Data Platform and AI for Precision Farming

Soil Moisture Modeling and Assessment as a Case Study

whoami

Matteo Francia

- Email: m.francia@unibo.it
- Research fellow @ UniBO

Research topics

- Spatio-temporal analytics
- Big data platforms

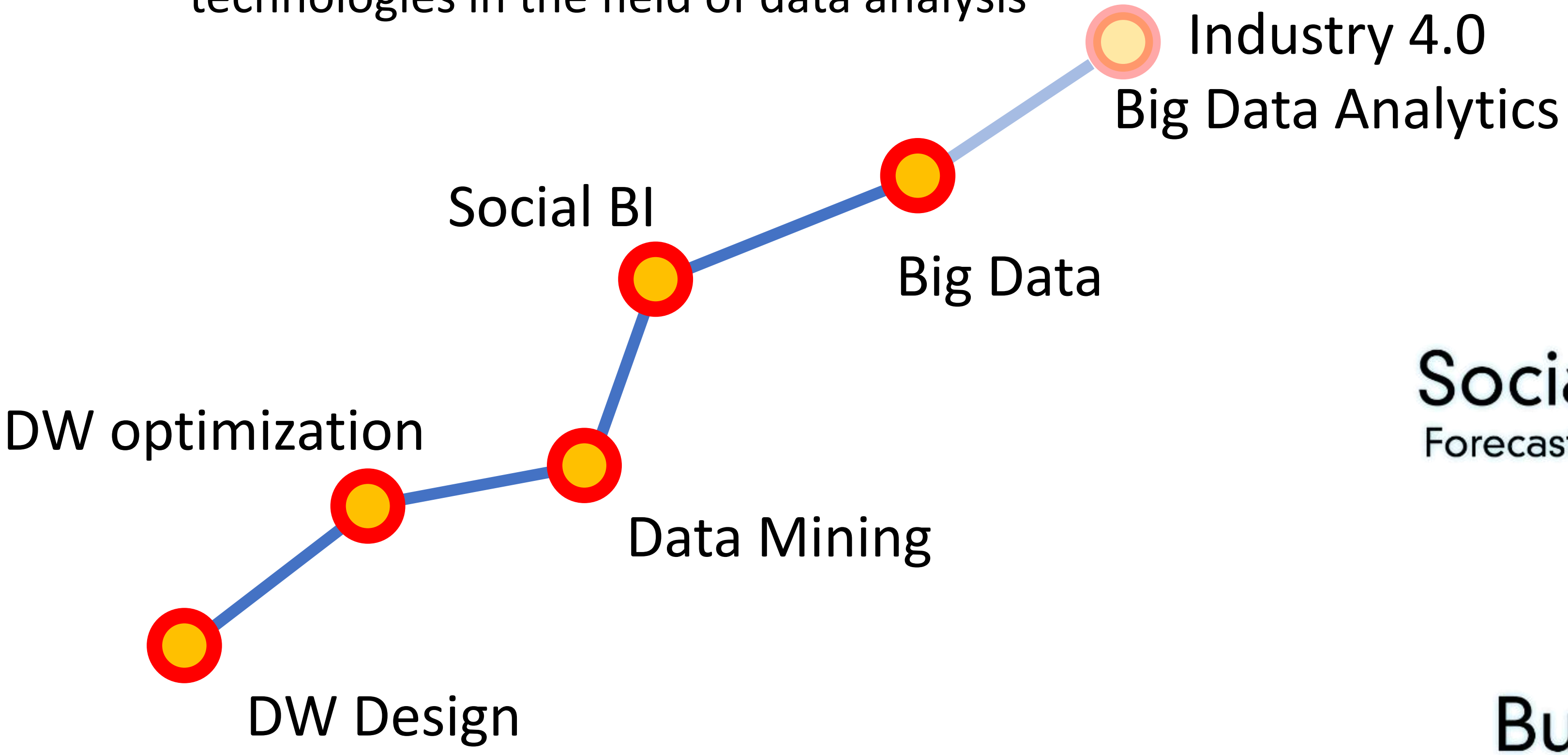
Website

- https://big.csr.unibo.it/

BUSINESS INTELLIGENCE GROUP

The Business Intelligence Group

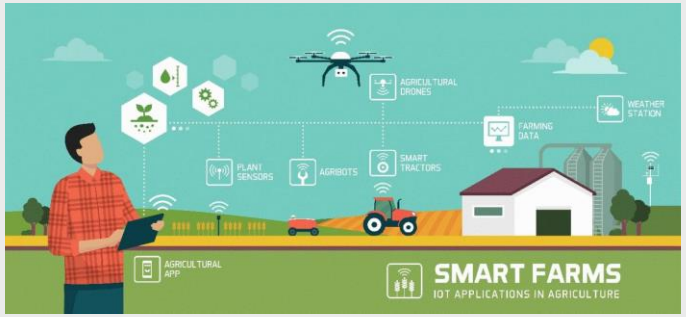
The Business Intelligence Group carries out researches on methodologies, techniques and technologies in the field of data analysis



Roadmap

Introduction to Data Platforms

Soil Moisture Modeling and Assessment as a Case Study



Acknowledgments

Part of this talk is the result of a joint work

- DISI @ UniBO --- Department of Computer Science and Engineering
  - Prof. Matteo Golfarelli
  - Dr. Matteo Francia
  - Dr. Joseph Giovanelli
- DISTAL @ UniBO --- Department of Agricultural and Food Sciences
  - Prof. Moreno Toselli

How did we get here?

Data-Driven Innovation

- Use of data and analytics to foster new products, processes and markets
- Drive discovery and execution of innovation, achieving new services with a business value

Analytics

- A catch-all term for different business intelligence (BI)- and application-related initiatives
- E.g., of analyzing information from a particular domain
- E.g., applying BI capabilities to a specific content area (e.g., sales, service, supply chain)

Advanced Analytics

- (Semi-)Autonomous examination of data to discover deeper insights, make predictions, or generate recommendations (e.g., through data/text mining and machine learning)

https://www.gartner.com/en/information-technology/glossary (accessed 2022-08-01)

Towards a unifying data platform

Data-driven company refers to companies where decisions and processes are supported by data

- Decisions are based on quantitative rather than qualitative knowledge
- Processes & Knowledge are an asset of the company and are not lost if managers change
- The gap between a data-driven decision and a good decision is a good manager

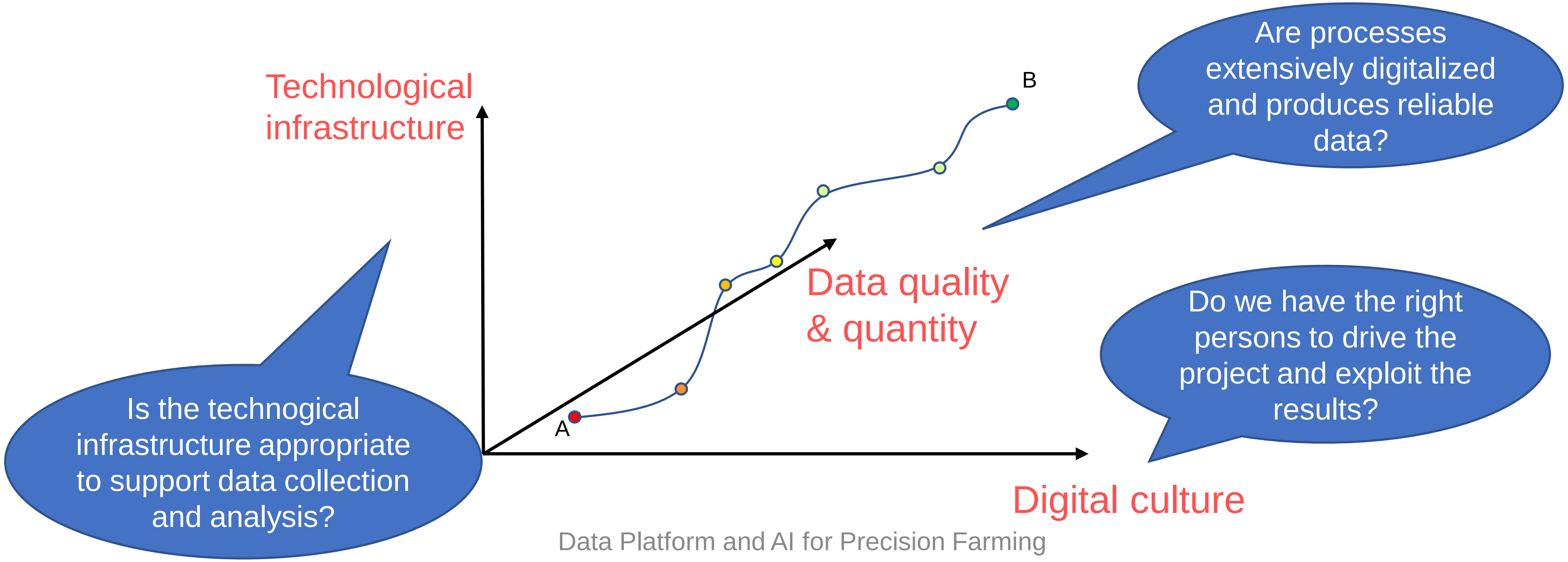
Adopting a data-driven mindset goes far beyond adopting a business intelligence solution and entails:

- Create a data culture
- Change the mindset of managers
- Change processes
- Improve the quality of all the data

Towards a unifying data platform

Digitalization is a journey that involves three main dimensions.
Moving from A to B is a multi-year process made of intermediate goals, each of which must be feasible:

- Solves a company pain and brings value
- Can be accomplished in a limited time range (typically less than one year)
- Costs must be economically related to gains



Reference scenario

Companies are collecting tons of data to enable advanced analytics

- Raw data are difficult to obtain and more and more heterogeneous and complex
- … making harder their collection, integration, transformation, and maintenance
- There is a need for describing/curating the data to make them consumable

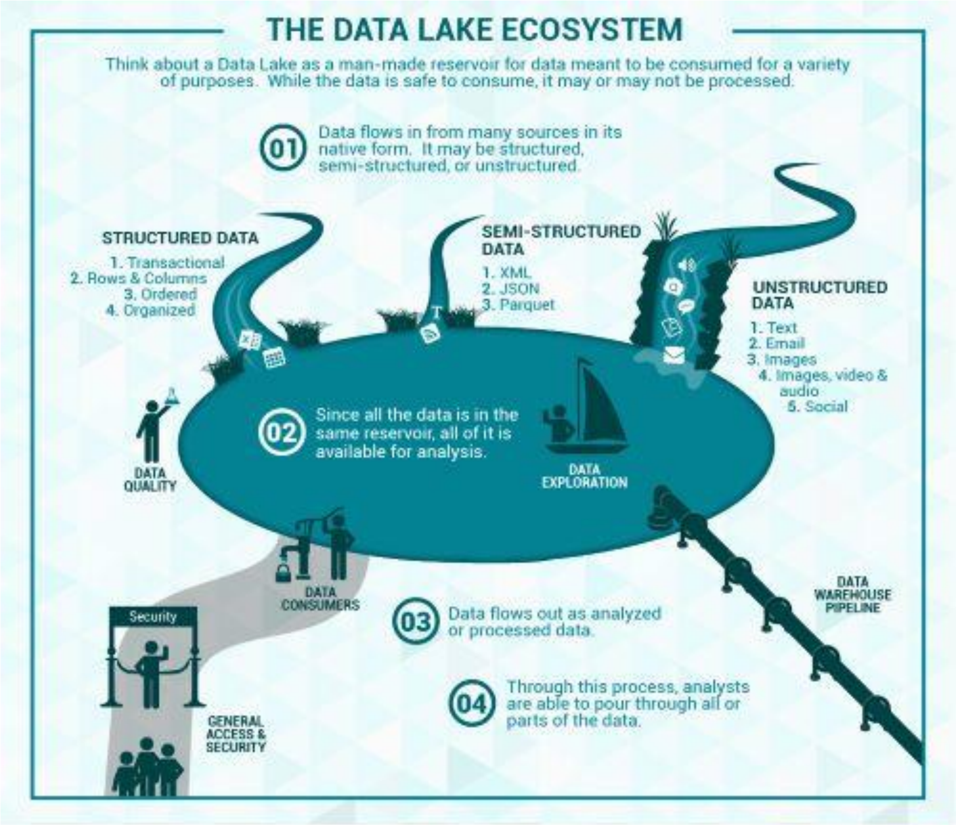
Assuming we got the data, where are we collecting and processing data?

