This project has received funding from the European Union's Horizon 2022 Research & Innovation Actions Project. 101086355 – HORIZON-CL6-2022-GOVERNANCE-01

SCALEAGDATA
MAKING DATA COUNT

SENSOR AND EDGE PROCESSING SELECTION, DEVELOPMENT, SPATIAL PLANNING AND DATA COLLECTION, V2

30 September 2025

Authors		
Name	**Organization**	**Draft release date**
Heikki Astola	VTT	25/09/2025
Alexander Kokka	VTT	25/09/2025
Giorgos Tsilimanis	ICCS	25/09/2025
Isabelle Piccard	VITO	25/09/2025
Nick Berkvens	EV ILVO	25/09/2025
Sarah Verbeke	UGENT	25/09/2025
Idan Kopler	MIGAL	25/09/2025
Dainis Jakovels	IES	25/09/2025
Theocharis Ampas	AUTH	25/09/2025
Maria P. González Dugo	IFAPA	25/09/2025
Mariapina Castelli	EURAC	25/09/2025
Paolo Cosmo Silvestro	DM	25/09/2025
Harald Sundmaeker	ATB	25/09/2025
Thomas Papakosmas	NP	25/09/2025
Valentina Manstretta	HORTA	25/09/2025
Adam Fojud	WODR	25/09/2025
Mateusz Krzyżanek	PSNC	25/09/2025
Gilles Orazi	EGM	25/09/2025

Approval on behalf of the Executive Board		
Name	**Organization**	**Date of approval**
Giorgos Tsilimanis	ICCS	29/09/2025
Nick Berkvens	EV ILVO	29/09/2025
Koen Van Rossum	VITO	29/09/2025

Revision records			
Version	**Date**	**Changes**	**Authors**
Draft 00	12/12/2023	Original document (draft)	All Partners
Draft 0.1	17/01/2024	Review draft	ICCS, ILVO, VTT
Issue 1.0	31/01/2024	Version 1	All Partners
Issue 1.1	27/01/2025	Revised version that takes the comments of the EC and external reviewers into account (D3.2)	All Partners
Issue 2.0	30/09/2025	Update of D3.2 to D3.5	All Partners

Acronyms and Abbreviations

Acronyms and Abbreviations	
ADC	Analogue-to-Digital Converter
AGINS	AgroInsurance International
AI	Artificial Intelligence
AIM	Agricultural Information Model
API	Application Programming Interface
ATB	Institut für angewandte Systemtechnik Bremen GmbH
AUTH	Aristotle University of Thessaloniki
AVR	AVR BVBA (Belgium)
B	Biomass
BTLE	Bluetooth Low Energy
CA	Consortium Agreement
CC	Chlorophyll Content
CDC	Capacitance to Digital Converter
CNH	CNH INDUSTRIAL BELGIUM
CNN	Convolutional Neural Network
CO2	CO2 flux
DAC	Digital-to-Analogue Converter
DES	Deimos Spain
DME	DEIMOS ENGENHARIA SA
DMK	DMK Deutsches Milchkontor GmbH
DSS	Decision Support System
EC	European Commission
EEAB	External Expert Advisory Board
EGM	Easy Global Market SAS
EnMAP	Environmental Mapping and Analysis Program
EO	Earth Observation
ET	Evapotranspiration
EURAC	Accademia Europea di Bolzano (Eurac Research)
EV ILVO	Eigen Vermogen van het Instituut voor Landbouw en Visserij Onderzoek
ExBo	Executive Board
FEU	Farm Europe
FPI	Fabry-Pérot interferometer
FUOTA	Firmware Update Over-The-Air
ICCS	Institute of Communication and Computer Systems
G	Soil heat flux

GA	General Assembly
GDPR	General Data Protection Regulation
GIS	Geographic Information System
GMM	Gaussian Mixture Model
GNSS	Global Navigation Satellite System
GPP	Gross Primary Production
H	Sensible heat
HE	Homomorphic Encryption
HPC	High Performance Computing
IES	Institute for Environmental Solutions
IFAPA	Instituto Andaluz de Investigación y Formación Agraria, Pesquera y Alimentaria
InGaAs	Indium Gallium Arsenide
IoT	Internet of Things
IPR	Intellectual Property Rights
KUVA	Kuva Space Oy
LAI	Leaf Area Index
LAU	Local administrative units
LE	Latent Heat
LPIS	Land Parcel Identification System
LUE	Light Use Efficiency
LUKE	Natural Resources Institute Finland
MCU	Microcontroller Unit
MIGAL	MIGAL Galilee Research Institute
ML	Machine Learning
MPC	Multi-Party Computation
MSE	Mean Squared Error
MST	Management Support Team
NDMI	Normalized Difference Moisture Index
NDRE	Normalized Difference NIR/Rededge Normalized Difference Red-Edge
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NP	Neuropublic SA
NPP	Net Primary Production
NUTS	Nomenclature of territorial units for statistics
ODM	OpenDroneMap
OEM	Original Equipment Manufacturer
OHB DS	OHB Digital Services GmbH, Bremen, Germany

PAR	Photosynthetically Active Radiation
PBMA	Poly (butyl methacrylate)
PCB	Printed Circuit Board
PEMA	Poly (ethyl methacrylate)
PET	Privacy Enhancing Technologies
PHEMA	Poly (2-hydroxyethyl methacrylate)
PI	Proportional-Integral
PIBMA	Poly (isobutyl methacrylate)
PISEC	Platform for Integrated Sensing and Edge Computing
PO	Project Officer
PSNC	Instytut Chemii Bioorganicznej Polskiej Akademii Nauk
PSRI	Plant Senescence Reflectance Index
Pt	Platinum
PTZ	Pan, Tilt, Zoom
R&D	Research and Development
RGB	Red Green Blue
RIL	Research and Innovation Lab
RMSE	Root Mean Squared Error
Rn	net Radiation
SERA	Squared Error Relevance Area
SDK	Software Development Kit
SM	Soil Moisture
SME	Small and Mid-size Enterprise
SOC	Soil Organic Carbon
SVM	Support Vector Machine
SWAP	Company: SWAP instruments (https://swapinstruments.com/?v=75778bf8fde7)
TBC	To Be Confirmed
UAV	Unmanned Arial Vehicle
UGent	Universiteit Gent
VITO	Vlaamse Instelling voor Technologische Onderzoek
VNIR	Visible-Near Infra-Red
VRI IES	Foundation "Institute for Environmental Solutions"
VTT	Technical Research Centre of Finland Ltd.
WASM	WebAssembly
WODR	Wielkopolski Osrodek Doradztwa Rolniczego w Poznaniu
WP	Work Package

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Project overview

ScaleAgData is a response to the call HORIZON-CL6-2022-GOVERNANCE-01-11 Upscaling (real-time) sensor data for EU-wide monitoring of production and agri-environmental conditions. The ScaleAgData project will run from January 2023 until December 2026 and consists of a consortium of twenty-six partners from fourteen countries. The vision of ScaleAgData is twofold. On one hand, it wants to obtain insights into how the complex data streams related to sensors and Earth Observation (EO) should be governed and organized (governance call). On the other hand, it aims to develop the data technology needed to scale data collected at the farm level to regional datasets, agri-environmental monitoring, and the management of agricultural production.

To do so, ScaleAgData has five objectives:

- Developing innovative approaches for collecting in-situ data and applying data technologies.
- Enabling and promoting data sharing along the entire data value chain.
- Demonstrating how the sensor data can be scaled to agri-environmental data products at the national, regional, or European level.
- Demonstrating the benefit of the improved monitoring capacities in a precision farming context.
- Demonstrating the benefit of upscaled regional datasets for the agricultural sector in general.

During its lifecycle, the project will explore seven innovation areas: innovative sensor technology, edge processing, data sharing architecture and data governance, satellite data augmentation, from data assimilation to service development, privacy-preserving technology, and data integration methodologies.

Six Research and Innovation Labs (RIL) have been identified within the project, across various bio-geographical regions of Europe, where different data upscaling and integration models or approaches will be evaluated and demonstrated. The six RI Labs each have their own thematic focus, being Water productivity, Crop management, Yield monitoring, Soil Health, Grasslands, and Sustain Dairy. With this extensive thematic coverage, sensor data will be collected on a wide variety of agri-environmental conditions, allowing ScaleAgData to cover the four dimensions of the environment: Soil, Water, Air and Living organisms (crops and livestock in this case).

1.2 Scope and evolution of the document

This deliverable summarizes the work done in WP3 Task 3.1 (later: T3.1), including the sensor selection and their data acquisition planning in ScaleAgData within the respective RILs (i.e. which data will be collected, where and when), and the development of new sensor technology and edge processing enabling technologies. This is the second issue of the document (delivery in M33). The first issue was delivered in M13 (D3.2 Jan 2024), and its revision in M25 (D3.2 Rev 1 Jan 2025).

A major part of the document will be the Sensor Data Catalogue, which will list all the sensor data sets collected during the project, and the related metadata to the extent available. This catalogue is a living document that will be updated constantly along the sensor data acquisition campaigns by the RIL leaders. The Sensor Data Catalogue will be implemented as an Excel spreadsheet document, and it will be added as an annex to this deliverable.

1.3 Document structure

This document is structured as follows:

- Chapter 1 provides a project overview and then goes on to describe the scope, responsibilities, and structure of this deliverable.
- Chapter 2 presents the data acquisition planning in each RIL.
- Chapter 3 describes the development of new sensor technology.
- Chapter 4 presents the ICCS approach to optimal sensor placement.
- Chapter 5 presents the developments in edge processing technologies.
- Chapter 6 presents the sensor data catalogue structure and gives a link to the catalogue that is a separate document (Excel file).
- Chapter 7 draws the conclusions of the work in WP 3 T3.1.

1.4 Evolution of the document

Version 1.0 of this document, submitted on 31 January 2024, provides a description of the sensor selection and their data acquisition planning in ScaleAgData within the respective RILs, and the development of new sensor technology and edge processing enabling technologies.

The revised version of the document, D3.2 Revision 1, submitted on 27 January 2025, included minor changes, taking the comments of the EC and external reviewers on the deliverable D3.2 into account.

This is Version 2.0 of the document (deliverable D3.5). This version will provide an update on sensor data acquisitions, the new sensor developments and technology implementations and on the results obtained.

Additional updates will take place if necessary, due to changed circumstances that require alterations to the approaches presented herein.

2 Sensor Data Acquisition Planning

The different data upscaling and integration models or approaches that will be developed in ScaleAgdata, will be evaluated and demonstrated in the six RILs. The chapters below will present the plan for collecting the essential sensor data for the RIL at hand, listing the data sources, the data acquisition schedule and the resources needed for the actual data acquisition work.

2.1 Approach

The work in WP3 T3.1 inherits from the work done in the co-design workshops, and their outcomes, which are documented in the deliverable D2.1 Vision scenarios, requirements, and innovative governance models. WP2 produced a high-level backlog for each RIL with the list of requirements and their time-related prioritization, as well as the rolling plan for facilitating the upcoming cooperation in developing the services and products in the following work packages: WP3, WP4 and WP5.

The outcomes of WP2 T2.1 are the starting point for the selection of the set of sensors needed in each RIL to fulfill the objectives and requirements obtained from the co-design phase. The RIL leaders and technology providers will verify the requirement lists and derive/refine the specifications and technical requirements and will adapt to this information in planning the data acquisition/development work. Figure 1 depicts the overall approach of sensor data acquisition planning and sensor data collection in ScaleAgdata.

This work in WP3 T3.1 will follow the principles and aspects described in D1.3 Open Science and Data Management Plan, where applicable.

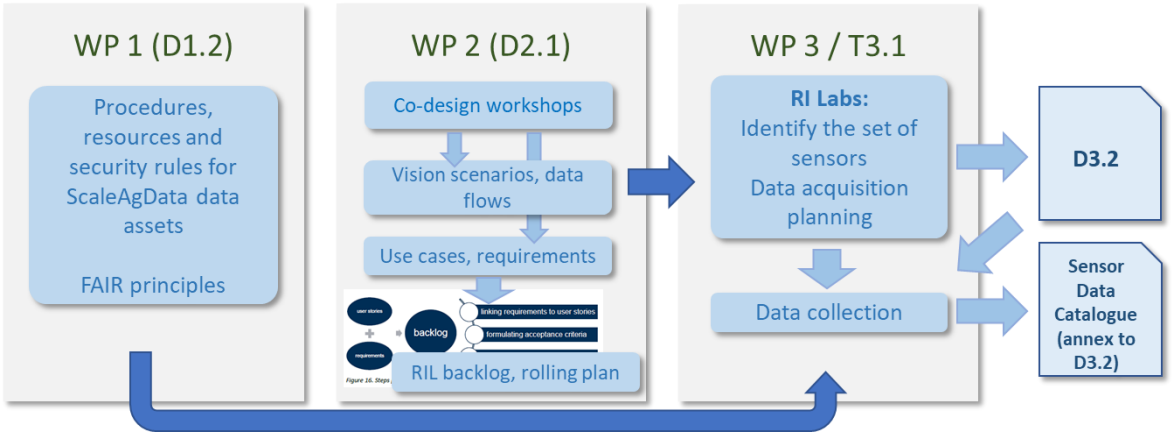

Figure 1. The overall approach of the sensor data acquisition planning and sensor data collection in ScaleAgData

2.2 RIL Yield monitoring

2.2.1 Sensor data

The following sensors will be used in the Yield monitoring Lab:

- AVR harvester yield sensor (see 2.2.2 for more information)
- AVR harvester machine parameters (traction, fuel usage)
- CNH harvester yield sensor (see 2.2.3 for more information)
- CNH harvester machine parameters
- Weather station (for some test fields)
- Soil scanner (for some test fields)
- Crop scanner (for some test fields)
- RGB camera
- LAI sensor

2.2.2 AVR yield sensor

2.2.2.1 Sensor Selection

AVR has implemented a yield measurement system on its AVR potato harvester. This consists of 2 weighing cells mounted on the top belt of the harvester connected to a signal amplifier, after which the data is sent each second to the cloud platform (AVRConnect). On this cloud platform, the yield information is visualized on a field view showing the variability of the yield measured on the field (see Figure 2).

