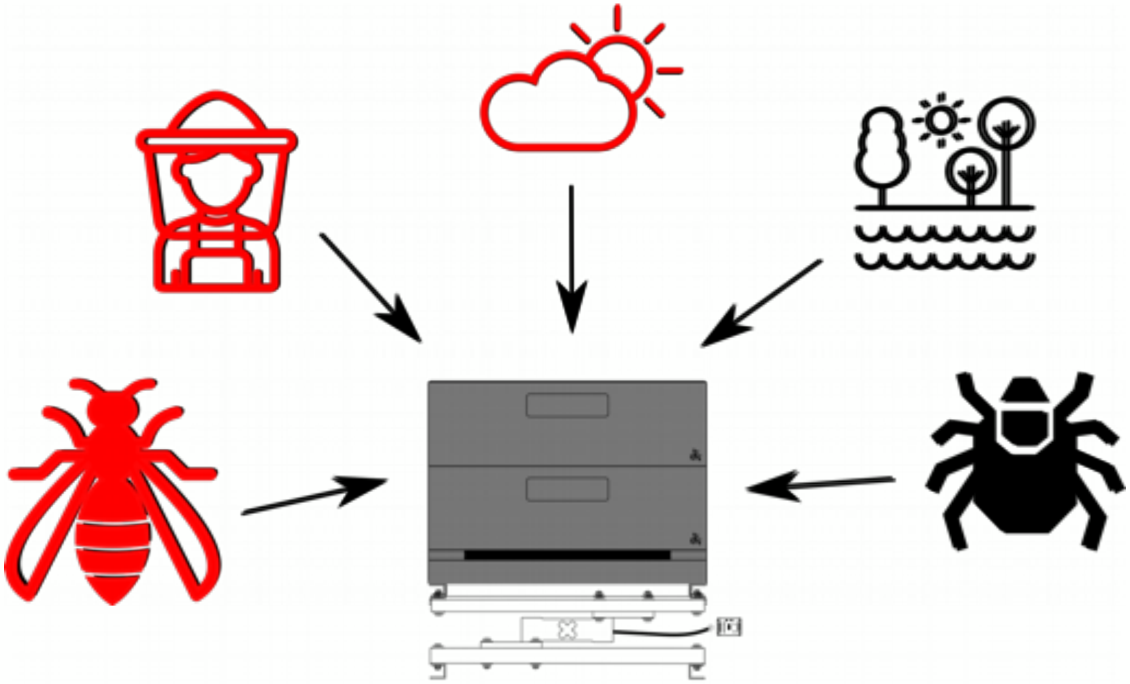
Knowledge Discovery in Datamining



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

parameter of moving average part

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

error term

constant

parameter of autoregressive part

ARIMA – full definition

integrated measurement (d)

