



Exploiting *big-data* for small ruminants

Riccardo Negrini L. Pascarella, C. Melilli, F. Luisi, M. Tolone, M.
Milanesi, D. Pietrucci, M. Fioretti, L. C. Leo Consortium

negrini.r@aia.it
riccardo.negrini@unicatt.it



Presentation Outline

Short introduction

Leo project and open data

Showcase of Leo contributions to
Small Ruminants 2.0

```

      10101
     110110
    1100110
   011100100
  100101000
 0011000011
1001100100110111
0000100000111000100
110011100100110001101
001001101001000001110100011001010111100000001

```

Welcome in the era of *big-data*

```

1110110111010010 10001101110101011011010
01111001001110 10011001010111001001107
0001 1100 10010000 101 0111101101101
011/ 101 10010 00000 11011/
100 10 00100 0000 0100
01/ 00 0000 0000 100
11 100 111/ 0 1 100
10/ 1/ 100 1 10 00
00 1 110 0 01 01
00 0 100 1 11 011
10/ 0/ 001 11 110
001 101 010

```



3.5 billion search and 20 petabytes of data processed daily

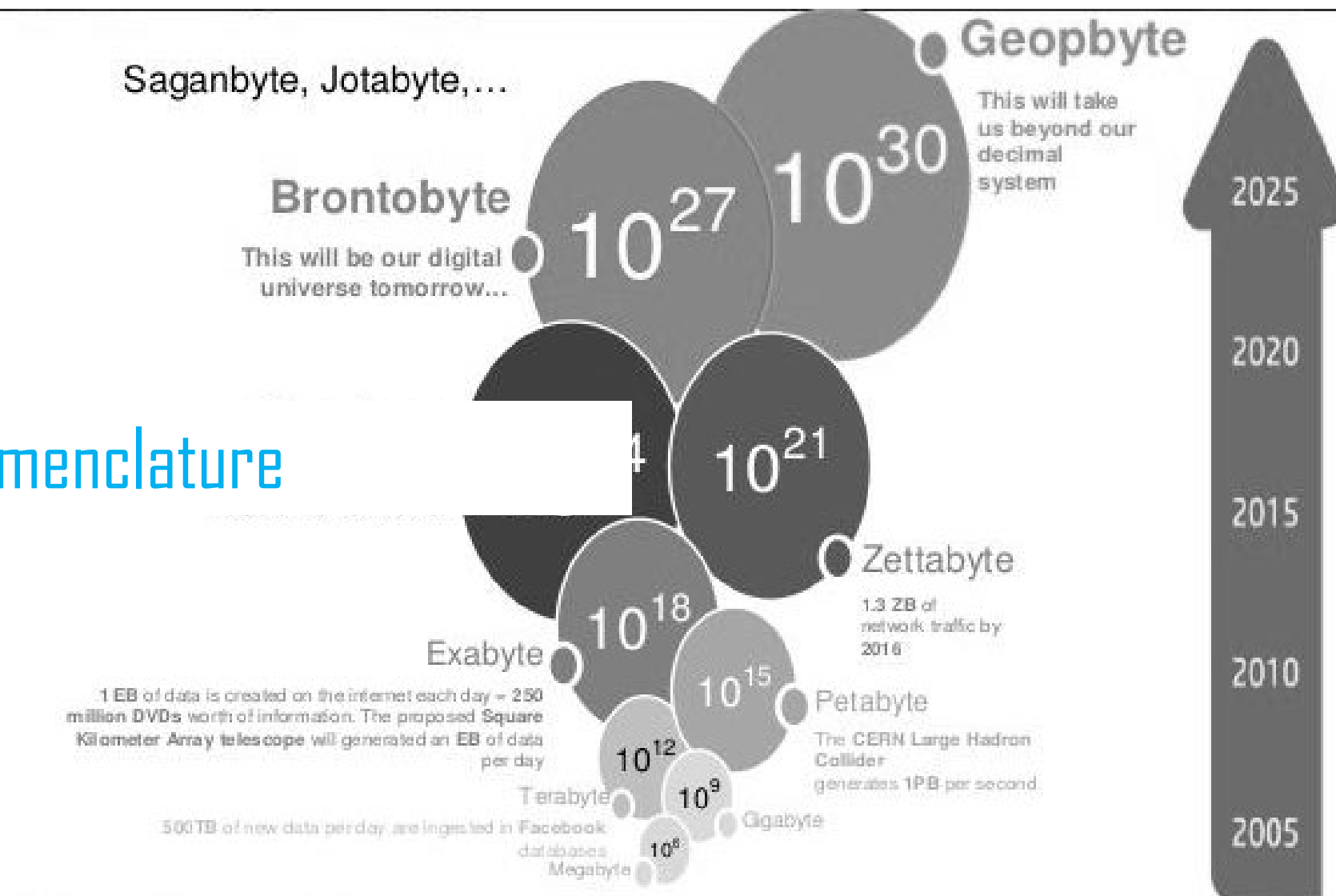


20 terabytes of data collected daily
40 petabyte of weather data stored

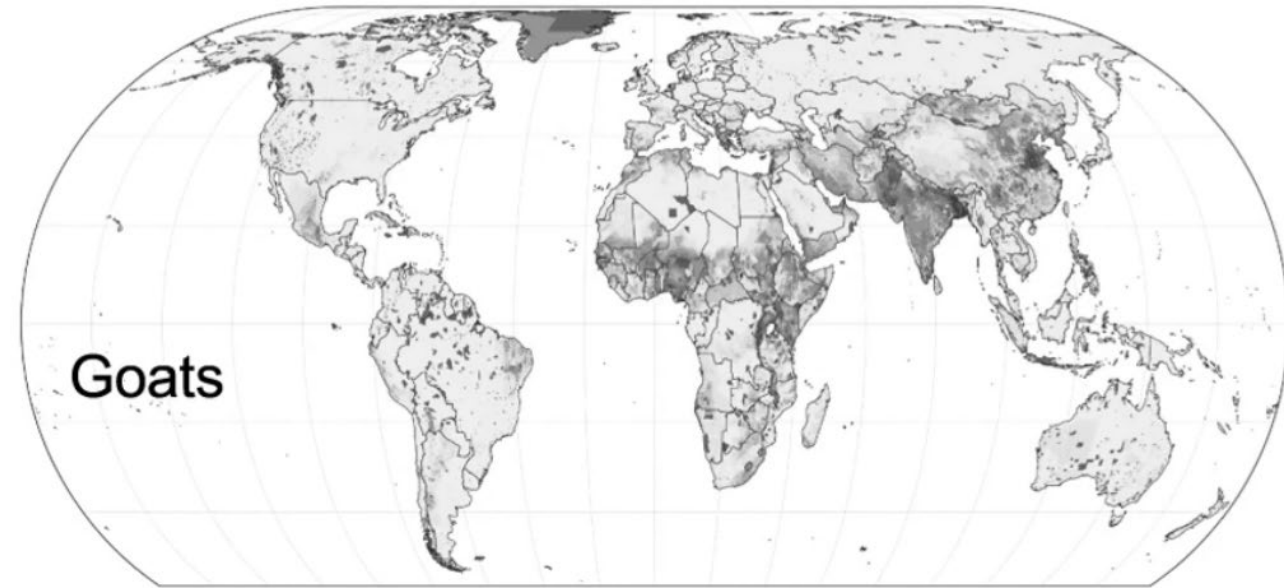
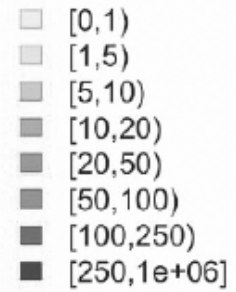


LHC: 1 petabyte of raw data per second
Storage capacity estimated in exabyte

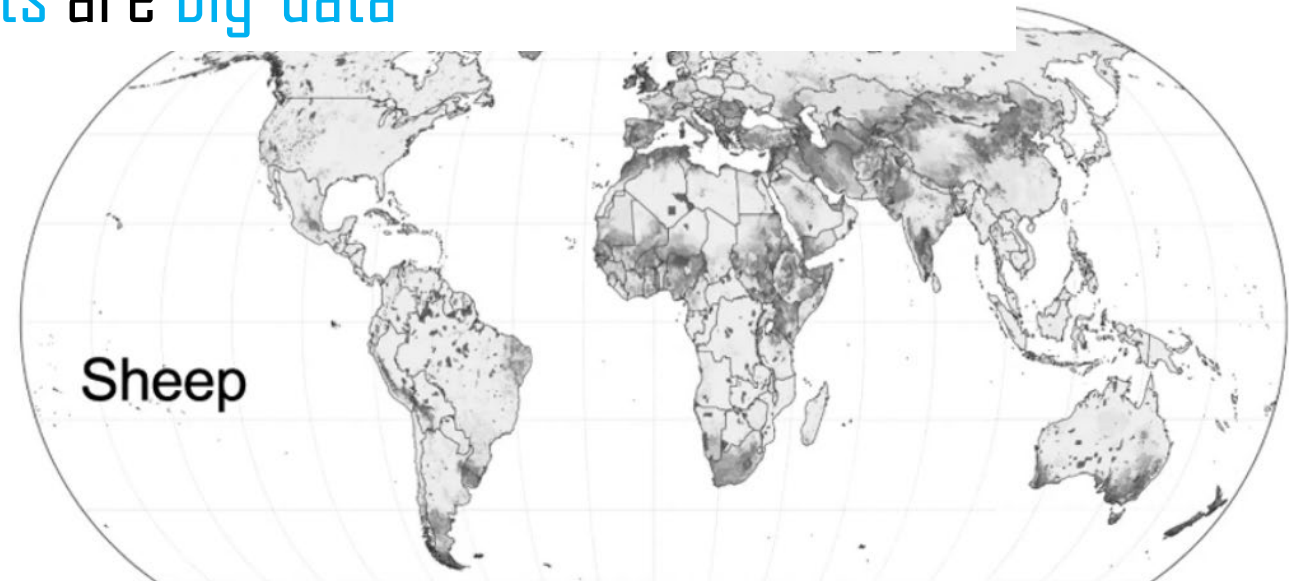
Time to refresh our nomenclature



Sheep & goats



Globally 1.1 Billion **Sheep** and 1 Billion **Goats** are **big-data**





5.770.142.822 data archived (h 11:09:04, September 4th, 2024)

4.3 Terabyte of data

18 billion triplets (ID-date-measure) in open source

259.928.575 data on 8.245.661 small ruminants

Innovative data type and collection mode

Data type

Parameters

Health and welfare

31

Extension services

35

Farm and animal
identification

6

Environmental
sustainability

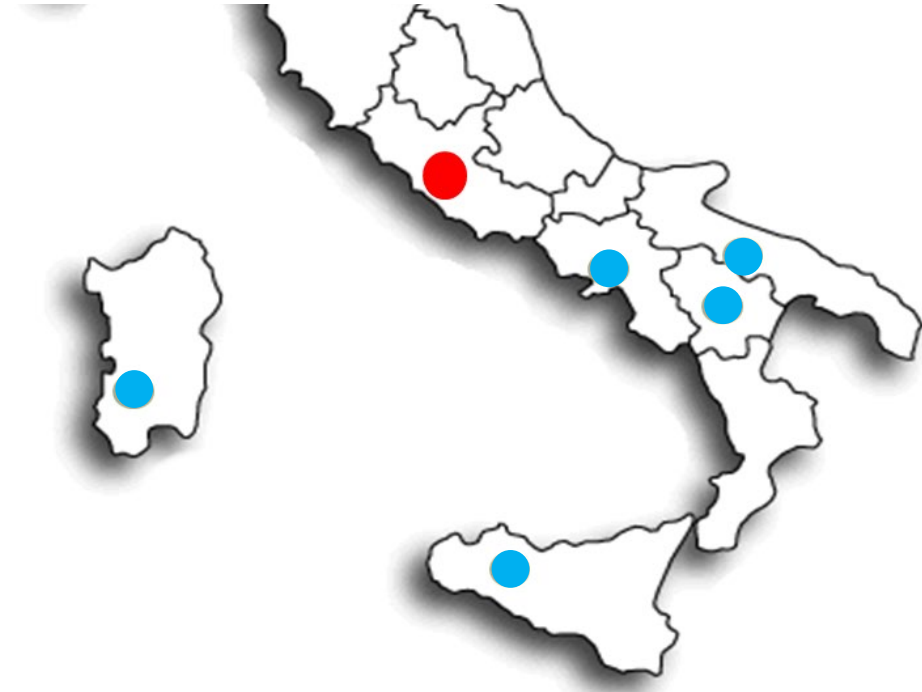
6

TOTAL

78

11 labs and 28 milk analysis lines
connected through Mosaic Clouds

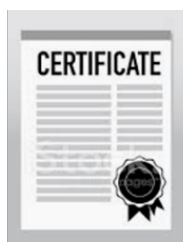
Stored 45 million MIR spectra
including small ruminants



CUS

BDN

Siall CF



A.M.

