Challenges of integration and validation of farm and sensor data for dairy herd management

K. Schodl, B. Fuerst-Waltl, H. Schwarzenbacher, F. Steininger, M. Suntinger, F. Papst, O. Saukh, L. Lemmens, D4Dairy-Consortium, C. Egger-Danner

Federal Ministry Republic of Austria Climate Action, Environment, Energy, Mobility, Innovation and Technology

Federal Ministry Republic of Austria Digital and Economic Affairs

FFG Promoting Innovation.

vienna business agency

Introduction

- **Precision livestock farming (PLF)**
  - Sensor technology on dairy farms
  - Herd (health) monitoring

- **Sensor data: large amount of current data**
  - Early detection of specific diseases or disease complexes
  - Integration of additional farm data
  - Analyses across farms and sensor systems
    → development of herd health management tools
  - Proxies for functional or health traits in breeding
    → possibility of largescale phenotyping

- **Possibility? Prerequisites? We need a project to tackle this!**

THE STORY OF D4DAIRY, OR HOW TO TAME THESE SENSOR DATA

D4Dairy – Sensor data

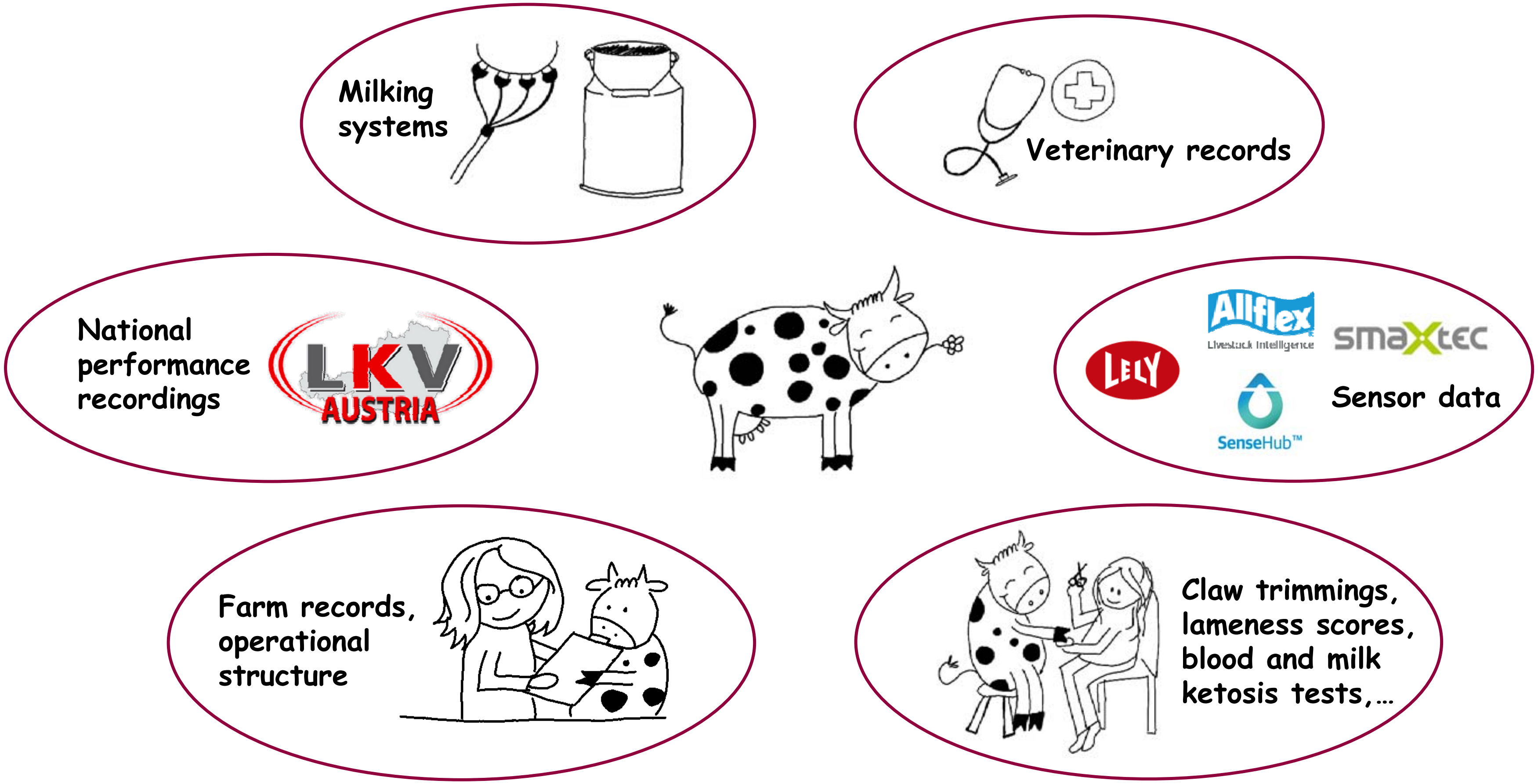
- Farms with sensor systems were motivated to participate
- Data collection January 2020 – August 2021