- Getting value from data is not only a matter of storage
- Need integrated and multilevel analytical skills and techniques

Data platform

Data lake

Couto et al.: *“A DL is a central repository system for storage, processing, and analysis of raw data, in which the data is kept in its original format and is processed to be queried only when needed. It can store a varied amount of formats in big data ecosystems, from unstructured, semi-structured, to structured data sources”*



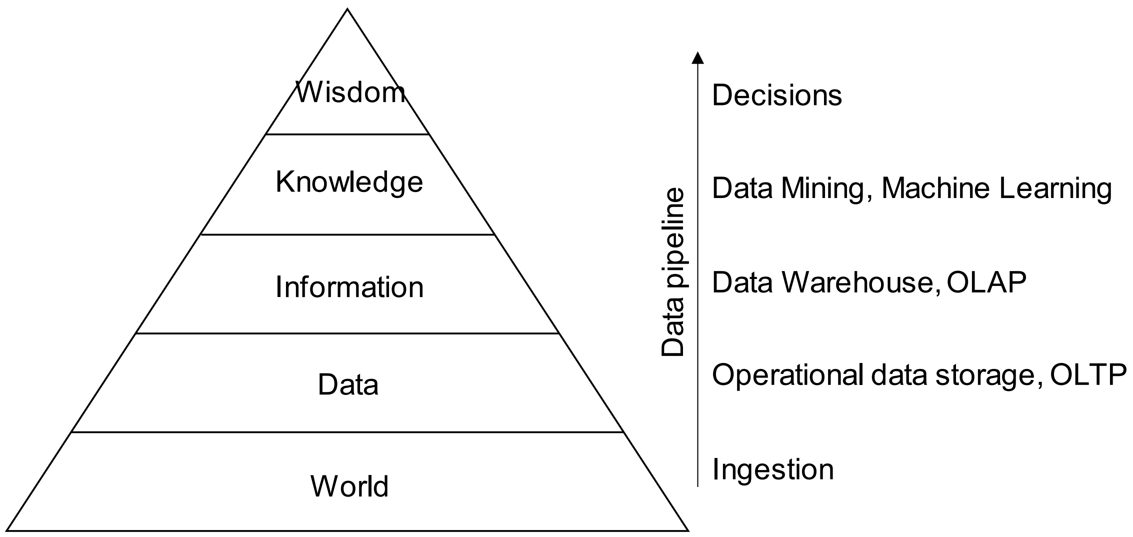
Couto, Julia, et al. "A Mapping Study about Data Lakes: An Improved Definition and Possible Architectures." *SEKE*. 2019.
https://dunnsolutions.com/business-analytics/big-data-analytics/data-lake-consulting

Data platform

Data platform

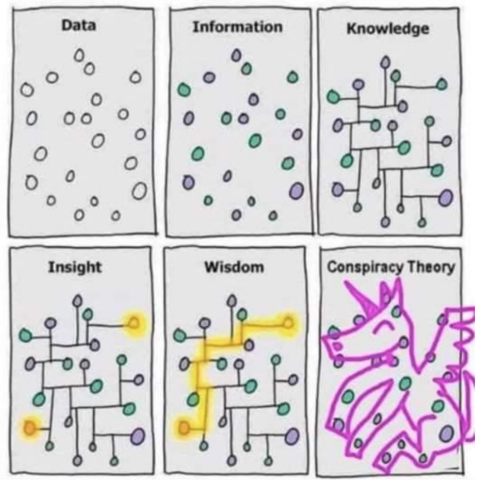
- An integrated set of technologies that collectively meets an organization’s end-to-end data needs such as acquisition, storage, preparation, delivery, and governance, as well as a security layer for users and applications
- Rationale: relieve users from complexity of administration and provision
  - Not only technological skills, but also privacy, access control, etc.
  - Users should only focus on functional aspects

Modeling data pipelines



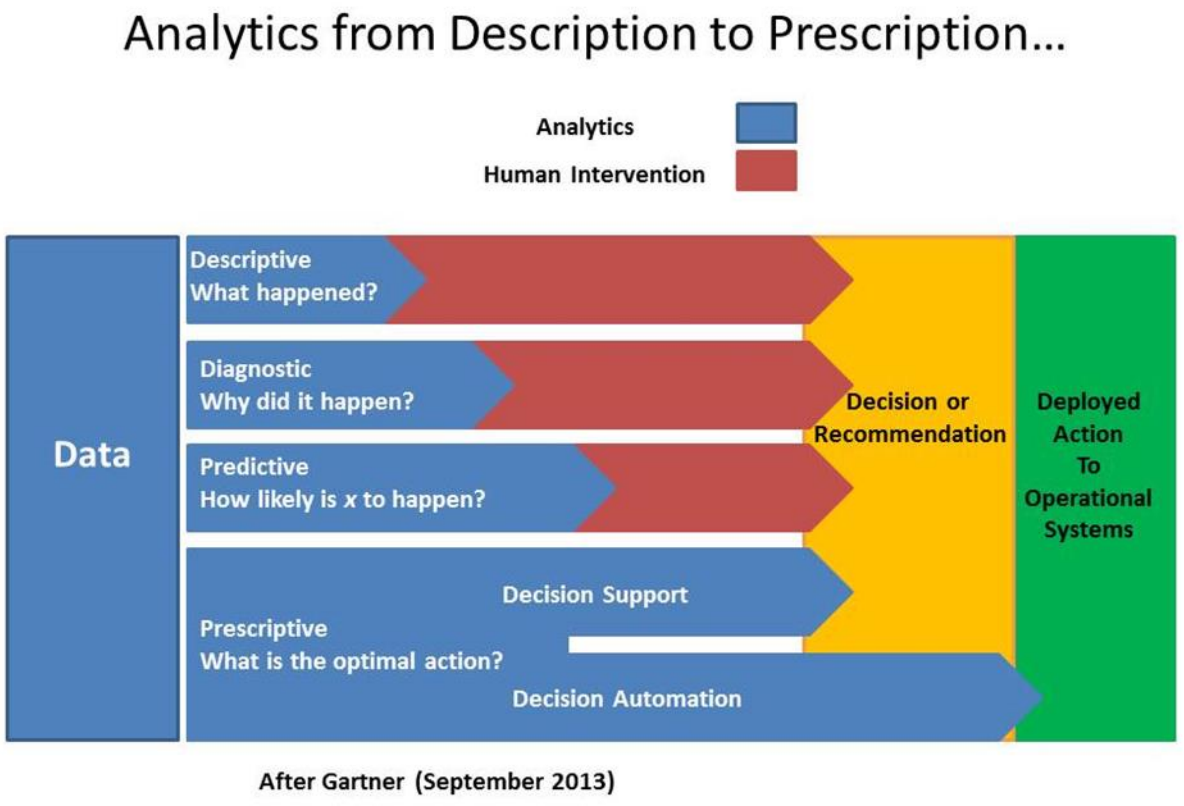
DIKW hierarchy

- Layers representing structural relationships between data, information, knowledge, and wisdom

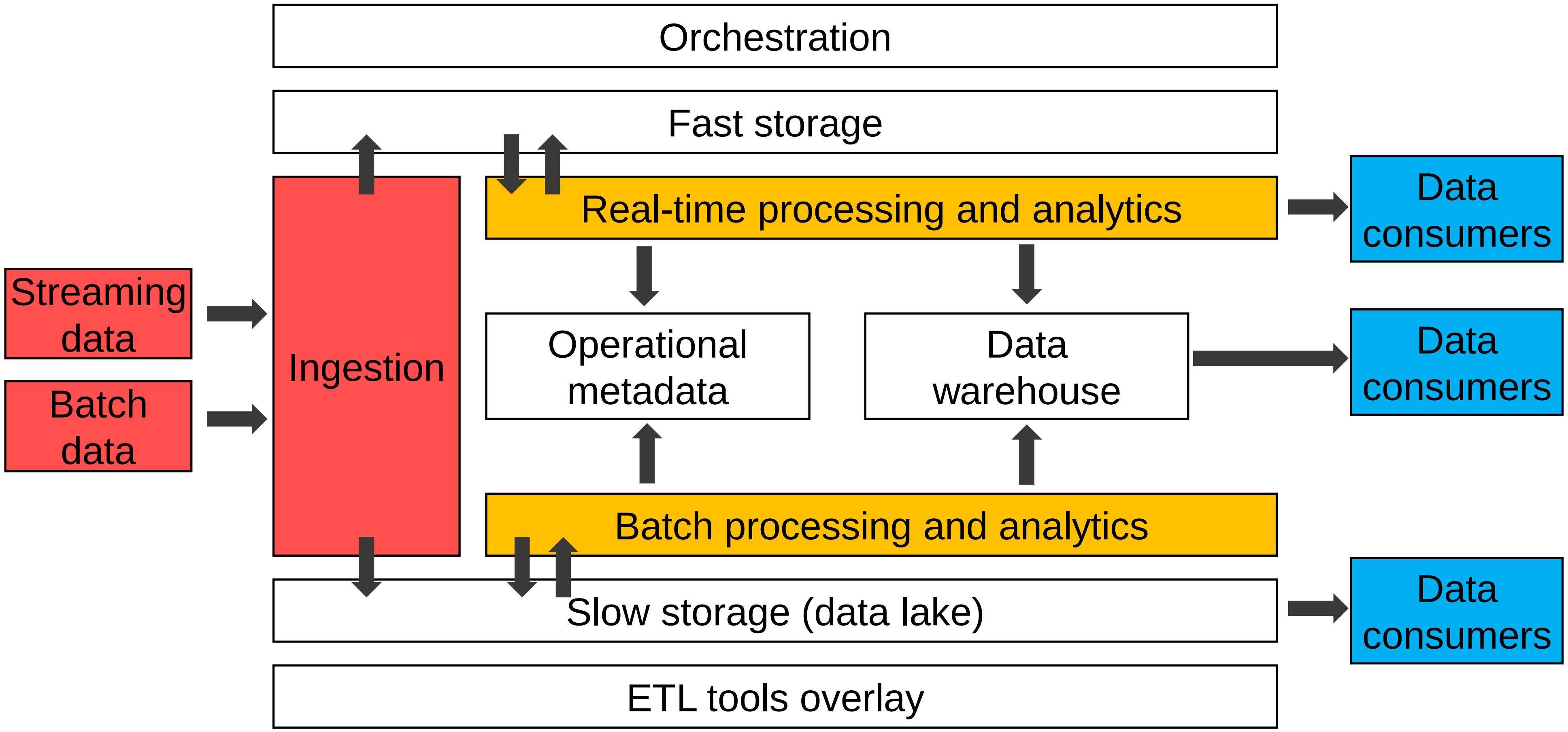


Ackoff, Russell L. "From data to wisdom." Journal of applied systems analysis 16.1 (1989): 3-9.

Modeling data pipelines



A tentative (service) organization



Ingestion + Processing + Serving

Data platform in the agritech domain

The synergy of internet of things (IoT) and precision farming is producing valuable applications in the AgriTech domain [1]

- AgriTech: use of technology for farming to improve efficiency and profitability

Some of our recent projects

- Smart watering management to produce Kiwifruit with higher quality while saving water [2]
- Sustainable weed management in with laser-based autonomous tools [4]
- Agro.Big.Data.Science [5]

[1] Vitali, Giuliano, et al. "Crop management with the iot: An interdisciplinary survey." Agronomy 11.1 (2021): 181.
[2] Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.
[3] Francia, Matteo, et al. "Making data platforms smarter with MOSES." Future Generation Computer Systems 125 (2021): 299-313.
[4] https://cordis.europa.eu/project/id/101000256
[5] http://agrobigdatascience.it

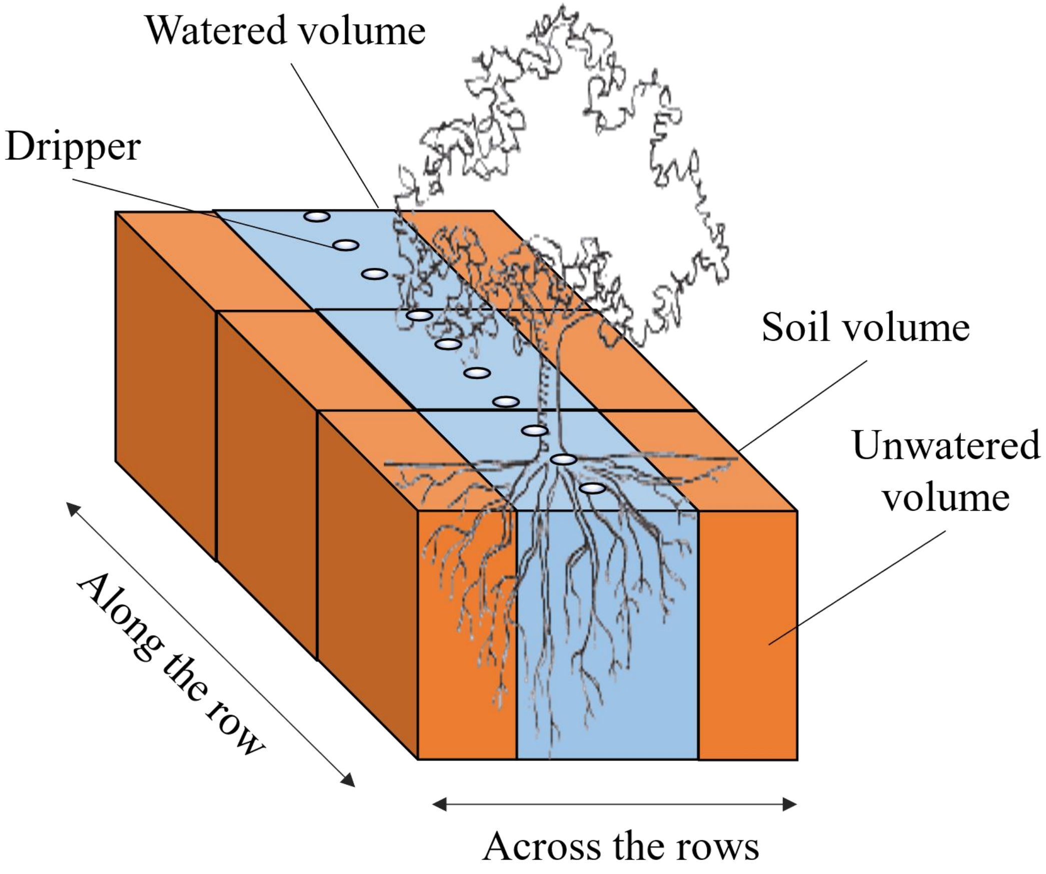
Soil moisture monitoring as a case study

Optimizing soil moisture is crucial for watering and crop performance [1]

- **GOAL**: saving water while improving fruit quality (i.e., provide a recommendation of the amount necessary water)
- Soils have different water retention
- Watering systems have different behaviors (e.g., drippers and sprinklers)
- Plants have different water demand (e.g., Kiwi [2] vs Grapes)
- Sensors produce different measurements with different precisions

[1] Turkeltaub et al., Real-time monitoring of nitrate transport in the deep vadose zone under a crop field–implications for groundwater protection, Hydrology and Earth System Sciences 20 (8) (2016) 3099–3108.
[2] M. Judd, et al., Water use by sheltered kiwifruit under advective conditions, New Zealand journal of agricultural research 29 (1) (1986) 83–92.