DQA

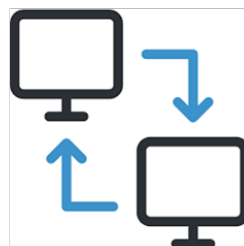
PSRN 10.2

E.U. projects



+ n...

Accordi di condivisione dei dati



Collaborative data sharing

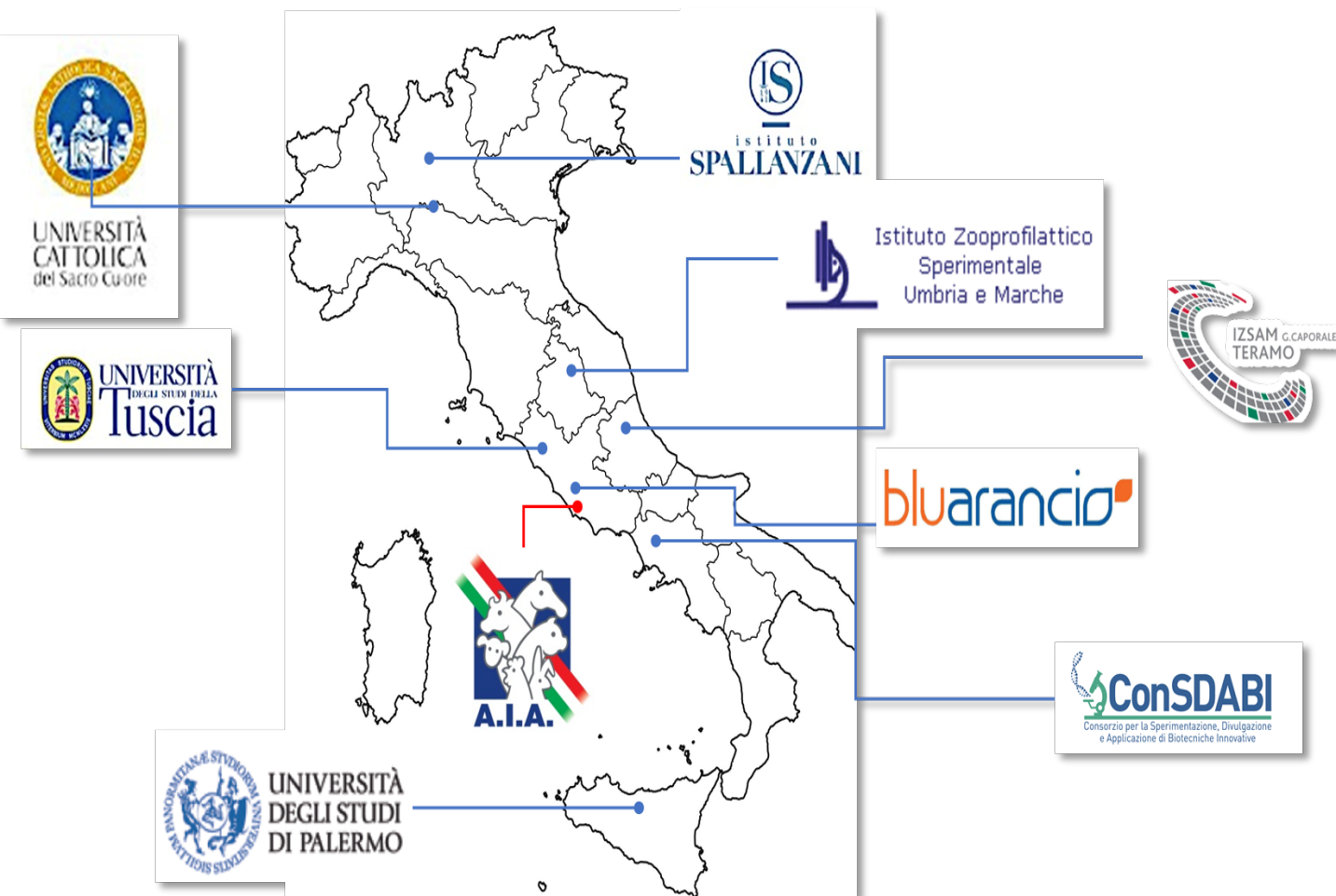
BIG DATA

A satellite map of Italy and its surrounding islands, with numerous small white dots scattered across the landmass, representing the locations of farms. The dots are most densely packed in the northern and central regions of the peninsula.

LEO farmer's engagement

Up to now almost **18,000** geo-located farms involved

More than **100 breeds**



LEO Consortium and Donors



**Fondo Europeo Agricolo
per lo Sviluppo Rurale:**
l'Europa investe
nelle zone rurali



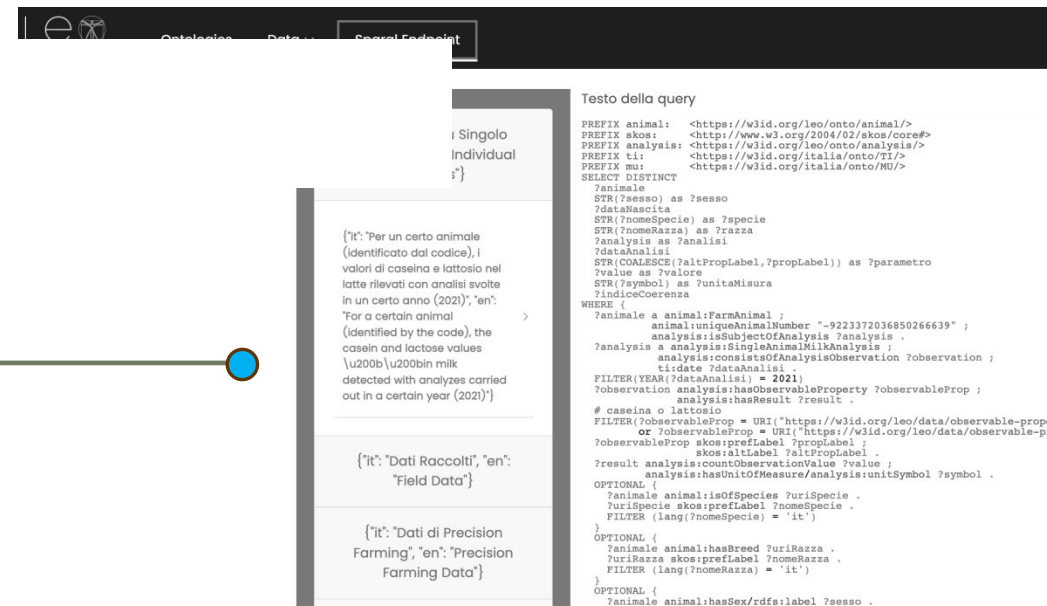
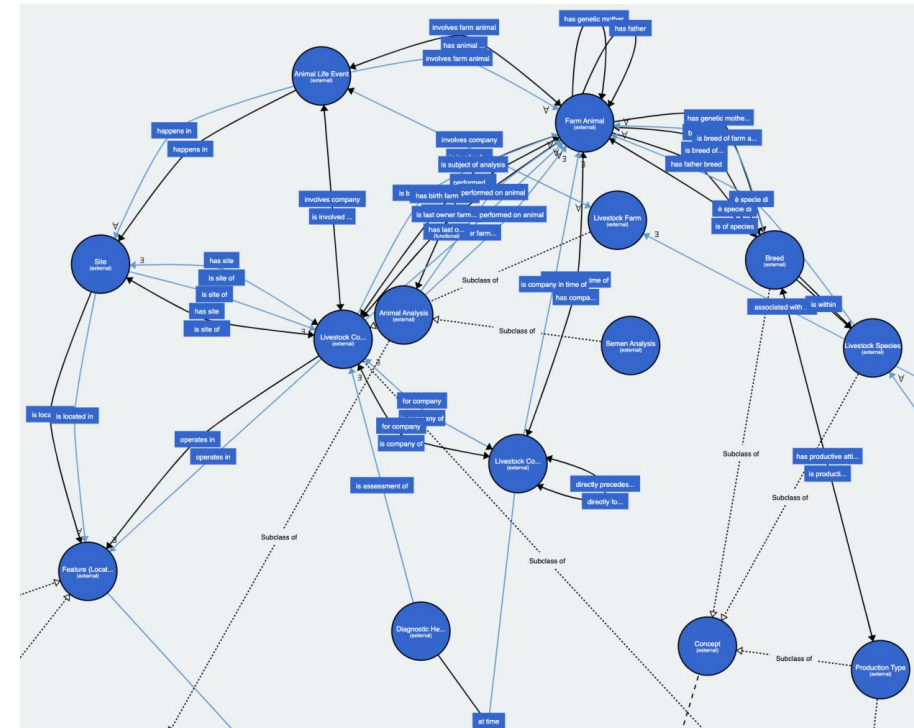
**MINISTERO DELL'AGRICOLTURA
DELLA SOVRANITÀ ALIMENTARE
E DELLE FORESTE**



Interactive navigation

Free Download open data level 5

Data and info directly delivered in farm



Leo URL: www.leo-italy.eu



Il patrimonio zootecnico italiano in Big Data

Una risorsa a supporto di allevatori, operatori di settore e comunità scientifica