Sensor systems

- smaXtec **25 farms**, activity, ruminal temperature and drink cycles, 10 min intervals
- LELY **35 farms**, activity and rumination time, 2 hour intervals
- Allflex Livestock Intelligence SenseHub™ **0 farms**, activity, rumination and feeding time, 2 hour intervals
- DeLaval **14 farms**, activity alarms (not yet processed)
- GEA GEA Farm Technologies **9 farms**, activity (not yet processed)

- **Desired outcomes**
  - Disease predictions for lameness, ketosis and mastitis
    → **foundation for herd management tools**
  - Auxiliary traits for claw, metabolic and udder health
    → **foundation for genetic health indices**

Integration with other farm data



Milking systems

Veterinary records

National performance recordings

Sensor data

Farm records, operational structure

Claw trimmings, lameness scores, blood and milk ketosis tests,…

Once upon a time there was a princess named Validation…

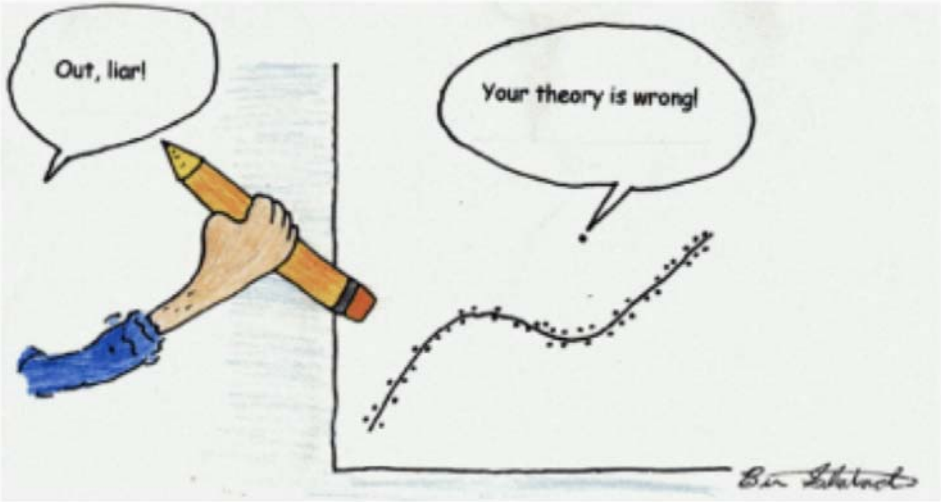
Data validation – sensor data

- **Validation of the sensor systems** by the companies based on (scientific) studies
  - Comparison to gold standard (e.g. behavioural observations)
  - Tailored to the company's purpose or application
    ➔ check validation study before further use of sensor variables!

- **Validation of output data for own purpose**
  - Clean data for outliers and implausible values, which may interfere with any further analyses
  - Depending on the purpose excluding expectable deviations (e.g. time around calving) may be necessary