Figure 2. Yield (tons/ha) measured on the field by an AVR potato harvester equipped with a yield measurement system. Lower than average yields are shown in red, average yields in green, and higher than average yields in blue.

2.2.2.2 Schedule

Yield information (tons/ha) is measured every second by the harvester. This data is only collected during yield measurement in the field while the potato harvester is harvesting potatoes on the agricultural fields.

2.2.2.3 Resources

Sensor data is collected from multiple harvesters. AVR has a coverage of 10 machines in Western-Europe.

2.2.2.4 Special data constraints

Special attention is needed to have correct data. Calibration is frequently needed to have “realistic” results. So, we need to be very conscious of the quality of the data. A common source of error, for example, arises from soil clumps that resemble potatoes. Additionally, in wet conditions, the absolute yield may be higher due to soil covering the tubers.

2.2.3 CNH yield sensor

2.2.3.1 Sensor selection

A yield sensor is available in the CNH portfolio. Each data point is linked to a geospatial location. The data is recorded in 1-second intervals and sent in batches to the cloud (see Figure 3).

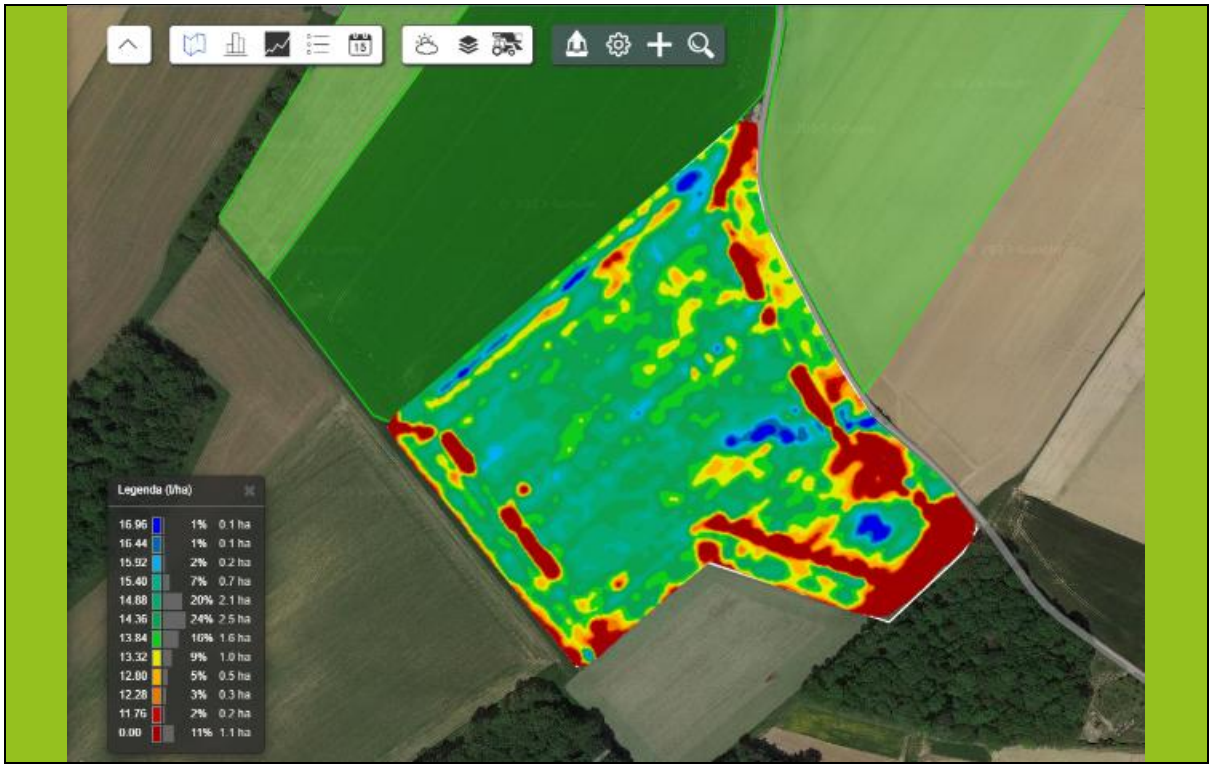

Figure 3. Yield measure in the field with CNH sensor

2.2.3.2 General information

CNH is working with a test farmer located in Braine-le-Château, Belgium. He records data on some of his 2000ha of winter wheat fields.

2.2.3.3 Schedule

Yield data is collected only during harvest time (July-August). This is highly dependent on the weather.

2.2.3.4 Resources

4 test machines collect data in this region.

2.2.3.5 Special data constraints

Data is collected in cn1 formats, this is not a directly readable format, so conversions need to be made. Data quality is also highly dependent on sensor calibration. If this is not done correctly the absolute values will be off.

2.2.4 Other field sensors

CNH's test farmer also collects the following data on selected fields:

- weather data (~10 weather stations, https://smartfarm.nl/en/sensors-en/)
- soil data (electrical conductivity at 4 depths) (https://www.topsoil-mapper.com/en/)
- crop data (augmenta scan ~ biomass index) (https://www.augmenta.ag/)

These data are collected throughout the season.

2.2.5 Sensor use cases

2.2.5.1 Tare weight estimation in potatoes (UGent in collaboration with AVR)

Currently, the tare weight of the potato harvest, i.e., the soil / dirt attached to the tubers and the harvester's conveyer belt, is not taken into account in the yield measurements, and hence the yield maps presented in section 2.2.2.

In this use case, RGB cameras mounted on the harvester will be used to estimate the tare weight. In image analysis, the pixels assigned to soil clumps need to be linked to the tare weight by calibration experiments. RGB images are recorded every 2 seconds during harvest. The harvester is halted between 10 and 15 times per field to sample the potatoes and soil under the RGB camera to have a 100% match between and RGB photo and a sample that can be weighed. To make sure the match is correct, manual photos (with a phone) are taken as well. Soil samples are taken as well to see whether the soil moisture content influences image analysis and the relation between the amount of 'soil pixels' detected and the tare weight. The soil moisture data can be used to assess the accuracy of EO soil moisture data and to determine whether EO data can be a good alternative to use as input in the image analysis.

An experiment was performed in 2024 for two fields in Flanders, Belgium (sandy loam and sand soil) following the methodology described above. On one field, a soil scan was also performed as additional comparison of the soil moisture. Data analysis showed no correlation between the different soil moisture methods, but also that the soil moisture had a non-significant influence on the image analysis of the tare weight.

The experiment will be repeated for an additional six fields in Flanders, Belgium (4 loam soils and 2 sand soils) in Sept-Oct 2025. In the 2025 season, the hyperspectral camera from VTT will also be tested (not on the field) to see whether certain wavelengths can distinguish between the potato tuber and soil clumps more easily. This could provide a basis for future research where a hyperspectral camera is installed on the harvester (alongside or instead of the RGB camera).

2.2.5.2 Improved winter wheat yield maps from CNH harvesters (UGent and Luke in collaboration with CNH)

Yield mapping by harvesters documents the success of cultivation at high spatial detail and therefore serves as a key sensor-based data source in precision farming. When the harvester instrumentation can extract variables such as yield weight and protein content from the grain stream and geo-locate it accurately, the resulting dataset offers highly valuable feedback for crop management. However, common challenges include data gaps caused by recording or geo-positioning failures, calibration drift, and variability between harvesters. To fill in the gaps in the yield information, a digital twin model (developed by LUKE) and EO data are being used to simulate the crop growth process and the respective spatial variation in yield, constrained by the physiological limits in the crop models. The spatial variation of previous years (in both EO data and harvester data) can be used to estimate variation in soil quality. The goal is to compare yield variation in a field simulated with the model and measured with the harvester sensors. If this is comparable, then the model can be used to fill in the missing data from the harvesters. However, care must be taken to distinguish the variations caused by other factors, like seasonal variations in weather, management practices, or seed variety from the (more stable) soil quality variation itself.

In the first iteration, sensor data (yield from harvester sensors, soil electrical conductivity - EC) and EO data were gathered for the seasons 2023 and 2024. Additional manual measurements were performed for the 2024 season to improve the calibration of the digital twin model. This included sensor data: soil moisture measured with a soil moisture sensor and plant parameters (leaf area index, leaf chlorophyll content, flavonol index, and nitrogen balance index). These measurements were gathered on four fields (4-6 subplots) located in Flanders, Belgium, on four different days in 2024 (14/3, 18/4, 15/5 and 4/7/2024).

In the second iteration (2025 season), new data was gathered for validation purposes. Due to unforeseen circumstances, only one field was sampled, instead of the planned five. For eight subplots in this field, leaf area index was measured at the end of the growing season on 23/7/2025. Chlorophyll content was not measured as the plants were mature and did not contain chlorophyll anymore.

2.2.5.3 Potato and winter wheat yield (variability) maps and statistics (VITO in collaboration with AVR and CNH)

The yield data collected by the AVR and CNH harvesters are used to train a ML model that uses EO, weather and soil data to estimate potato and winter wheat yield, respectively. This will enable farmers – also farmers who do not have access to a harvester equipped with a yield sensor – to analyze yield variability within their field and make informed decisions to adjust field practices accordingly. Other target groups are agricultural service providers, machinery producers, insurance companies or seed companies. They could use the yield (variability) maps for improving or extending their current products or services or for research purposes. Aggregated yield estimates may be of interest to public authorities for crop production monitoring or policy evaluation.

2.2.5.4 Schedule

- Oct-Nov 2023: tare calibration experiment (1 soil type, RGB camera)
- Summer 2024: prepare the sensors on the fields of CNH, AVR
- Summer 2024: prepare the digital twin model for wheat.
- Summer 2024: gather EO data of previous years of CNH fields and use it for testing the digital twin model.
- Oct-Nov 2024: tare measurements on 2 AVR potato fields (soil samples + potato samples)
- Oct-Nov 2024: first soil moisture calibration and tare estimation experiment (1 soil type)
- July 2025: wheat measurements on 5 fields from CNH (validation digital twin model)
- July-Oct 2025: cleaning the harvester dataset to be used for yield modelling

- Sept-Oct 2025: tare measurements on 6 AVR potato fields

2.2.5.5 Resources

The following human resources will be necessary:

- Farm technicians: to perform the tare calibration experiments.
- Data scientists: to read and analyze all the data, to make and assess the models

Support we expect to receive:

- VITO: EO data for a few fields, to test the models
- LUKE: digital twin for the wheat fields (crop model)
- VTT: support for hyperspectral camera mounting, usage, data analysis

2.2.5.6 Special data constraints

Tare measurements have been performed on two fields so far (sandy loam and sand). In the next iteration four loam soils and two sand soils will be added to the dataset. This is still a very small dataset for calibration of image-based tare estimates.

2.2.6 Summary of the first 30 months of the project / RIL Yield Monitoring

Yield data collected by AVR harvesters have been used to train a ML model to estimate potato yield. Therefore, harvester data that coincide with potato fields registered in the government-reported Land Parcel Identification System (LPIS) dataset are resampled to a grid of 20m x 20m subfields (see Figure 4), corresponding to the Sentinel-2 pixel grid, whilst removing any calculated subfields that are not fully contained within the harvester dataset.

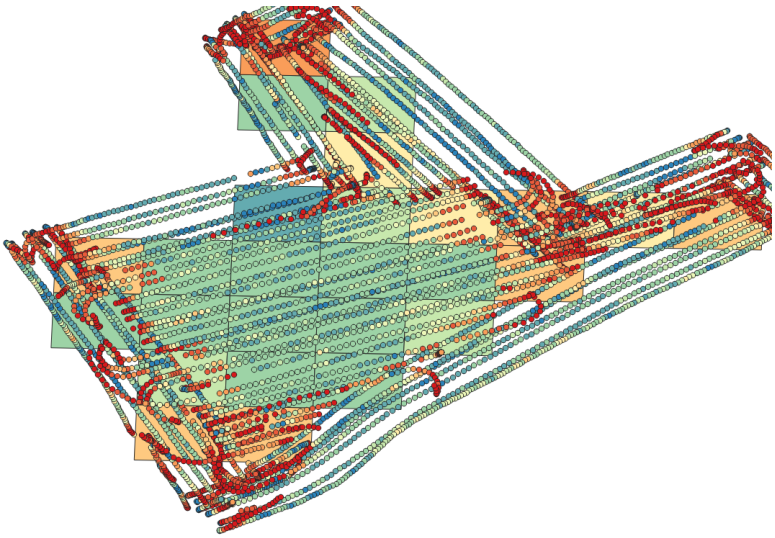

Figure 4: Potato yield data recorded by an AVR harvester, resampled to a 20x20m grid

As illustrated in Figure 5, there is a large variation in yield values within these 20m x 20m subfields. Low yield values as well as overlapping low and normal yield values are recorded, especially at the turning zones at the end of the field.
Further effort is needed to clean the harvester datasets. This is planned for the second iteration. To develop appropriate cleaning methods, it is important to determine the causes of the variation in yield values. Therefore, we will have a closer look at the harvester data from a subset of fields together with AVR and the farmer. Depending on the outcome, simple methods may be used to remove extreme

values, or smarter methods may need to be developed, e.g., reducing harvester data to only data points that are clearly recognized to be collection data points (i.e., when the harvester is moving across a field at the predominant angle signifying it is collecting and at the predominant speed that is used when collecting). This may also reduce noise caused by false collections such as during harvester movement or traversing the field following collection.