number of lags for autoregressive part

parameter of moving average part

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

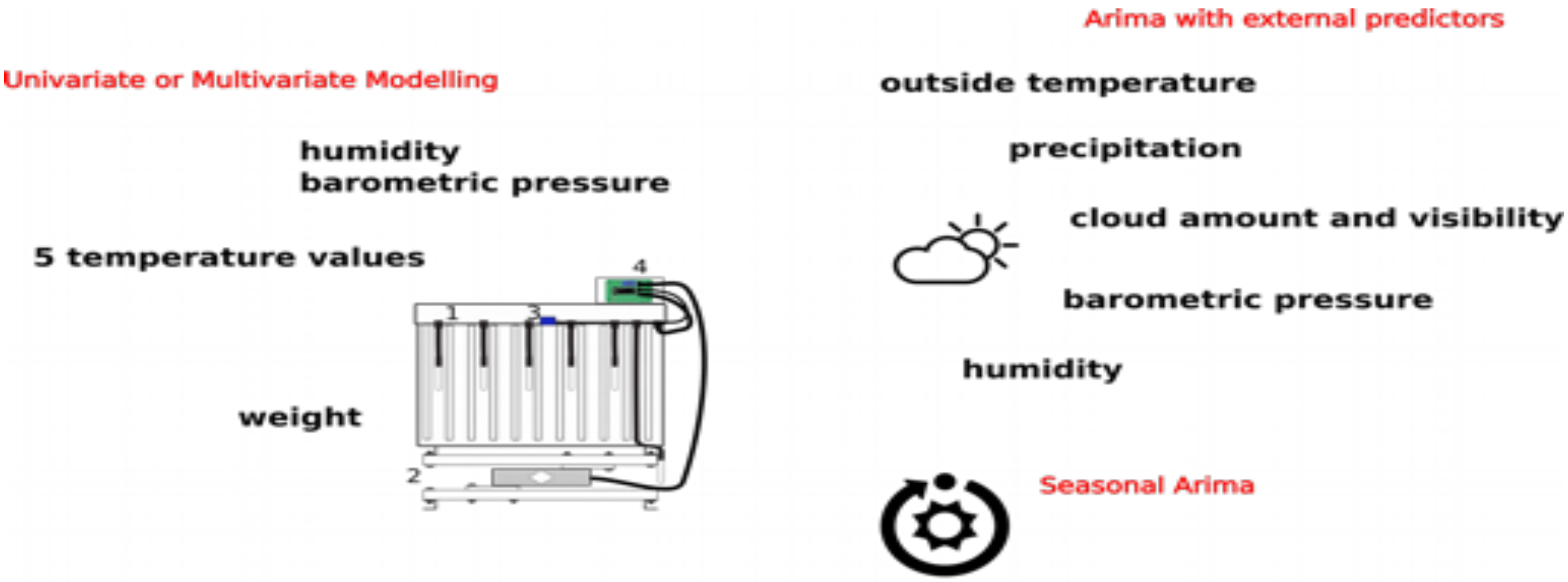
error term

constant

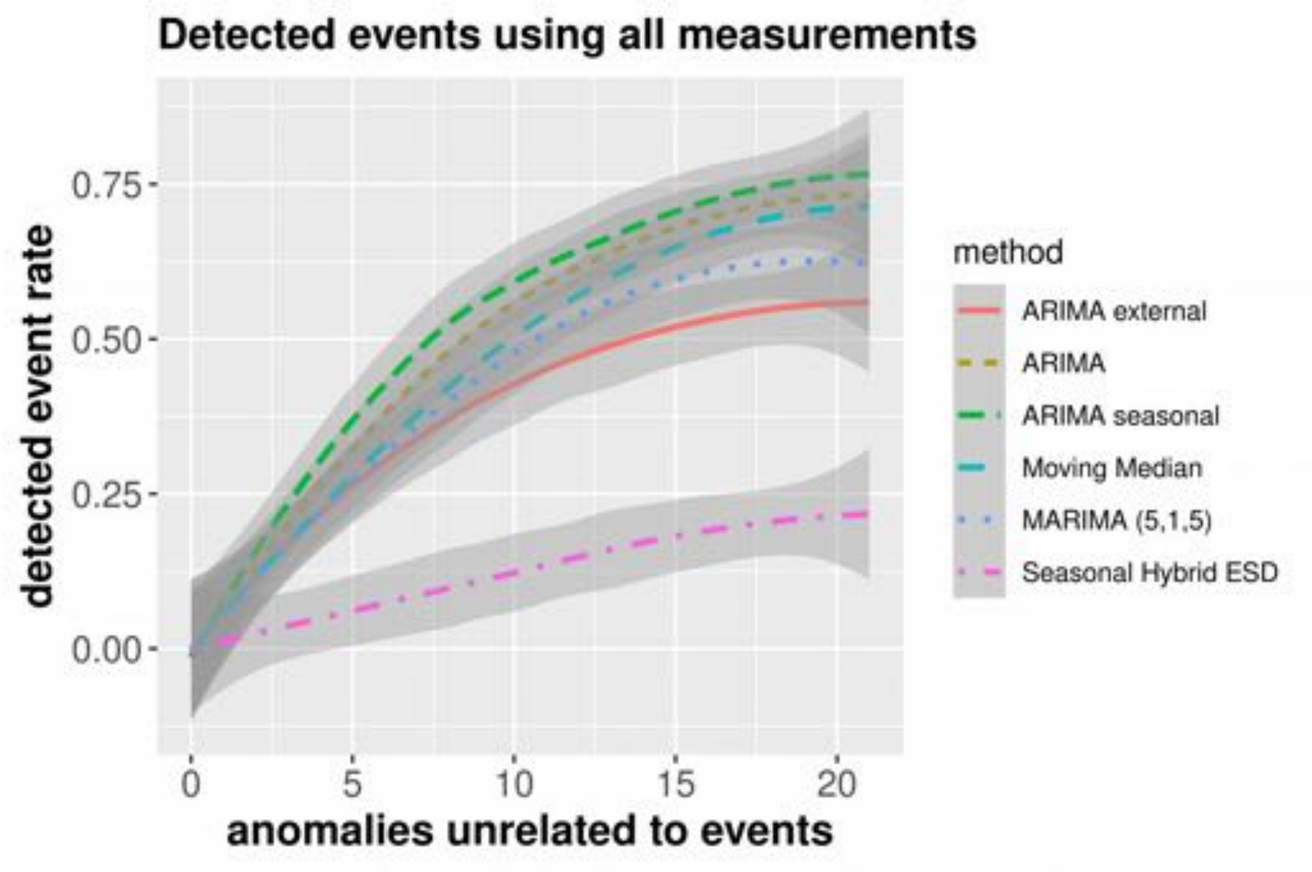
parameter of autoregressive part

number of lags for moving average part

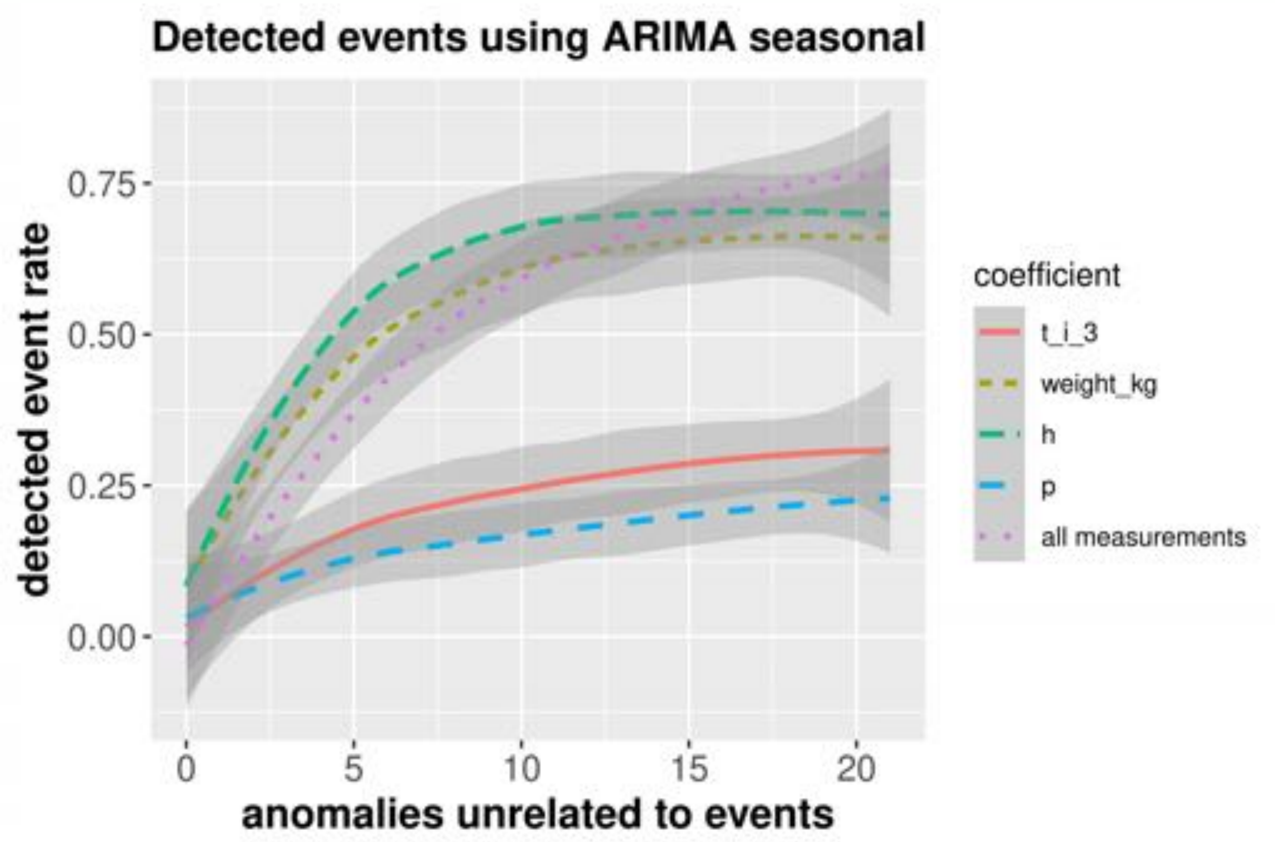
ARIMA modelling