Data validation – data quality and robustness

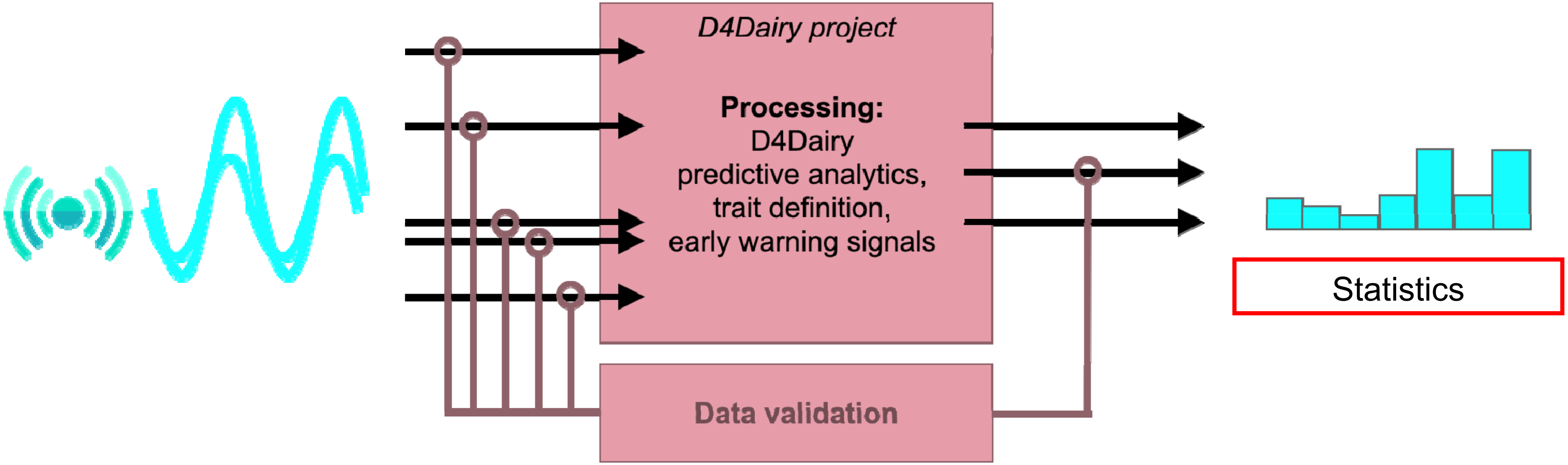
- **Outliers**
  - Outliers cause surprise in relation to the majority of the samples
  - **BUT** outliers may be correct and a physiological response we are looking for!
  - It can be difficult or even impossible to spot ‘real’ outliers in multivariate or highly structured data



- **Using statistical approaches to detect outliers**
- **Spotting outliers using domain expert knowledge**

Data validation – automatic outlier detection

- Data quality assurance using Isolation Forest anomaly score



F. Papst, K. Schodl, O. Saukh
Exploring Co-dependency of IoT Data Quality and Model Robustness in Precision Cattle Farming
International Workshop on Challenges in Artificial Intelligence and Machine Learning for Internet of Things (AIChallengeIoT 2021)

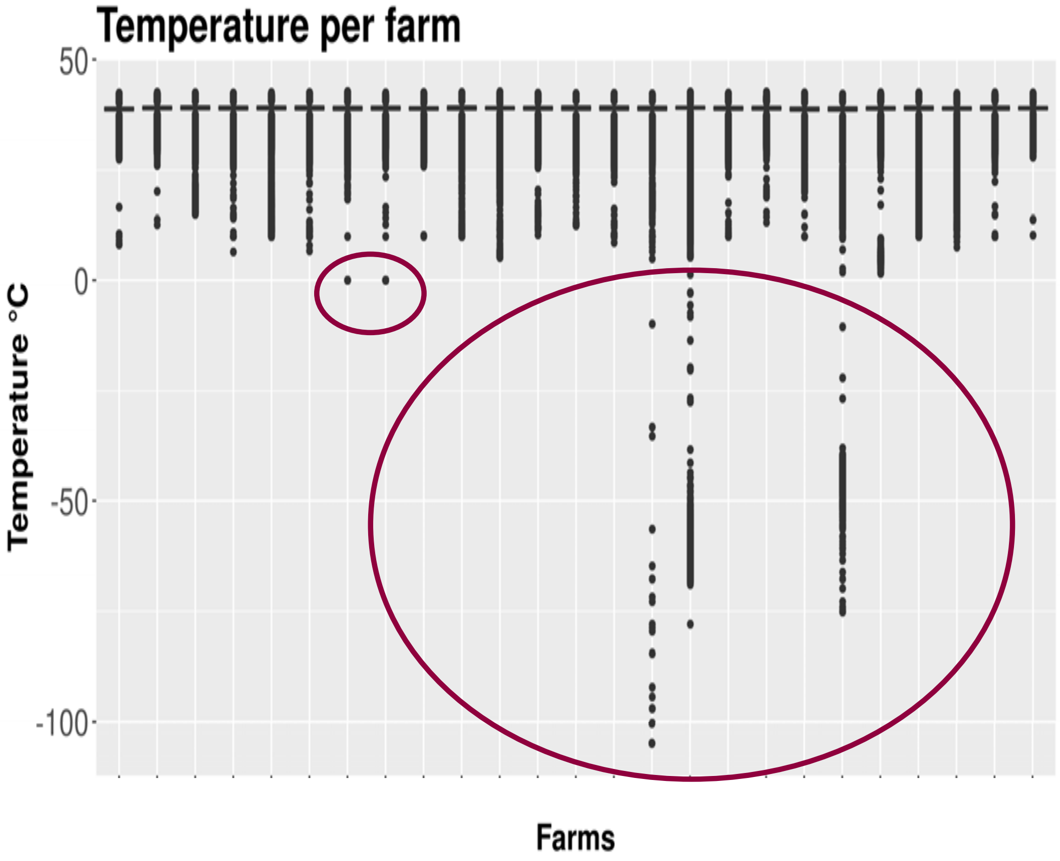
Data validation – manual outlier detection