Reference scenario

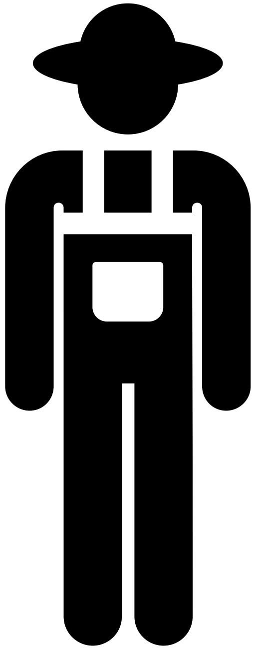


We consider an orchard where

- Kiwi plants are aligned along rows
- Each row has many drippers (e.g., 1 every meter)
- Drippers can water a limited soil volume

Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.

Soil moisture monitoring as a case study



(Example) Scenarios of digital transformation in agriculture

Scenario #1

- The farmer/technician controls the watering system based only on the experience
- No digital data/KPIs/automation

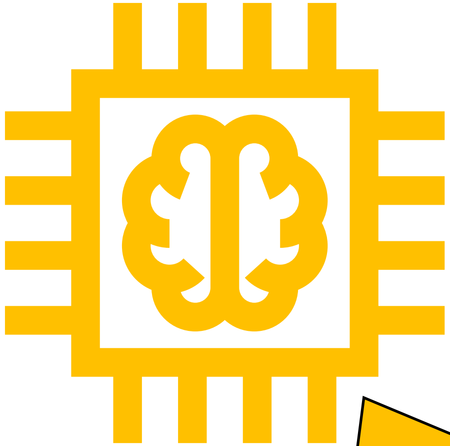
Scenario #2

- The control of the watering system is refined by observing sensor data
- Sensor data is digitalized, no KPIs/automatic

Scenario #3

- Sensor data feeds a decision support system that, knowing how to optimize KPIs, controls the watering system

Soil moisture monitoring as a case study



(Example) Scenarios of digital transformation in agriculture

Scenario #1

- The farmer/technician controls the watering system based only on the experience
- No digital data/KPIs/automation

Scenario #2

- The control of the watering system is refined by observing sensor data
- Sensor data is digitalized, no KPIs/automatic

Artificial intelligence (AI) is intelligence demonstrated by machines. AI research has been defined as the field of study of intelligent agents, which refers to any system that perceives its environment and takes actions that maximize its chance of achieving its goals.

a decision support system that, knowing how to
rols the watering system

Soil moisture monitoring as a case study

To achieve our goal, it necessary to understand of the soil behaves

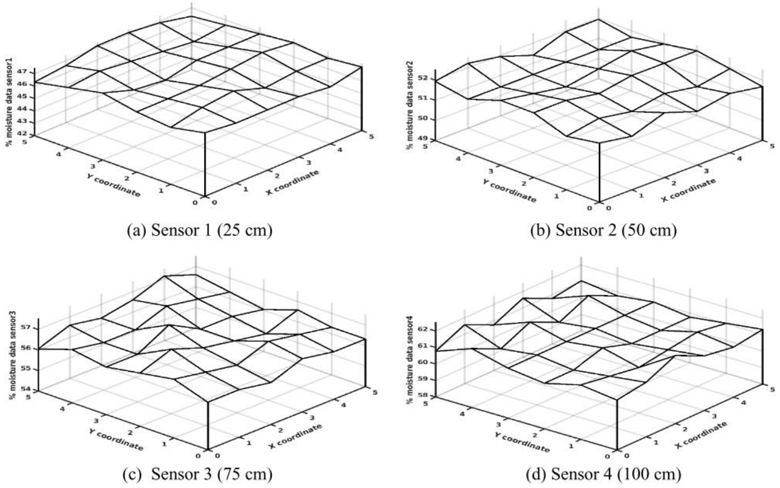
Simulate [1, 2] the soil behavior according to physical models [3]

- However a fine tuning is required, we need to know/parametrize everything
  - Soil (e.g., retention curve, hysteresis [3])
  - Plant (e.g., roots, LAI)
  - Weather conditions (temperature, humidity, wind, precipitations)
  - Watering system (e.g., capacity, distance between drippers)
- Tuning can take months (of human interactions)!
  - Need to collect samples from the field… if some parameter is incorrect we need to trace back
  - Need to implement/code all these features into the simulator [1, 2]
  - Hyper-parameter tuning with machine learning can help, but it is not a silver bullet

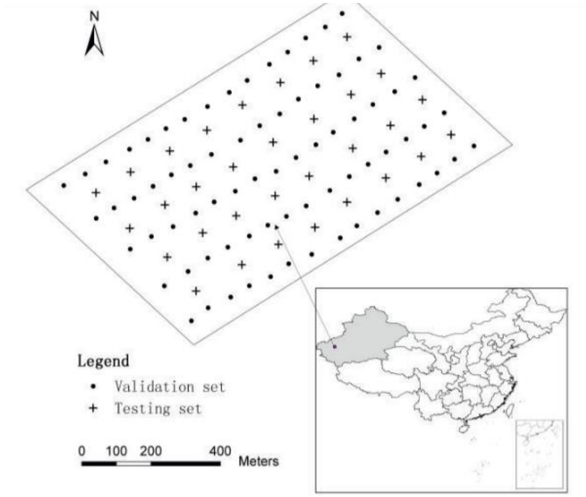
[1] Šimunek, J., et al. "HYDRUS: Model use, calibration, and validation." Transactions of the ASABE 55.4 (2012): 1263-1274.
[2] Bittelli, Marco, et al. Soil physics with Python: transport in the soil-plant-atmosphere system. OUP Oxford, 2015.
[3] Van Genuchten, M. Th. "A closed-form equation for predicting the hydraulic conductivity of unsaturated soils." Soil science society of America journal 44.5 (1980): 892-898.
[4] Pham, Hung Q., Delwyn G. Fredlund, and S. Lee Barbour. "A study of hysteresis models for soil-water characteristic curves." Canadian Geotechnical Journal 42.6 (2005): 1548-1568.

Soil moisture monitoring as a case study

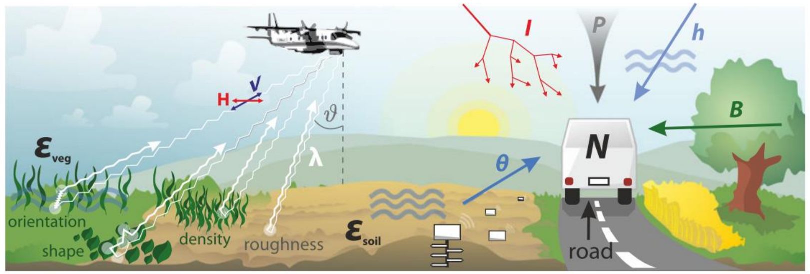
But… we have sensors!



[1]



[2]

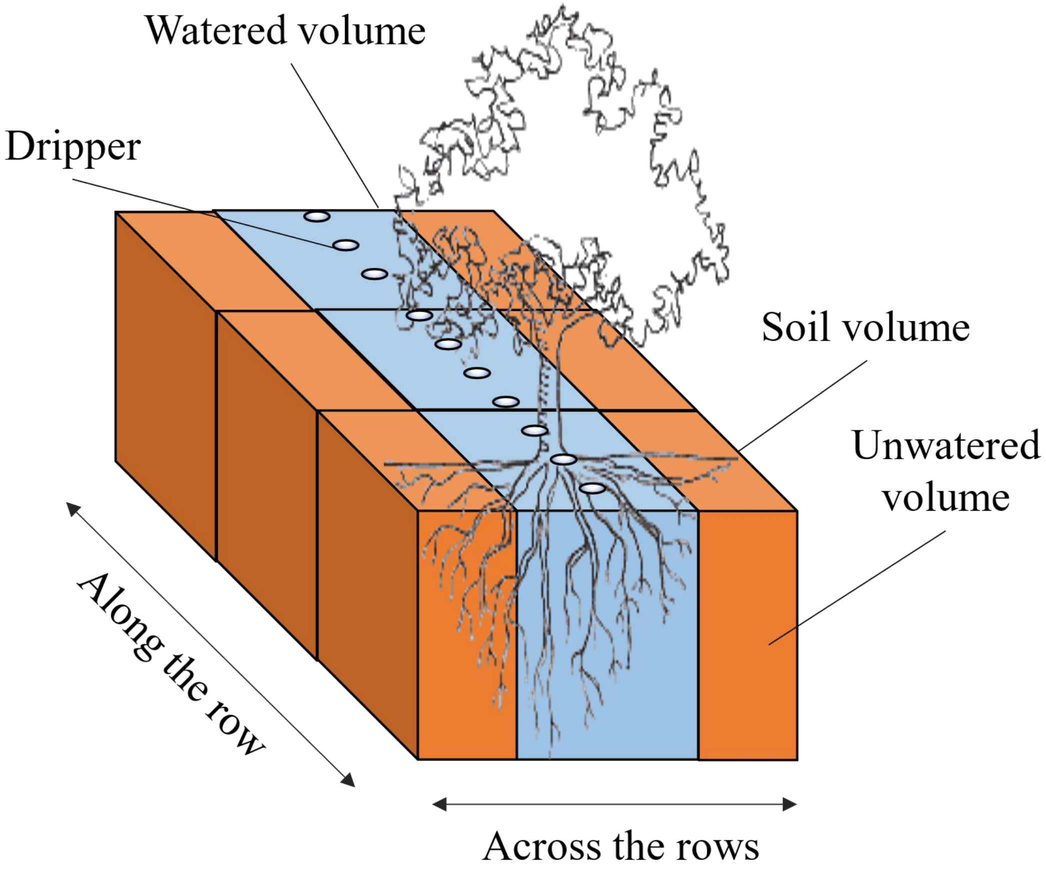


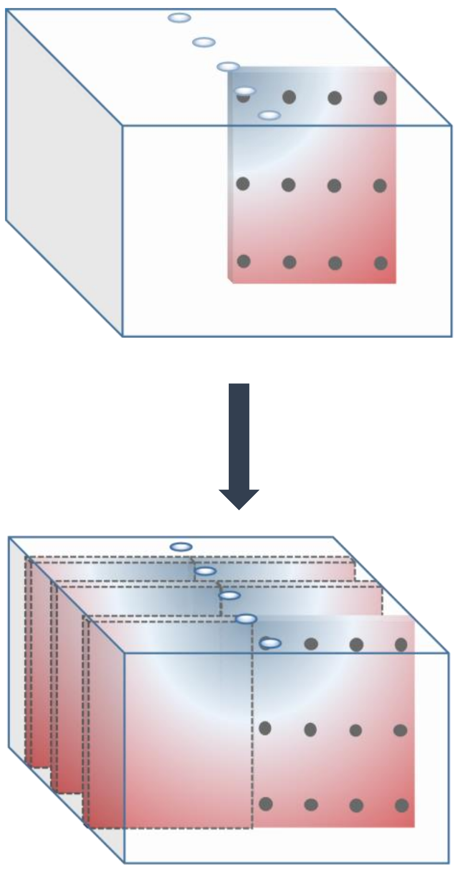
[3]

These settings are too coarse to monitor soil moisture with precision, and they require many sensors

[1] Koyuncu, Hakan, et al. "Construction of 3D soil moisture maps in agricultural fields by using wireless sensor communication." Gazi University Journal of Science 34.1 (2021): 84-98.
[2] Zheng, Zhong, et al. "Spatial estimation of soil moisture and salinity with neural kriging." International Conference on Computer and Computing Technologies in Agriculture. Springer, Boston, MA, 2008.
[3] Fersch, Benjamin, et al. "Synergies for soil moisture retrieval across scales from airborne polarimetric SAR, cosmic ray neutron roving, and an in situ sensor network." Water Resources Research 54.11 (2018): 9364-9383.

Sensor layouts and symmetry





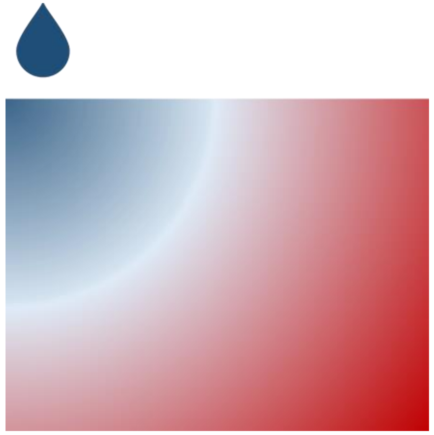
If watered volume is symmetric along the row, a 2D sensor grid is sufficient to represent the soil volume

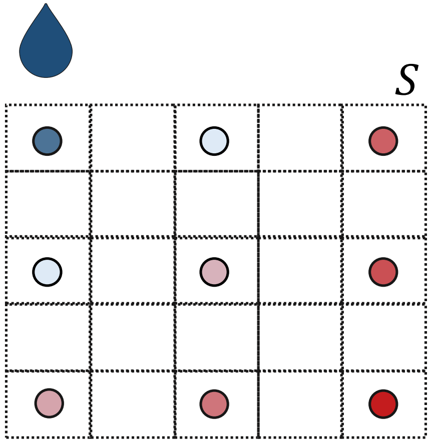
If moisture variations take place along the row too, a 3D grid of sensors is required

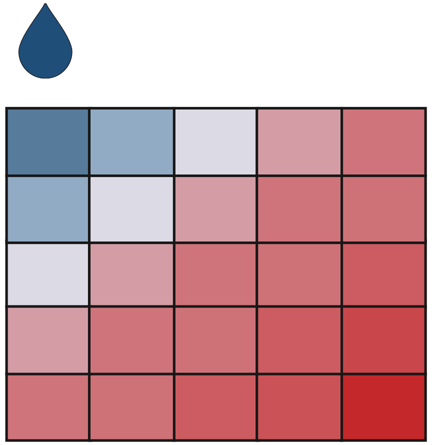
- E.g., too sparse drippers

Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.