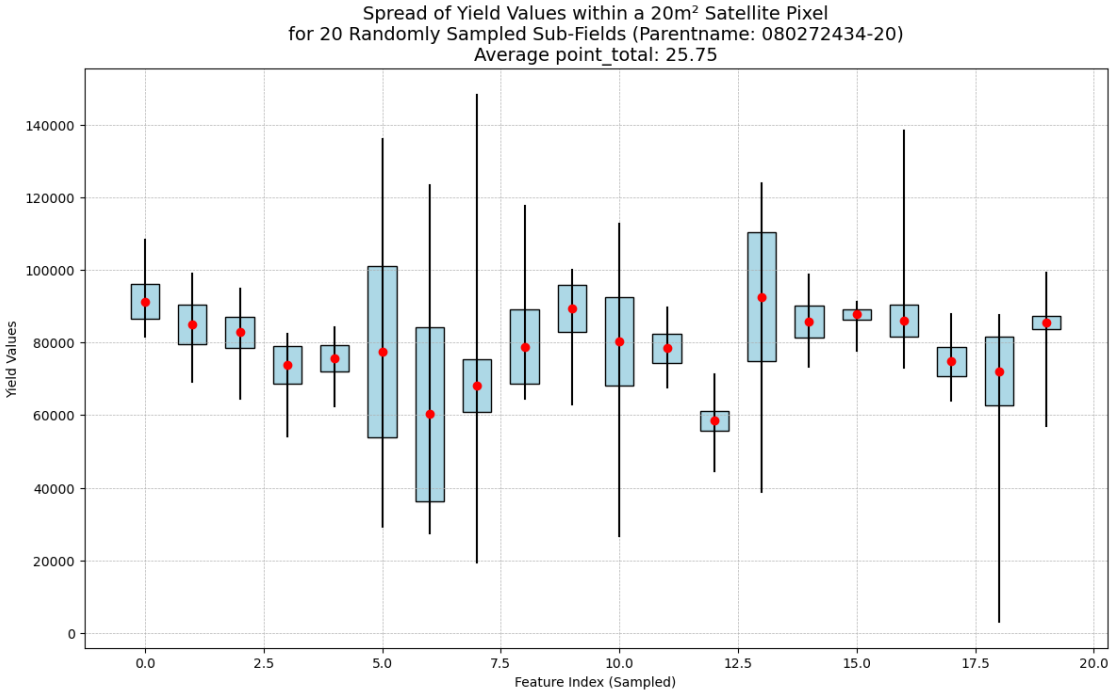

Figure 5: Variation of harvester yield values within a 20x20m satellite pixel, for 20 randomly sampled sub-fields. The red dots represent the median yield value of the sub-field.

Another issue with the dataset is that harvesters are susceptible to errors induced by different adherence to calibration procedures by the farmers/contractors who operate the harvesters. The data provider has provided a list of harvesters operated by farmers considered to be strictly adherent to required calibration procedures and thus minimize risk of error in the dataset.

Finally, environmental conditions such as soil moisture or soil type may influence the yield measurement by the harvester. This issue could be tackled by applying the tare weight correction described in section 2.2.5.1.

2.3 RIL Water

2.3.1 Data Acquisition Plan

2.3.1.1 Sensor data

Local (or regional) IoT meteorological station data:

- Precipitation
- Air temperature (Max, Min)
- Humidity (Max, Min)
- Wind speed and direction
- Air pressure
- Solar radiation
- Evapotranspiration

Local IoT soil sensors:

- Soil moisture
- Soil temperature
- Water pressure

Local airborne data:

- Spectral data in VIS-NIR (400-1000 nm) spectral range
- Thermal data

Regional satellite data products:

- Evapotranspiration data products (provided by DHI)
- Soil moisture data products (provided by DHI)

2.3.1.2 Schedule

Local IoT meteorological and soil sensor data were acquired for fields of interest at least once every 30 min for the whole cropping season to ensure a near-real-time data flow. Local airborne data were acquired for fields of interest upon request when the situation in the field had significantly changed to update the spatial distribution of water status and predicted yield potentially. The expected update frequency was once per 1-4 weeks. Regional satellite data products (soil moisture and ET) were obtained with the highest possible frequency (daily) to upscale developed models for water status assessment and yield prediction on a regional scale for fields where local IoT sensor data is not available.

2.3.1.3 Resources

IoT meteorological and soil sensors used in the study are the property of project partners – IES (Latvia) and MIGAL (Israel) and were placed in chosen test fields. Airborne data acquisition equipment is a property of project partners – IES (Latvia) and MIGAL (Israel) —and was used for data acquisition in chosen test fields. The RIL received support from project partners in satellite data processing:

- DHI – satellite data-based evapotranspiration and soil moisture data products
- VITO – satellite data-based vegetation and other indices

2.3.1.4 Special data constraints

Cropping seasons vary in time in Latvia and Israel, therefore, data acquisition was adapted to the agricultural schedule.

2.3.2 Summary of the first 30 months of the project / RIL Water

It was agreed that maximally similar data acquisition setups will be used in both Latvia and Israel despite different climatic conditions and crops. IoT meteorological station and soil sensors were planned as the primary sensors for field water status assessment complemented with airborne data for increased spatial resolution. The secondary data source was EO data products provided by DHI for

potential model upscaling to other fields where local IoT sensor data isn't available. A summary of acquired data from each test site is provided below.

2.3.2.1 Acquired data from peppermint fields in Latvia

The experimental setup with different irrigation applications to peppermint in Latvia is shown in Figure 6. The peppermint field was divided into four subfields (each at least 1 ha large) where different irrigation regimes (100%, 67%, 30% and 0%) were planned. The experiment was implemented in collaboration with a local organic farming company "Field and Forest". Initially, it was planned that peppermint seedlings will be received and planted early in May 2024, but the delivery from Poland was delayed by one month due to unfavorable climatic conditions which resulted in the planting of peppermint seedlings only in early in June 2024.

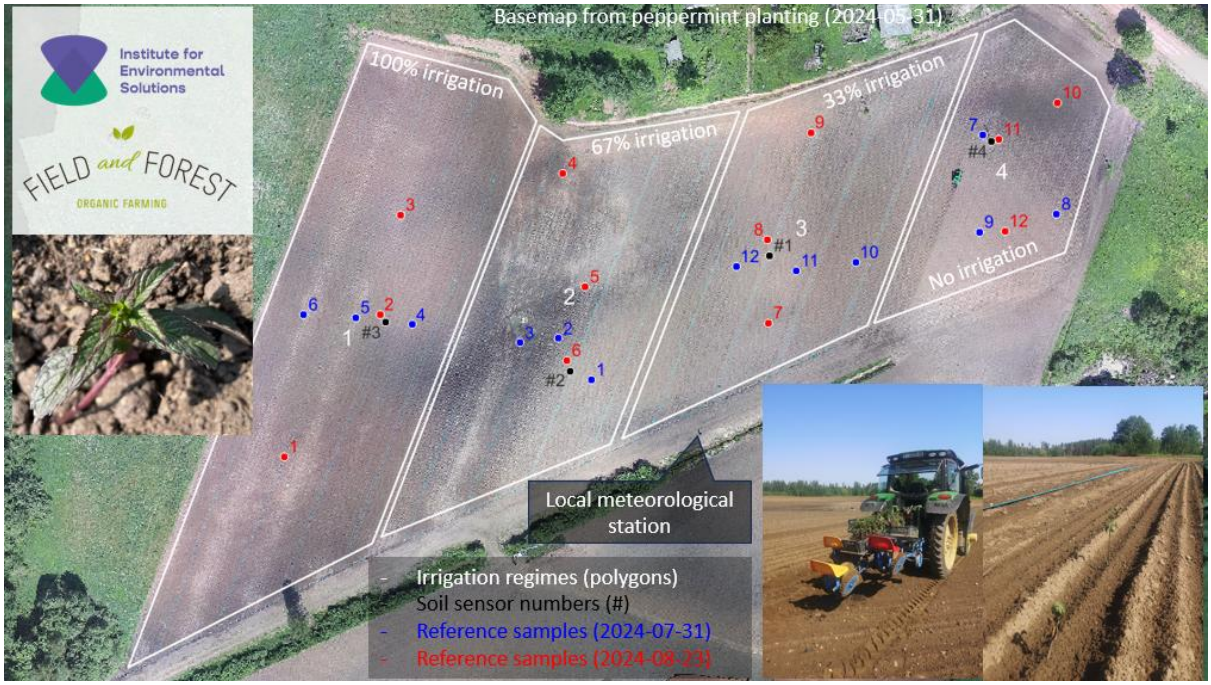

Figure 6. Peppermint pilot site in Latvia.

The local meteorological station was installed near the peppermint field and meteorological data acquisition was performed once every 15 minutes since the beginning of May 2024, see Figure 7. A month -long spring/early summer drought period could be seen in the precipitation graph with monthly precipitation rates significantly below the climatic norm as well as significantly increased air temperature above the climatic norm. Such phenomena whereas often observed in the recent decade in Latvia affecting the development of seedlings. This was the main motivation for the local farmer to test irrigation of peppermint in Latvia. Unfortunately, seedlings arrived during the drought period and were planted during relatively harsh weather conditions. Therefore, the farmer decided to apply 100% irrigation to all fields during the growing phase. Differentiated irrigation was applied five times in July 2024 when precipitation was below the climatic norm. In total, 100% of the irrigation fields received an additional 60 mm of precipitation, 67% of the irrigation fields – 40 mm, and 33% of the irrigation fields – 20 mm. At the end of July rainfall started to exceed the climatic norm and application of irrigation wasn't needed anymore.

Soil sensors were installed in sub-fields on June 18 and acquired data every 15 minutes till September 18, 2024, see Figure 8.

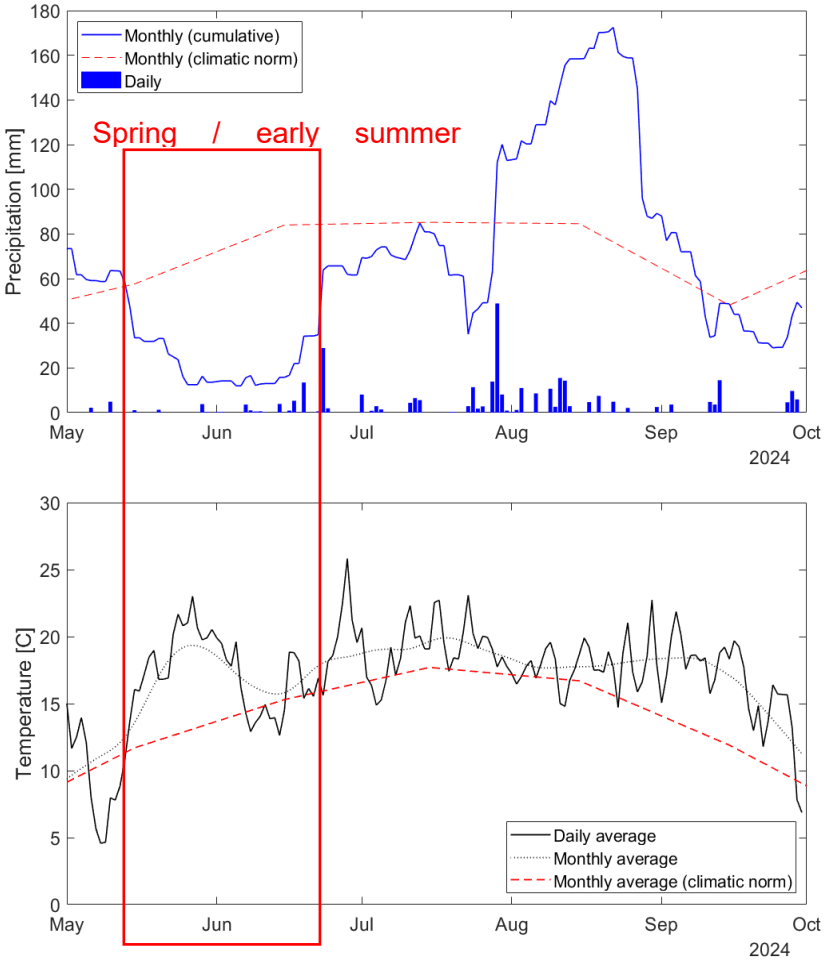

Figure 7. Meteorological data from the local meteorological station located near the peppermint field.

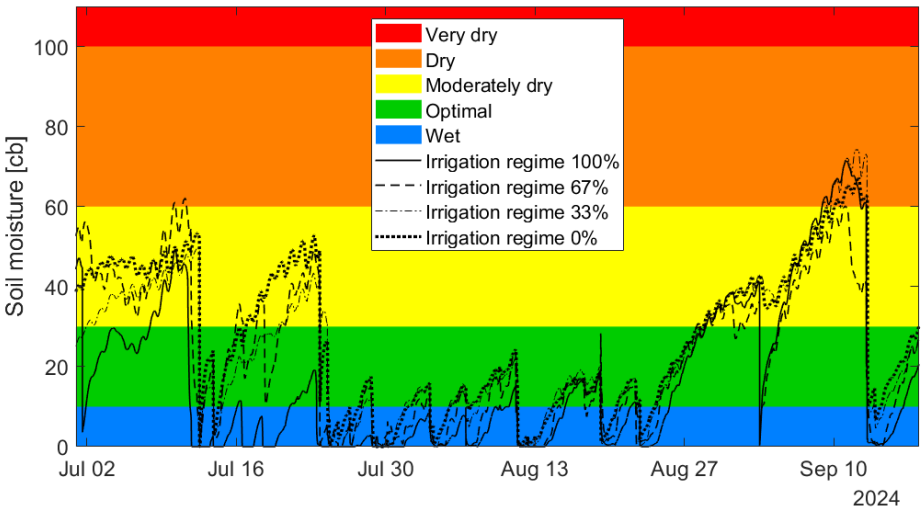

Figure 8. Soil moisture data from IoT soil sensors located in peppermint sub-fields.