cloud and sun, by Mihail Zhukov, Creative Commons, year, by Zach Bogart, US, Creative Commons

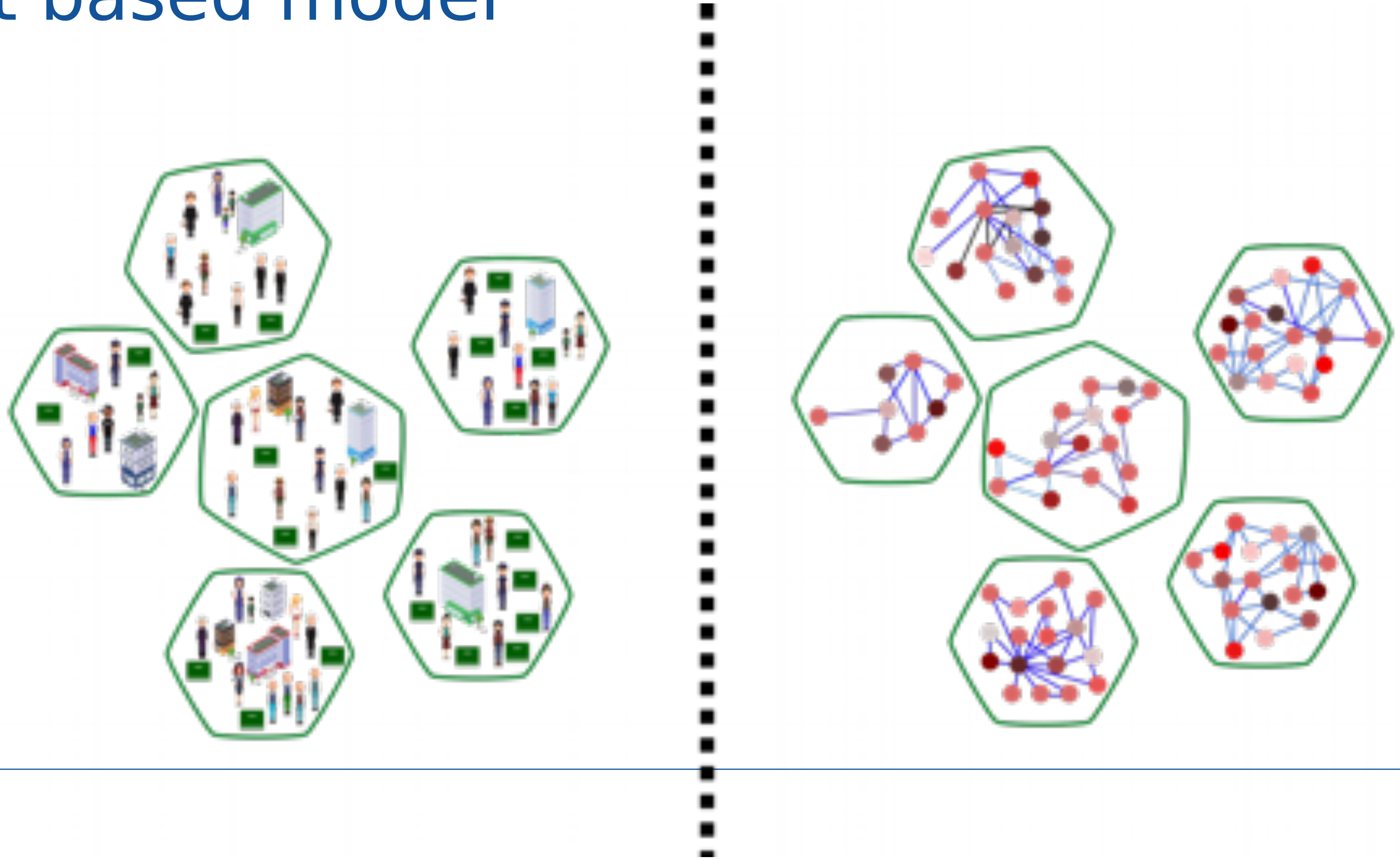


Detected events using all measurements
detected event rate
0.75
0.50
0.25
0.00
0
5
10
15
20
anomalies unrelated to events
method
ARIMA external
ARIMA
ARIMA seasonal
Moving Median
MARIMA (5,1,5)
Seasonal Hybrid ESD