From sensors to soil profiles







(a) Soil moisture is a continuum in the soil

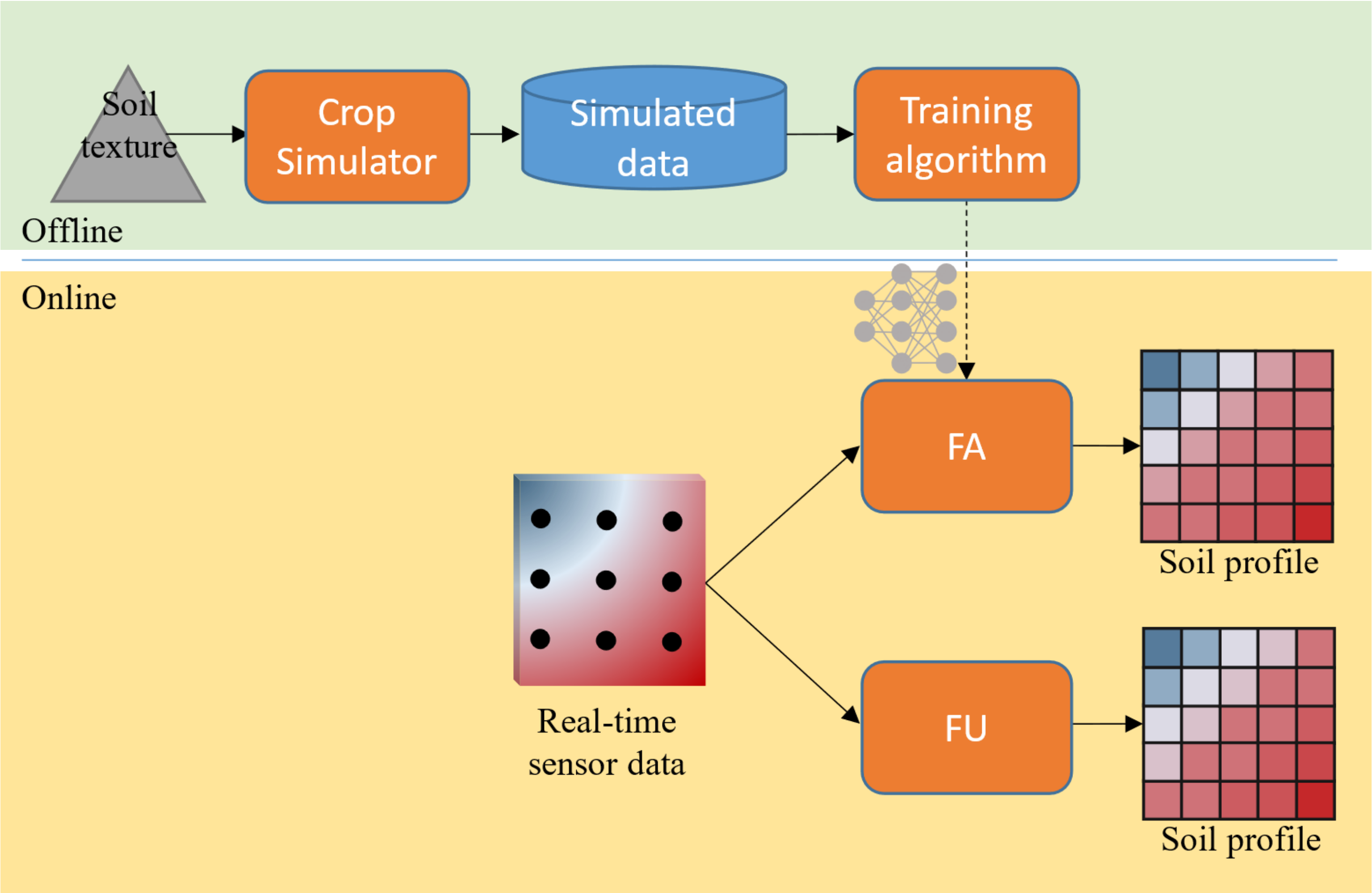
(b) Sensors return a discretized representation of soil moisture

- Depending on the number of sensors and on their layout the monitoring accuracy changes

(c) **Goal:** produce fine-grained soil profiles out of coarse-grained layouts

Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.

Overview



Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.

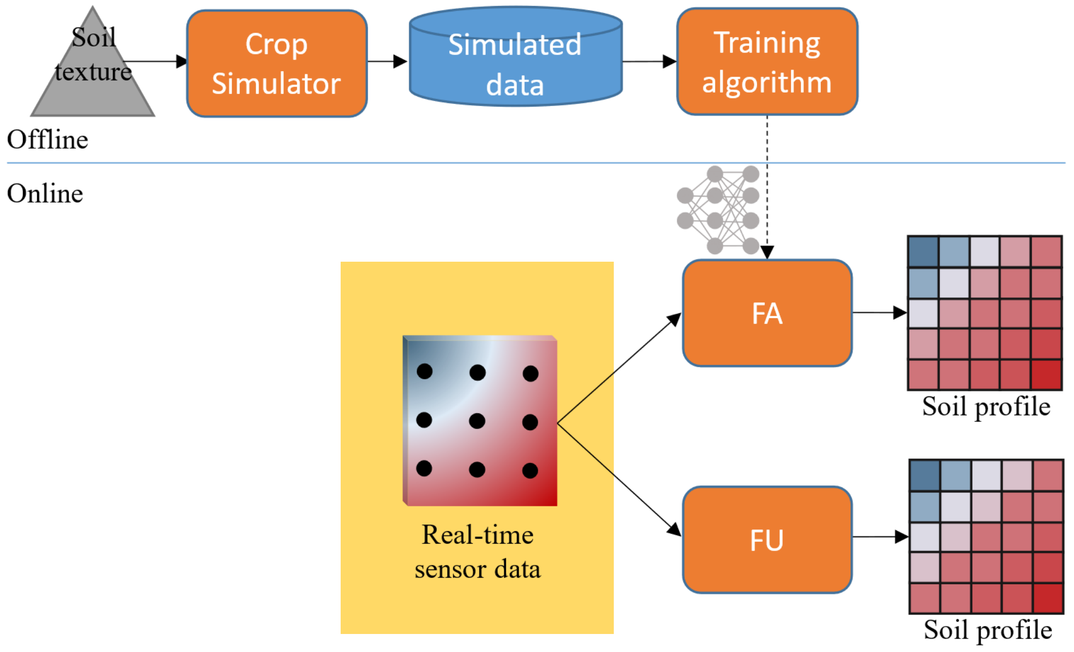
Ingestion

Setup

- We install a 2D/3D grid of sensors

For instance, in the 2D setting

- 4 columns of sensors are located across the row (e.g., 0/30/60/90cm)
  - The column (0, *) is under the dripper
- Each columns has 3 sensors located at 3 depths (e.g., 20/40/60cm)



Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.

Ingestion

In the 2D setting (3 x 4 gypsum block sensors)

- Sample soil moisture-sensor data every 15 minutes
- Collect dripper and weather data (humidity, temperature, solar radiation, wind) every hour

How many data does each monitored field produces every season?

- $\left(12 \cdot 4 \frac{samples}{hour} + 5 \frac{samples}{hour}\right) \cdot 24 \frac{hour}{day} \cdot 30 \frac{day}{month} \cdot 5 \frac{month}{year} \cong 200 \cdot 10^3 \frac{samples}{year}$

We monitored 6 fields for 2 years

- $200 \cdot 10^3 \frac{samples}{year} \cdot 2\ years \cdot 6 = 2.4 \cdot 10^6\ samples$

We should consider accessory data for storage and optimization structures

- In two years, we collected/generated 16GB data (as of 2022-08-30)

Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.

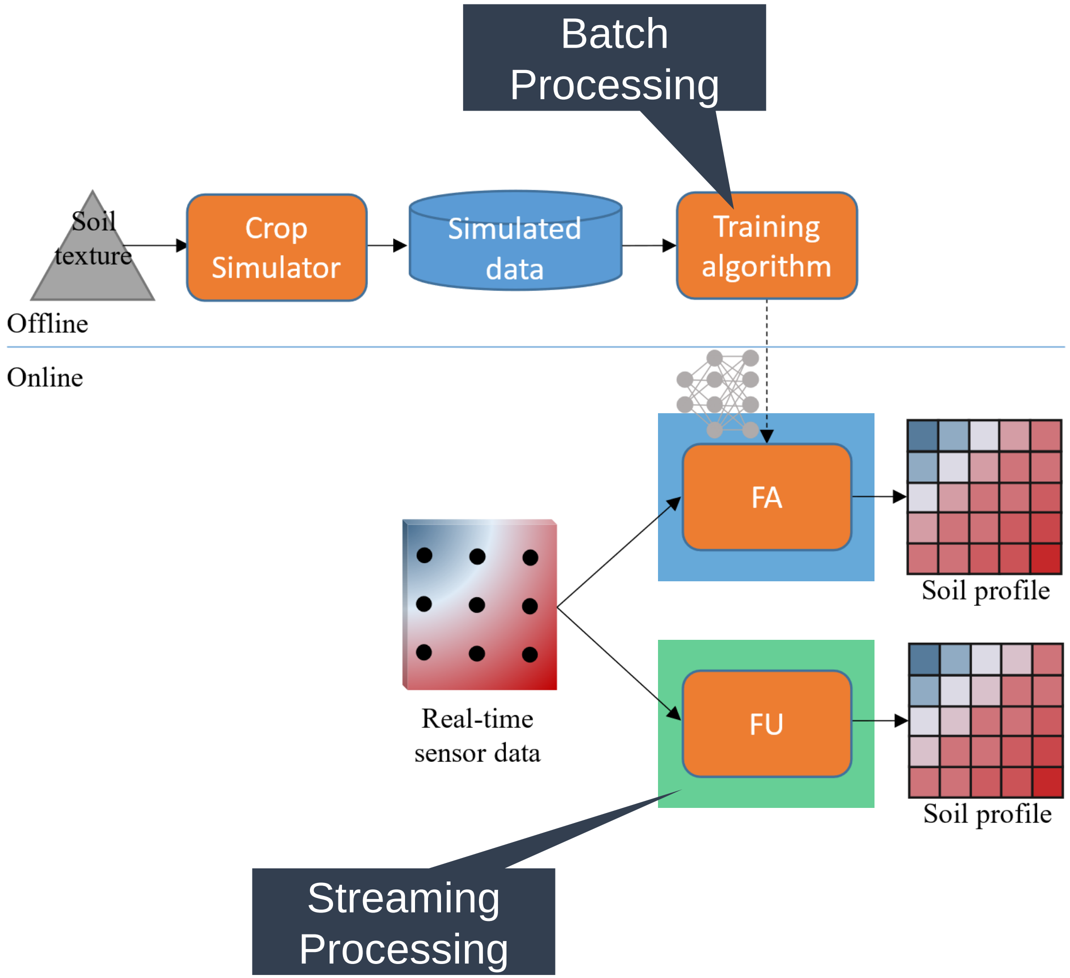
Processing

Feature unaware (FU)

- Plug-and-play
- Create a linear interpolation of the real-time sensor data

Feature aware (FA)

- Require time for data collection and training/testing
- Create an interpolation of the real-time sensor data through machine learning

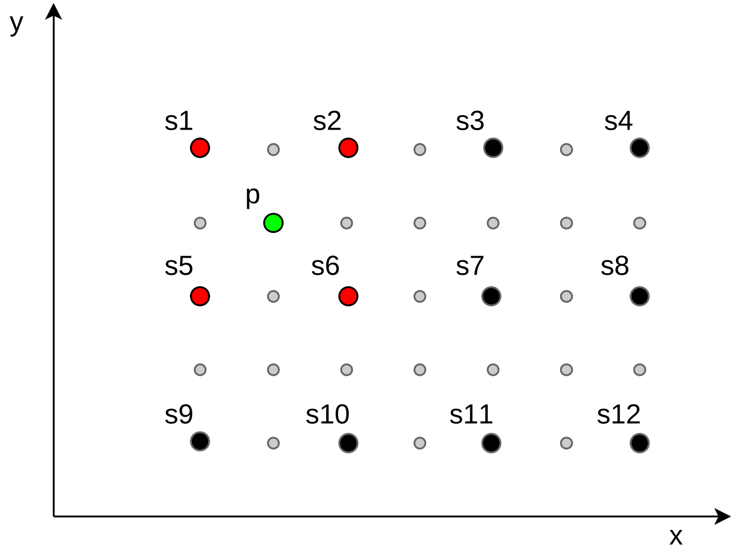


Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.

