Stepwise Modelling for Improved Bovine Health

Caspar Matzhold[1], Katharina Schodl[1], Franz Steininger[1], Peter Klimek[2], Christa Egger-Danner[1]

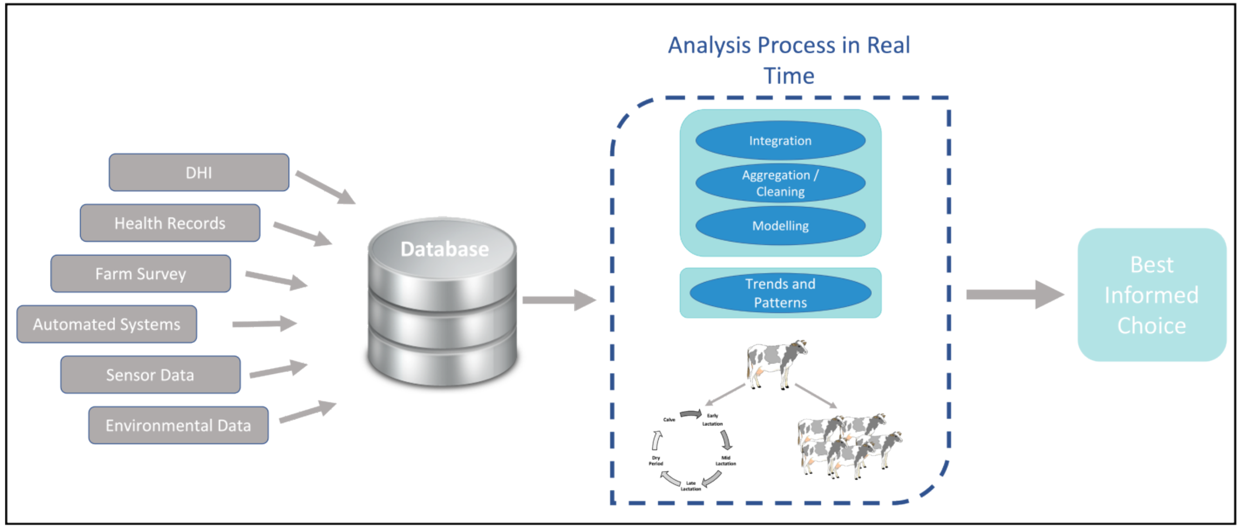
[1]ZuchtData EDV-Dienstleistungen GmbH, Dresdner Str. 89, 1200 Vienna, Austria

[2]Complexity Science Hub Vienna, 1080 Vienna, Austria

Background

- **Objective:** To improve animal health, welfare, and farm efficiency.
- **Strategy:** By implementing a data warehouse that integrates and analyses data in real time and supports farmers in their decision-making process.
- **Challenges:** Creating a system for data integration, aggregation, and developing methods for turning data into knowledge is a complex task.
- **Proposed Solution:** Stepwise Modelling to reduce the analysis complexity.

Data Warehouse



Stepwise Modelling

Step 1: Farm Management Risk Factors Identify farm-specific risks including infrastructure and management practices.

Expected Outcome: Knowledge on the relationship within and between groups of clusters and disease risks.

Step 2: Cow Risk Scores Gather cow health history and DHI records. Create individual risk scores based on historical data.

Expected Outcome: Enhanced understanding of each cow's health profile, enabling targeted interventions and reduce health risks.

Step 3: Alert System Integration of sensor and AMS data for real-time health monitoring. Predict health issues and trigger alerts for proactive disease prevention.

Expected Outcome: Early detection of health issues, timely interventions, and minimized disease impact through predictive alerts.

Farm Management Risk Factors: Objectives

1. **Identification of clusters**, sets of similar features describing a specific type of farm management.
2. **Inter-Cluster-Group** Analysis:
   - Disease risk differences **between distinct groups** of clusters, such as sensor and non-sensor clusters.

   – Insights on the impact of farm management defining features (sensor systems) on bovine health.
3. **Intra-Cluster-Group** Analysis:
   - Disease risk differences **within a group** of clusters (sensor farms only).

   – Insights on cluster-specific risk-factors and how they interact.