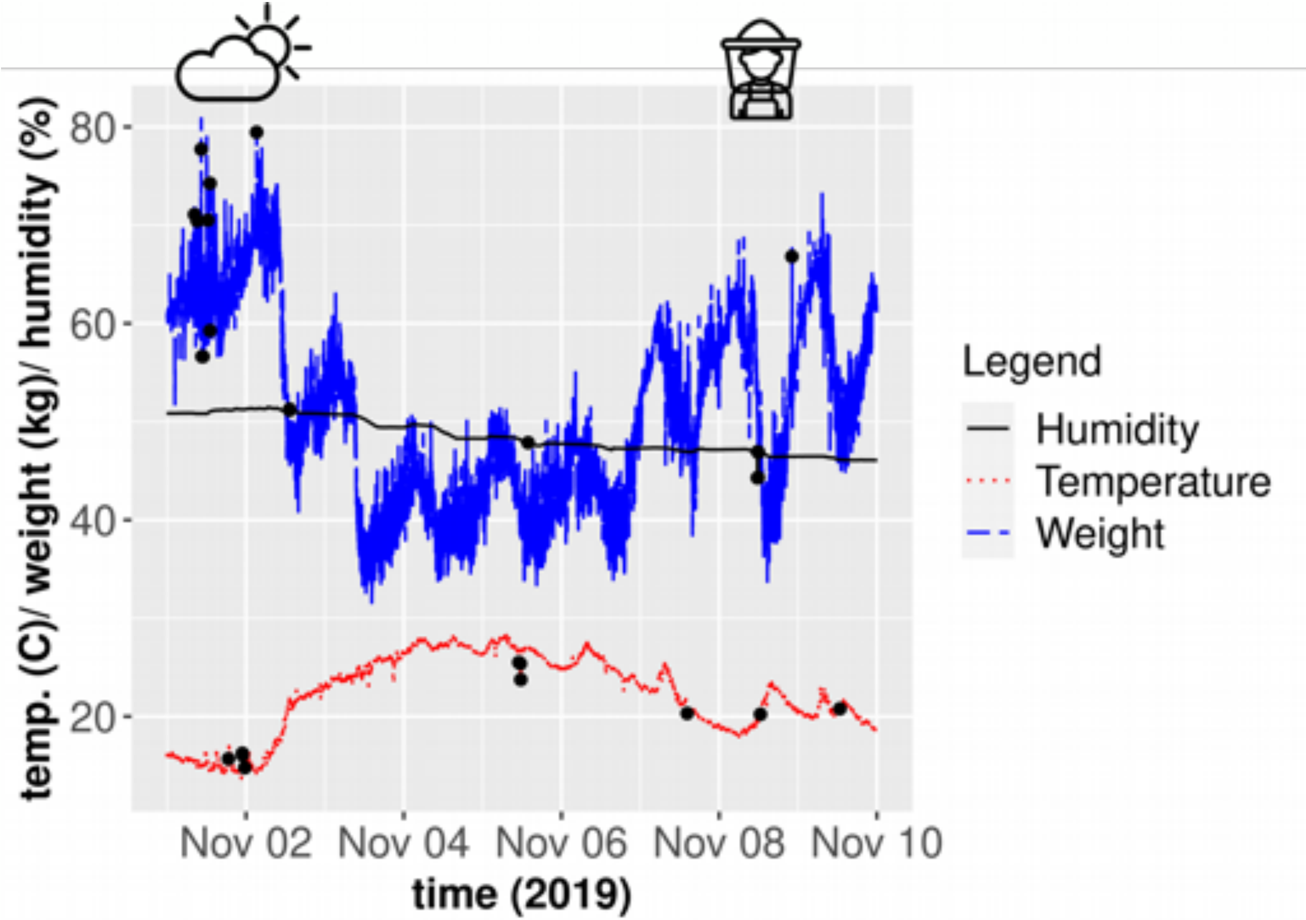
CORRELAID
GOOD CAUSES. BETTER EFFECTS.

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 then a fit one


80
60
40
20
temp. (C)/ weight (kg)/ humidity (%)
Nov 02
Nov 04
Nov 06
Nov 08
Nov 10
time (2019)
Legend
Humidity
Temperature
Weight