Processing

Feature unaware

- Create a linear interpolation of the real-time sensor data

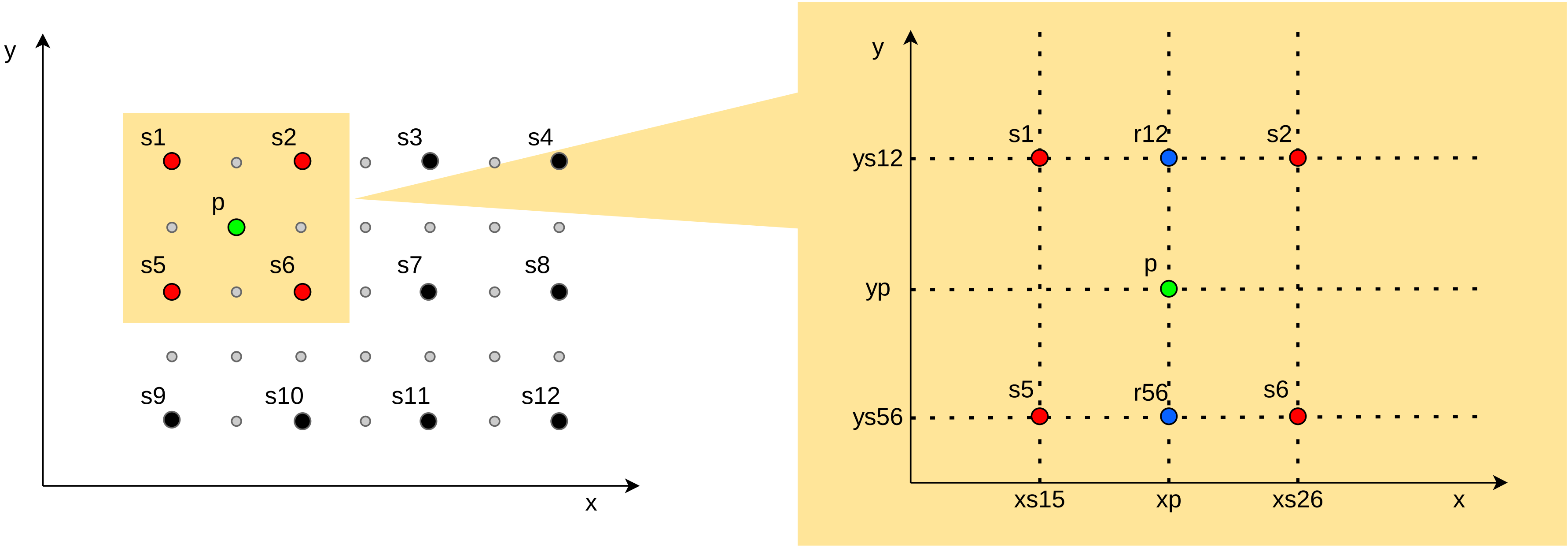


Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.

Processing

Feature unaware

- Create a linear interpolation of the real-time sensor data



Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.

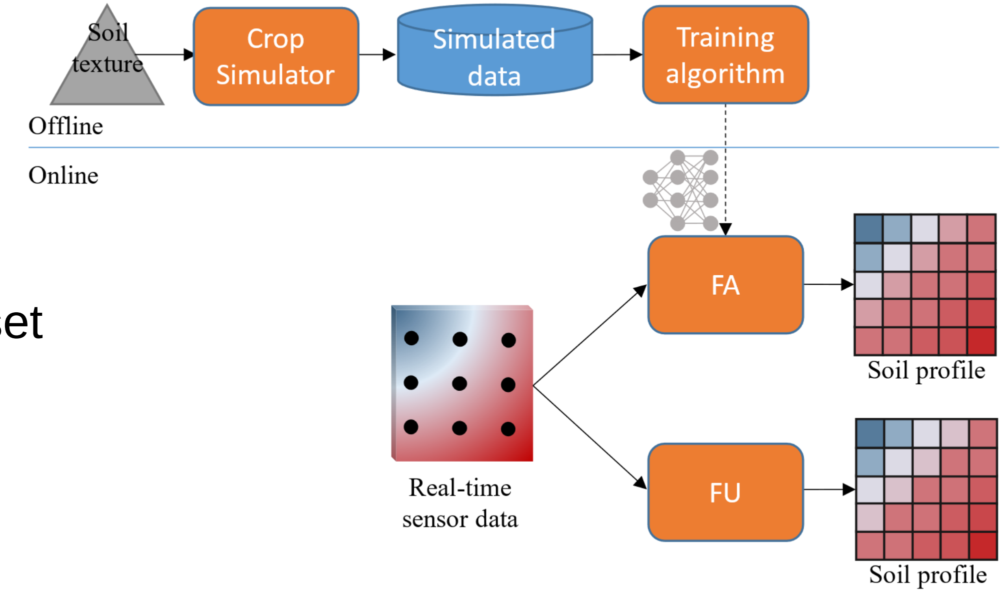
Processing

Feature aware

- Create an interpolation of the real-time sensor data through machine learning

Offline pipeline

- Given the soil texture as input
- Simulate different patterns of SM to produce a dataset of simulated SM
- Train a machine learning model on such data
- Deploy the model to estimate the soil profile

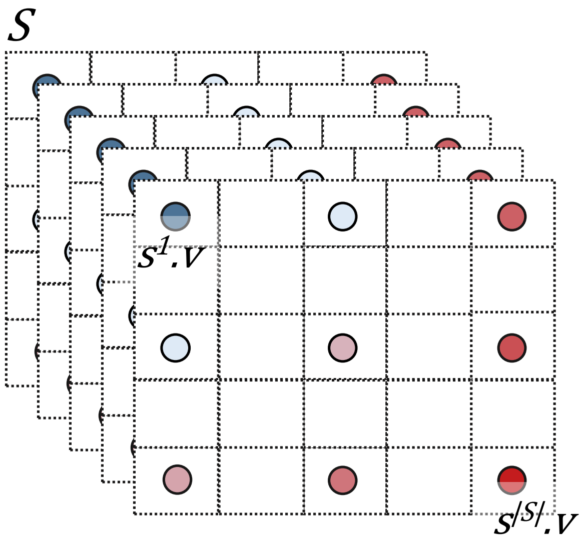


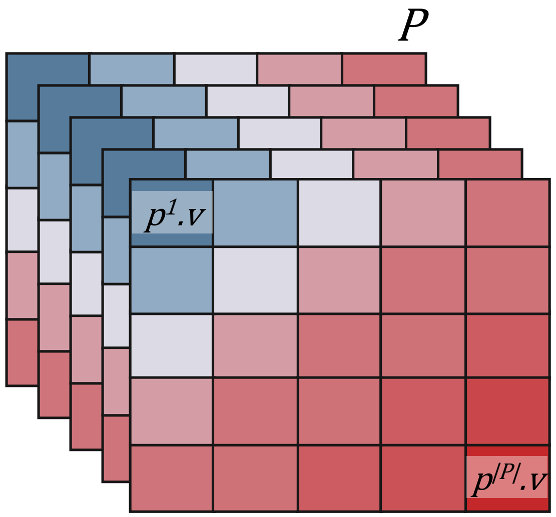
Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.

Processing

Input

Output





How do you classify this task?
Which issues do you foresee?

Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.

Processing

Data generation and augmentation

- CRITERIA 3D to simulate the hydrological fluxes in the soil following Richard's equations
- Inputs
  - The soil texture
  - "Default" settings for the kiwi-plant (e.g., shape of the tree roots / LAI)
  - Watering system based on a single dripper
  - Weather conditions from ARPAE
  - Different watering patterns (by changing watering intervals and the amount of supplied water)
- Output
  - Training set = $\left(12\frac{samples}{hour}\right) \cdot 24\frac{hour}{day} \cdot 30\frac{day}{month} \cdot 4\ \text{months} \cong 35 \cdot 10^3$ training samples
  - Validation set = same as training set, but we simulate with different weather/irrigation patterns
  - Test set = 4 month from the real field
- Different weather conditions & watering patterns to enable generalization and avoid overfitting

ARPAE: Agenzia regionale per la prevenzione, l'ambiente e l'energia dell'Emilia-Romagna

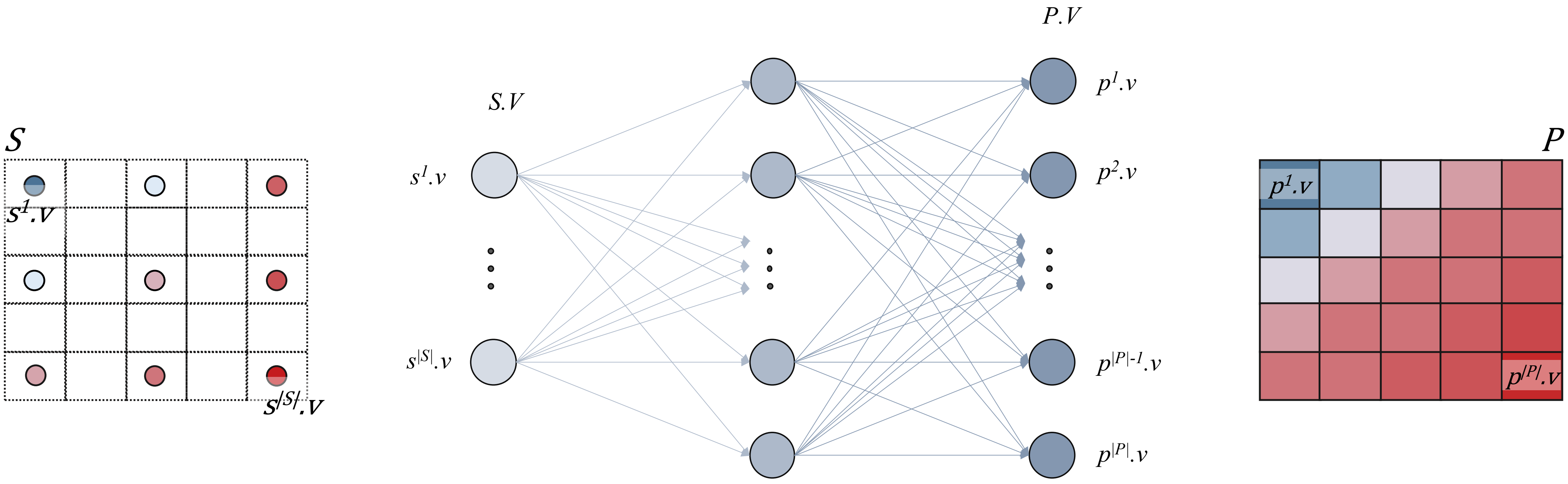
Processing

This is a (multi-output) regression problem

- The task is to learn the function mapping the input to the continuous output
- We tried several machine learning models
  - SVR, Random Forest Regression, Linear Regression, and ANN
  - (A simple) ANN is the best performing model
  - The hyper parameters (structure/learning rates) are set through a hyper-parameter tuning process
    - HyperOpt: state-of-the-art optimization technique to explore the huge search space of hyper-parameters

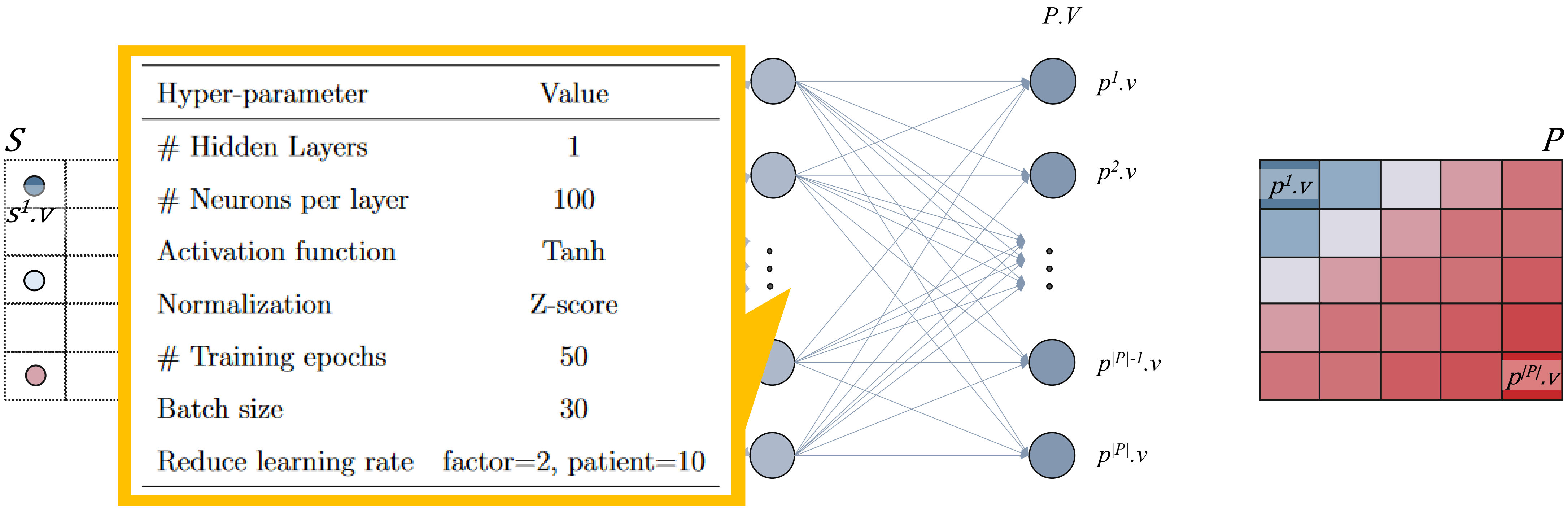
https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html
https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html
https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
https://keras.io

Processing



Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.
https://keras.io/api/callbacks/reduce_lr_on_plateau

Processing

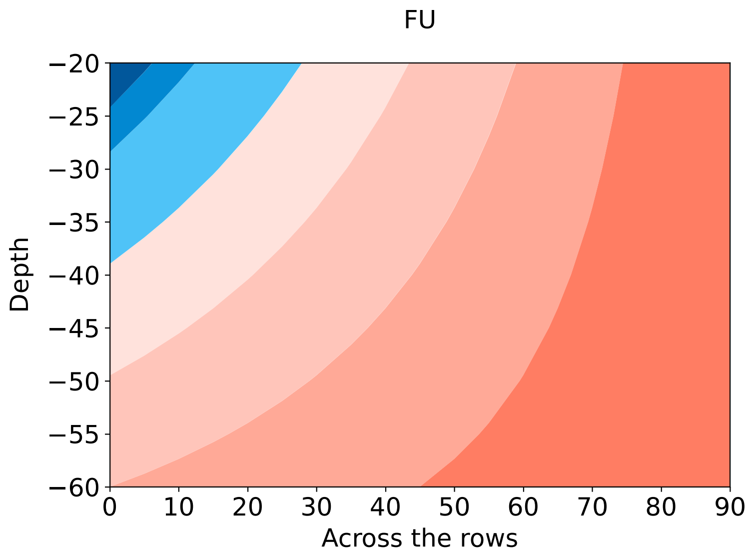


Hyper-parameter	Value
# Hidden Layers	1
# Neurons per layer	100
Activation function	Tanh
Normalization	Z-score
# Training epochs	50
Batch size	30
Reduce learning rate	factor=2, patient=10