Data

1) Farm dataset:

- 3.855 (8.000) farms described by 107 Hot-Encoded Features describing:
  - *Feeding-, Milking-, Housing-Systems,..*
  - *Husbandry management, pasture strategies,..*
  - *Smart Technology Applications like AMS and Sensor System,..*

2) Health dataset:

- Diagnoses were categorized into 3 groups:
  1. Udder diseases (Chronic and Acute Mastitis)
  2. Fertility disorder (Anestrous, Uterus Inflammation, Ovarian cysts)
  3. Milk fever (Hypocalcemia)
- Unique counts of diagnoses – cow observations
  - Multiple health observations of a cow within a diagnosis group counted only once to avoid bias

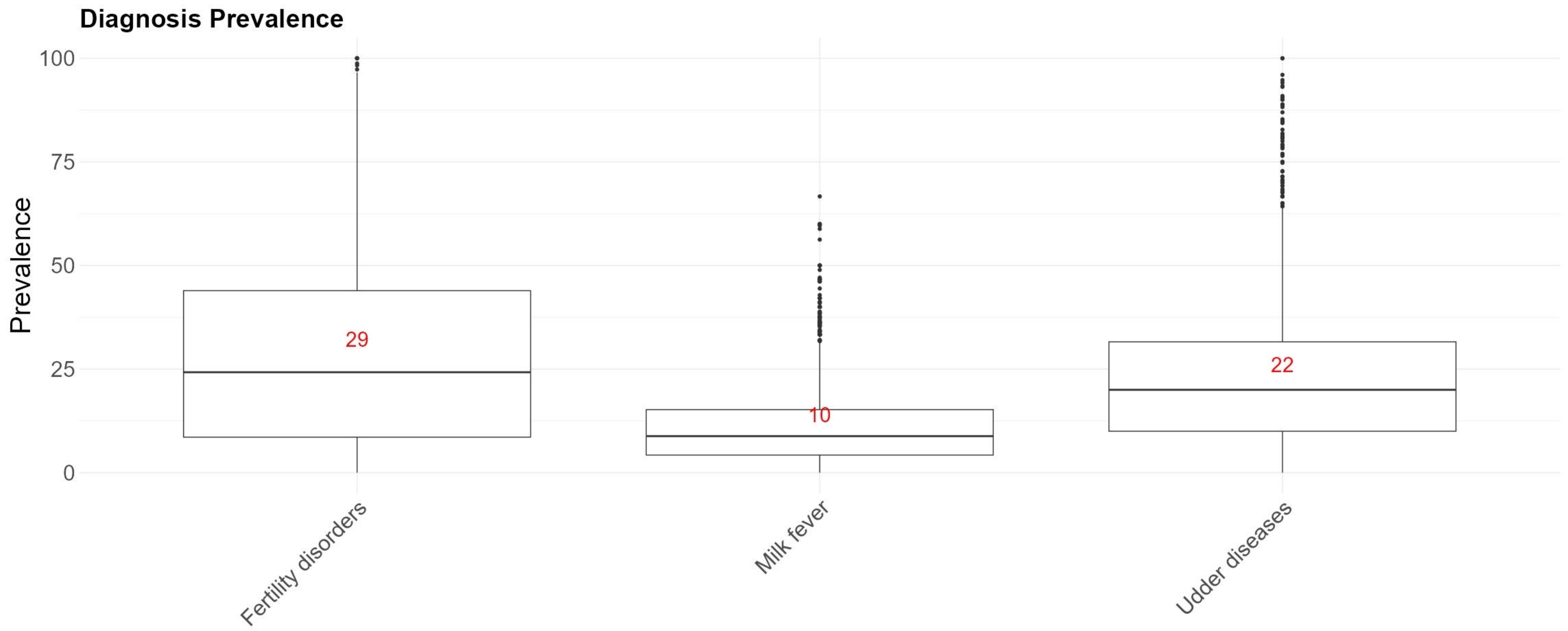
➢**Cleaning & Validation & Filtering & Aggregation**

➢**Observation period: 2 years**

Data Descriptives

Farm Dataset	Counts
Farm	3.855
Feature	107
Cow	122.783

Diagnose Group	Unique Counts over 2 year
Udder diseases	26.582
Milk fever	12.822
Fertility disorders	34.337
Sum:	**122.475**



Methods: Key-Feature-based Clustering Procedure

- What is a key-feature?
  - A key-feature is known to have a significant impact on farm management.
- Why defining key-features?
  - Defining the pattern for the similarity measurements allows for more detailed analysis of the effects of features of interest, such as sensor systems.