Airborne data (VNIR hyperspectral and thermal, 1 m/pixelspatial resolution) were acquired four times during the vegetation season:

- 5-June-2024

- 25-June-2024
- 16-July-2024
- 14-August-2024

Additionally, two drone data (RGB and thermal, 1 cm/pix spatial resolution) were acquired twice:

- 31-May-2024
- 05-June-2024

Therefore, different resolution airborne (1 m/pixel) and drone (1 cm/pixel) data was acquired on June 5, 2024, for the barren field before peppermint was planted.

Unfortunately, peppermint stands didn't develop large enough to be harvested at the end of the season. Vegetation cover was dominated by weed rather than peppermint, therefore, application of spectral data for peppermint yield or biomass assessment wasn't possible. Nevertheless, two reference data gathering campaigns were organized (July 31 and August 23, 2024) to collect peppermint samples for further chemical composition analysis in the laboratory. The main reference parameter of interest was concentration of essential oils indicating the quality of the potential yield.
EO data products (soil moisture and evapotranspiration) were provided by DHI at the end of the season on request for the area of interest resulting in 153 data scenes for each parameter for period May 2024 –September 2024; see

Figure 9. Provided data was used to test model upscaling from fields with IoT sensors to other potential fields of interest.

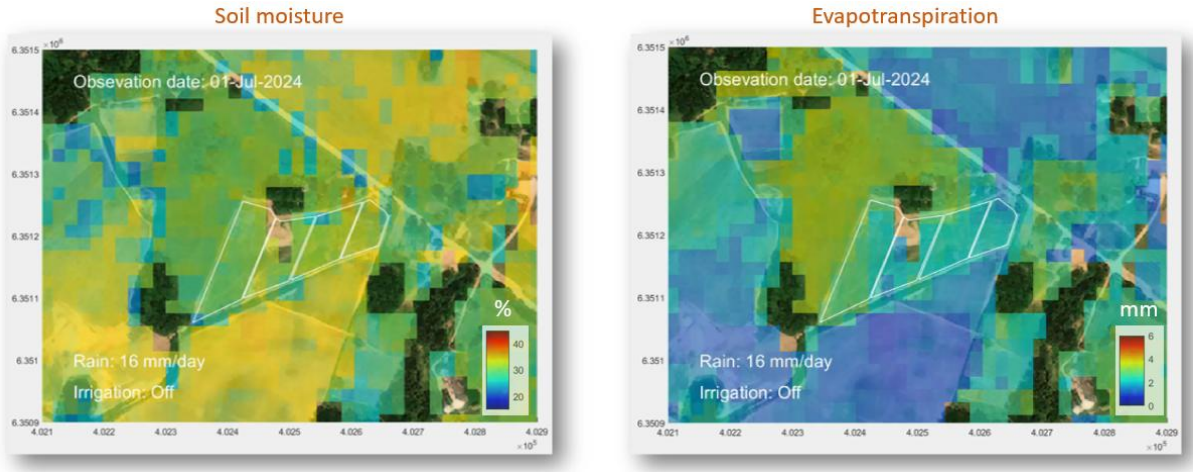

Figure 9. Examples of soil moisture and evapotranspiration data products provided by DHI group.

It was planned to continue irrigation experiments with above mentioned data acquisition during the vegetation season in 2025. However, the monthly precipitation exceeded climatic norm nearly twice since during the whole vegetation season in 2025, see the precipitation data from meteorological station located near peppermint fields in Figure 10. Irrigation application wasn't required, and no heat waves / droughts were observed. Thus, Water productivity RIL demonstration with different irrigation regimes in peppermint fields for its impact assessment on yield wasn't possible. It should be noted that emergency situation in agriculture was announced in the whole country by the Ministry of Agriculture of the Republic of Latvia due to continuous rainfalls during the whole vegetation season.

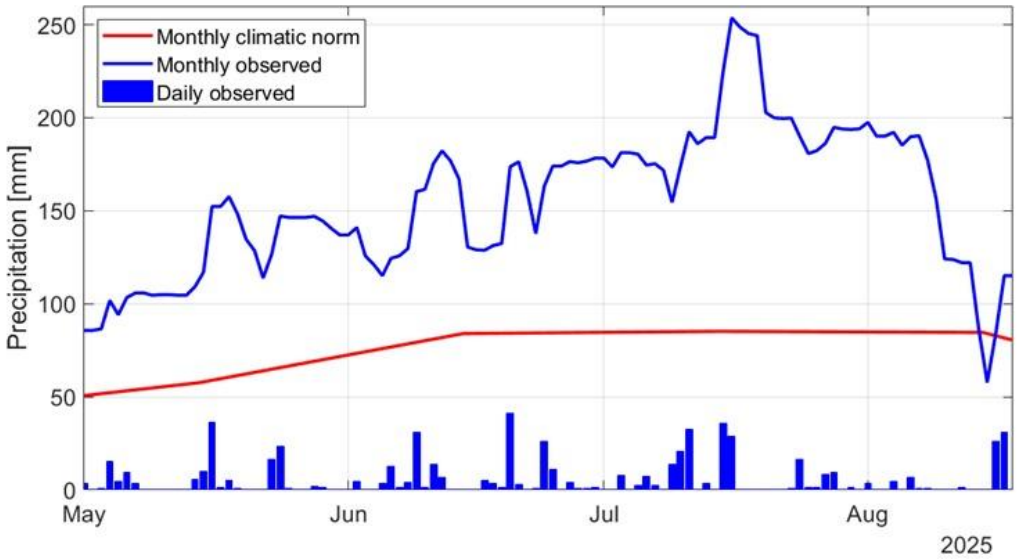

Figure 10. Precipitation in peppermint fields in Latvia during the vegetation season in 2025.

Nevertheless, IoT sensor data acquisition continued during the whole vegetation season. Additionally, repeated soil sampling in 36 points was performed to cross-check previously mapped soil granulometric composition. Peppermint leave samples were collected at the same 36 points at the end of the season (Sep 12, 2025) for chemical composition analysis and the assessment of concentration of essential oils (yield quality). Several repeated airborne data acquisitions were performed over peppermint fields, but it was observed that vegetation cover was dominated by weed as peppermint development was disturbed.

2.3.2.2 Quinoa fields in Israel

As a lesson learned from the unsuccessful cultivation in the previous season (April 2024 - September 2024), the experimental plot was transferred from the MIGAL Institute's experimental farm to the plot of a private farmer. The farmer has experience with the cultivation of quinoa, a decisive factor in choosing him, since quinoa cultivation during the summer in Israel had never before been carried out on such a large scale.

The irrigation scheme contains 3 treatments—control treatment of ET-dependent irrigation (100%), deficit irrigation of 80%, and deficit irrigation of 60%. Each treatment contains 3 repetitions of 0.36 ha each, totaling 1.08 ha, thus an overall size of 3.24 ha (Figure 10).

Figure 10. Quinoa experimental plot with the division into 3 irrigation regimes – from right to left – 100%, 80%, and 60%, respectively.

Quinoa sowing took place on 28 May 2025. Meteorological data is extracted from a nearby (500 m) meteorological station of the Israeli Ministry of Agriculture. The data is free to the public via an internet site and a cell phone application. Concomitantly, 6 sensors stations were hired from Phytec AgTech company for the duration of the growing season. Each station is composed of a soil moisture sensor (Electronic Capacitance Continuous Logging Probe based on Capacitance Technology (FDR) - SMP6 PHY90 CLP5 Phytech Wired Probe), which measures volumetric soil moisture and soil temperature up to 90 cm depth, 3 plant stem diameter sensors, and a sensor for measuring the water pressure in the irrigation pipe. The 6 stations were distributed evenly between the treatments (2 for each treatment) and located in the middle row of each treatment. All measured data is logged at the designated cloud of Phytec company and is presented at the company's site where only approved users can access the visualization, the analyses and can retrieve the data (Figure 11).

Two acquisitions of RGB, thermal and multispectral, (1 cm/pixel spatial resolution) imagery with MIGAL drones were acquired on 21 July 2025 and 28 August 2025. The NDVI index produced, and the thermal image of the 21 July 2025 flight are presented in Figure 12.

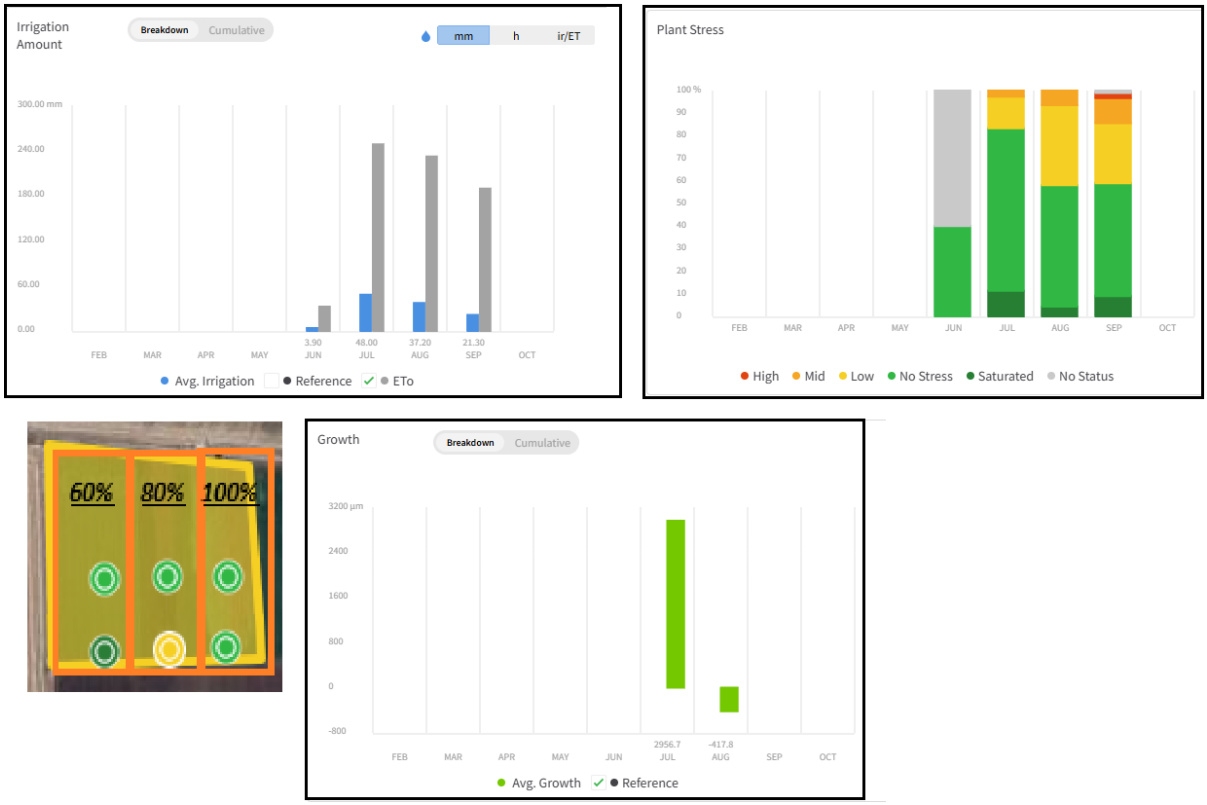

Figure 11. An example of cumulative data of irrigation amount, plant stress and growth provided by Phytec' dashboard.

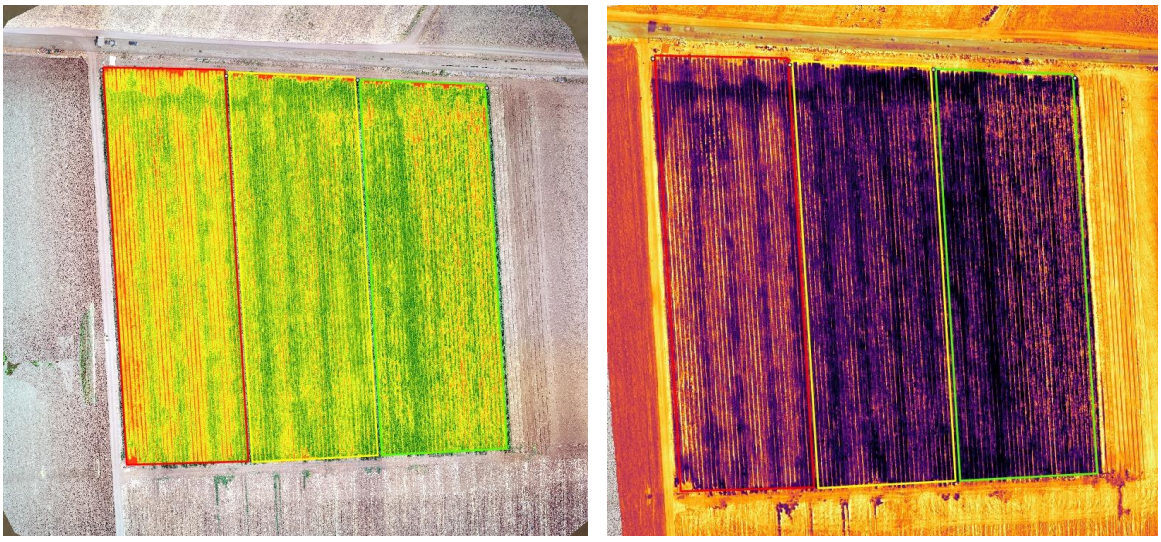

Figure 12. NDVI Index image (left) and thermal image (right) of the quinoa experimental plot with the division into 3 irrigation regimes – from right to left – 100%, 80%, and 60%, respectively.

2.4 RIL Soil

The RIL Soil will focuses on using innovative hyperspectral imaging monitoring tools to improve on the contemporary products that rely on remote sensing data of multi-spectral data. The goal is to both i) provide enhanced estimation accuracy, as the more detailed hyperspectral signal can assist AI models to better estimate the soil properties by exposing more absorption bands, and ii) provide soil maps at unprecedented detail using UAV data (i.e., <1m) which is particularly important for more effective land management practices and to provide maps in orchards where the current spatial resolution (10 m) is insufficient to differentiate between tree and exposed soil.