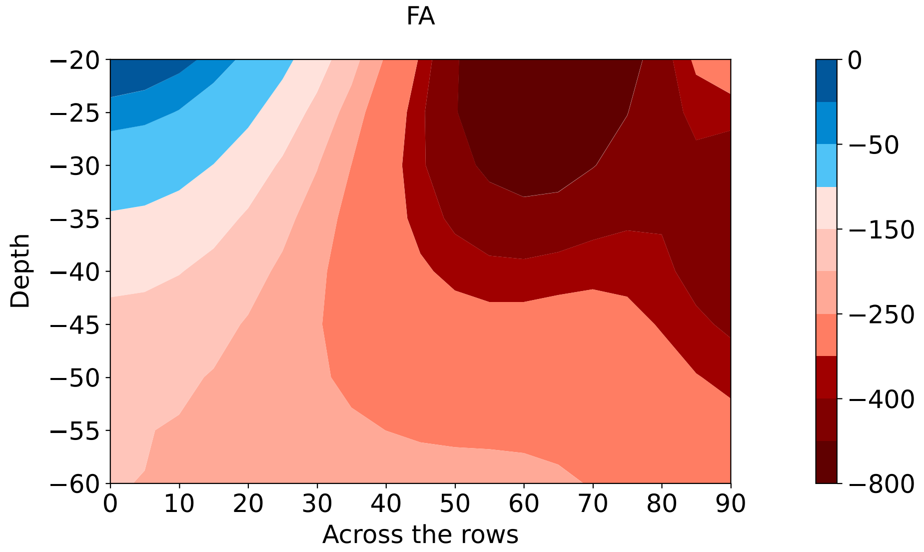
Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.
https://keras.io/api/callbacks/reduce_lr_on_plateau

Processing

Feature unaware

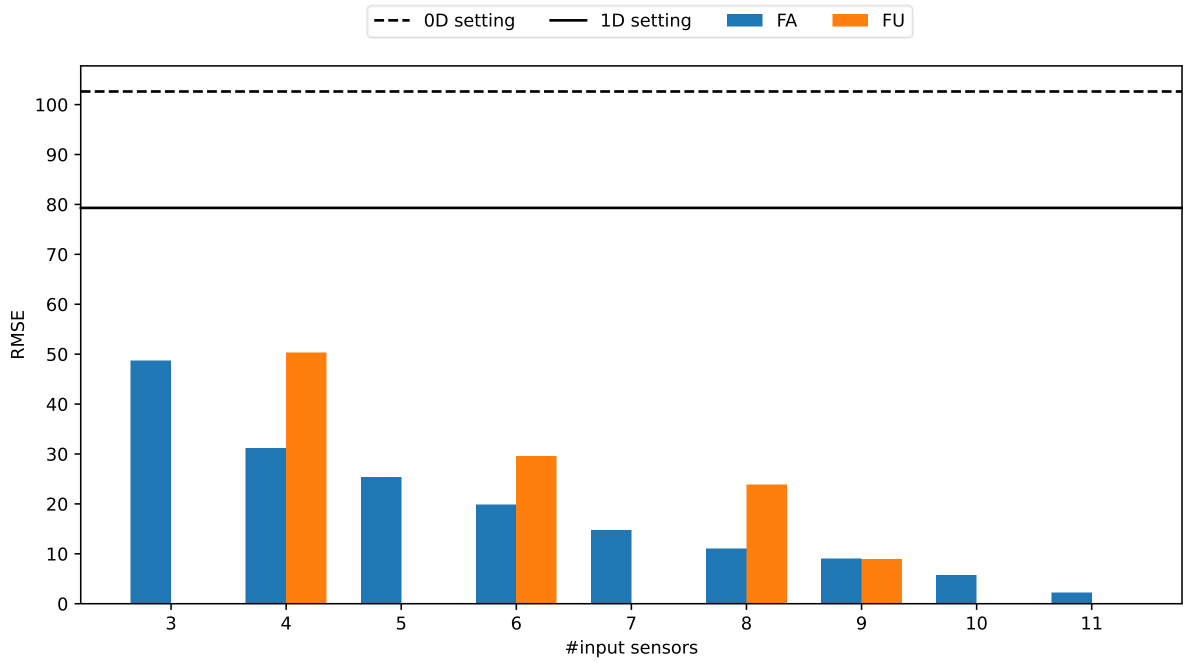


Feature aware



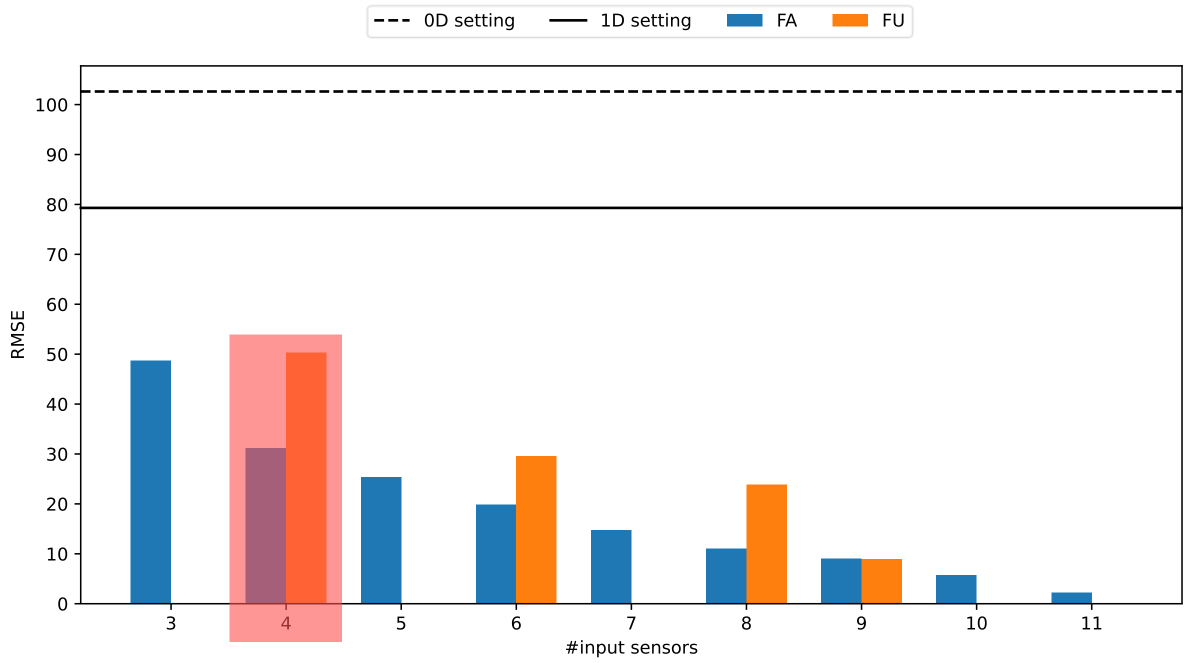
Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.

Processing: empirical evaluation



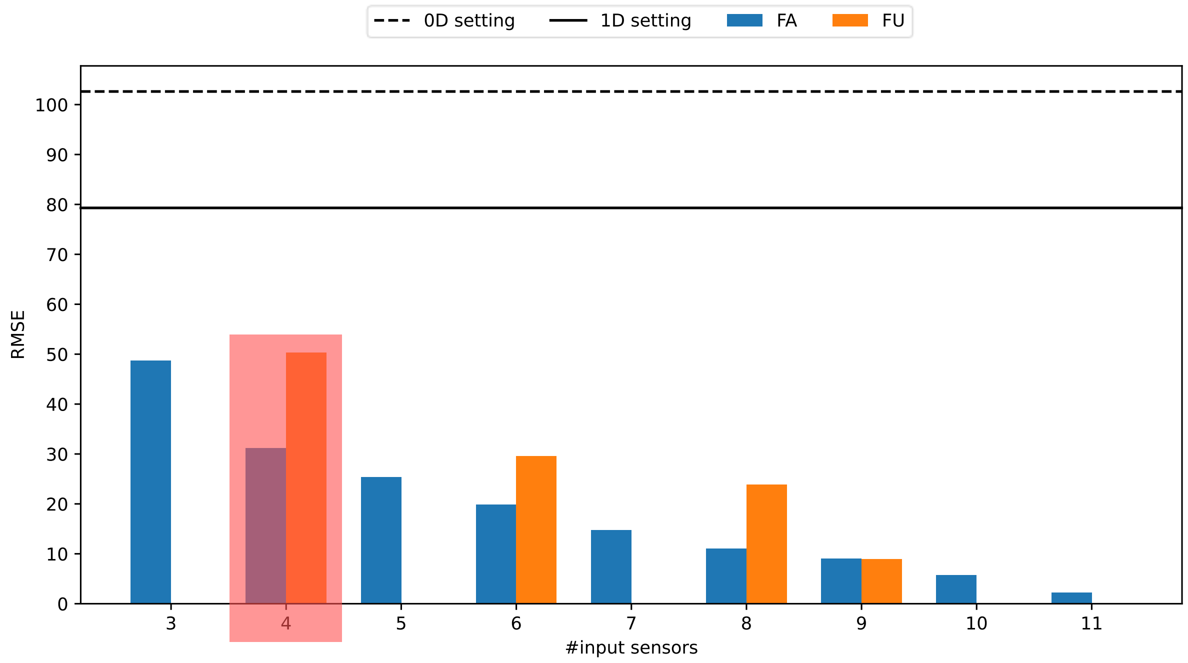
Processing: empirical evaluation

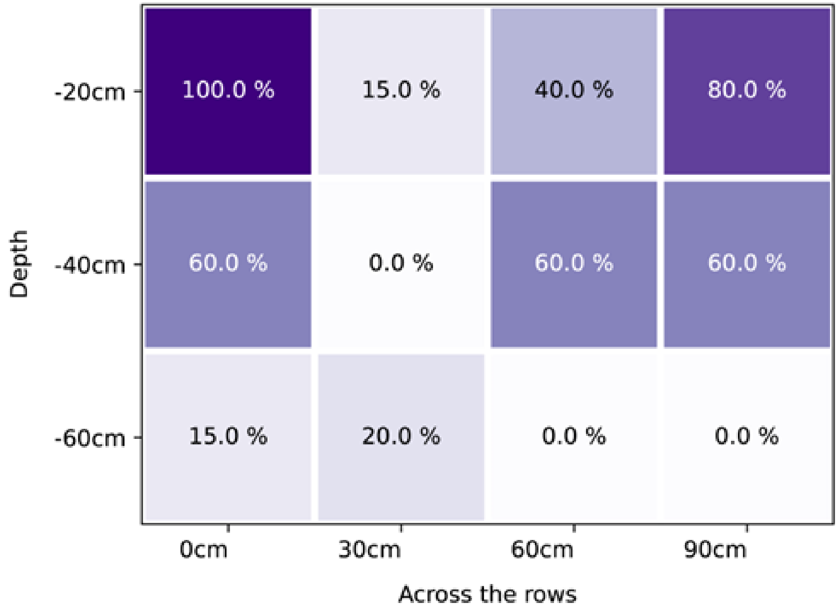
If I got 4 sensors, what layout should I choose?



Processing: empirical evaluation

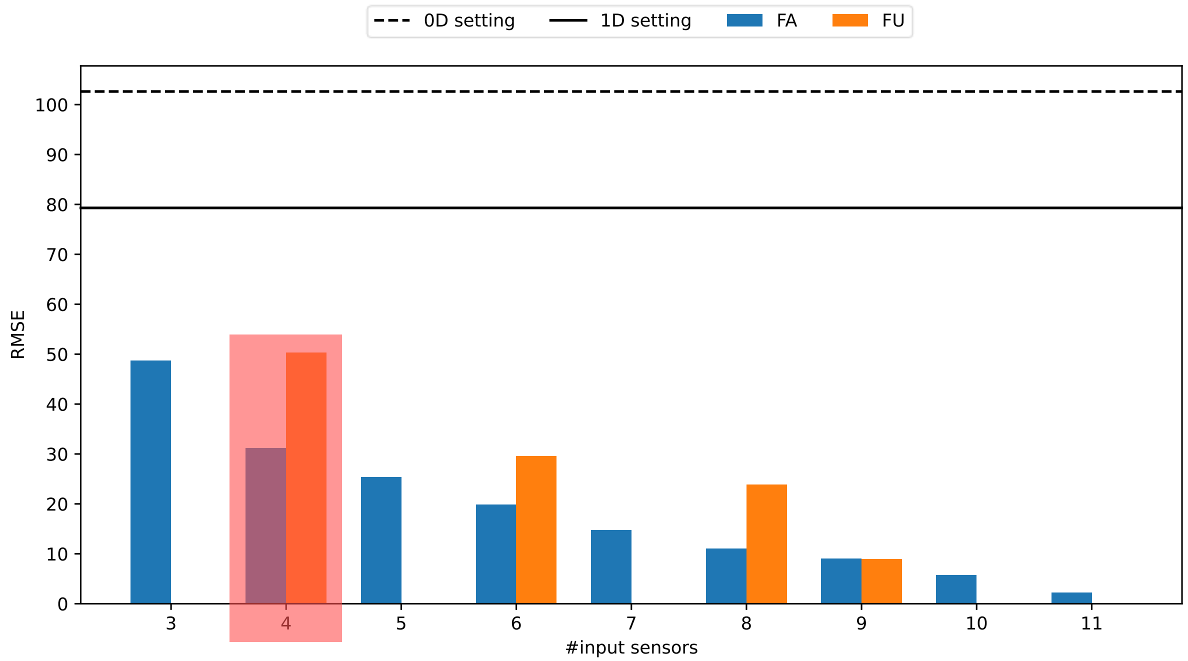
If I got 4 sensors, what layout should I choose?