Key Features
Herd size
Milk yield average per year
Altitude
Pasture
Access to open areas
Free-stall housing system
Automatic milking system
Sensor

Methods: Key-Feature-based Clustering Procedure

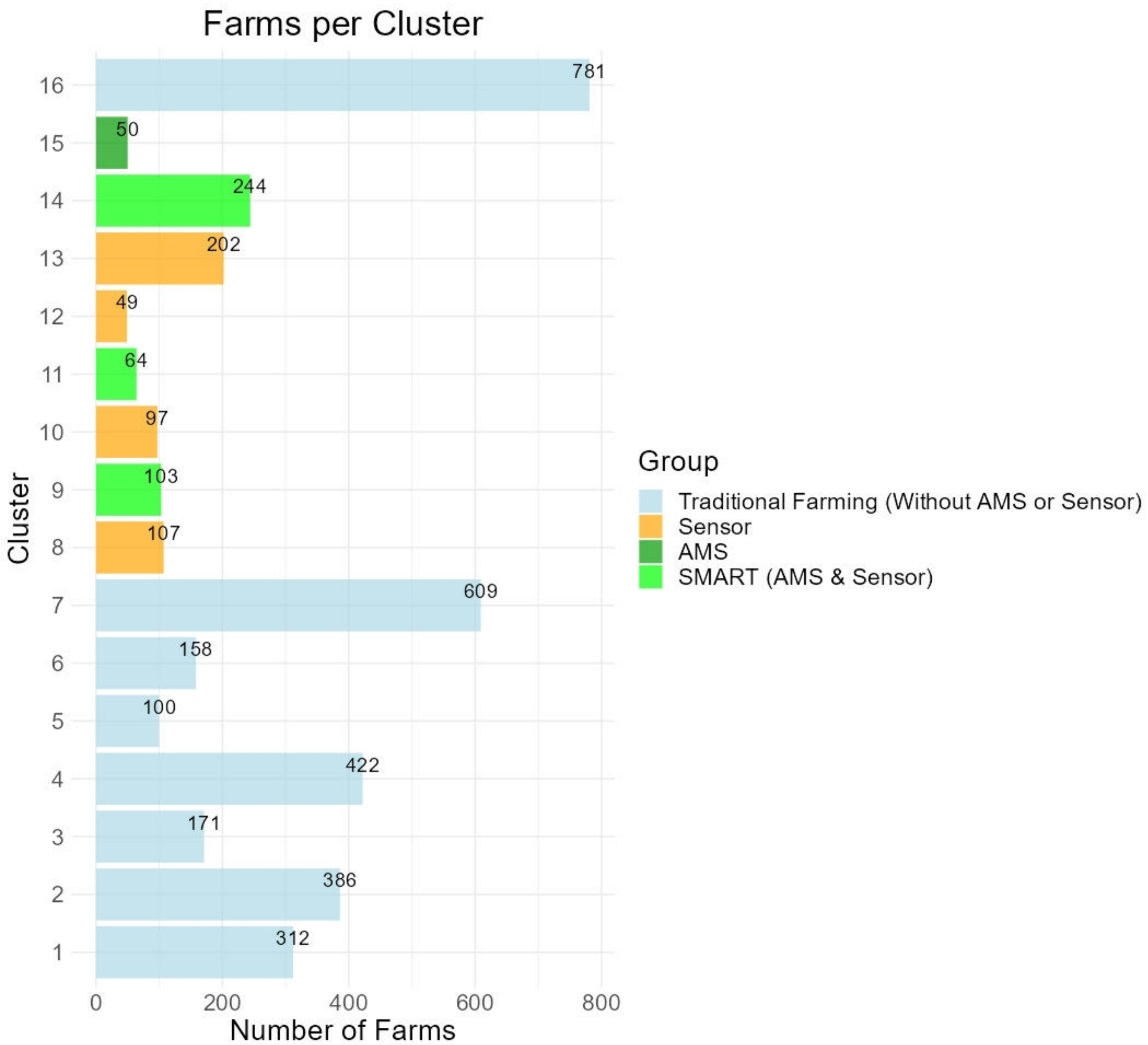
1. **Define Key-Features & Assess Similarity**
   – Similarity is assessed among the key-features using similarity measurements such as Cosine distance for numeric features and Jaccard distance for binary features.