Scopri

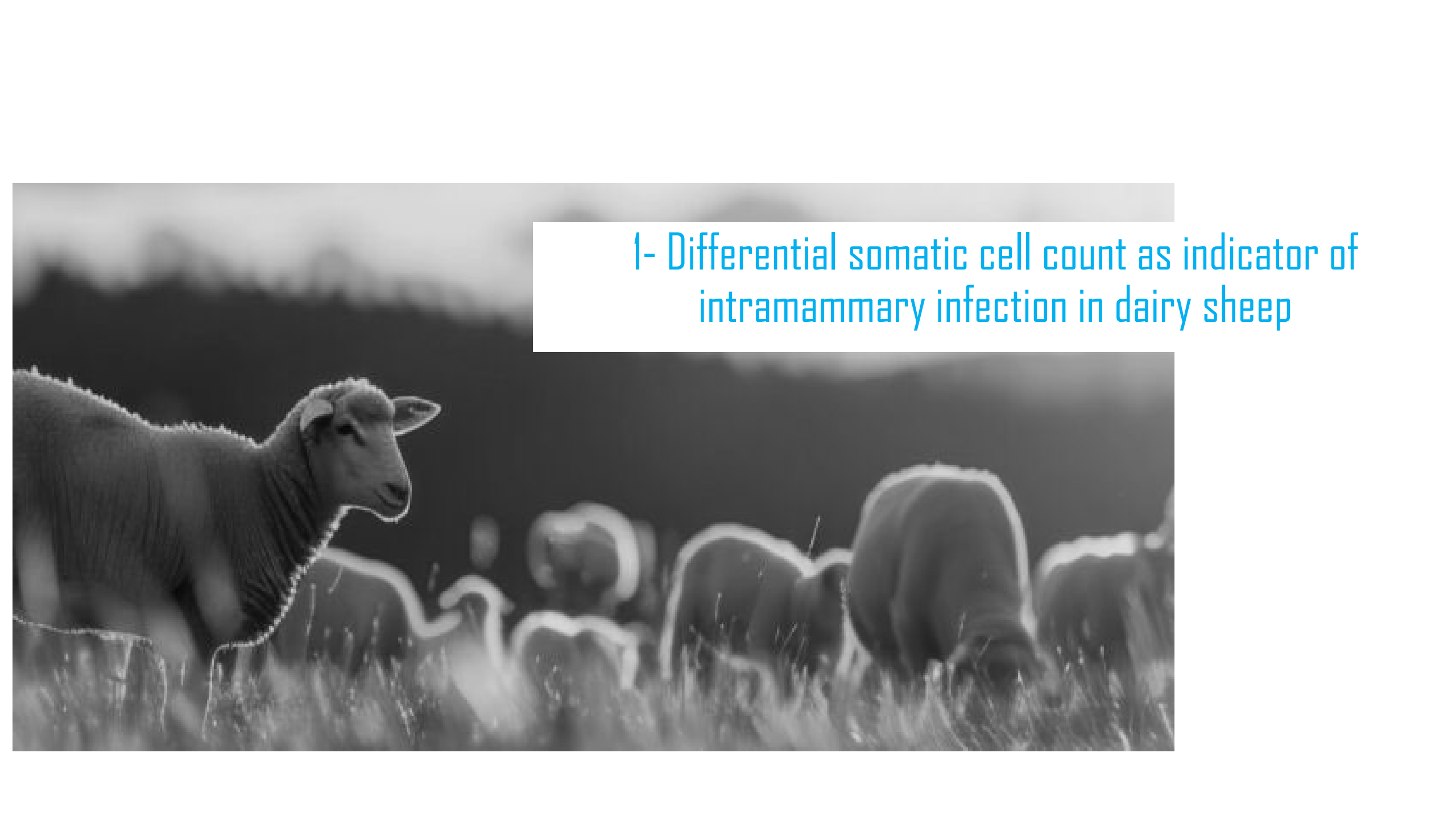


Showcasing **three outputs** fostered by
Leo Big Data

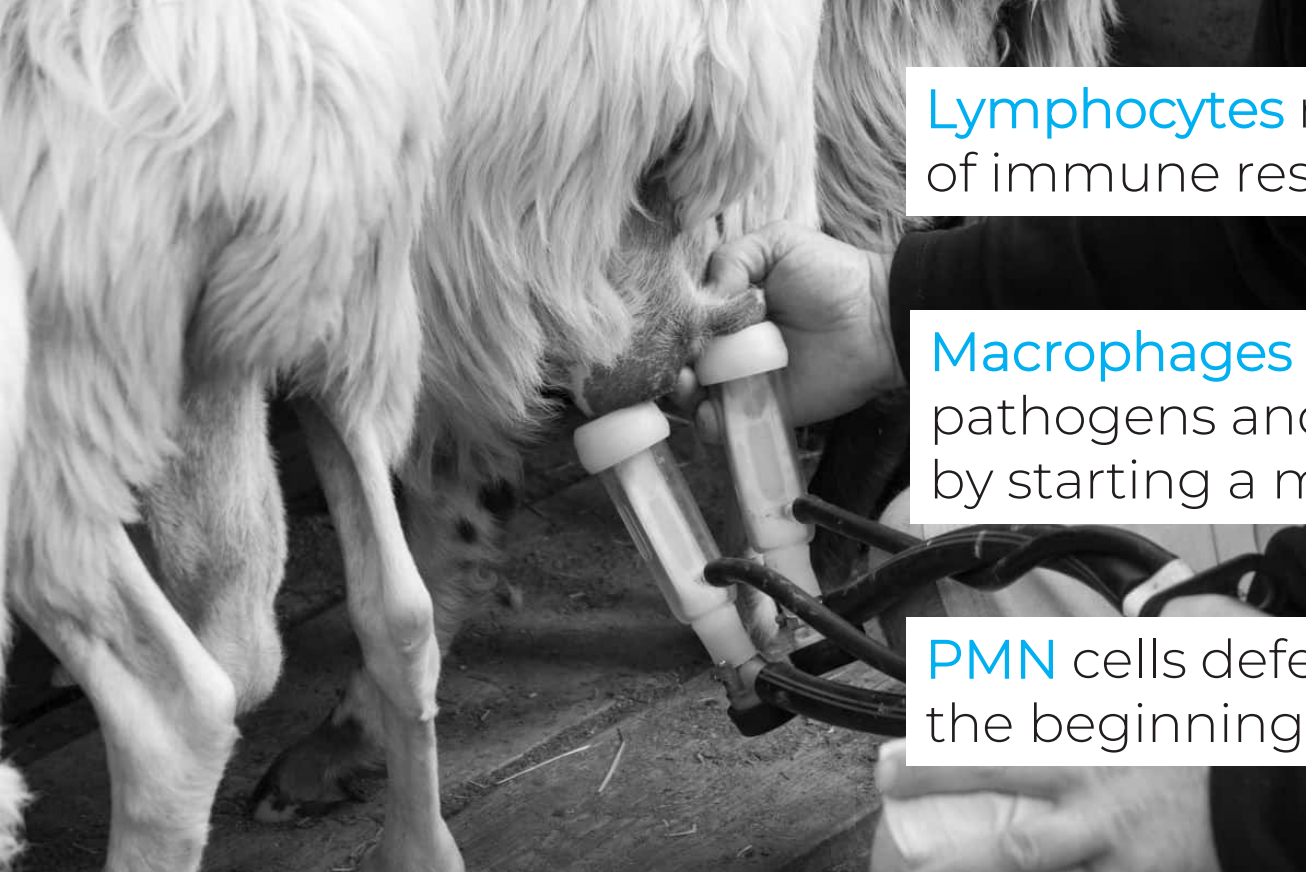
- Health and welfare in sheep

- Advanced services for small ruminant farmers





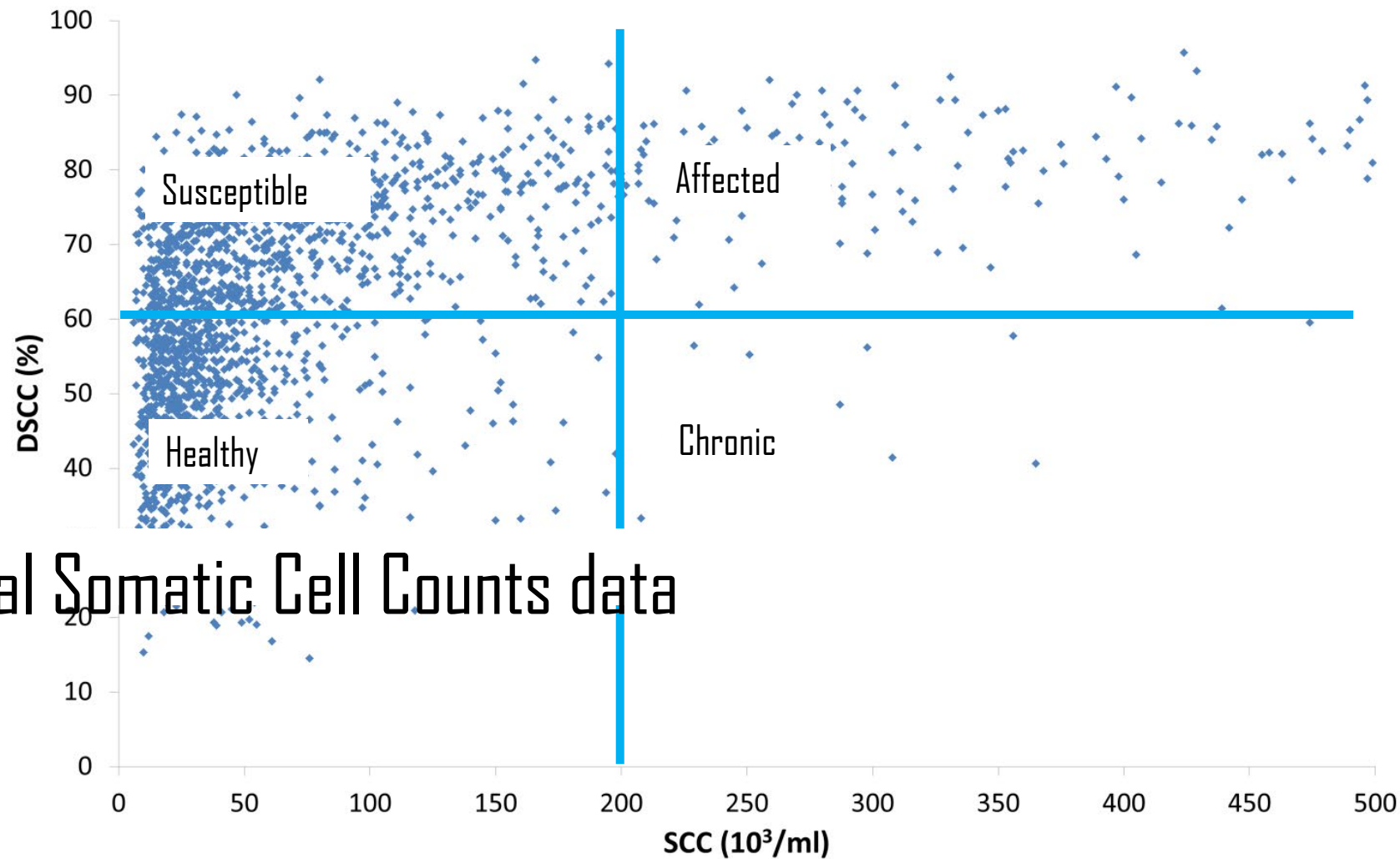
1- Differential somatic cell count as indicator of intramammary infection in dairy sheep



Lymphocytes regulate the induction and suppression of immune responses.

Macrophages recognize invading mastitis pathogens and initiate the immune response by starting a massive influx of PMN.

PMN cells defend against invading bacteria at the beginning of an infection.