The overall approach plans on using two forthcoming hyperspectral products, namely the VTT HSI system and the KUVA HS satellite mission of Hyperfield-1. Recognizing that there is an inherent risk in relying on future sensors whose delivery might be delayed, and to ensure that there will be adequate data for the RIL Soil, a contingency plan was developed as follows:

- For UAV data, both ILVO and AUTH have their own UAV platforms and previously procured hyperspectral cameras, albeit at a lower spectral range than the VTT-developed HSI. The plan is to use these devices if, for any reason, the VTT HSI fails to become operational.
- With respect to spaceborne data, in case Hyperfield-1 for any reason does not become operational or due to the limitations of the revisit time and weather conditions (e.g., consistent cloud coverage at each satellite overpass during the critical period when bare soil is visible from above), the plan is to use other HS satellite data. To this end, both ILVO and AUTH will make sure to place orders for PRISMA and/or EnMAP. It should be noted that the same limitations apply to PRISMA and EnMAP data as well.

2.4.1 Sensor selection

Two sensors will be used to further improve existing Soil Organic Carbon (SOC) models, i.e., that predict the topsoil SOC content from visible and infrared reflectance spectra.

- VTT HSI camera
- KUVA HS satellite Hyperfield-1

Currently, multispectral EO (Sentinel 2) is used for the SOC models; with hyperspectral imagery the plan is to gain more fine-grained spectral measurements and a higher spectral resolution for the agricultural soils we are monitoring.

The spaceborne data from Hyperfield-1 provide much larger coverage which is particularly suitable for annual crops which exhibit wide areas of bare soil in the critical period. The UAV data, on the other hand, are more useful to detect bare soil in orchards and tree crops (e.g., olive trees, vineyards, that are very prevalent in the Mediterranean and apple trees in Northern Europe). Therefore, the aim is to use both data sources to provide maps of SOC.

In addition, we combine spectral data with datasets providing more contextual information about the soils of the agricultural fields, e.g., the 'bodemassociatiekaart' of Flanders (https://www.dov.vlaanderen.be/page/bodemkaarten, a digital vectorial dataset with an overview of the occurrence and classification of soil associations in Flanders), various soil property maps provided by the European Soil Data Centre (ESDAC, https://esdac.jrc.ec.europa.eu/content/topsoil-physical-properties-europe-based-lucas-topsoil-data).

2.4.2 Data Acquisition Plan

2.4.2.1 General information

VTT HSI sensor: One sensor, was attached to a UAV (Matrice 600) , the first version of the sensor was ready for experiments in 2024. In the second phase of the project, VTT has further optimized the camera by integrating an enhanced Fabry-Pérot interferometer, which will extend the hyperspectral range of the sensor. In addition, the sensor now records GNSS positional data with each data cube. The upgraded sensor will be transferred to both ILVO and AUTH to conduct experiments in the second iteration of the project during 2025 Q3-Q4 and 2026 Q1-Q2.Kuva Space successfully launched two hyperspectral satellites (Hyperfield-1A in August 2024 and Hyperfield-1B in June 2025) as part of the ScaleAgData project, both operating in VIS-NIR wavelength ranges from 500km altitude. The satellites provide 50x50km hyperspectral images with ~50GB daily data collection, offering multiple product levels from raw radiance to atmospherically corrected surface reflectance. Kuva Space has implemented a data access platform and API, with processing capabilities delivering products. The company will provide access to Scaleagdata project member s to download data through the Kuva Sense™ platform, access to this platform will be available in September 2025.

2.4.2.2 Schedule

Bare soil is needed in agricultural fields to be able to use satellite/sensor images to predict the topsoil SOC (VTT HSI sensor). In Flanders we plan to perform experiments on potatoes fields before harvest (March-May) or after harvest (August-October) of crops, in Greece we plan to perform experiments in tree crops (e.g., grapes, olives, apricot, peaches) in which case bare soil may be identified throughout the year (in best cases). Experiments will be in fields of 1-2ha.

- KUVA HS satellite: results of an already executed soil campaign collecting and analyzing soil samples throughout the whole of Flanders and Central Macedonia will be used, subsequently satellite imagery for the whole or large regions of Flanders and Central Macedonia and available throughout the whole year will be used (we will use as many satellite images as available and suitable for spectral data related to bare soils).

2.4.2.3 Resources

The following human resources will be necessary:
Farm technicians, UAV and robotic engineers, data engineers, GIS scientists and data scientists.
We expect to receive support from project partners VTT and KUVA in satellite data processing.

2.4.2.4 Special data constraints

The weather conditions must be good for UAV flights and capturing data with the VTT sensor, in addition clouds prevent KUVA HSI satellite from capturing spectral measurements of the agricultural fields.

2.4.3 Summary of the first 30 months of the project / RIL Soil

In mid-June 2024, EV ILVO successfully received the first VTT HSI sensor prototype (as detailed in Section 3.2). The delivery timeline experienced delays due to unforeseen administrative challenges, including insurance requirements, export legislation compliance, and loan agreement negotiations. To address these issues, AUTH submitted a project amendment request to secure additional funding for insurance coverage and transport costs.

Following sensor delivery, test flights were conducted at EV ILVO facilities in July 2024 to configure the sensor mounting system, validate UAV operations, and establish protocols for hyperspectral image data acquisition and processing. The experimental phase commenced in mid-September 2024 when agricultural fields at EV ILVO were harvested, creating the required bare soil conditions for data collection.

On 20 September 2024, two comprehensive field trials were executed: Field S28A (1.09 hectares, previous crop: potato) and Field R13 (2.2 hectares, previous crop: winter barley), as illustrated in Figure 13. During these trials, soil samples were systematically collected to establish ground truth Soil Organic Carbon (SOC) content values. These reference measurements are essential for training machine learning algorithms to predict SOC distribution across entire fields using hyperspectral sensor data.

Laboratory analysis of the collected soil samples was completed in November 2024. The spatial sampling locations were determined using the sensor spatial planning algorithm developed by ICCS (described in Section 4), with sampling points depicted in Figure 14.

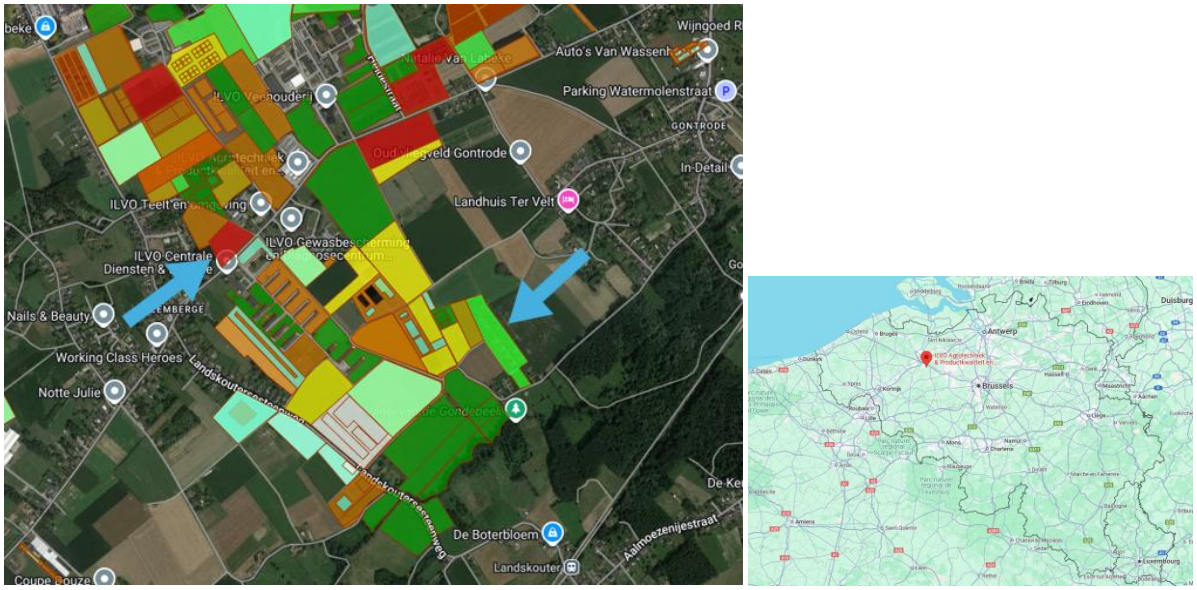

Figure 13. Locations of two EV ILVO fields used during first field trials

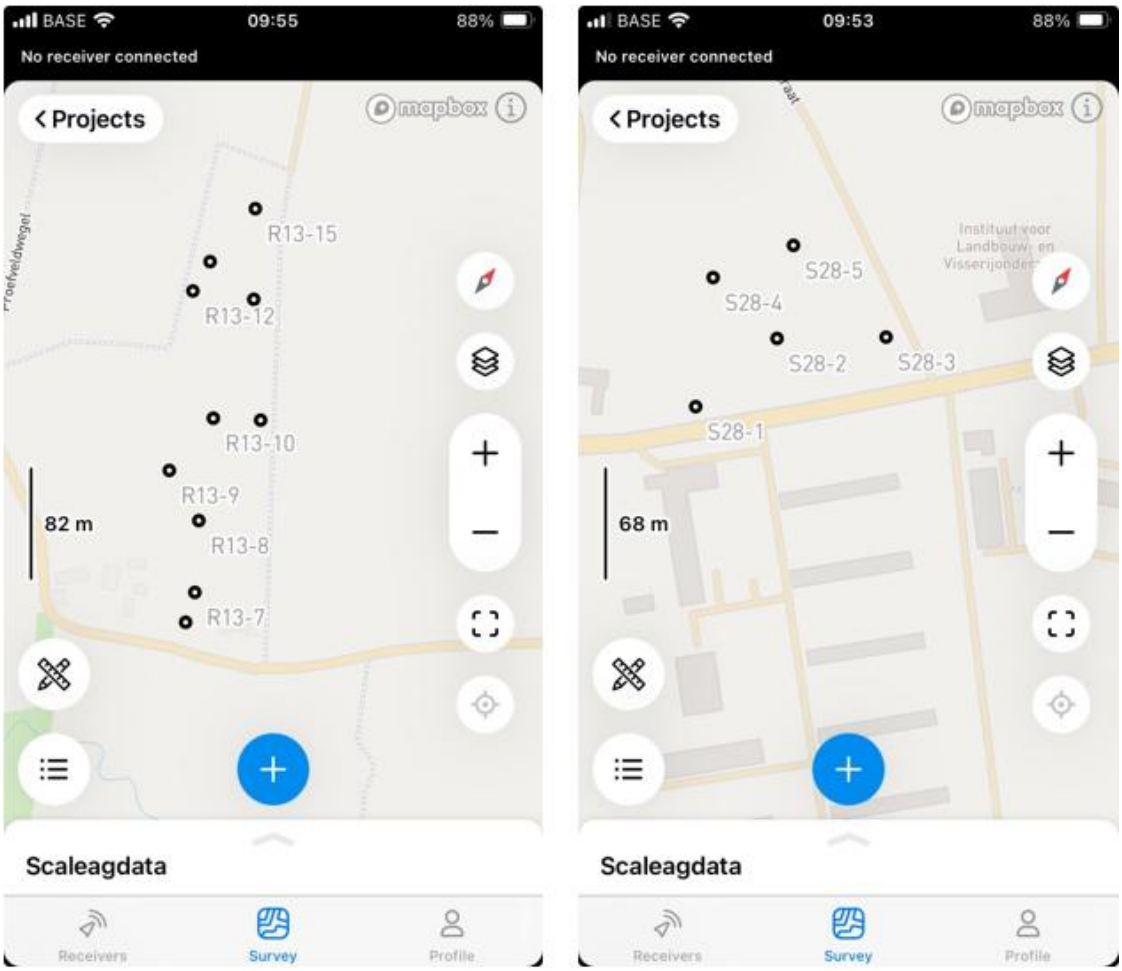

Figure 14. Soil sampling location determined by using the spatial sensor planning algorithm of ICCS

After the soil campaign and UAV flights with the VTT camera, the following workflow was followed to generate field SOC maps. Image Extraction and Stitching: Raw images were extracted from drone memory and processed using OpenDroneMap (ODM) software to create georeferenced mosaics, with GPS coordinates manually deduced from flight path data. A stitched image of one hyperspectral band from the first field is presented in Figure 15.

Band Registration and Alignment: Individual spectral bands were geometrically corrected using georeferenced ground control points to ensure precise spatial alignment, with all bands co-registered and stacked into a single multi-band image cube.

Band Reordering: The spectral bands were reorganized into the proper sequence to maintain consistency across the multispectral dataset for subsequent analysis.