Analytical Use Case for Supervised Learning - Swarm Detection

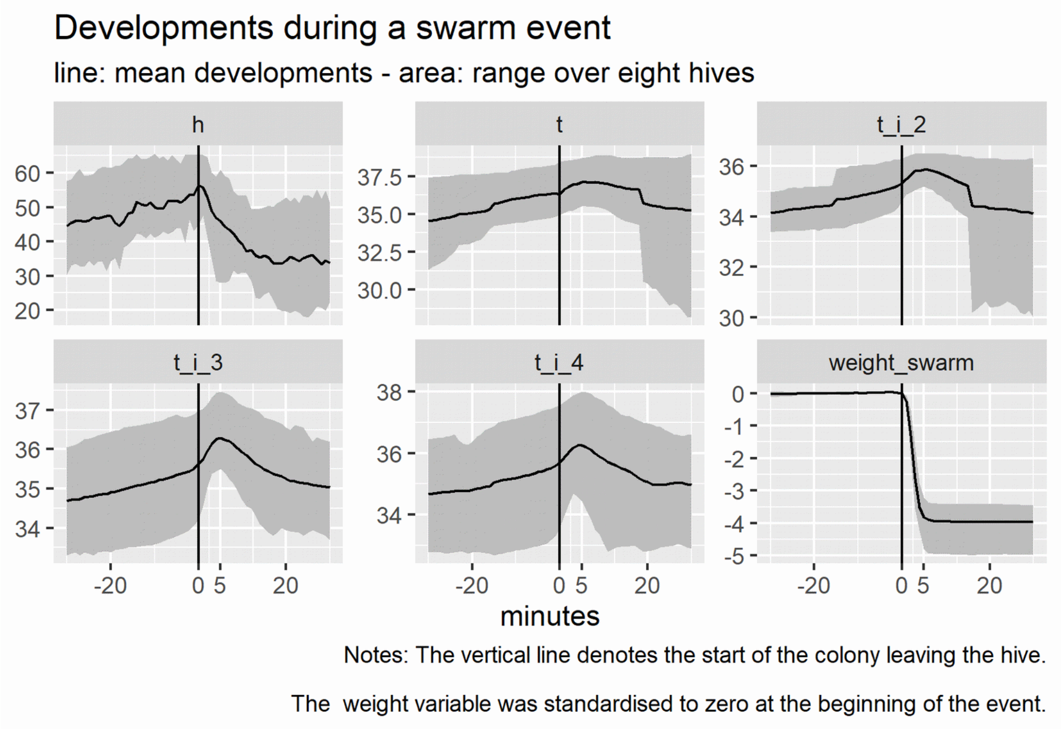
- During an swarm event half of the colony emerges from the hive with the old queen bee

- A beekeeper might loose half of their bees

©Thomas Söllner - stock.adobe.com

CORRELAID
GOOD CAUSES. BETTER EFFECTS.

What happens during a swarm?



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

CORRELAID
GOOD CAUSES. BETTER EFFECTS.


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.