Interpreting Differential Somatic Cell Counts data

[modified from Zecconi et al. Animal, 2017 for illustrative purpose only]



Materials and methods

5.556 individual milk samples from Valle del Belice breed

60 flock in Sicily

SCC e DSCC (Fossomatic 7DC)

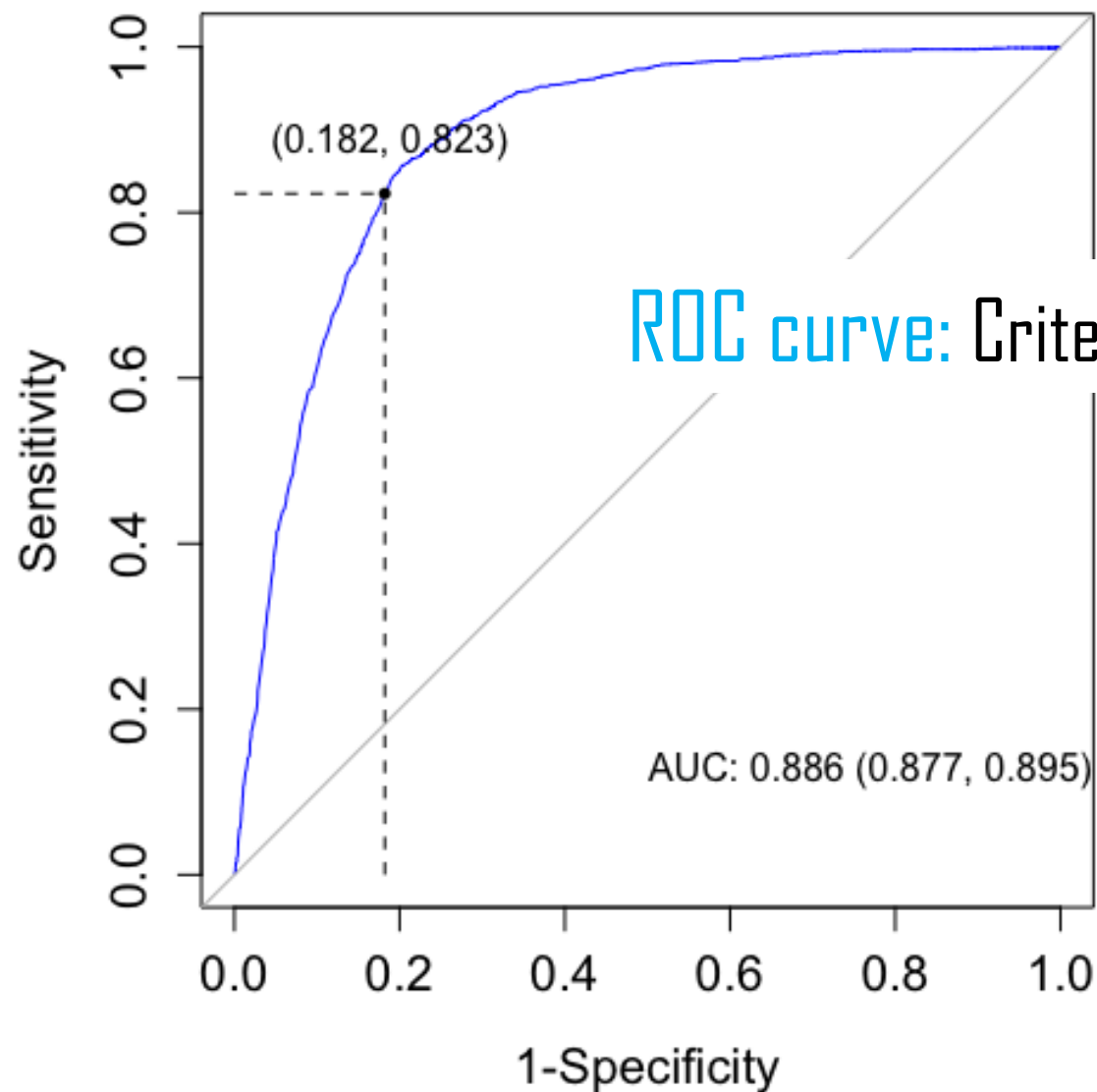
SCC threshold 650.000 cells/ml [Paape et al., 2007]

Data QC rules $SCC < 50.000$ e Gose = 0

ROC curves (R packages e.g. OptimalCutpoints)

Main results

SCC class (x 1000)	n	1 st Qu.	3 rd Qu.	Max	DSSC (%)	SE
50 < SCC ≤ 250	2,601	47.8	75.3	97.6	61.16 ^a	0.27
250 < SCC ≤ 500	826	59.9	79.5	95.3	68.27 ^b	0.56
500 < SCC ≤ 750	413	70.8	83.6	95.3	75.20 ^c	0.74
750 < SCC ≤ 1,000	251	77.2	85.6	93.0	80.14 ^d	0.92
1,000 < SCC ≤ 1,500	376	80.2	87.4	95.60	82.97 ^e	0.77
SCC > 1,500	1,089	83.5	88.8	97.80	85.75 ^f	0.50



ROC curve: Criterion Max Specificity-Sensitivity

Threshold	80.7 %
Sensitivity	0.82 (0.80-0.84)
Specificity	0.82 (0.80-0.83)
Positive predicted	0.67(0.65-0.70)
Negative Predicted	0.91 (0.90-0.92)
AUC	0.89 (0.88-0.90)

Expectation in line with epidemiological data

7.2 %

Susceptible

34.5 %

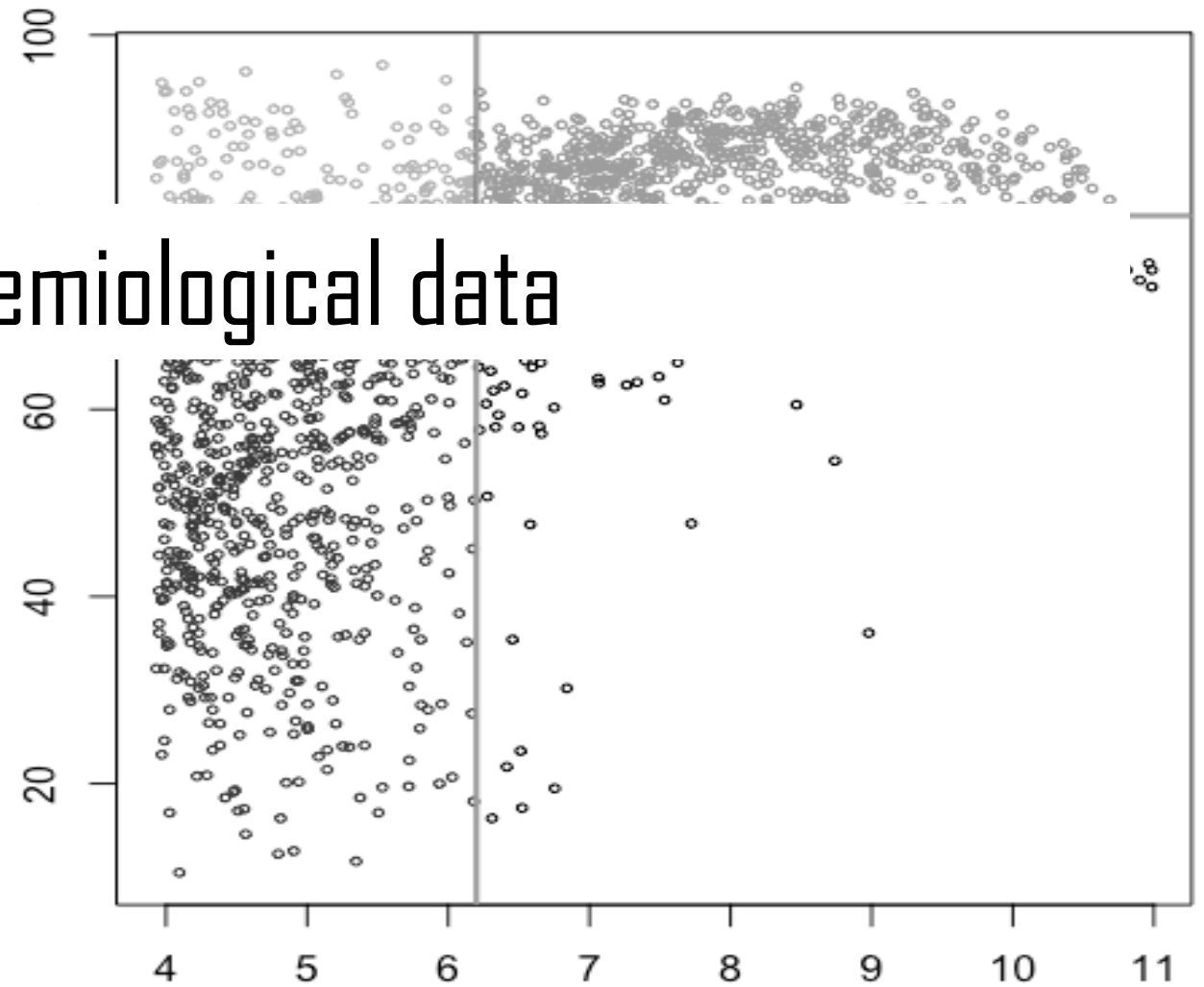
Affected

46.5 %

Healthy

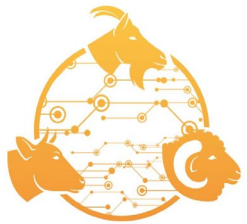
11.7 %

Cronic





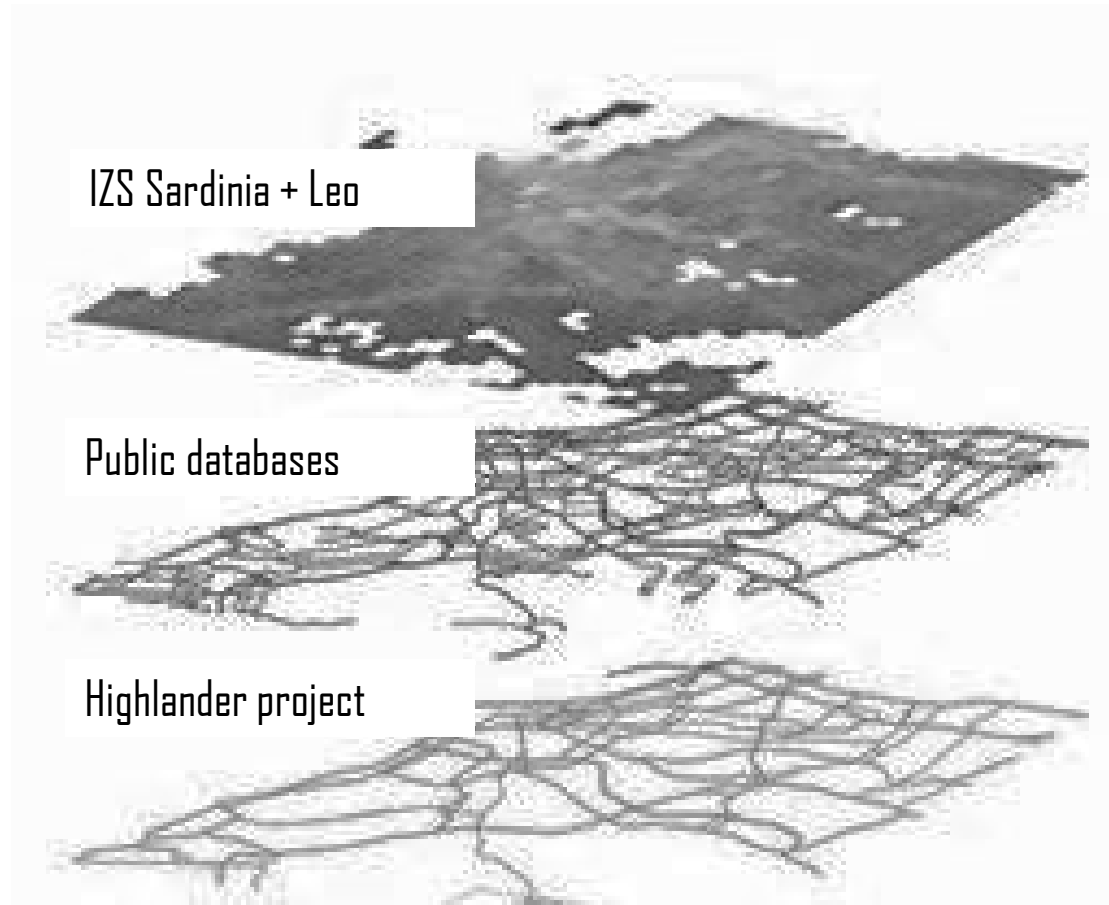
2. Advanced risk-maps for Blue Tongue in sheep



SEBASTIEN



Materials and methods



Data from 2013 **Blue Tongue epidemic** collected from **5,600 farms**

Information on soil characteristics, and the presence or absence of water near the farm.