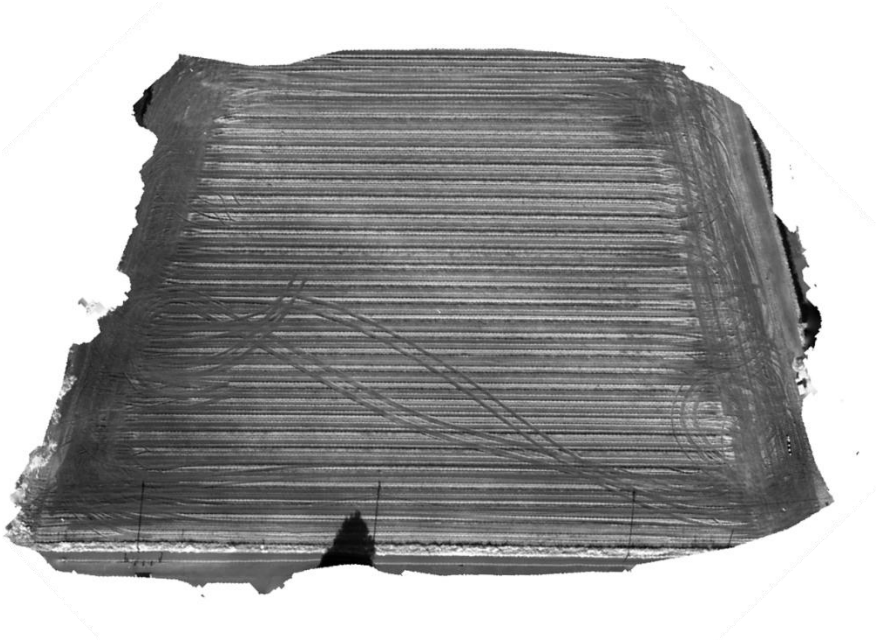

Figure 15. Stitched image of first field for one of the hyperspectral bands of VTT HSI sensor

Radiometric Calibration: Raw digital numbers from multispectral images were converted to reflectance values using reference panels with known reflectance (50%, 80%, 95%) through linear regression calibration models applied band-by-band.

Ground Truth Collection: Soil samples with measured organic carbon content were spatially matched to their corresponding pixel locations in the calibrated reflectance images to create training datasets.

CNN Model Training: A 1D convolutional neural network (Figure 16) was trained using 6-band reflectance values as input features and measured OC values as targets, with the model learning spectral-chemical relationships from the limited sample points. The trained DL SOC model achieved a mean square error of 0.8954 and an R^2 of -42.2772. While these performance metrics are clearly inadequate (Figure 17), they reflect the limited number of soil samples available for training. This issue will be addressed in the second iteration through expanded soil sampling campaigns planned for autumn and winter 2025.

Layer (type)	Output Shape	Param #
conv1d_2 (Conv1D)	(None, 6, 64)	512
leaky_re_lu_5 (LeakyReLU)	(None, 6, 64)	0
max_pooling1d_2 (MaxPooling1D)	(None, 3, 64)	0
conv1d_3 (Conv1D)	(None, 3, 32)	14,368
leaky_re_lu_6 (LeakyReLU)	(None, 3, 32)	0
max_pooling1d_3 (MaxPooling1D)	(None, 1, 32)	0
flatten_1 (Flatten)	(None, 32)	0
dense_4 (Dense)	(None, 512)	16,896
leaky_re_lu_7 (LeakyReLU)	(None, 512)	0
dense_5 (Dense)	(None, 256)	131,328
leaky_re_lu_8 (LeakyReLU)	(None, 256)	0
dense_6 (Dense)	(None, 64)	16,448
leaky_re_lu_9 (LeakyReLU)	(None, 64)	0
dense_7 (Dense)	(None, 1)	65

Figure 16. Architecture of the deep learning model used to make SOC predictions at field scale

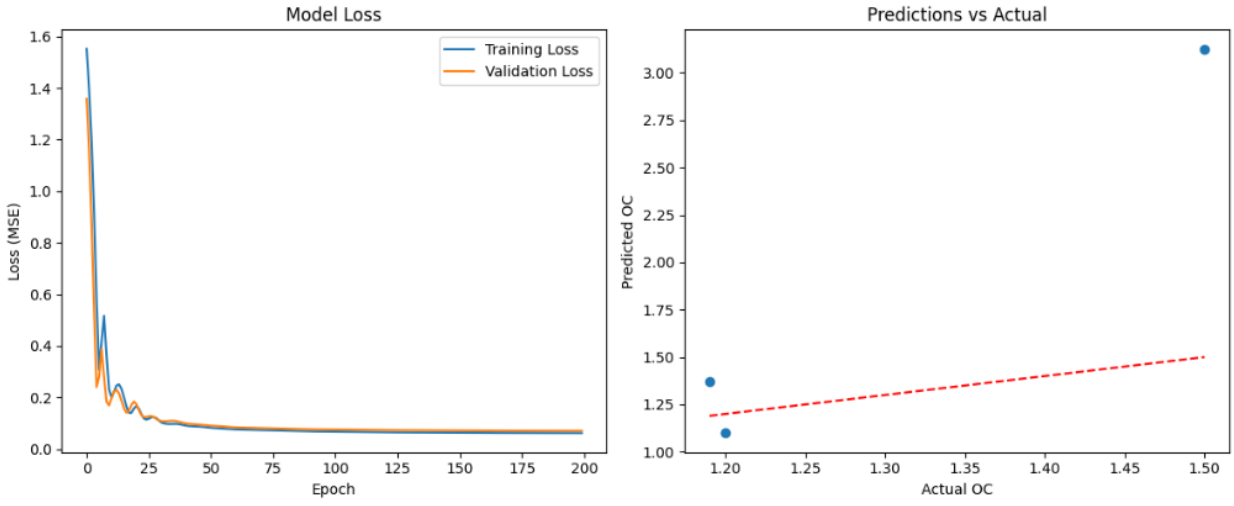

Figure 17. Graphical visualization of loss during training of the model and prediction accuracy on the testing dataset during training.

Spatial Prediction: The trained CNN model was applied pixel-by-pixel across entire calibrated images to generate spatially continuous maps of predicted organic carbon content for the surveyed fields (Figure 18).

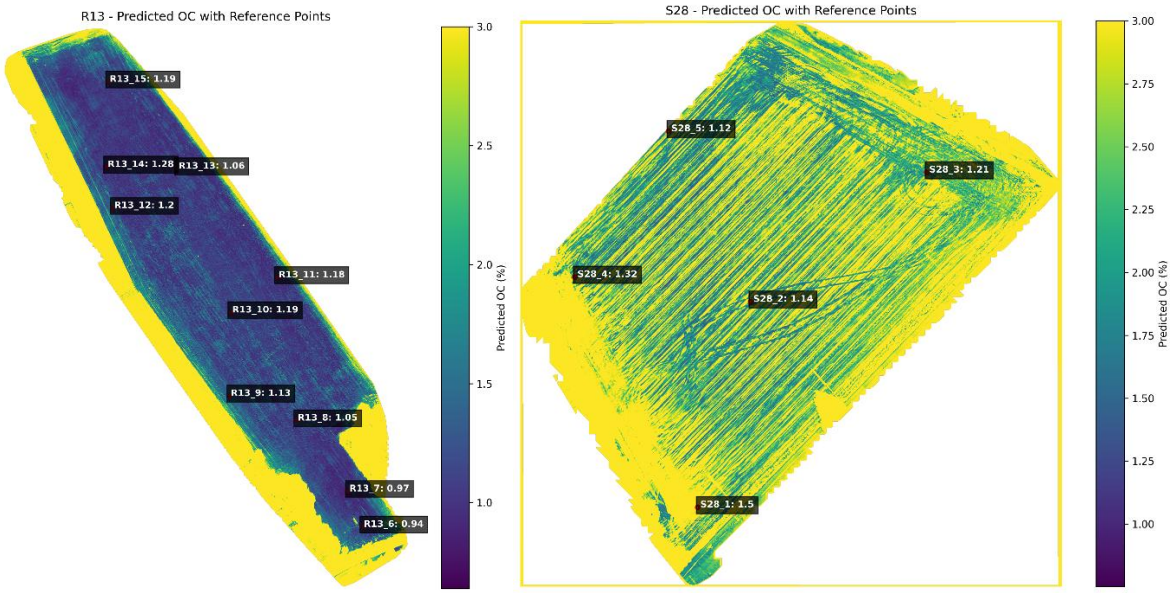

Figure 18. Predicted SOC values by trained DL SOC model for 2 EV ILVO fields tested during first soil campaign in 2024

In general, the first field trials went very well with the VTT sensor. The following minor comments were provided to VTT:

- The flight was conducted at a higher altitude to accommodate the limited field of view, allowing the entire field to be covered with one battery set.
- The pressure exceeded the 60 mBar threshold during the first flight at second field.
- The frames per second fluctuated significantly during the flights at the second field (a long, narrow field), with fps being lowest at the farthest point and returning to normal (6-7 fps) when close to the ground station.
- Autofocus functionality was not always working.
- Linking GPS tracking to camera images would improve image stitching during data processing.

Following the handover of the VTT HSI sensor in mid-November 2024, AUTH conducted its first data acquisition trial in the vineyard Ktima Gerovassiliou in Thessaloniki Central Macedonia, Greece, on January 23, 2025.

The sensor was mounted onto AUTH's custom-built drone platform. Due to weather conditions the UAV flight was conducted in manual mode, and no detailed flight logs were recorded. Since the VTT HSI sensor does not include an onboard GPS module, no GPS information was embedded in the captured data by default. As a result, GPS coordinates were manually interpolated to enable image geotagging.

While a full ortho-mosaic or machine learning modeling was not performed in this phase, a straight line of geotagged images was stitched together to validate the data processing workflow (Figure 19) following the workflow previously tested by EV ILVO.

In parallel, field spectra were collected using the PSR+3500 spectrometer, covering all Sentinel-2 pixel centers across the vineyard (Figure 20). A number of soil samples were also taken for lab analysis. The combination of spectra and soil measurements will support future modeling tasks and contribute to calibrating the ICCS spatial planning algorithm.

This trial served primarily as a technical validation of our HSI data acquisition and preparation workflow and will inform larger-scale efforts in future campaigns. Based on our experience with this first deployment, technical feedback was provided to VTT regarding sensor performance and integration.

Figure 19. Stitched straight line of geotagged hyperspectral images acquired with the VTT sensor at Ktima Gerovassiliou.

Figure 20. Distribution of PSR+3500 field spectra acquisition points across the vineyard. Sampling locations correspond to Sentinel-2 pixel centers and were complemented by soil sample collection for subsequent laboratory analysis.

Both EV ILVO and AUTH will conduct similar trials in the second phase 2025 Q3 – 2026 Q2, but on a larger scale (20 ha at EV ILVO and 30 ha at AUTH), incorporating the findings from the first trials. EV ILVO received the upgraded VTT camera on 12/06/2025 in Flanders and has started preliminary experiments in view of the setup and application of the sensor with the UAV.

We will investigate the feasibility of combining hyperspectral imagery from the KUVA HSI satellite with VTT drone-based HSI sensor data to produce optimal field-level SOC estimates. This data fusion approach requires spatial overlap (same geographical areas) and temporal overlap (synchronized or closely timed data collection) between the satellite and drone datasets.

Regarding the KUVA HSI satellite, the Hyperfield-1 satellite was successfully launched via SpaceX on 16th August 2024, the second satellite Hyperfield-1B was successfully launched on June 2025. Hyperfield-1B is functioning as expected, but extensive testing, calibration, and validation remain ongoing. The commissioning phase continued until the beginning of 2025. Meetings were held in March 2025 between KUVA, EV ILVO and AUTH to discuss the collection and access of HSI images of large regions in Flanders and Central Macedonia during 2025 Q2-Q4. Access to the data via the platform KUVA Sense is planned for September 2025.

2.5 RIL Grasslands

Monitoring key biophysical parameters and exchange processes in grasslands allows for a comprehensive assessment of grassland ecosystem health. Changes in LAI, PAR, biomass, chlorophyll content, soil moisture, and CO2, energy and water fluxes can indicate stress, environmental changes, or the impact of management practices. The data obtained from these sensors can inform land managers and researchers about the need for adjustments in land management practices, such as altering grazing patterns, adjusting irrigation schedules, or addressing nutrient deficiencies. In the study areas of the RIL Grasslands the focus was on the following parameters (see Table 1):

Table 1. The key biophysical parameters selected for RIL Grasslands

Parameter	Importance	Use
Leaf Area Index (LAI)	LAI is significant for understanding the density and vigor of vegetation.	LAI helps estimate the amount of solar radiation intercepted by the vegetation, which is critical for understanding photosynthesis, evapotranspiration, and overall plant health.
Photosynthetically Active Radiation (PAR)	PAR represents the portion of sunlight that plants use for photosynthesis. It is crucial for understanding the energy available for plant growth.	It helps assess the efficiency of light captured by the vegetation, which, in turn, relates to plant growth, productivity, and ecosystem health.
Chlorophyll Content (CC)	Chlorophyll is a pigment crucial for photosynthesis, and its content is indicative of the plant's ability to convert light into energy.	Provides information about the physiological status of plants. It helps assess plant stress, nutrient deficiencies, and overall plant health.
Soil Moisture (SM)	SM is critical for plant growth as it directly affects water availability to roots.	It enables efficient water use and prevents over-irrigation or drought stress.
Biomass (B)	Essential for understanding plant growing patterns and productivity.	Provides information about the growth of the vegetation and the aboveground Net Primary Production (NPP)
CO2 flux (CO2)	Important for understanding plant growth patterns and productivity	Provides information about the growth of the vegetation and the Gross Primary Production (GPP)
Energy fluxes: latent heat (LE), sensible heat (H), net radiation (Rn) and soil heat flux (G)	It is required to assess water use and system functioning	It enables the measurement of the energy balance of the surface and plant water consumption
Meteorological data (Rad, Ta, u, DPV, P)	Required to evaluate the influence of climate on plant growth and water status	Data input to growth and water models

2.5.1 Sensor Selection

The planned sensors are shown in Table 2.