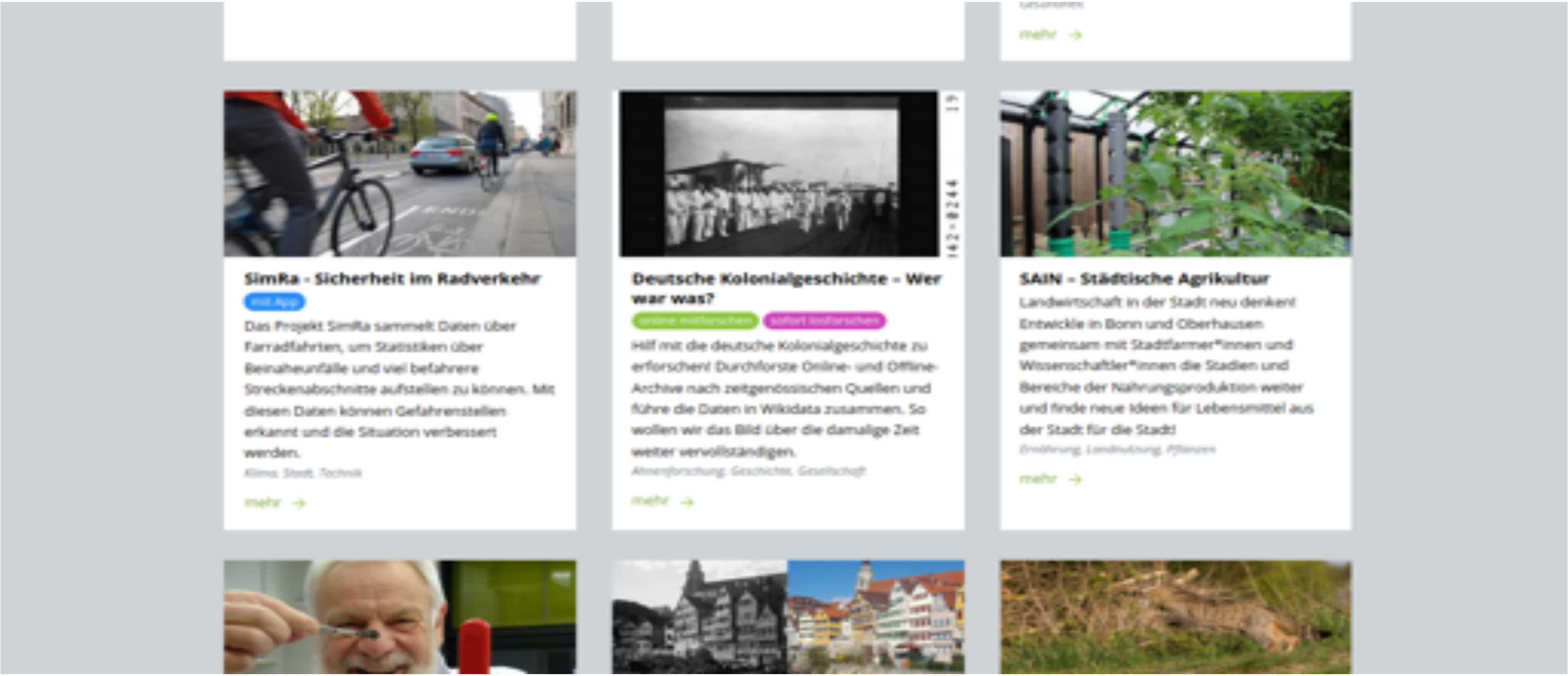

Superorganism
Honey bee colony


Environmental influences
Parasites
Living conditions

Citizen Science – Lead Your Project

Science by nonprofessionals, citizens as data collectors or thinkers

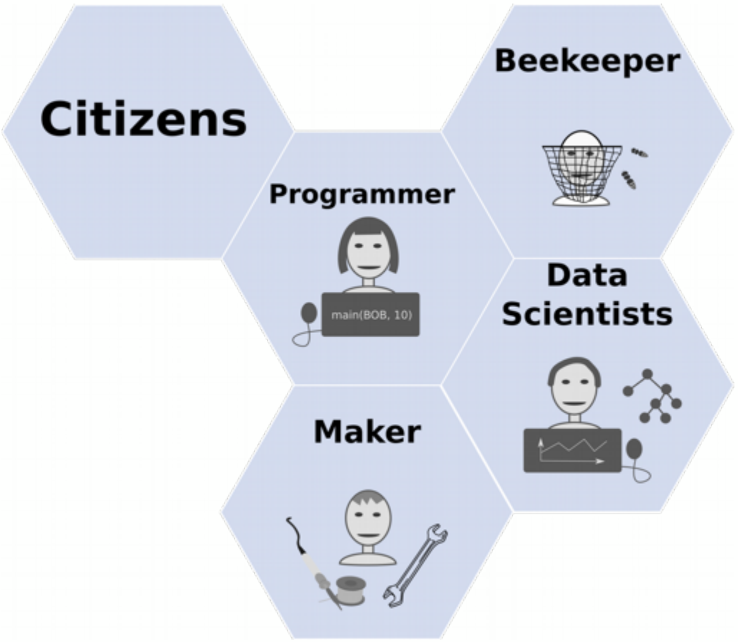
Citizen Science – Find Your Project



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

2018

First Meetings
&
Hardware
Testing

2019

20 Sensor Kits
&
90 Beekeepers
across
Germany

CorrelAid joins
the Project
&
Data Collection
and
Workshops

2020

Data Cleaning
&
Outlier
Detection

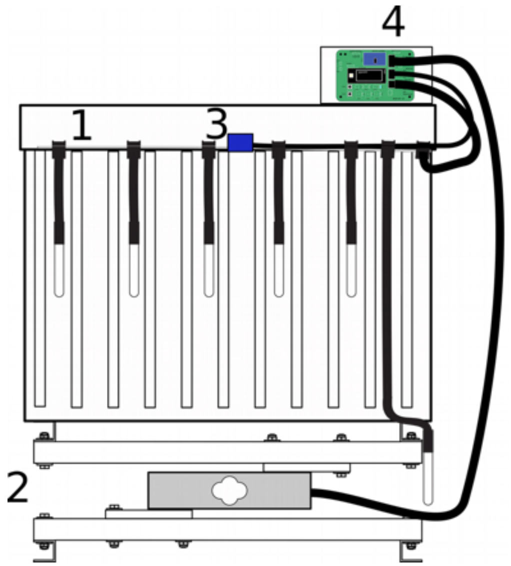
Research
Papers
Publication
&
ML Testing

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)

CORRELAID
GOOD CAUSES. BETTER EFFECTS.

Bee Observer - Inspections App

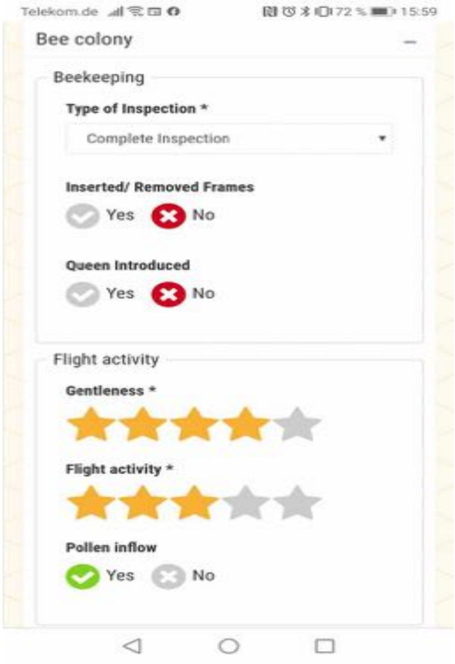
Sensor data is complemented by

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)