2. **Cluster Formation:**
   – The obtained vectors of similarity are then used as inputs for the unsupervised clustering algorithm HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) to group the clusters into sets of similarity.

3. **Cluster Evaluation:**
   – The disease risk for a specific cluster is then computed, assessed as Odds Ratios (ORs) between the individual clusters.
   – The pertinent features of a cluster are assessed by computing z-scores comparatively.

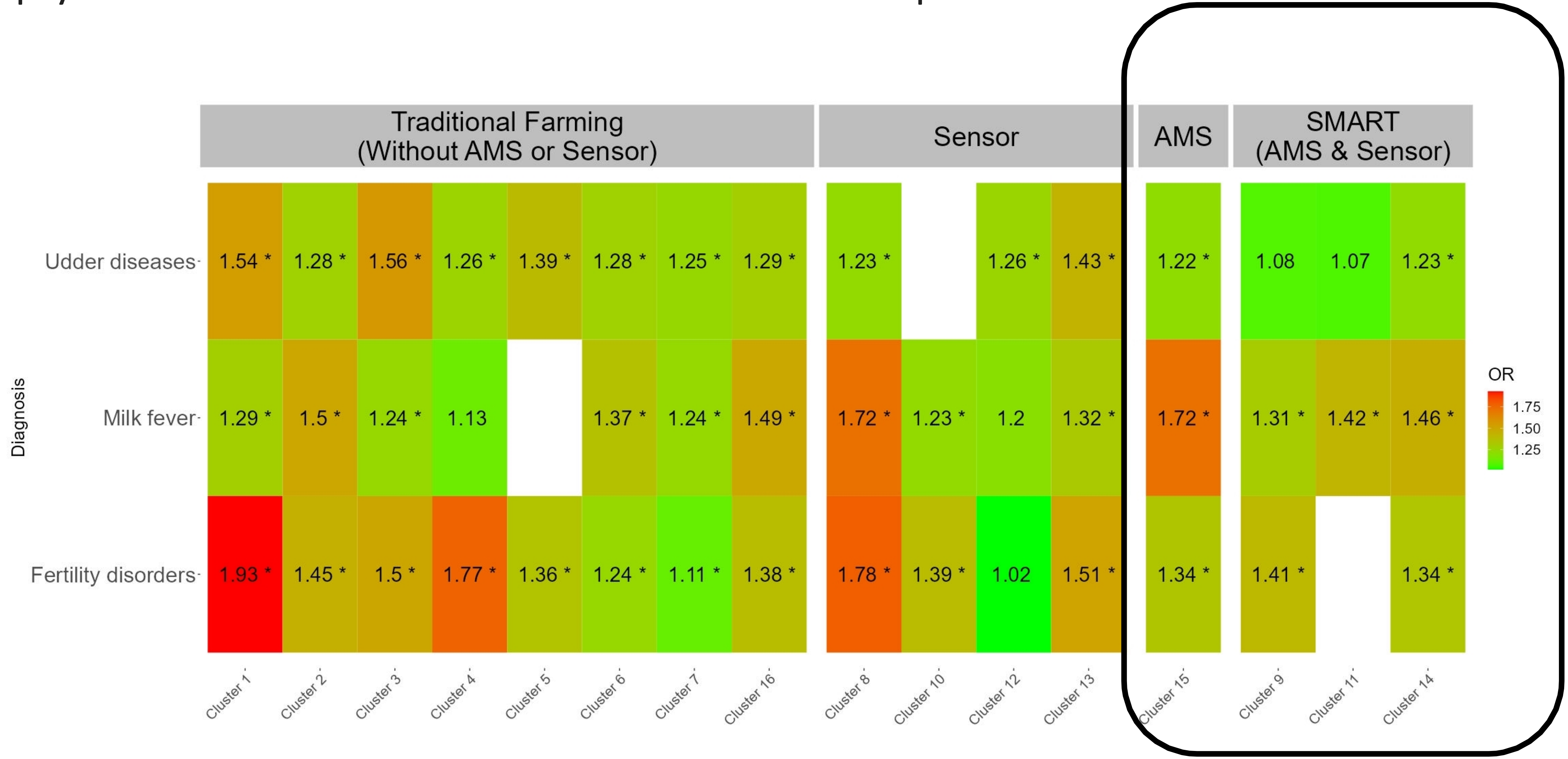
Inter-Cluster Results: Technology Usage

Inter-Cluster



Inter-Cluster: Disease Risk

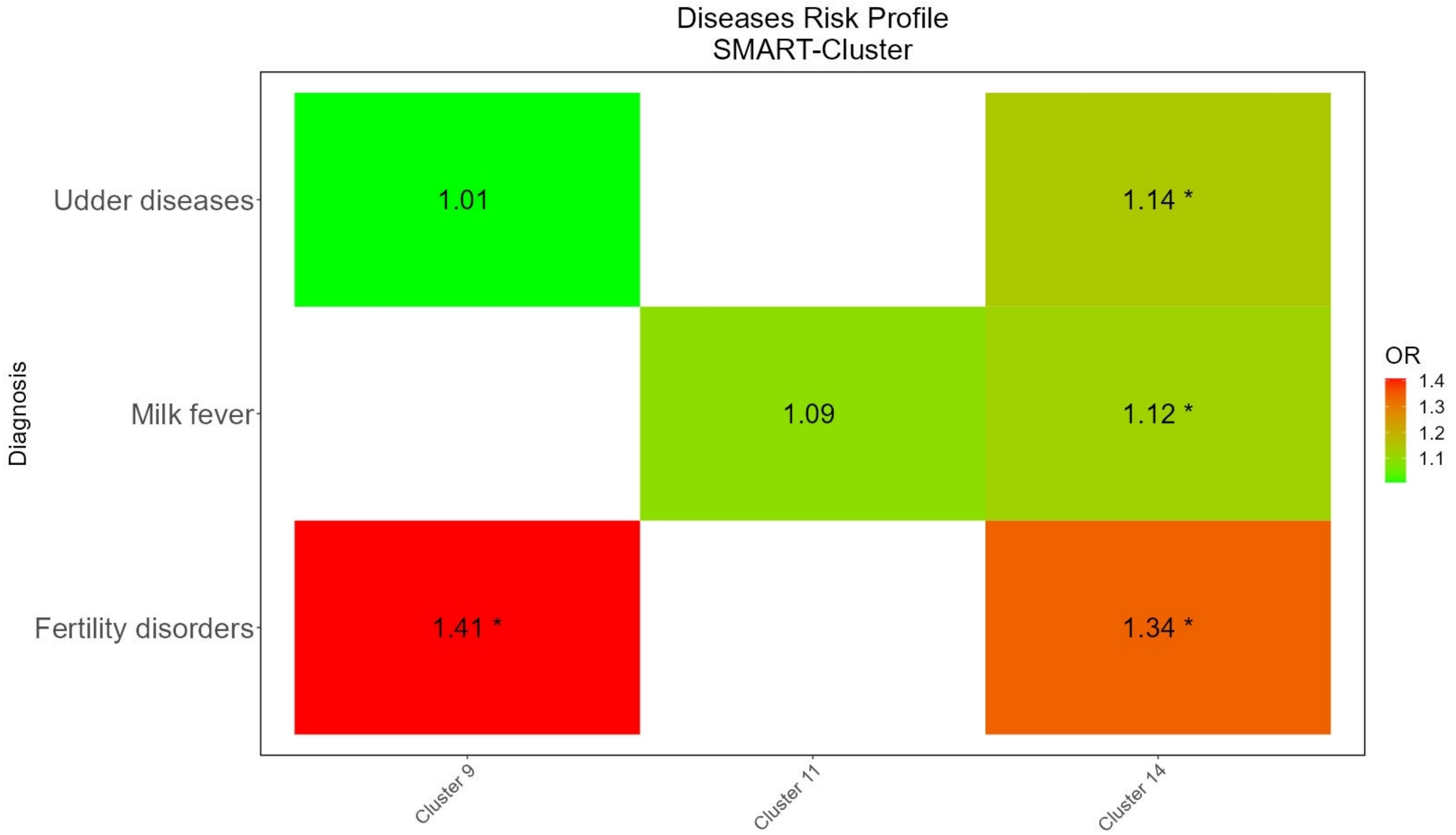
Odds Ratio of the individual cluster.
Empty fields indicate cluster with the lowest disease prevalence.