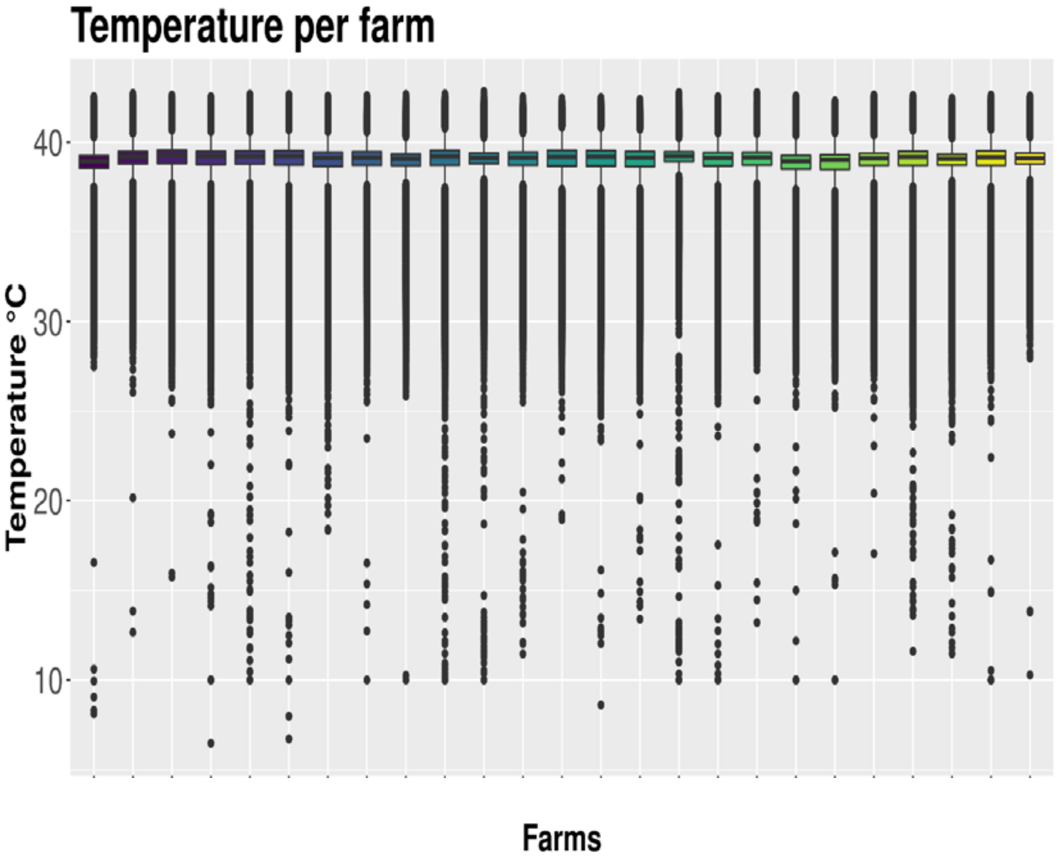
- **Validation based on data availability & frequency**
  - **Removal of duplicates & missing data**
  - **Potentially erroneous measurements**:
    - Consultation with sensor company
    - Isolated measurements (time to next/previous data point exceeded regular frequency of data retrieval from the sensor)
    - Multiple measurements within the regular time window of data retrieval
  - **Time alignments:**
    - Variables calculated based on sensor measurements were not aligned and had to be shifted to match the correct time of data retrieval