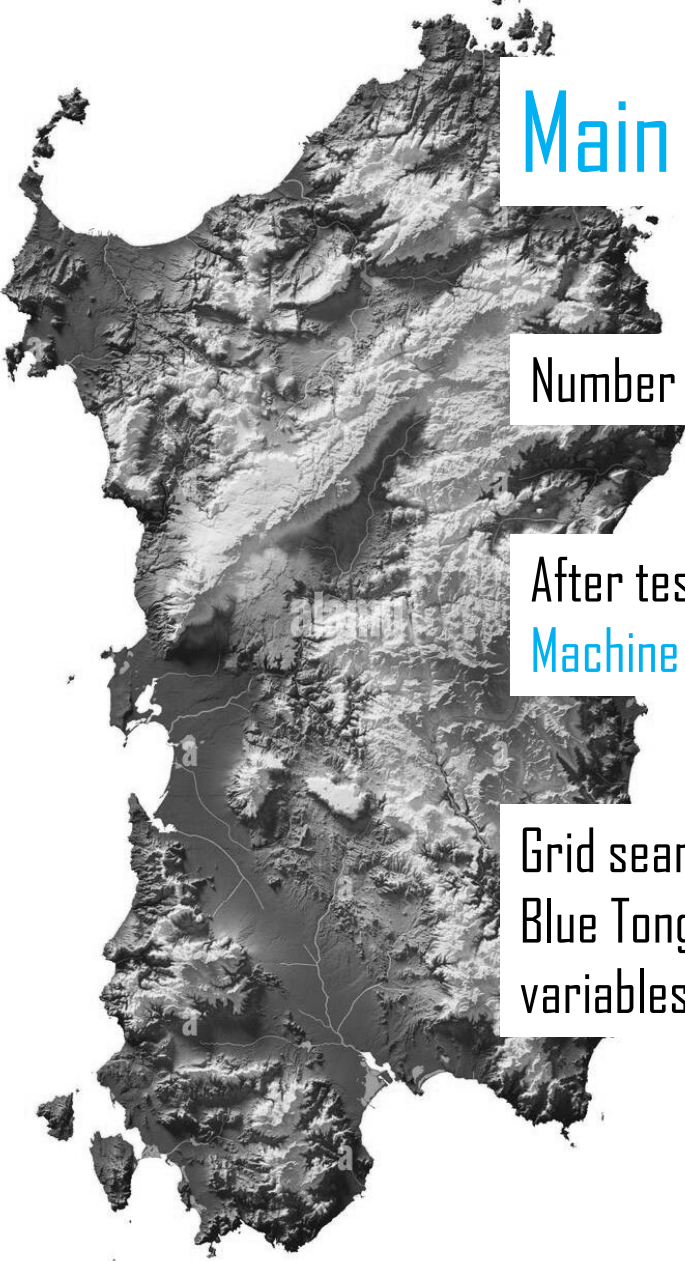
Climatic variables from VHR-REA_IT dataset with a spatial resolution of 2.2 km collected up to 60 days before the clinical case (historical 1989-2023)



Remove collinear climatic variables via Pearson correlation coefficient

Identify the best algorithm family among 1) Distributed Random Forest; 2) GLM; 3) Extreme Gradient Boost and 4) Gradient Boosting Machine.

Optimize the parameters of the best algorithm and subset the informative variables



Main Outcomes

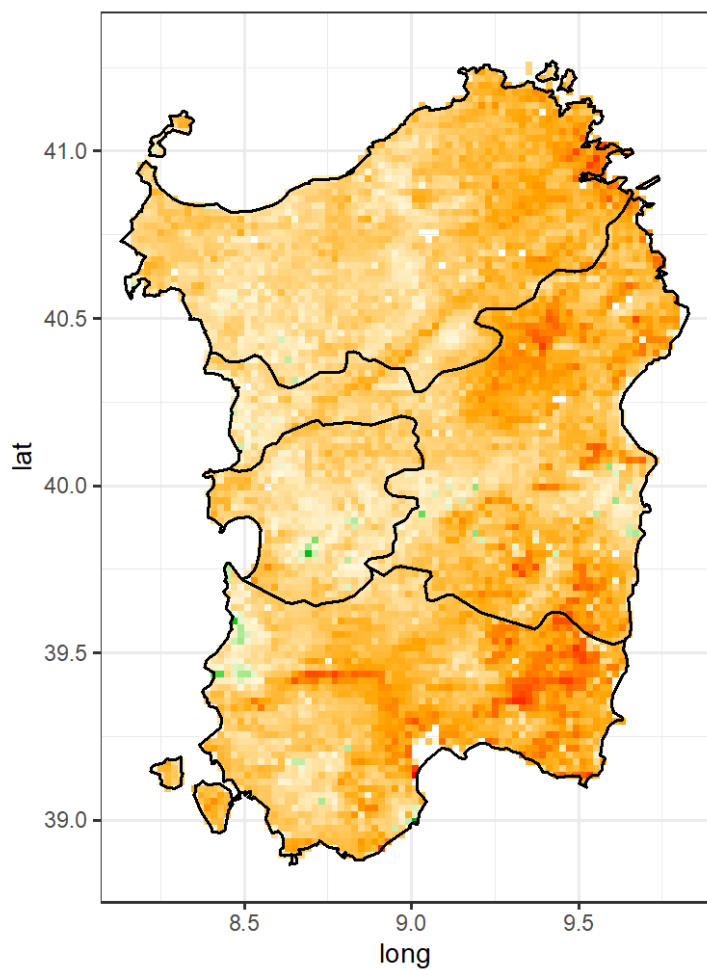
Number of climatic variables reduced from 120 to 58

After testing more than 1800 models, the most performant was the **Gradient Boosting Machine**

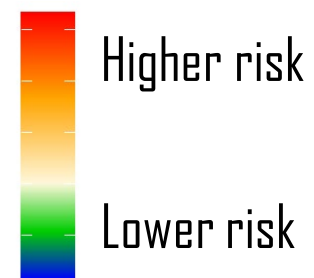
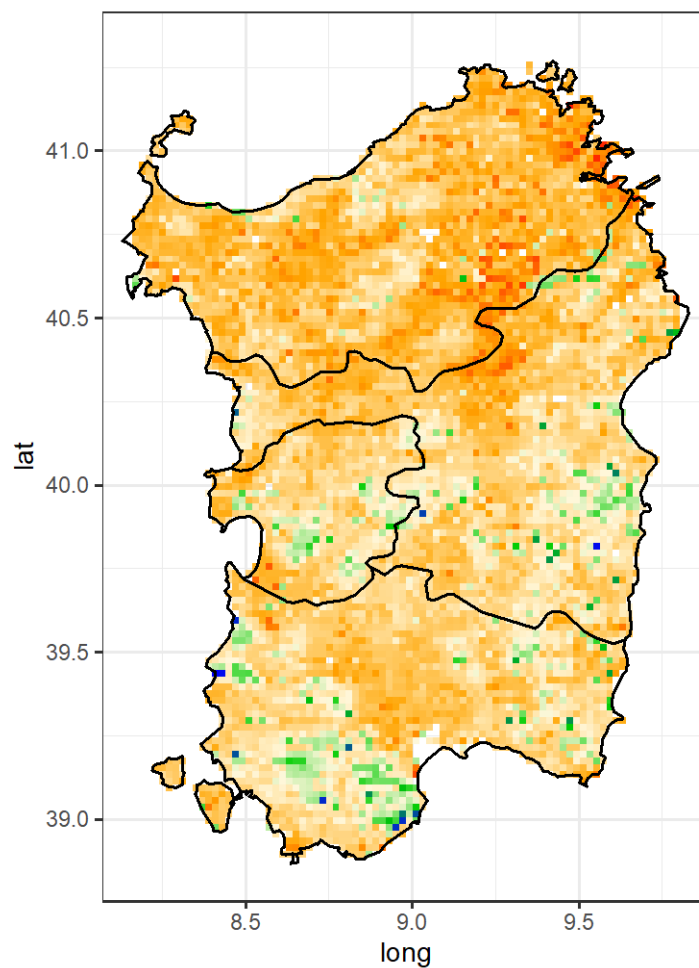
Grid search applied to GMB resulted in **43 variables significant** to predict the probability of Blue Tongue infection (number of animals + vaccination + climatic variables + environmental variables)

Model applied to
climatic **projection** data (from 2021 to 2050)