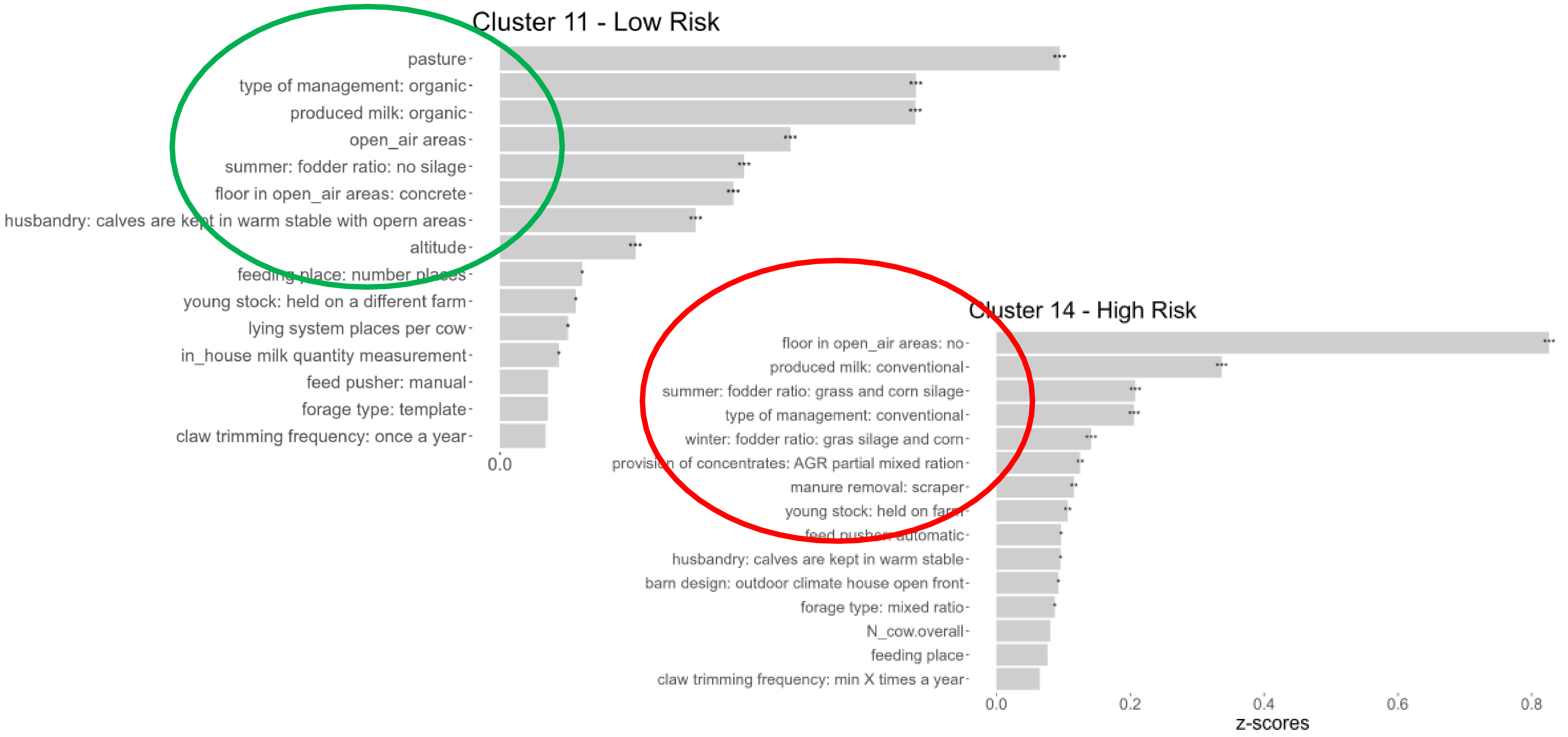
Intra-Cluster Results: Smart Farms

Intra-Cluster: Disease Risk