Data validation – Plausibility assessment

- **Plausibility assessment using a priori knowledge**
  Time and temperature
  - More than 12 hours with 0 activity or rumination time → sensor not yet administered
  - Temperature range in the data set: -104.9°C – 42.9°C
  - Negative temperature values physiologically impossible
    → faulty data due to low sensor battery was discarded (6.3°C)

Data validation – Plausibility assessment





Data validation – Plausibility assessment

- **Plausibility assessment using a priori knowledge Time and temperature**
  - More than 12 hours with 0 activity or rumination time → sensor not yet administered
  - Temperature range in the data set: -104.9°C – 42.9°C
  - Negative temperature values physiologically impossible → faulty data due to low sensor battery was discarded (6.3°C)
- **Depending on the level of aggregation and pre-processing**
  - **Lely & Sensehub**: pre-processed and aggregated → plausibility checks OK
  - **smaXtec**: data as provided by the bolus prior to any processing for analysis
    - **~5% of the data had to be discarded based on the above criteria**

Data validation – other farm data

- **Validation of data from automatic milking system (AMS)**
  - Milking intervals should be >60 min or <24 hours
  - Single milkings should exceed 1 kg
  - First milking of lactation was discarded
  - Milk production per hour: discard if >50% higher than ±10-day average of the animal in the current lactation
- **~2% of the data did not fulfil the criteria**

- **Cross-validation of data sets**
  - Calving dates in the sensor data compared to data from AMS and official calving records (multivariate plausibility checks)

VALIDATION SUCCESSFULLY MASTERED THE FIRST PART OF THE QUEST, BUT THERE IS MORE TO COME...

Data validation – challenges to tackle

- **Comparisons of sensor systems for detection of diseases**
  - Compare results of disease detection between sensor systems with respect to data patterns (e.g. activity patterns)
  - Genetic analyses: auxiliary traits from different sensor systems for the same disease complex
- **Routine application**
  - Stable test data set → use in routine genetic evaluation and herd management requires automatization
  - Algorithms for data quality assurance based on machine learning approaches
  - Concepts for (automatic) multivariate plausibility assessment of data
  - Who takes the final decision – human or artificial intelligence?

THE END (OF PART 1).

Thank you for your attention!

Research partners:

AGES

MEDIZINISCHE UNIVERSITÄT WIEN

BUNDESMINISTERIUM FÜR NACHHALTIGKEIT UND TOURISMUS
HBLFA RAUMBERG-GUMPENSTEIN LANDWIRTSCHAFT

TU Graz

research & testing BLT wieselburg
BUNDESMINISTERIUM FÜR NACHHALTIGKEIT UND TOURISMUS
HBLFA FRANCISCO JOSEPHINUM WIESELBURG

vetmeduni vienna

BOKU

FFoQSI

COMPLEXITY SCIENCE HUB VIENNA

LIÈGE université Gembloux Agro-Bio Tech

HOCHSCHULE FÜR Agrar- und Umweltpädagogik

LfL Tierzucht

Wallonie recherche CRA-W

Company partners:

PLF Agritech EU Ltd
Technology Enhancing Production

Allflex Livestock Intelligence

AMA AGRARMARKT AUSTRIA

ARGE RIND
arbeitsgemeinschaft rind

Berglandmilch eGen

EMR

BIOMEDICA

BIOMÉRIEUX

Biomin

FiH Muhhht zu Handeln

LKV Baden-Württemberg

LKV BAYERN

LKV

NOEGENETIK RINDERZUCHTVERBAND

lk Futtermittellabor Rosenau Landwirtschaftskammer Niederösterreich

MILCH PRÜFRING

OFK

ÖSTERREICHISCHE FLEISCH KONTROLLE

QUALITÄTSLABOR Niederösterreich

Pessl Instruments

lk Landwirtschaftskammer Österreich

TIER ARZT
Dr. Walter Obritzhauser

OBER-ÖSTERREICHISCHER tgd

TGD Tiergesundheitsdienst Burgenland

tgd NÖ TIERGESUNDHEITSDIENST

lk Landwirtschaftskammer Oberösterreich

TIERGESUNDHEITSDIENST TGD Salzburg

smaXtec INSIDE MONITORING

WASSERBAUER

RINDERZUCHT AUSTRIA Innovation

ZUCHT DATA

Cooperation partners:

LELY

GEA
GEA Farm Technologies

DeLaval

The end (of part 1).