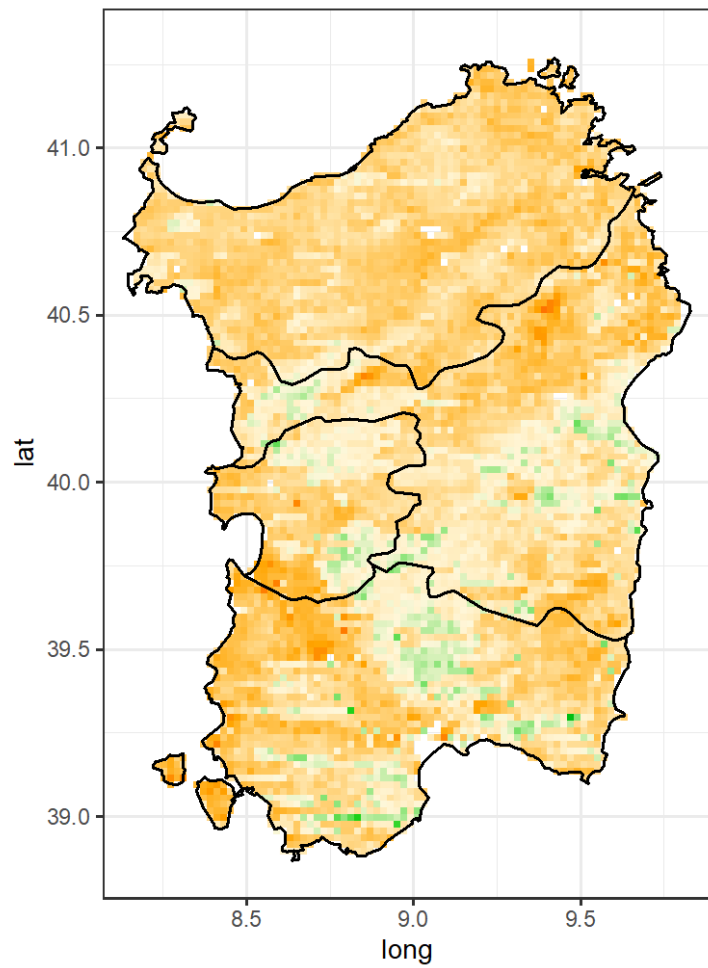
July



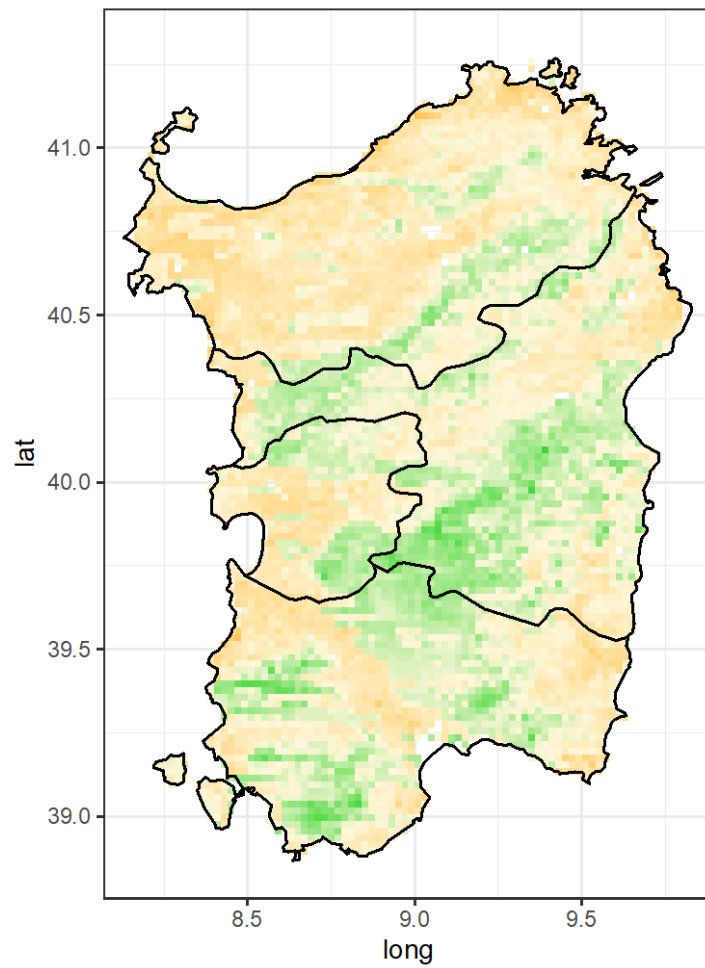
August



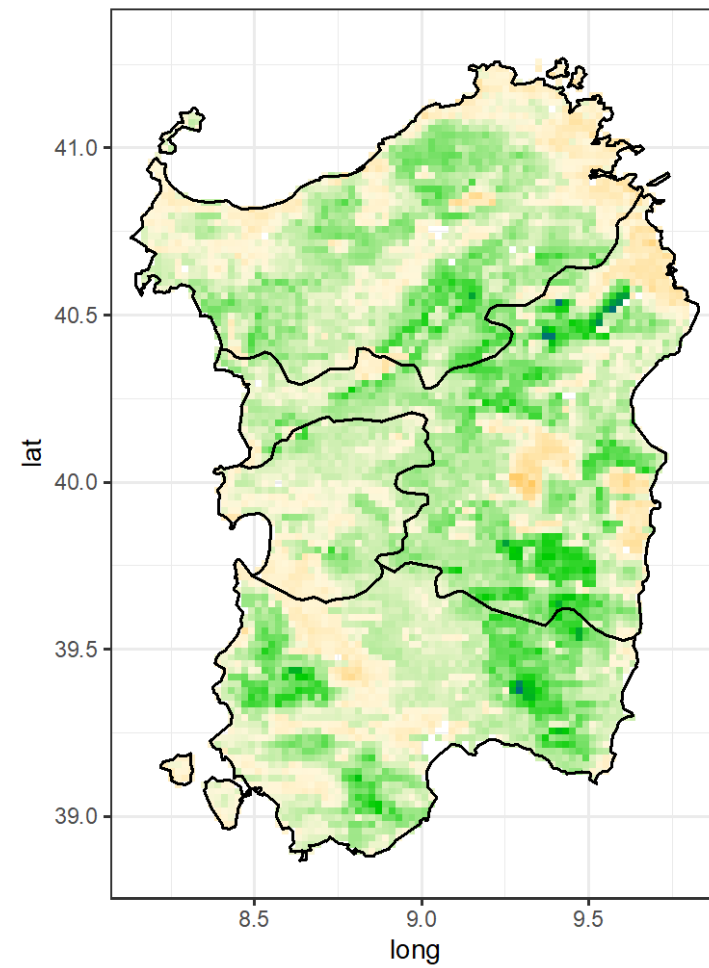
September



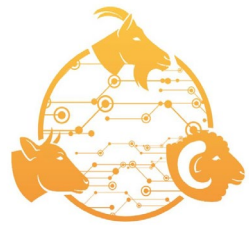
October



November



3-Decision Support System predicting biomass availability



SEBASTIEN



agritech

National Center for
Technology in Agriculture

Plan. Plant. Planet.



Materials and methods



Direct Data on pasture collected from June 2023 to May 2024 at three farms in northern Lazio (Italy)

9 lab parameters [Dry matter, Ashes, Crude protein, ADF, ADL ...]

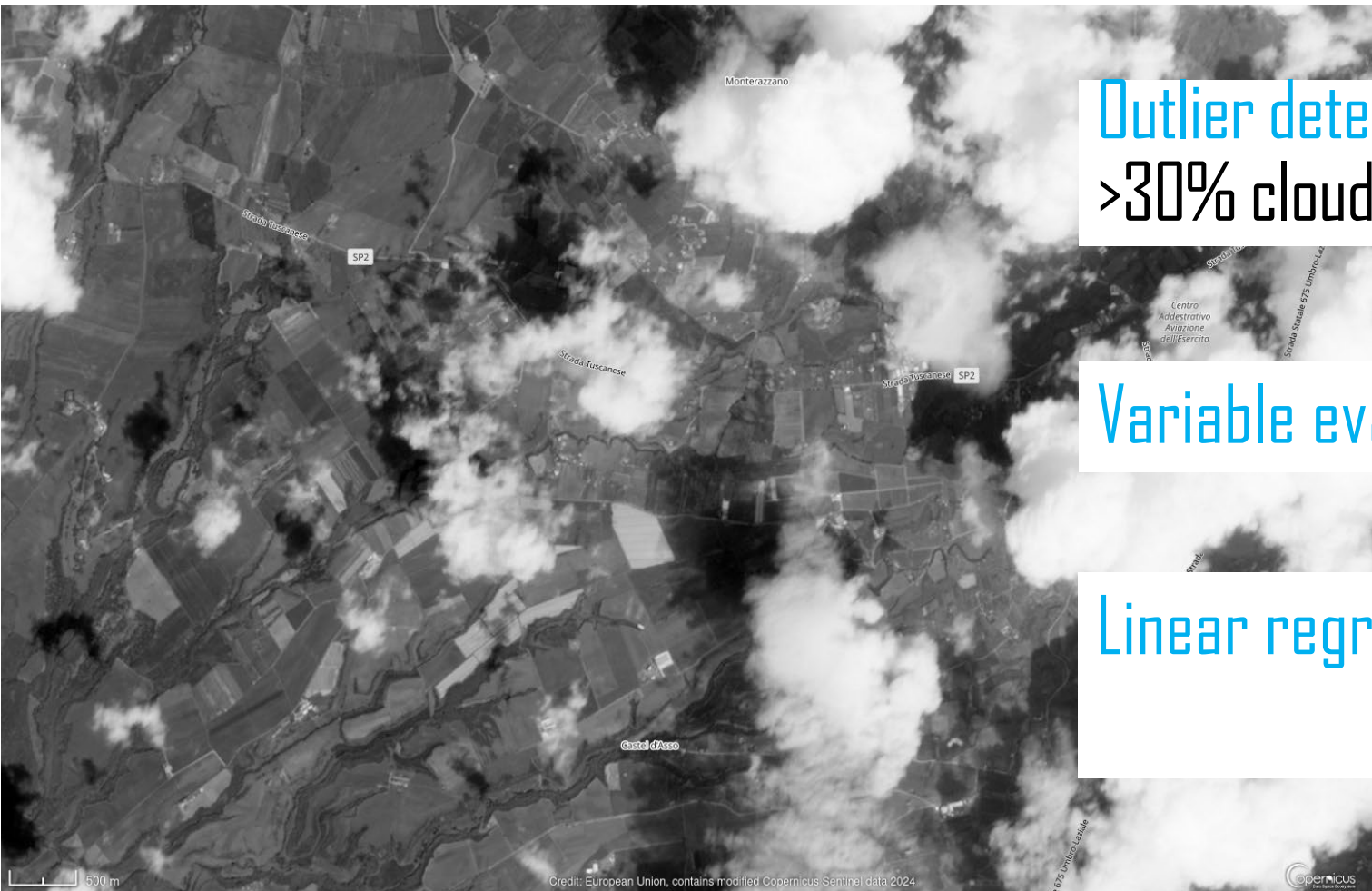


Sentinel-2 spectral bands:

B02, B03, B04, B08, B08A, B12

9 Vegetation indices

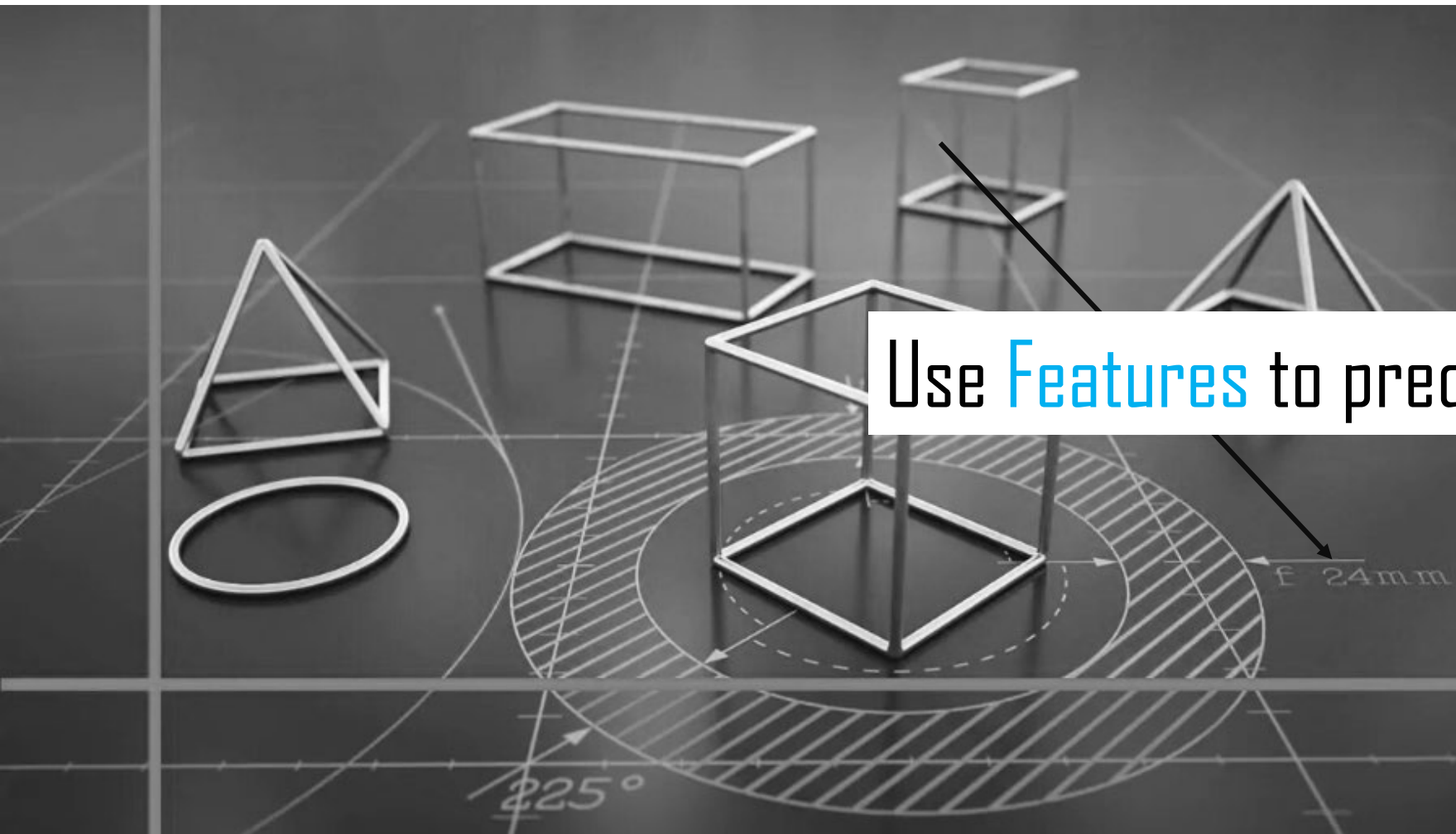
NDVI, NDWI, GCI, ARVI, GLI



Outlier detection and removal ($> 3 \sigma$ or
>30% cloudy)