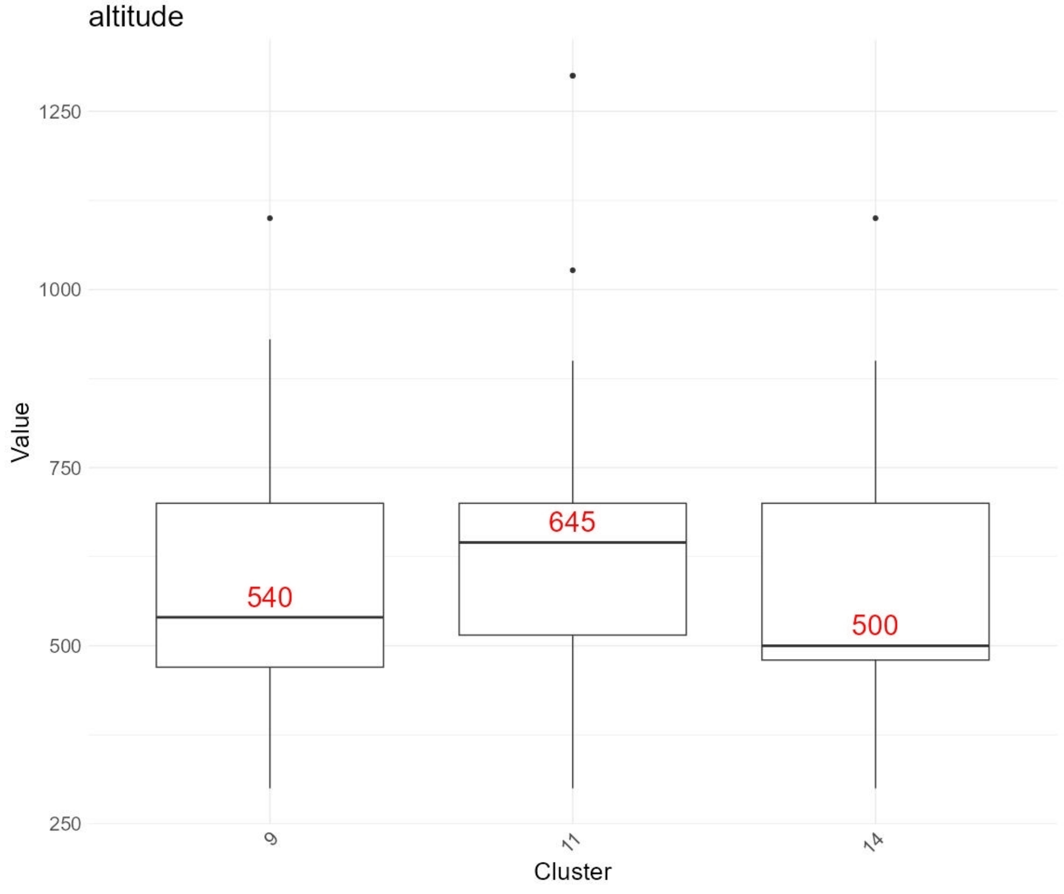
Odds Ratio of the individual cluster.
Empty fields indicate cluster with the lowest disease prevalence.