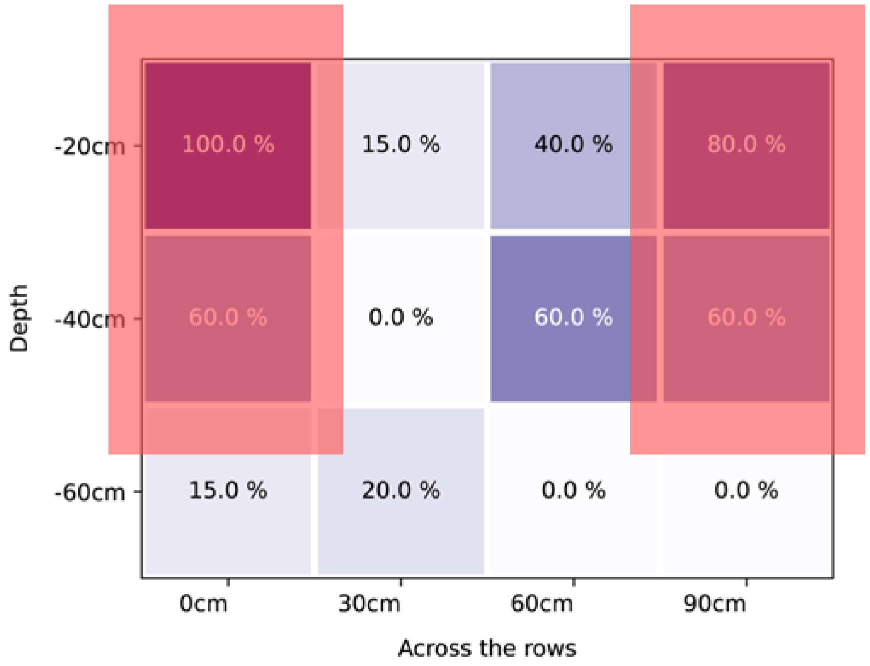




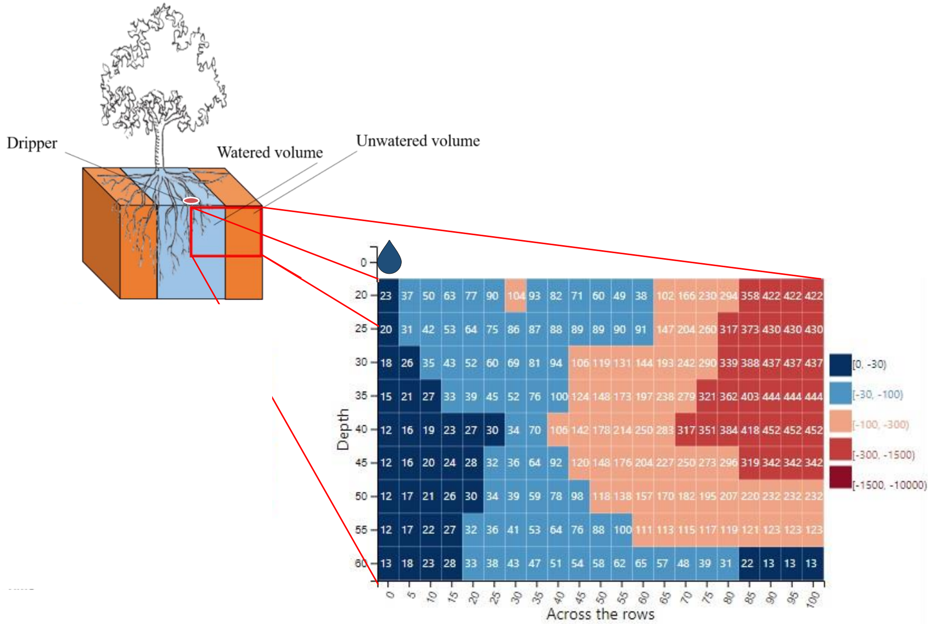
Processing: empirical evaluation

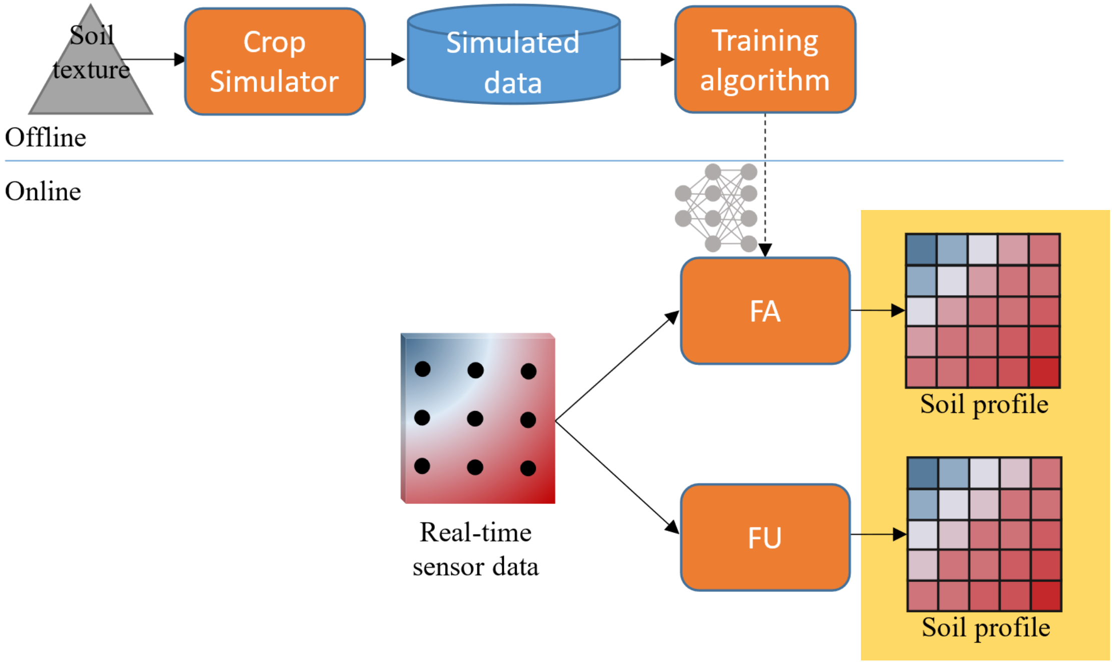
If I got 4 sensors, what layout should I choose?





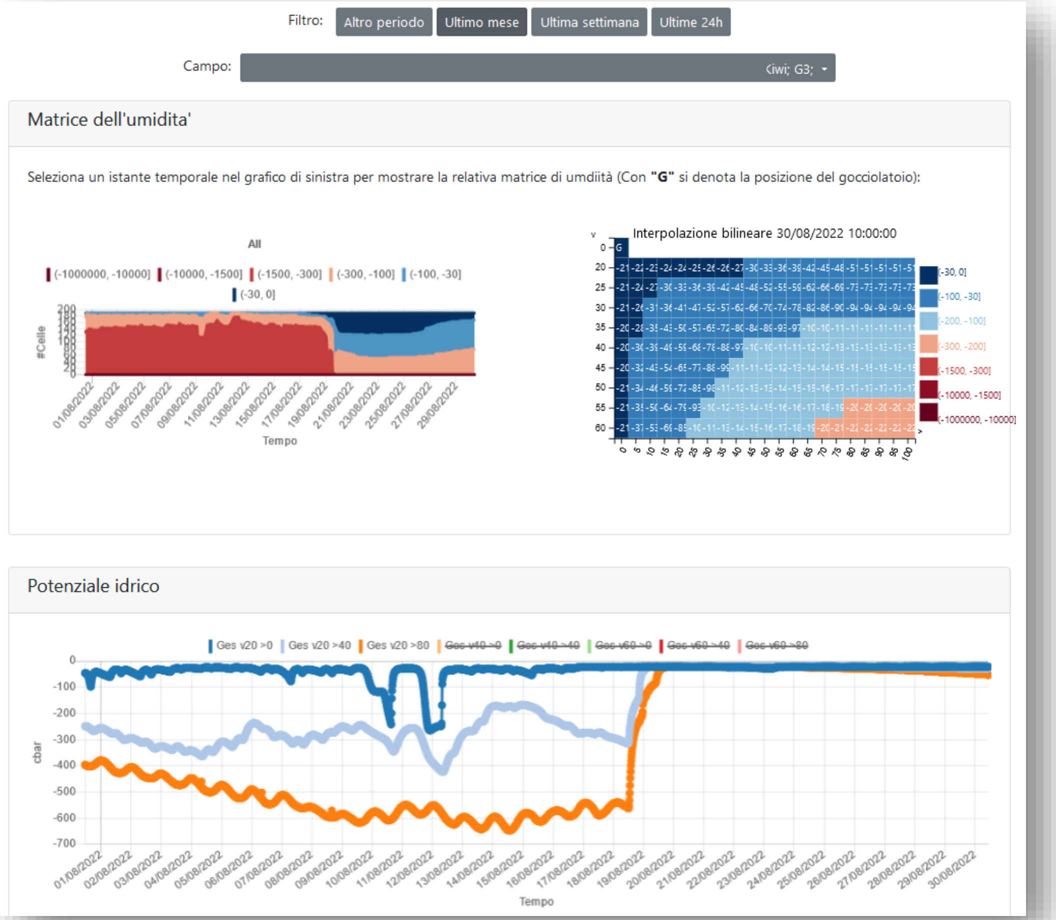
Serving





Francia, Matteo, et al. "Multi-sensor profiling for precision soil-moisture monitoring." Computers and Electronics in Agriculture 197 (2022): 106924.

Serving


Filtro:
Altro periodo
Ultimo mese
Ultima settimana
Ultime 24h
Campo:
Matrice dell'umidita'
Seleziona un istante temporale nel grafico di sinistra per mostrare la relativa matrice di umdiità (Con "G" si denota la posizione del gocciolatoio):
All
Interpolazione bilineare 30/08/2022 10:00:00
#Celle
Tempo
Potenziale idrico
cbar
Tempo

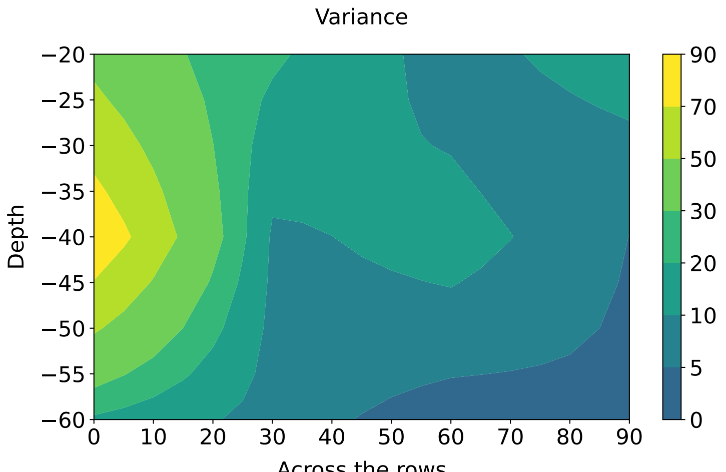
Descriptive Analytics

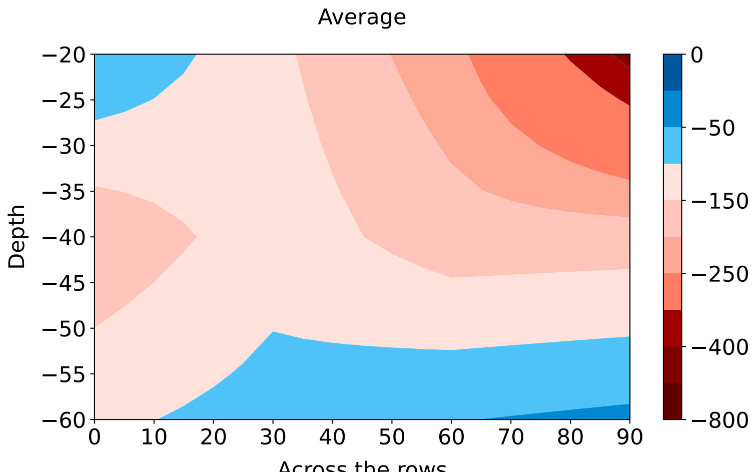
Starting from the profile, we derive meaningful visualizations/analysis

- SM variance (left; lighter areas are those where SM varies the most) and average (right)

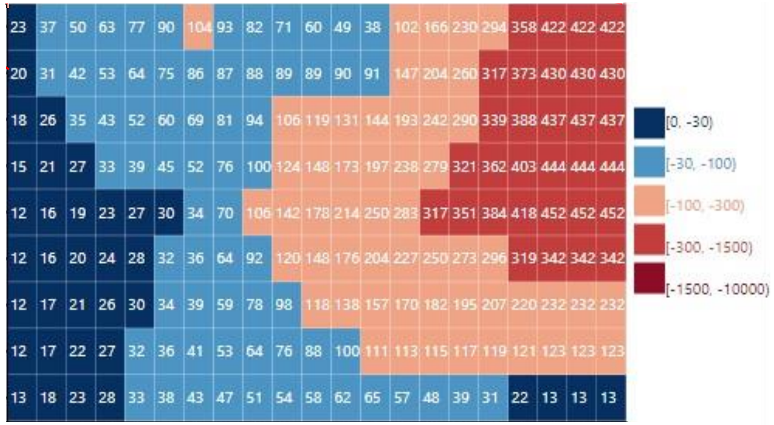
The charts support both agricultural technicians and farmers