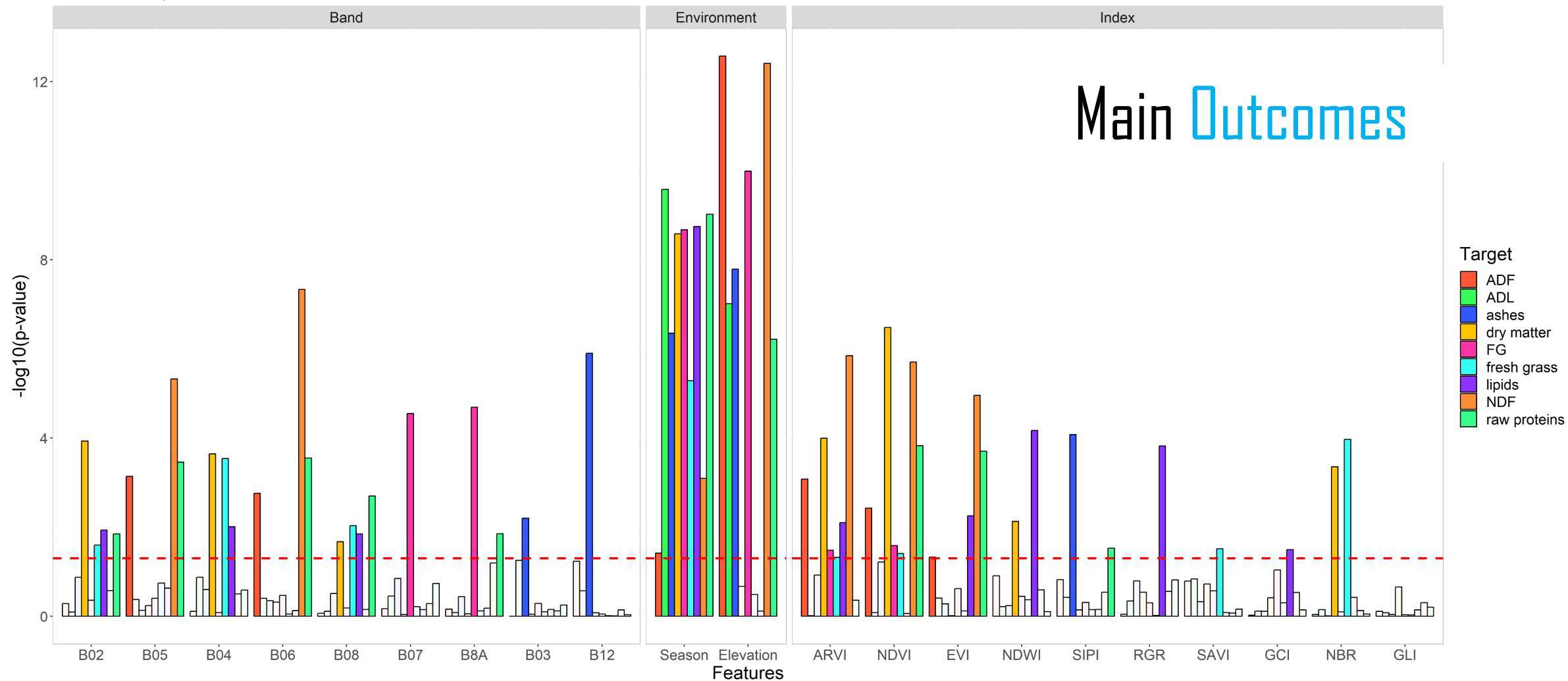
Variable evaluation and selection

Linear regression and model evaluation



Use **Features** to predict **Targets**

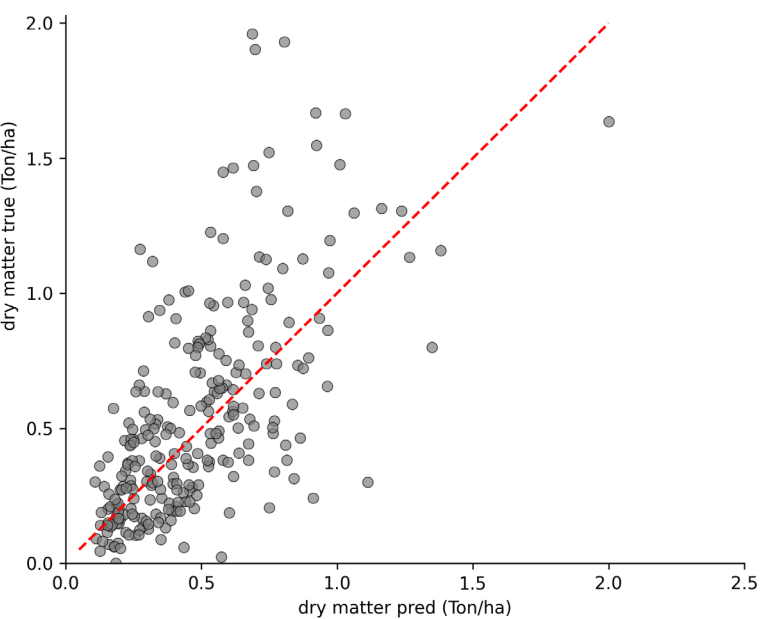
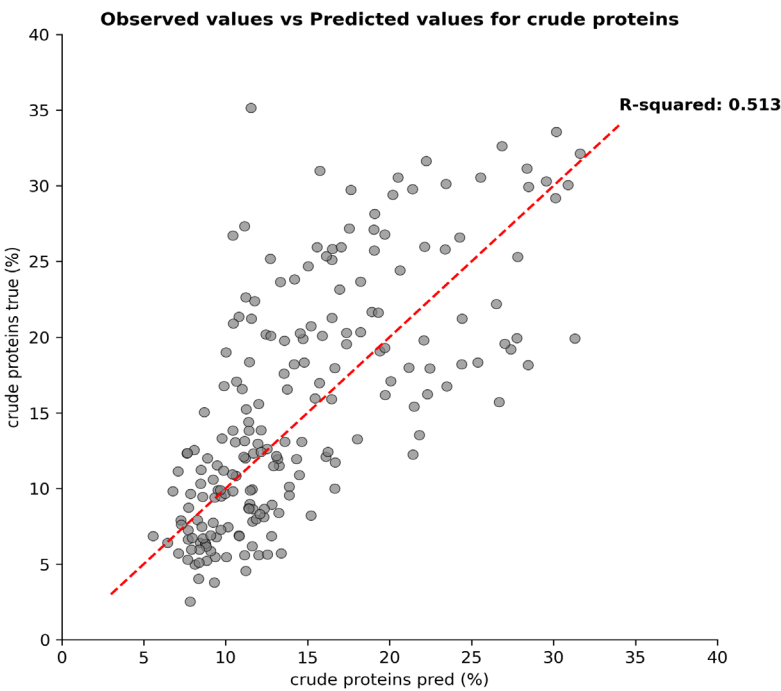
Variable importance



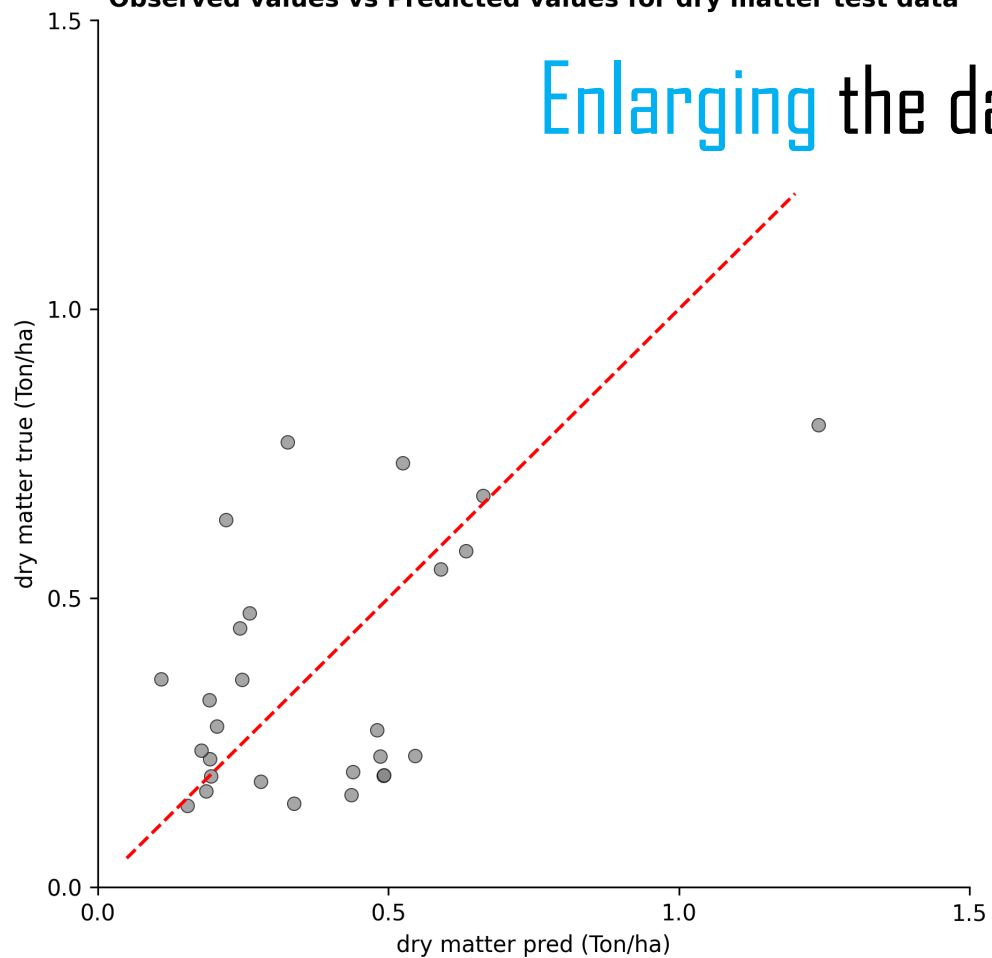
Main Outcomes

Seasonality and elevation effects are almost ubiquitous

Target	«real» MAE	«naive» MAE	R ²
Fresh grass	0.86	1.21	0.32
Dry matter	0.24	0.33	0.38
Ashes	2.08	2.58	0.33
Crude Proteins	4.46	6.74	0.51
Lipids	0.40	0.54	0.40
aNDFom	7.95	10.74	0.44
ADF	6.25	7.96	0.35
ADL	2.41	2.81	0.29
FIBER	5.30	7.07	0.46



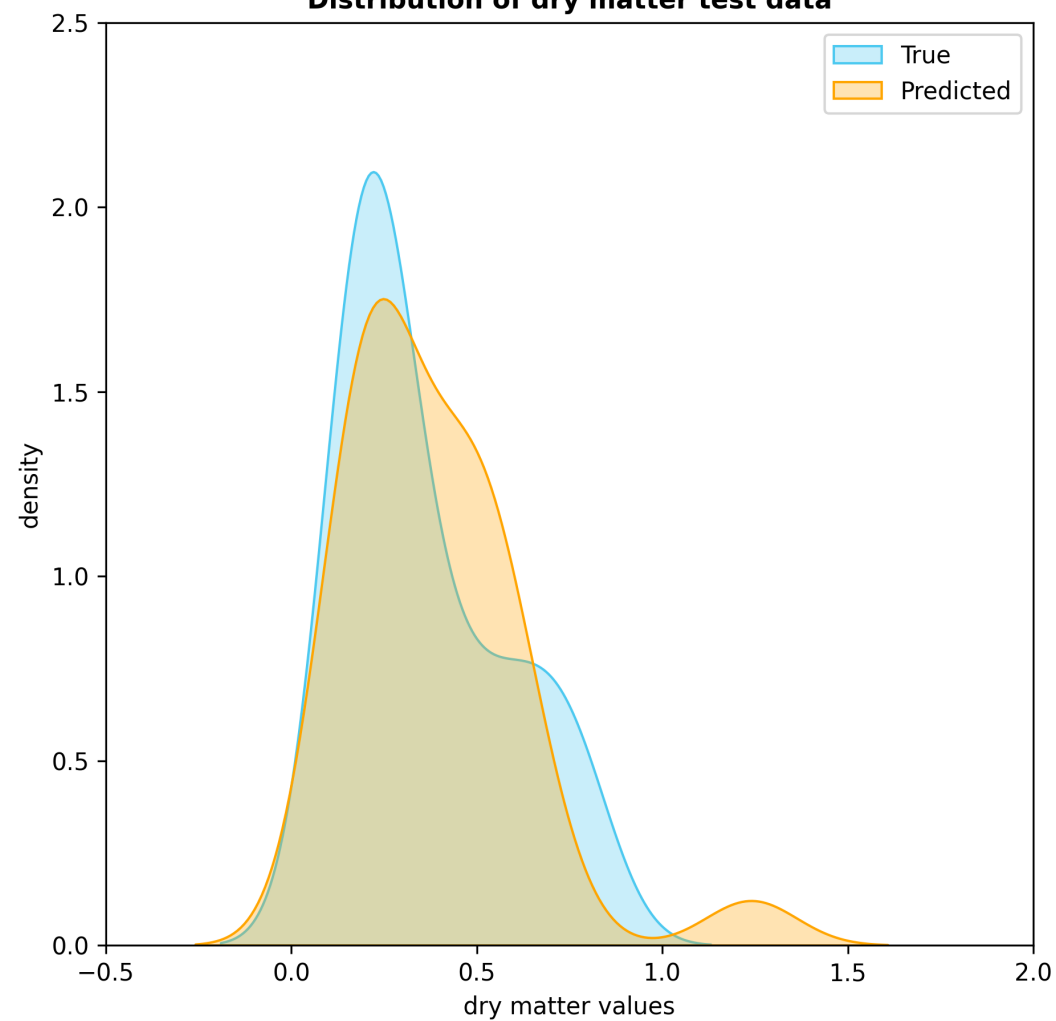
Observed values vs Predicted values for dry matter test data