Table 2. The sensors to be used in RIL Grasslands

Sensor	Producer/Model	Location	Sampling freq.	Repository
LAI	Licor, LAI 2200C	South Tyrol, Trento, Italy	Bi-weekly	DOI
PAR	Apogee, MQ-301X	South Tyrol, Trento, Italy	Bi-weekly	DOI
CC	Konica 502, SPAD 502	South Tyrol, Trento, Italy	Bi-weekly	DOI
SM	Campell, HydroSense II	South Tyrol, Trento, Italy	Bi-weekly	DOI
LAI	Licor, LAI 2200C	Cordoba, Spain	Monthly	DOI
fPAR	Accupar LP-80 Ceptometer	Cordoba, Spain	Monthly	DOI
B	Grasmasster Pro, Jenquip EC20 Platemeter	Cordoba, Spain	Monthly	DOI
CO2, H, LE	*Gas analyzers:* (1) LICOR 7500, (1) LI-7500DS, (2) Cambell IRGASON *3D sonic anemometer:* (1) Campbell CSAT, (1) Gill Windmaster	Cordoba, Spain	Continuous (half hour)	DOI
Rn, G	Rn: (1) Kipp&Zonen CNR4, (3) Campbell NR01 G: Hukseflux HFP01 (6 measurement point), (6) TCAV	Cordoba, Spain	Continuous (half hour)	DOI
SM	(7 probes, 3 depths) Sentek Environscan (3) Stevens Hydraprobe	Cordoba, Spain	Continuous (half hour)	DOI
Rad, Ta, u, DVP, P	(1) Campbell LP02, (3) Vaisala HMP155, (1) Vaisala45A, (1) Campbell, TR525, (2) Campbell ARG100	Cordoba, Spain	Continuous (half hour)	DOI

2.5.2 Data Acquisition Plan

2.5.2.1 General information

The data is acquired in two study areas located in Italy and Spain. Eight different alpine meadows have been selected in the provinces of Trentino and South Tyrol, North Italy. Ten Mediterranean pasture fields, corresponding to five oak savanna farms, were chosen in Córdoba, southern Spain (see Figure 21).

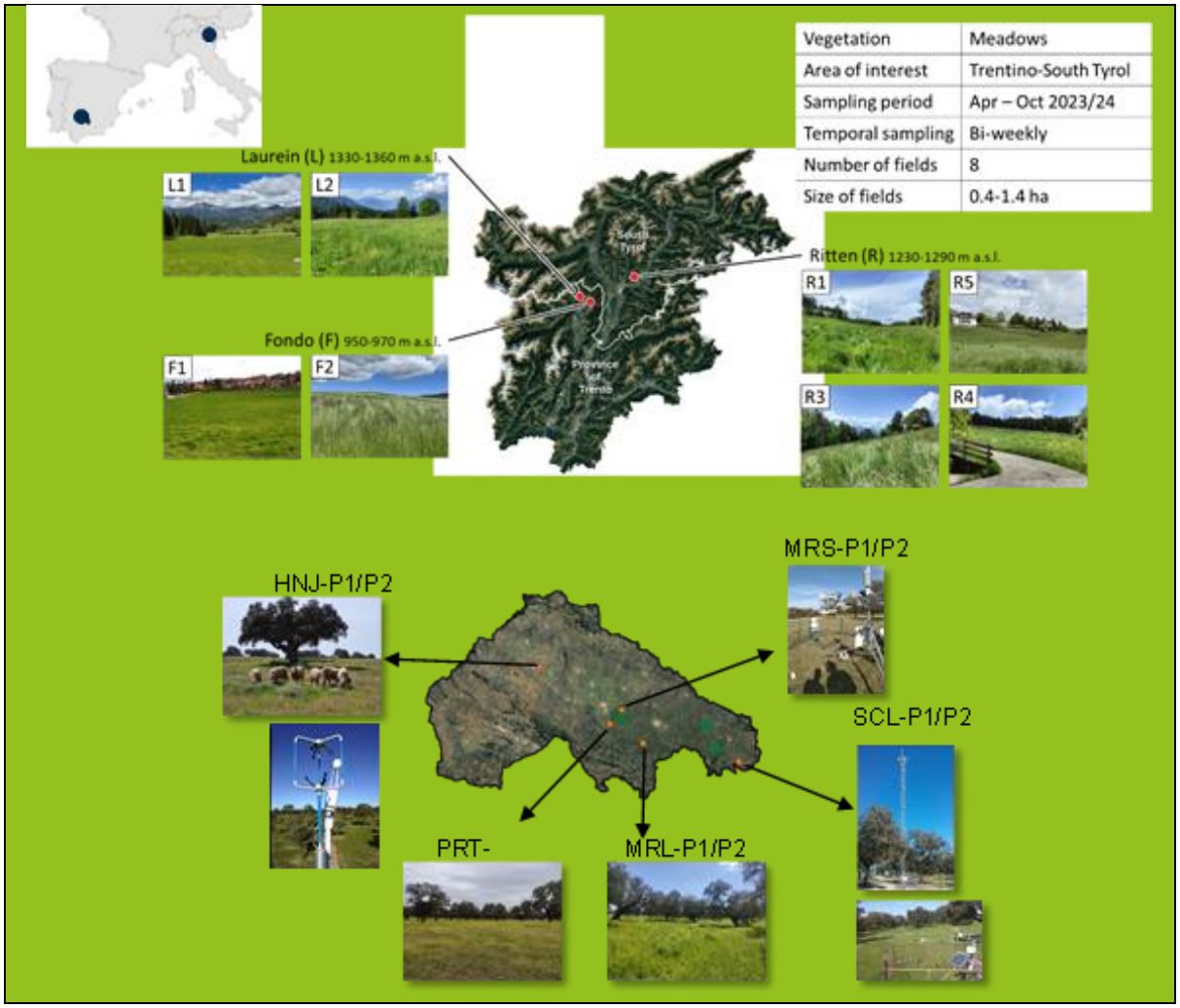

Figure 21. RIL Grasslands test sites

2.5.2.2 Schedule

The sampling frequency in Italy is every two weeks. Taking measurements over different pixel-distributed areas that correspond to the spatial resolution of Sentinel 2, to later assimilate such information (Figure 22). Apart from data collected from sensors, the height and composition are registered using a quadrat technique, reporting conditions (e.g., lodging, fertilization) and collecting biomass. A similar procedure is followed in Spain with a monthly frequency. Information about grazing activities is also collected (Figure 23).

Figure 22. Two weekly strategy and Data collected in Italy

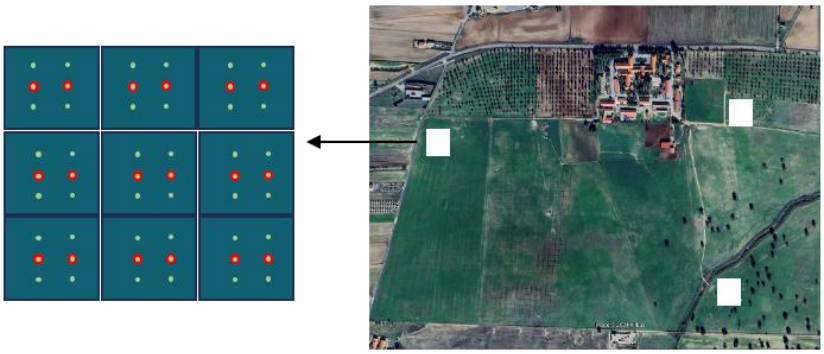

Figure 23. Sampling strategy in Spain

2.5.2.3 Resources

Instrumentation is described in Table 2. Two trained people in the use of instruments and field data collection.

2.5.2.4 Special data constraints

If the day(s) before the data campaign was rainy, the sampling is shifted for a few days to avoid water saturation in biomass and SM. In Spain, the duration and calendar of the growing cycle vary each year depending on climatic conditions, mainly rainfall distribution. Production can be partly or entirely lost during drought events (e.g., e.g. 2023). In those cases, the data collection period must be extended (for example, adding 2025 to complete two seasons).

2.5.3 Summary of the first 30 months of the project / RIL Grasslands

According to the planned schedule, in-situ data collection of Alpine grasslands' biophysical variables was completed by EURAC during the first iteration. Some results from the 2024 field campaign are presented in *Figure 24*. In Spain, because of the drought conditions during the first year of the project, data collection of environmental variables over the Mediterranean grassland

was extended to M29 to complete the monitoring of two full growing seasons. In addition, carbon fluxes have been continuously measured since the project began, with a new flux tower added as of M13.

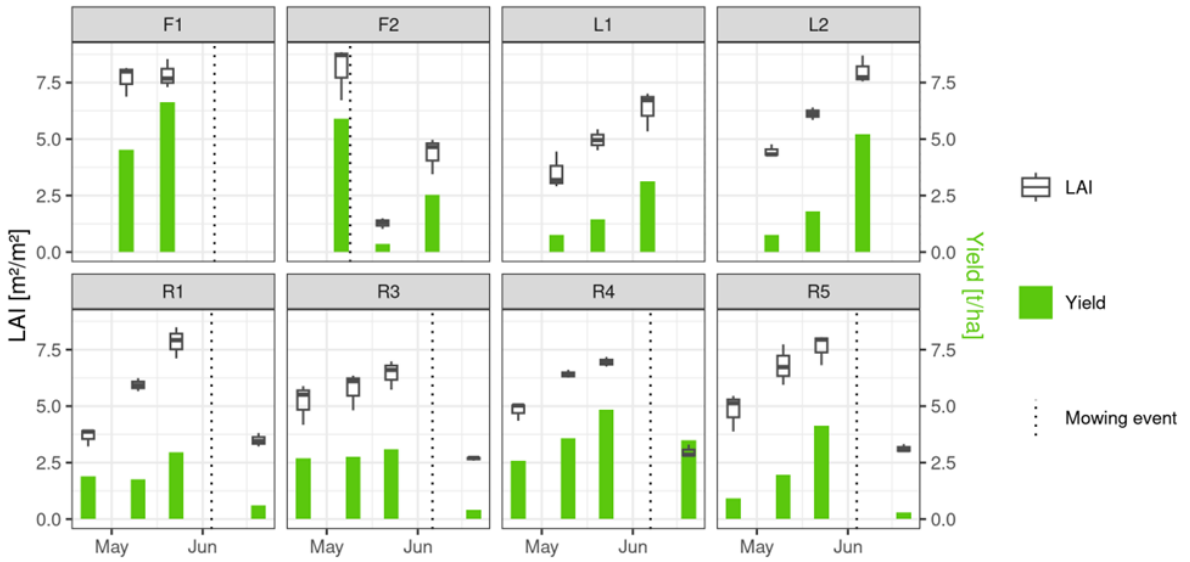

Figure 24. 2024 measurements data of grassland LAI and yield

Regarding the estimation of LAI at the pixel and parcel scales in Italy, integrating Sentinel-1-derived data, the first year of in situ data was used to test a fusion algorithm integrating Sentinel-1-derived data, following the workflow presented in Figure 25. However, extrapolating to the second season was not entirely successful. Therefore, during the second iteration, the entire dataset will be used to validate ML estimation methods at the pixel level and to conduct a sensitivity analysis of Sentinel-1 data with ground-measured variables at the parcel scale. Proper planning of the acquisition campaigns, including specific protocols, enabled the use of the collected data for these additional analyses, which will be carried out during the second iteration.

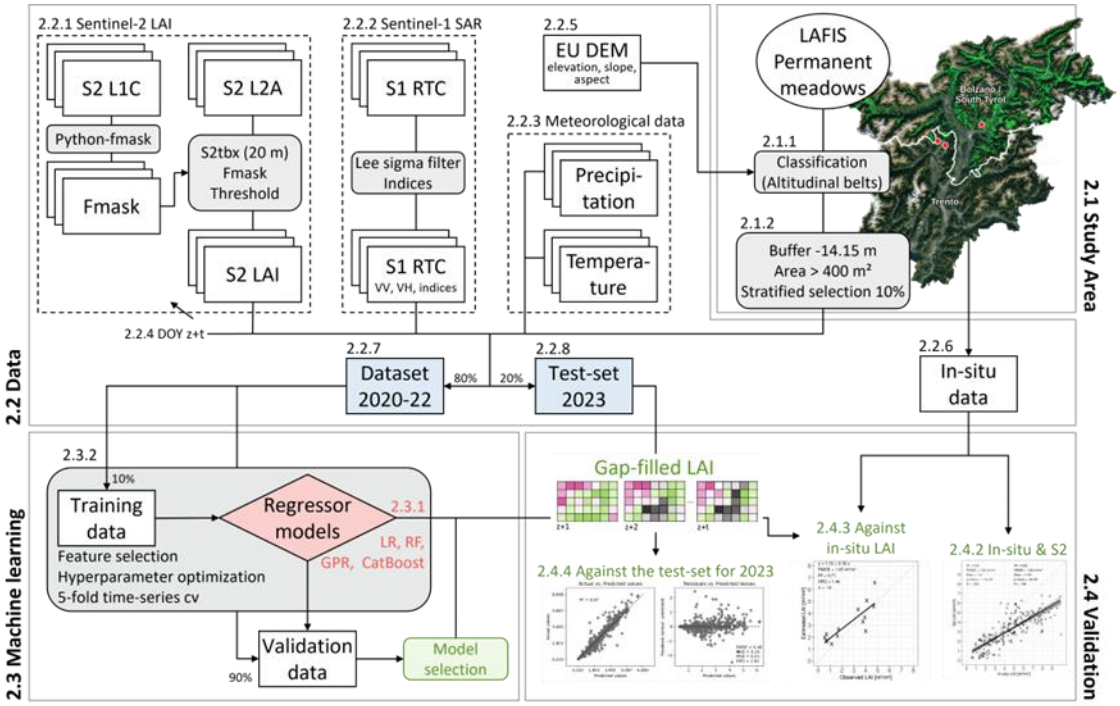

Figure 25. Workflow to model LAI of Alpine grassland using spatial gap-filling

The approach used by IFAPA to estimate grassland biomass in Spain, which adapts a Light Use Efficiency (LUE) model to Sentinel-2 inputs and the structure of Mediterranean oak savanna, was tested during the first iteration with field data collected during the initial complete growing season. Figure 26 shows the difference in net primary production (NPP) between the first two years of the project at one of the pilot farms. An issue with LAI measurements was identified after the first season. At the start of the season, the vegetation's size and shape (very short but with high ground coverage) made it challenging to obtain accurate data, similar to satellite estimates using the initially planned sensor, Licor-LAI 2200C. Therefore, during the second growing season, a sample of green leaves was cut to estimate biomass and scanned to calculate total leaf area and LAI. This highlights the challenge of applying standard measurement protocols to diverse types of vegetation.