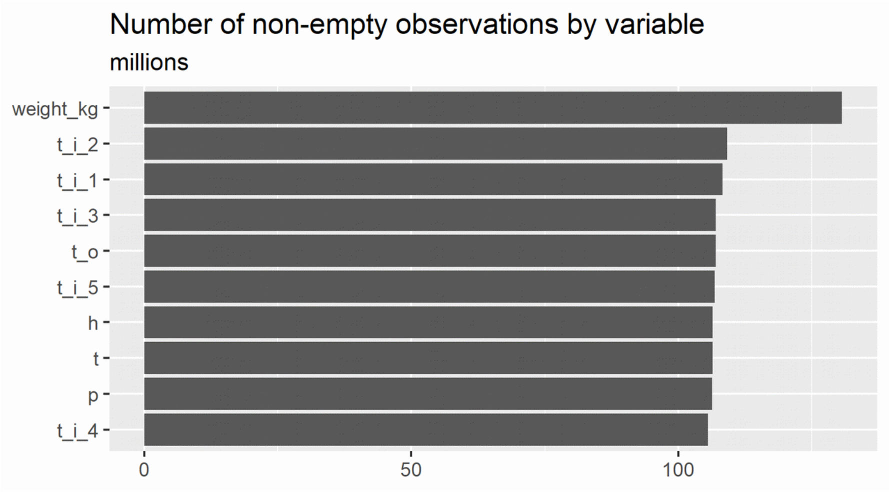


Data Profiling - Sensor Data

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

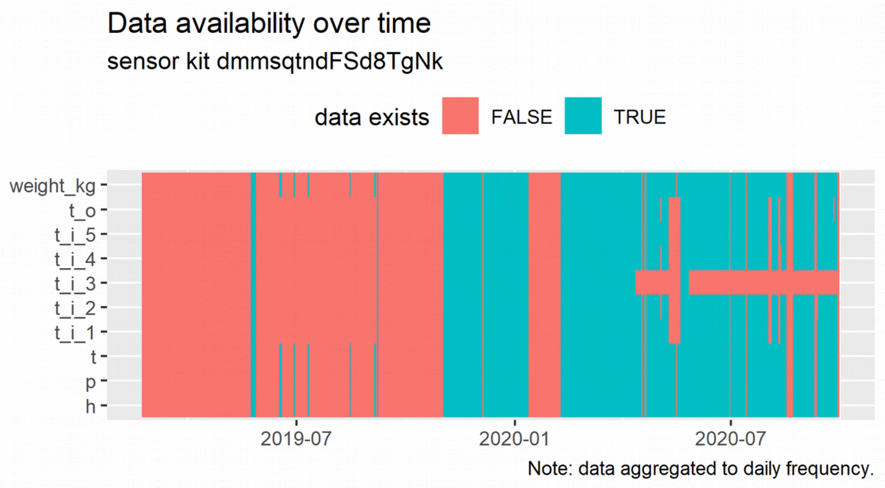


Note: Data until 30 September 2020.

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

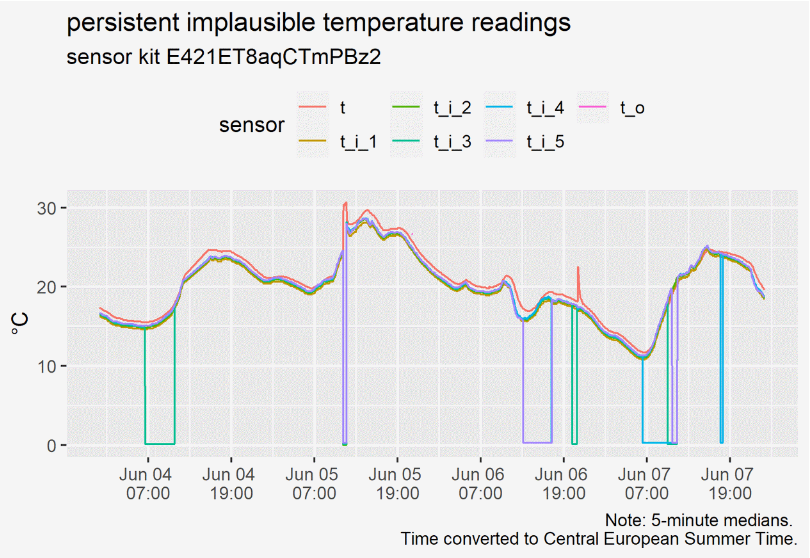
CORRELAID
GOOD CAUSES. BETTER EFFECTS.

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

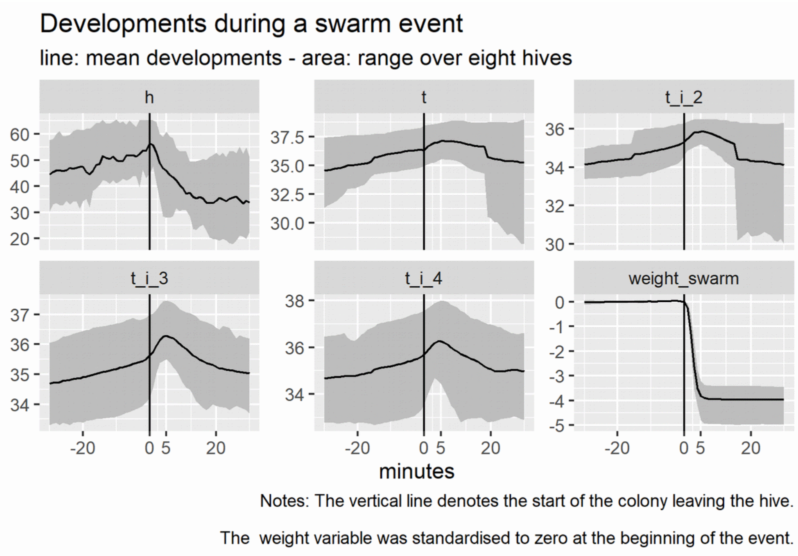
When colonies swarm, the beekeeper has around one day to react. Otherwise the swarmed part of the colony may die or at least be lost to the beekeeper. Also the “old” part of the colony may be weakened.

Swarms can be characterised by typical developments!

©Thomas Söllner - stock.adobe.com

CORRELAID
GOOD CAUSES. BETTER EFFECTS.

What happens during a swarm?



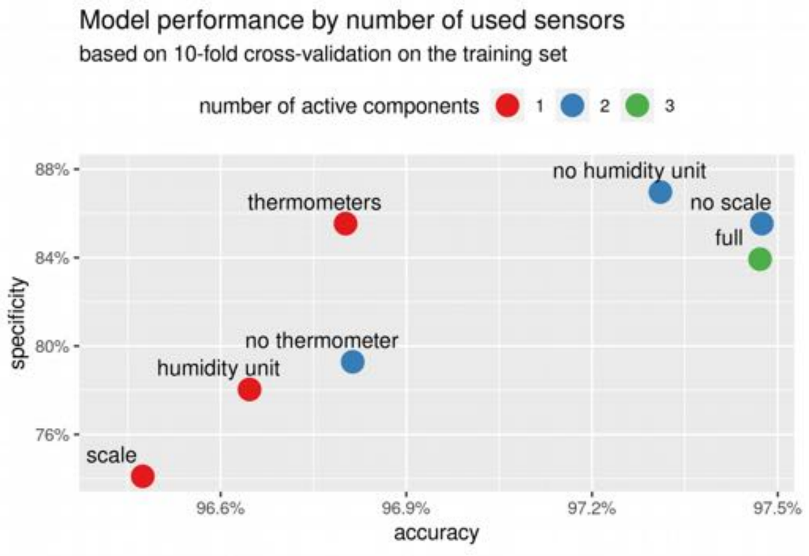
Developments in key variables during a swarm