Real dry matter 0.36 ± 0.21
Predicted dry matter 0.38 ± 0.23

Enlarging the dataset will improve correlations

Distribution of dry matter test data





Tolls available in a β -version app for farmer



Monitoraggio del pascolo

AREA DEL PASCOLO

☐ SHAPE file☒ Area

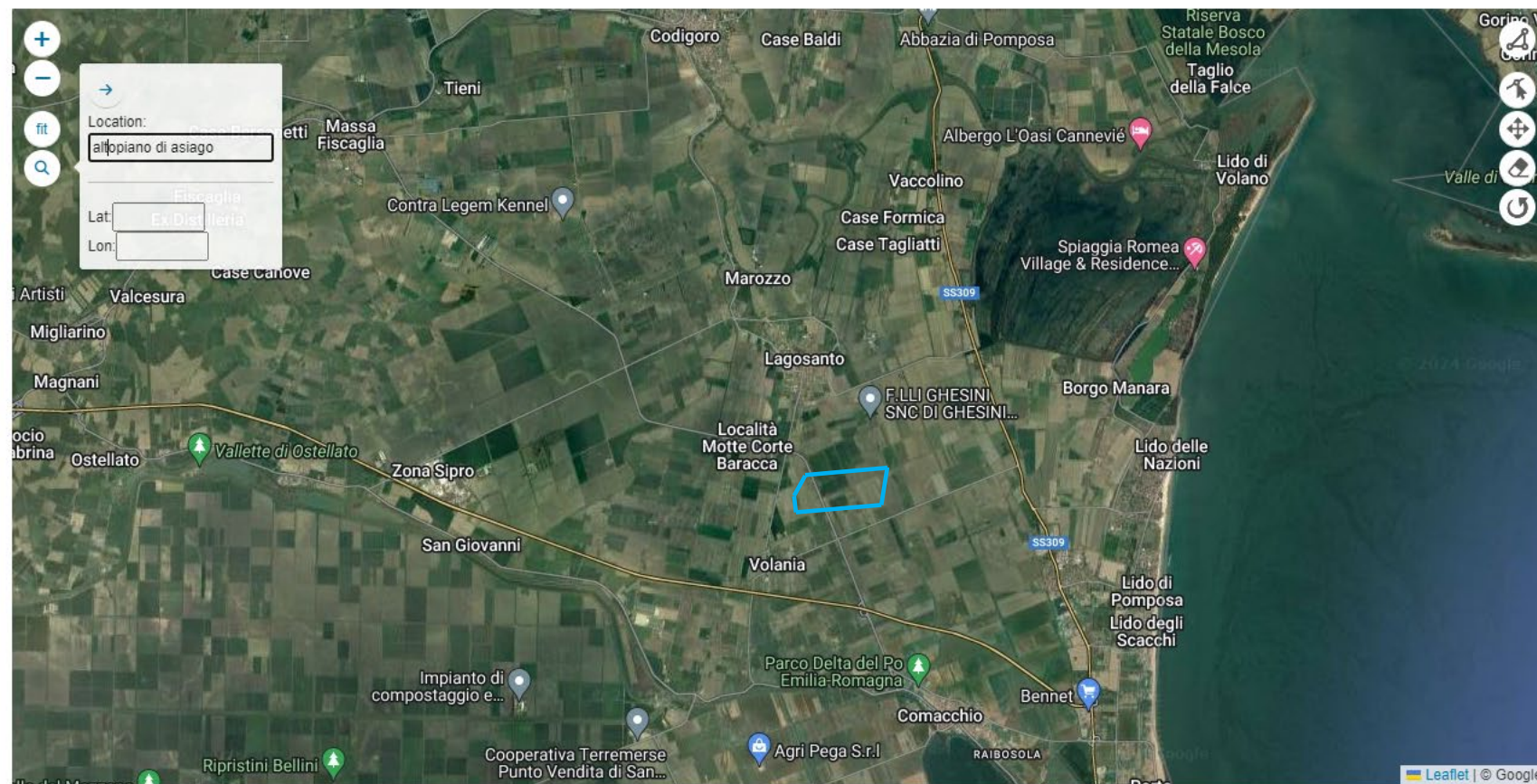
AREE SALVATE

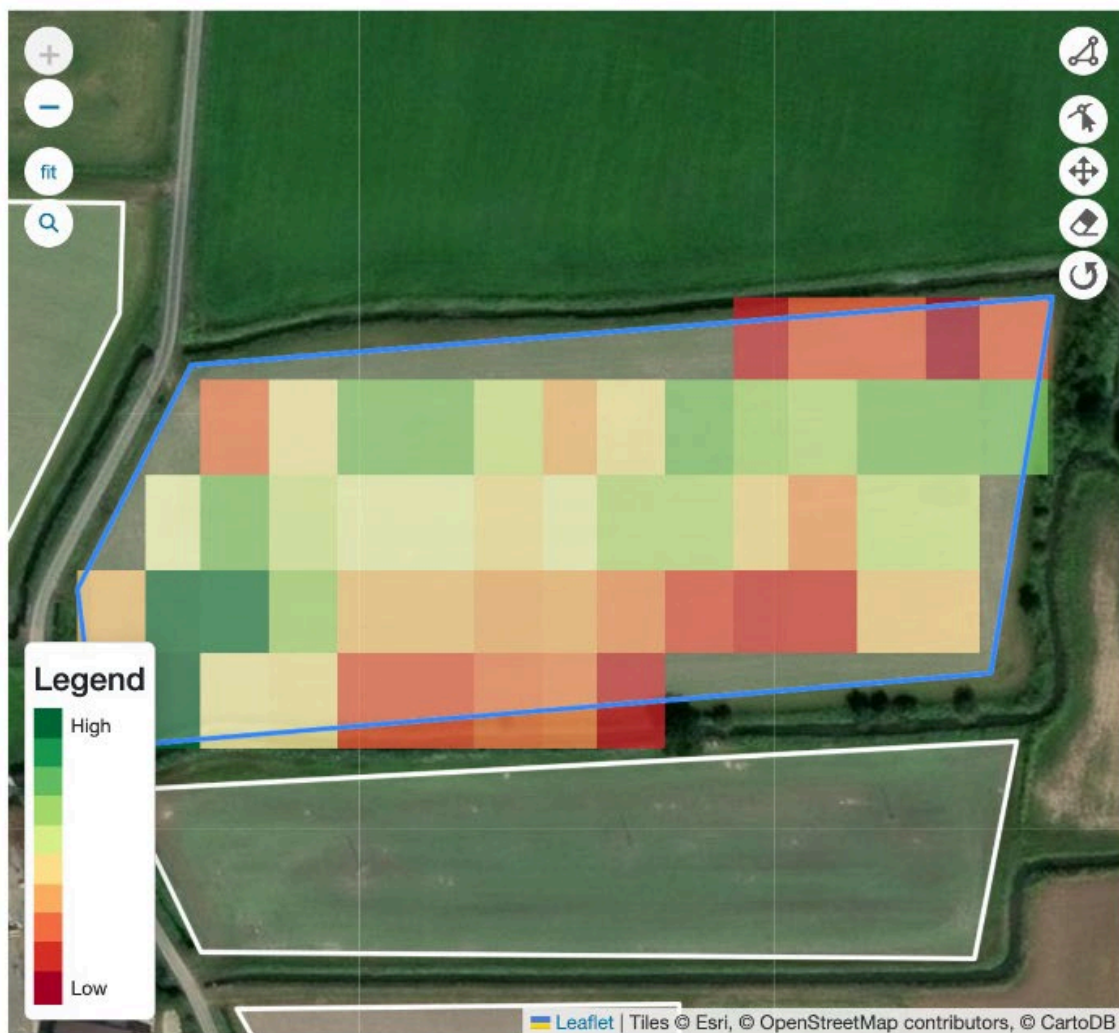


Pascolo Rubizzano



Pascolo San Benedetto





PASTURE AVAILABILITY



Data: 28/08/2024 10:20:21
Biomass As It Is: 446.27 ton
Biomass Dry: 46.96 ton
Surface: 18.75 ha

Density: 5.333 capi/ha

SPECIES

Sheep



NUMERO DI CAPI

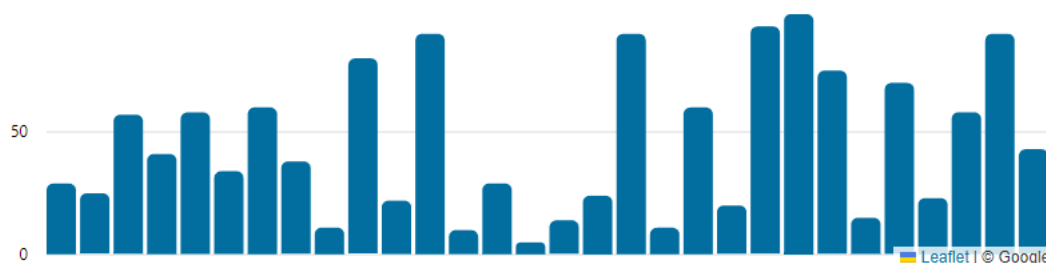
GIORNI

100

62

☒ Biomassa secca ☐ Biomassa talquale

Biomassa secca





LEO

vision on Big-Data

Taking complexity in livestock with data driven approaches

Networking data and pipelines national and international wise

Sharing knowledge across expertises, borders, production models and food cultures

Supporting the global transition to a more sustainable food production

Foster the One Health attitude

Develop scalable and stepwise solutions



A/LIVESTOCK-DATA HACKATHON

Launch : October 1^o 2024

Who are we looking for: "young" passionate scientists or with youthful spirit

What we ask for: ideas to shape the future of livestock using the power of data

What we offer: an exciting challenges and.....

More info soon on



<https://vhive.buzz>

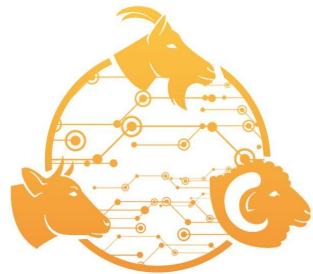
<https://leo-italy.eu>

Thanks to the large team of contributors

Thanks you all for listening

Happy to collaborate





SEBASTIEN

Smarter livEstock Breeding through Advanced Services Tailoring
Innovative and multi-sourcE data to users' Needs

<https://www.sebastien-project.eu>



agritech

National Center for
Technology in Agriculture

Plan. Plant. Planet.

Centro Nazionale Agritech, finanziato dall'Unione Europea –
NextGenerationEU

Spoke 5: Sustainable productivity and mitigation of environmental
impact in livestock systems