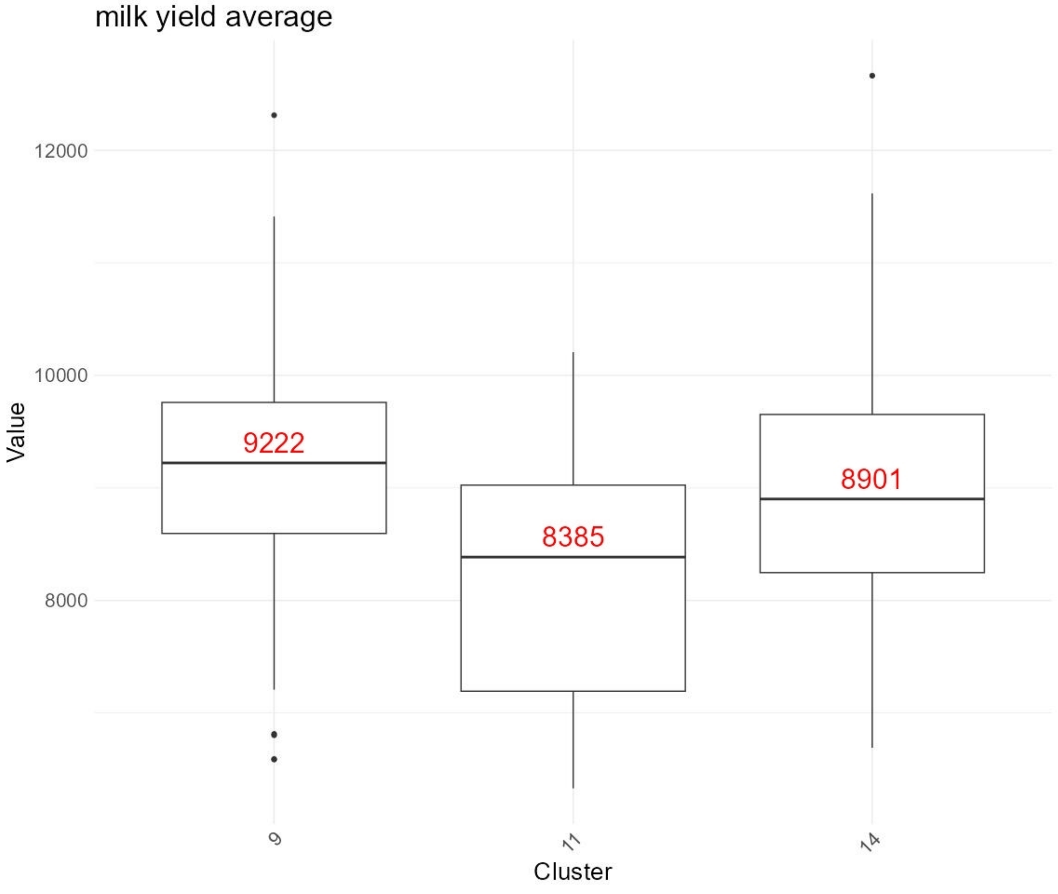


Highest vs. Lowest Smart Cluster Disease Risk: Binary Feature Comparison



Highest vs. Lowest Smart Cluster Disease Risk: Numeric Feature Comparison





Risk Factor Evaluation

Cluster 11 - Low Risk Characteristics	Cluster 14 - High Risk Characteristics	Cluster 14 Potential Risk Factors
AMS + Sensor-equipped	AMS + Sensor-equipped	
Organic milk production	Conventional farm management	
Access to pasture and open areas	No access to open areas	Less opportunities for physical activity
Summer fodder ration: no silage	Fodder ration: grass and corn silage (summer and winter)	Unbalanced diet with less fiber
High altitude location	Low altitude location	Heat stress from warmer climate
Lower milk yield	High milk yield	High milk yield

Conclusion

- We presented a clustering approach based on key-features that enables differential risk analysis between and within groups of clusters.
- We demonstrated how the approach can be used to study the influence of specific features, such as technological systems, on bovine health.
- Inter-cluster analysis shows a trend of reduced disease risk for more technologically advanced farms.
- Intra-cluster analysis, we find that farms that use smart technologies further benefit from an organic management that promotes regular physical exercise and a balanced milking and feeding management.

Thank You!

- Additional / Alternative Slides

Features indicating milking management intensity
- is interesting but not needed for argumentation …