- What is the watered volume?
  - This region is typically characterized by watering and high suction by the roots
  - Over-watering: high values in the average chart and low values in the variance chart
- Where is the root suction higher?
  - A high root suction quickly reduces the moisture in the soil and results in high soil moisture variance
- How soil moisture dynamics impact on the watered volume?
  - If, after increasing the water supplied, the profile does not change then the soil disperses water





Prescriptive Analytics



Watering recommendation

IF $\left(\frac{\#\text{BlueCells} + \#\text{LightBlueCells}}{\#\text{Cells}} < 0.5 \;\&\&\; \frac{\#\text{BlueCells}}{\#\text{Cells}} < 0.25\right.$ in the last 12h) && precipitations < 7mm in the last 12h

THEN

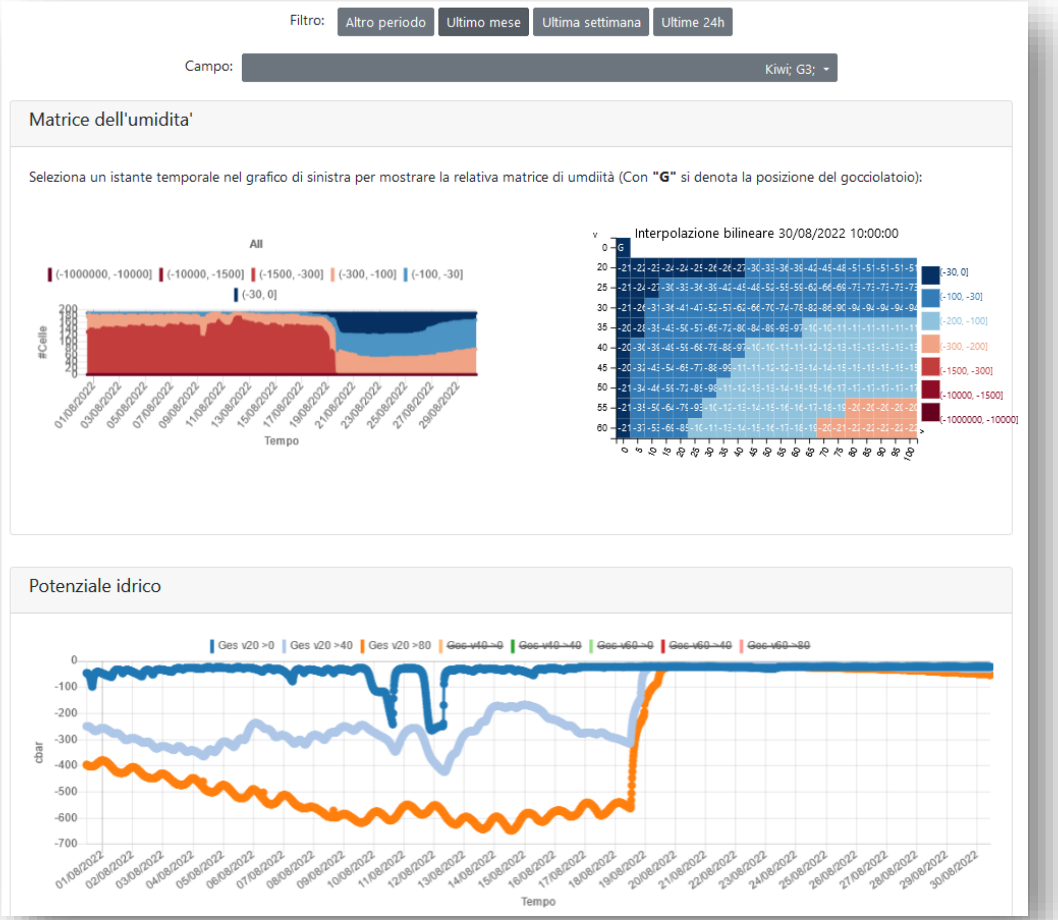
    Recommended water = Evapotranspiration (ET) of the day before

ELSE

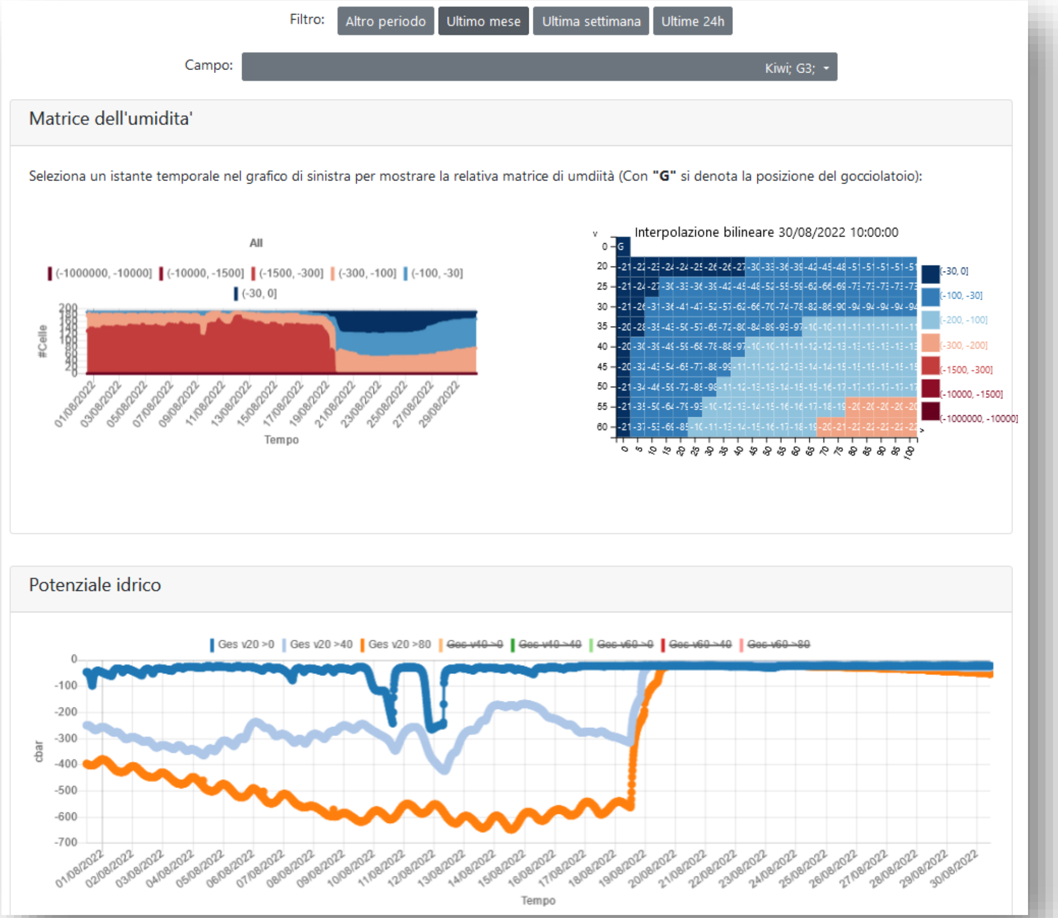
    Do nothing

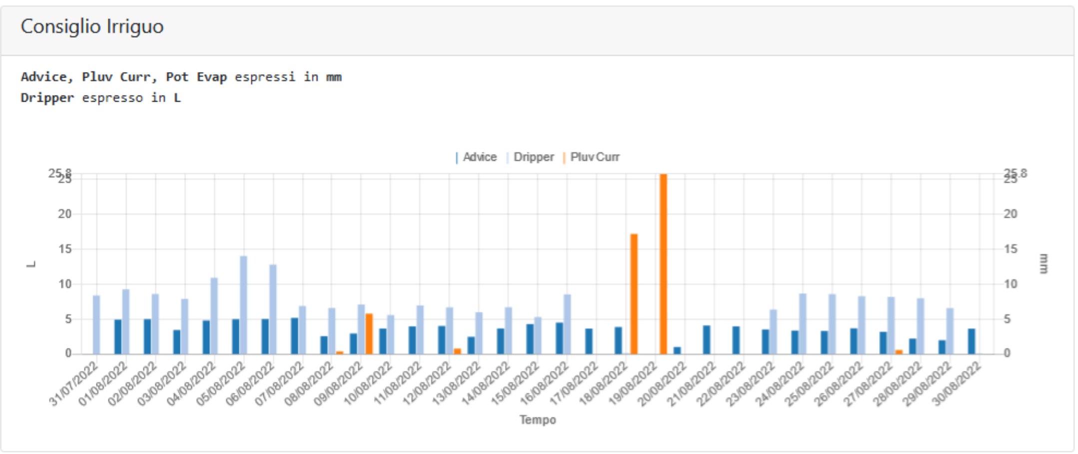
M. Quartieri, M. Toselli, E. Baldi, G. Polidori, M. Germani, M. Noferini, E. Xylogiannis, Effect of the method and volume of irrigation on yield and fruit quality of yellow fleshed kiwifruit in northern Italy, in: X International Symposium on Kiwifruit 1332, 2021, pp. 211–218.

Prescriptive Analytics



Prescriptive Analytics





Prescriptive Analytics

Two irrigation setups during the 2021 campaign (i.e., May/October) within the same orchard

- Managed Row: irrigation is *automatically* controlled using a 2D installation of 12 sensor
- Control Row: irrigation is *manually* controlled by the farmer

Results

- Water management
  - MR saved 44% of water during the whole campaign
  - Maximum saving in June and September: for the farmer is harder to estimate the SM level and water requirement
- Fruit quality
  - Productivity of vines was unaffected by the irrigation and ranged from 32 to 39 kg/vine (35-44 t/ha)
  - Fruits from CR appeared greener (hue angle of 105) than fruits from MR (hue angle of 102)
  - Fruits from CR had lower soluble solid concentration at harvest (12.7 brix) than fruits from MR (15.3 brix)
  - The gap has been maintained after 2 months of storage (and 1 day of shelf life)
    - The soluble solid concentration was 17.4 brix for the MR vs 16.1 brix for the CR

M. Quartieri, M. Toselli, E. Baldi, G. Polidori, M. Germani, M. Noferini, E. Xylogiannis, Effect of the method and volume of irrigation on yield and fruit quality of yellow fleshed kiwifruit in northern italy, in: X International Symposium on Kiwifruit 1332, 2021, pp. 211–218.

Future direction: soil moisture profiling

Continual learning to overcome the limitations of the simulation

- Adapting the model after its deployment
- Use the data that is coming into the production environment and retrain the model based
- Fit to unforeseen field conditions

Improving the recommendation

- Sometimes the soil does not behave as expected

Homogeneous water recommendation; however, we need to handle:

- The "water needs" of the plant
- The phenological growth stages
- Field conditions
- Latitude/longitude
- Availability of water

Future direction: forecasting

While profiling looks at the current state of soil moisture, how will soil moisture change --- for instance --- in a week?

Forecasting soil moisture

- Soil profiles are snapshots of soil moisture, we should learn from time series of snapshots
- Features to consider
  - Soil conditions
  - Weather conditions
  - Type of irrigation

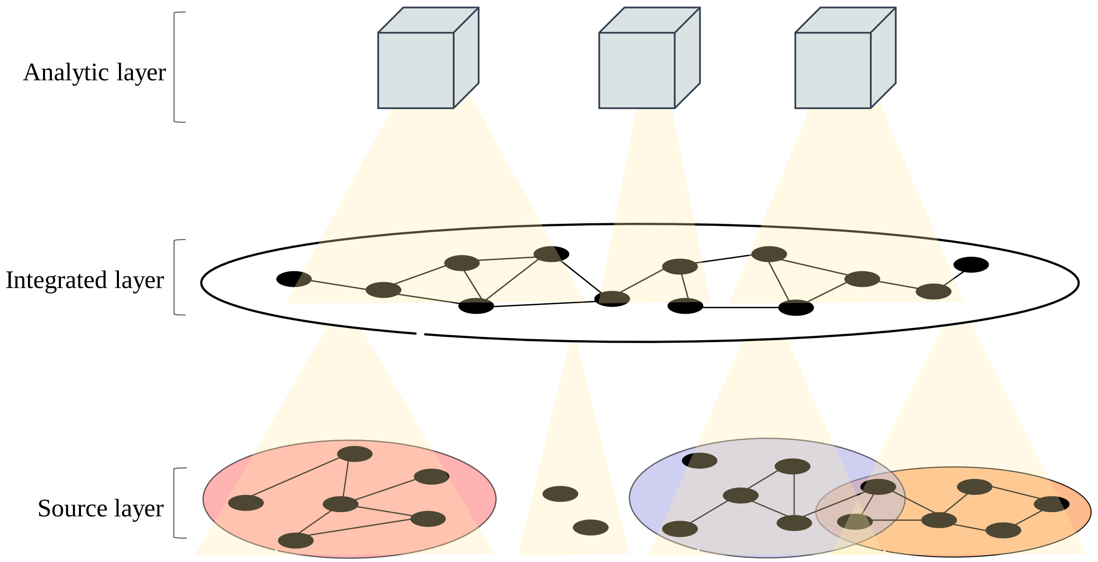
Future direction: unifying data platform

Soil monitoring is a possible application of data platforms for precision farming

- Robotics, tractors and implements
- Satellite images and remote sensing indexes
- Spatio-temporal analysis
- And many others!

Research issues

- Shared dictionary. Many sub-domains of precision agriculture, each with its dictionary
- Data integration. We need a common layer (storage+processing) to access data sources
- Heterogeneous analytics. Data have multiple natures, spatial and temporal data require ad-hoc techniques



Questions?

Thank you for your attention