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

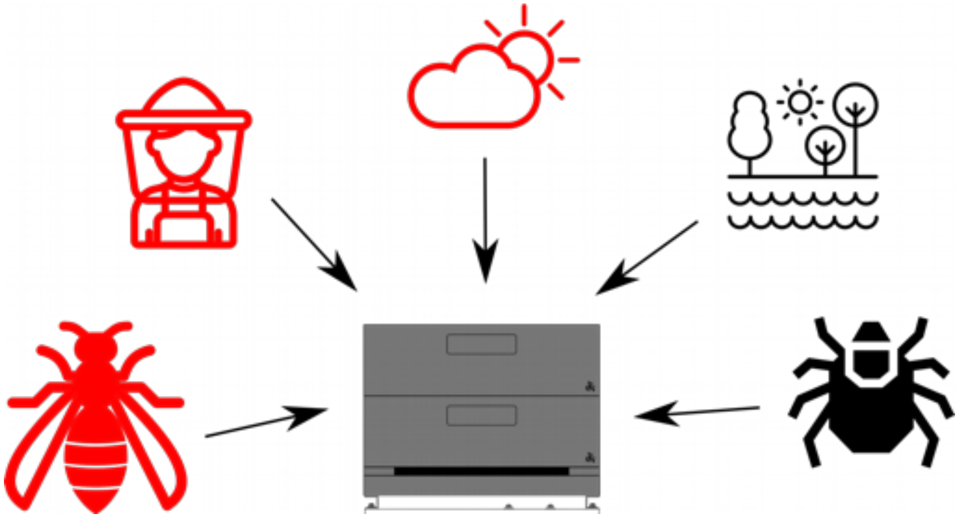


Performance metrics of RF models using subsets of sensor components

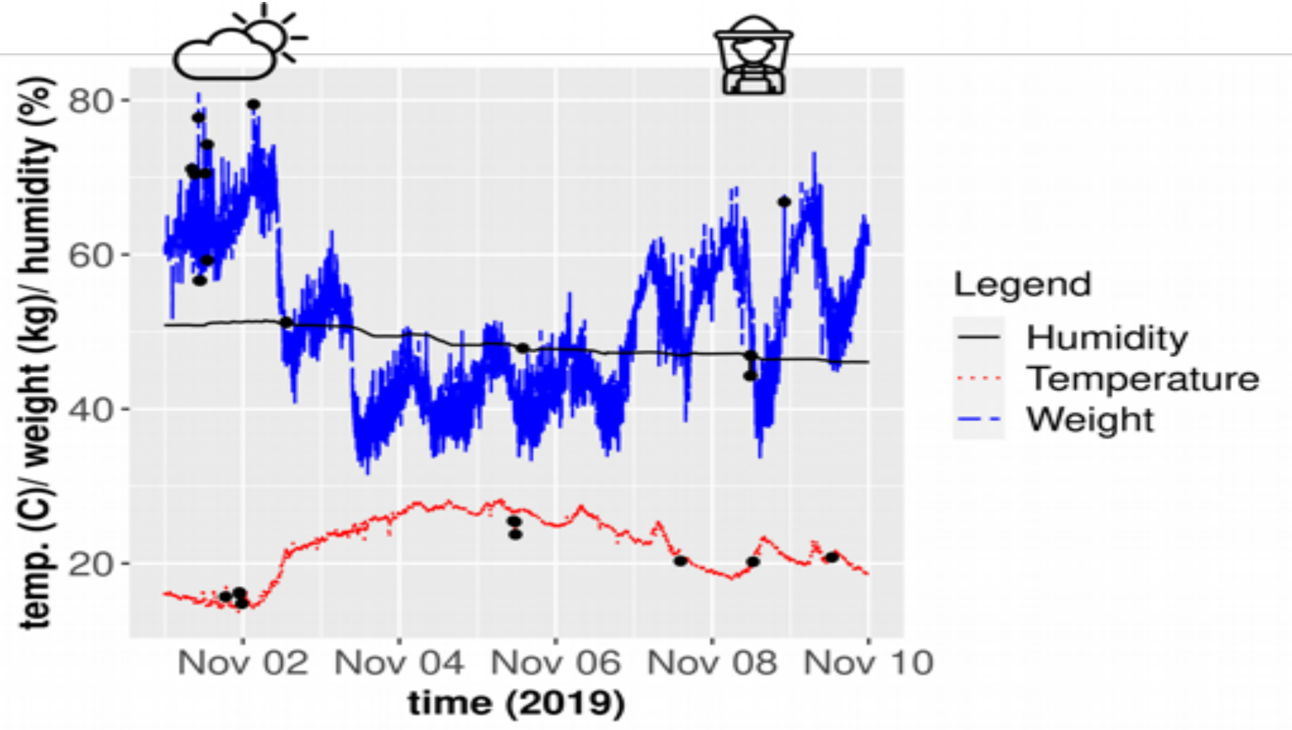
Local Outlier Detection

Data Modelling & Sensor Measurements Predictions

Difference between Prediction and the Actual Value



Local Outlier Detection



CORRELAID
GOOD CAUSES. BETTER EFFECTS.

Q&A

Bee Observer

CORRELAID

RHEIN-MAIN

CORRELAID
GOOD CAUSES. BETTER EFFECTS.

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

Outline

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