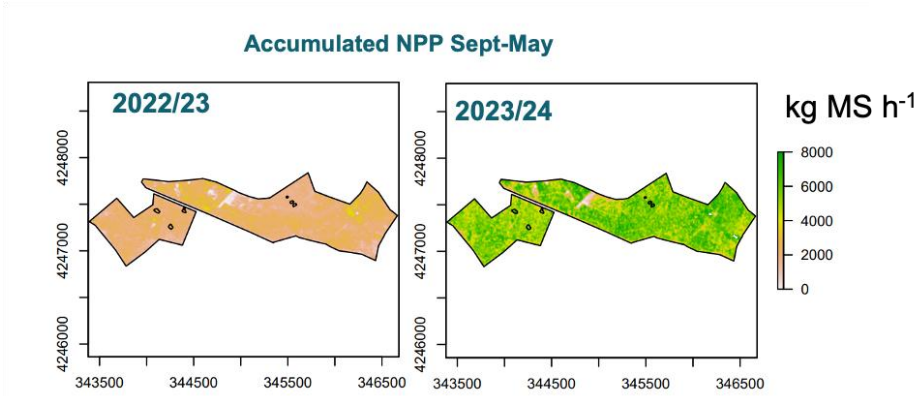

Figure 26. Estimated NPP for the first two seasons in one of the RIL pilot farms.

Deimos (Indra) and IFAPA collaborated on applying ML techniques to estimate grassland gross primary production (GPP) using data collected in situ by eddy covariance flux towers and Sentinel-2 and Sentinel-1 images. The results from the first iteration, though encouraging, relied on a database with two limitations: 1) a small number of samples and 2) an uneven distribution of samples throughout the growing cycle. This database is therefore preliminary, mainly serving as a test for the model architecture and code function. In the second iteration, a larger database of flux measurements is being used, and Sentinel-1 data will be employed during cloudy periods to better capture the growing season.

2.6 RIL Dairy

2.6.1 Sensor Selection

The main objective of the dairy lab is to investigate potential for improving the forecasting of data concerning milk quality and quantity, which is relevant for the production facilities of the dairy cooperative involved. If this is feasible, there might be future potential to provide additional feedback for farmers.

The lab aimed at combining the use of the following sensors/ data sources:

- Sentinel Data 10m archive
- Milk samples and subsequent laboratory devices/analysis
- Historical weather data and weather forecast data
- Sensors installed in forage harvesters

The Sentinel data was considered as a basis to analyze the NDVI. Further EO-data-based services were used to provide information about the available biomass. The combination of in-situ sensor data with EO data was analyzed, also deriving further requirements for the refinement of the sensor selection. It was further analyzed how to make use of meteorological data in relation to the regions addressed as well as how to integrate this into the analysis of milk quality and quantity.

Since the dairy cooperative aims at covering a large land area, due to their 20 production facilities and 4,700 milk producers, the aim is to combine in-situ data with EO data, where the in-situ data will serve for a calibration of EO data-based analysis of milk quality and quantities.

The list of sensors could grow as the project progresses, including operational satellites or institutional data (e.g., on the number of cows, or the performance of cows) where the data products contribute to the correlation analysis.

2.6.2 Data Acquisition Plan

2.6.2.1 General information

All milk that is delivered to the dairy cooperative is analyzed by taking samples from each milk delivery of each farmer. This is generally done for product safety as well as for determining the specific milk parameters that are also defining the revenue of farmers. The milk samples are taken at the farm when the milk is put into the milk transporter. Specific laboratories are used to perform the related analysis. Therefore, several milk samples per week are taken for each of the 4,700 farmers in the overall area covered by the dairy cooperative in Germany, Italy and the Netherlands.

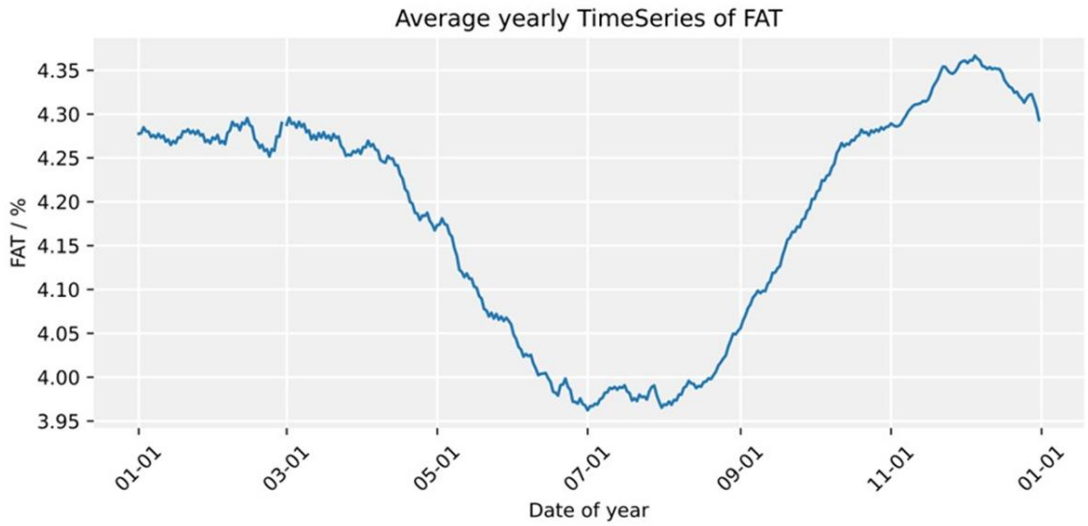

Figure 27. Typical variation of milk characteristics over the year (example of time series for fat).

In the first phase of the project, the dairy lab was investigating the available EO data that could be used to understand the development of the vegetation in the area of interest (using e.g., NDVI and NDMI) over the year. This work also analyzed the amount of data available (i.e., revisit frequency and resolution).

The in-situ data measured during harvest on the field by forage harvesters shall provide exact details about the e.g., harvested grass or maize. It was aimed at measuring the yield per hectare and moisture in dry matter. This shall be used to provide an overview of the amount and quality of available feed over time in specific geographical areas.

In a later phase after the project, it might also be analyzed if there are possibilities to measure amounts of supplementary feeding on a large scale, without burdening farmers, considering silage and concentrated feed or to focus on individual farms and farm management information system related data.

2.6.2.2 Schedule

In the first two years of the project, the aim was to use historic data available for the last 5 years to analyze potential correlations of EO data and in-situ data. The correlation analysis provided initial positive results. This is the baseline for analyzing further details and searching for optimization potentials in cooperation with the milk quantity and milk quality forecasting department as well as the production facilities in the second two-year cycle of the project.

2.6.2.3 Resources

The basic EO data analysis is based on available Sentinel data that can be accessed for the analyzed regions for free. However, a more detailed and fine-grained analysis might require the usage of paid services to facilitate real-time access to required EO data.

Based on collaboration with the project partners, there is also the offering by VITO with respect to satellite data processing, considering VITO's satellite-data-based vegetation and other indices.

The in-situ data was collected from forage harvesters that are widely used by most farmers in the area concerned and aggregated accordingly to comply with privacy/confidentiality matters. In general, the required data is gathered by the telemetry of the forage harvesters and collected by the OEM in a proprietary cloud-based system. This required related efforts to access related data, as well as to handle the data in accordance with the EU code of conduct on agricultural data sharing by contractual agreement.

2.6.2.4 Special Data Constraints

The satellite data was not available in the optimal frequency due to cloud cover. At the same time, due to specific limitations of data protection, there is no exhaustive list of farmers' field parcels that could serve for a one-to-one analysis of the fields based on EO data. Hence, the analysis was based on average results for each region.

The milk quality and quantity data are available in high detail at the dairy cooperative, covering every sample of every farmer for the related delivery days. However, a related analysis needs to consider requirements for data protection by aggregating data. It was analyzed how to balance the required level of detail and the required anonymization of data.

The data of forage harvesters was aggregated to enable the anonymization of data since they were provided by the OEM. This required further work to understand the needed level of aggregation of data as well as the level of detail that is required to enable a meaningful analysis. The lab started to analyze conditions for experimental use for research purposes as well as future potential for commercial exploitation. The analysis of the vegetation was also aligned with the related cropping seasons accordingly.

2.6.3 Summary of the first 30 months of the project / RIL Dairy

The RIL Dairy incorporates a large amount of in-situ data, generated by the laboratory analysis of the milk samples that are taken for each of the delivery events of every farmer, usually every other day. This in-situ data is available to the dairy cooperative as well as to the farmers to analyze and control the process at the farms as well as for milk processing. The current usage of that data is still limited, the data storage is rather focused on access for small amounts of data sets, not designed to scale for large and repeated amounts of data use by diverse stakeholders. At the same time, it is currently cumbersome to prepare an estimate of milk-related data. Hence, the data is rather used for administrative purposes by the cooperative and can be used as input for feeding control at the farm side. The usage of additional data, like from Sentinel, forage harvesters and weather data, focused on facilitating the correlation analysis of the farm context with milk quality and quantity.
The Dairy RIL was investigating different directions, aiming to identify correlations by combining different data sets that were not yet considered. Therefore, the team limited the work on general features of a potential ICT-based system but focused on the possibilities to realize the correlation analysis in a joint team effort. Therefore, the first phase included several iterations to analyze the different data sources and how to design a related approach for exchanging data and results. Also, the data quality was analyzed, showing that each data source requires a careful analysis in terms of data quality and timely availability.

The second iteration in the ScaleAgData project will focus on both the validation of the initial correlation analysis and the investigation of the usage of different types of context data. Moreover, potential overall system designs will be elaborated and analyzed in the scope of the different stakeholders involved and facilitating data access.

2.7 RIL Crop management – Sub lab Agri-environmental monitoring for Policy Makers

2.7.1 Sensor Selection

This sub-lab focuses on integrating in-situ sensor data with Earth Observation (EO) data and farm log records to support regional agri-environmental policy monitoring.

In addition to field-deployed IoT weather stations and a nanoparticle-based pesticide sensor, the lab uses Sentinel-2 satellite imagery processed through the Sentinel Hub platform. The following types of data sources are utilized:

A. In-situ Sensors

The sensors that will be utilized are IoT weather stations that provide measurements on an hourly basis regarding the following parameters listed in Table 3.

Table 3. The parameters measured in RIL Crop management (Sub-lab NP)

Parameter Group	Measurement Parameters	Units
Gaiatron		
Atmospheric	Air temperature, humidity, wind, solar radiation, pressure	°C, %, km/h, W/m², hPa
Soil	Moisture, temperature, salinity at 10–70 cm depths	%, °C, dS/m
Vegetation	Leaf wetness, foliage temp/humidity	—, °C, %
Pesticide Sensor		
Pesticide applications	Detects and classifies spray events	[no spray, water, pesticide]

B. Earth Observation (EO) Data: Sentinel-2

- **Provider**: Copernicus Data Space via Sentinel Hub
- **Indices computed**:
  - NDVI (pending: NDRE, LAI, NDWI, PSRI)
- **Processing method**:
  - Evalscript-based workflows define band selection, masking, and index generation
  - Quality filtering: Cloud masking, SCL filtering (e.g. , water), and dataMask validation
  - Output: Valid, cloud-free daily time series at 10–20m spatial resolution

C. Crop Classification Model

- **Algorithm**: Support Vector Machine (SVM) with RBF kernel
- **Training data**: Parcel-level polygons with crop labels and season dates
- **EO inputs**: Sentinel-2 bands B2–B6 (Blue, Green, Red, Red Edge)

Purpose: Automate crop-type mapping for indicator inference

2.7.2 Data Acquisition

2.7.2.1 General Information

- 8 IoT sensors were deployed in 2023 in pilot areas in Thessaly and Crete for tomato, cotton, and potato monitoring. In 2024, additional IoT sensor data from wheat and peach crops in Central Macedonia were incorporated into monitoring activities. In total, 18 pilot stations were used to gather data for years 2022-2024.

- Sentinel-2 EO imagery is continuously collected via Sentinel Hub using custom Evalscripts.
- Farm record (irrigation and pesticide spraying activities) integration through the GaiaSense API, harmonized via AIM-based semantic translator.
- A custom-developed pesticide sensor based on a nanoparticle gas-sensing array is used to validate pesticide spraying events in-field. This edge-deployable sensor classifies spray events (no spray, water spray, pesticide spray) using a residual-based classification algorithm and is integrated via a RESTful API.

2.7.2.2 Schedule

- **Field sensor operation**: For annual crops (e.g., tomato, cotton, potato, and wheat), sensors are active during their respective growing seasons (tomato, cotton, and potato: approximately March–October; wheat: September–June). For perennial peach orchards, sensors remain operational throughout the entire year to capture year-round environmental dynamics.
- **EO data ingestion**: Continuous, with daily revisit aggregation (P1D) throughout the season.
- **SVM crop model**: Trained on 2023 data, applied across new parcels for the 2024 iteration.

2.7.2.3 Resources

- EO data: Sentinel-2 Level-2A surface reflectance
- Sensor data processing: Edge filtering, PostgreSQL storage, semantic translation (AIM)
  - Integration: Geolocation API for LAU/NUTS mapping; REST API for access to calculated indicators

2.7.2.4 Special Data Constraints

- EO data pre-filtered to ensure cloud-free, non-water observations.
- Regular field inspections are needed to safeguard sensor function and calibration (especially during extreme weather).
- SVM classifier must be retrained as needed if land use or EO quality changes.
- A key constraint was the inconsistency and fragmentation of ground-level (field) data, especially when attempting to aggregate to LAU or regional levels:
  - In several cases, available parcels were scattered across multiple LAU regions or located far apart, reducing the statistical coherence of any aggregated result.
  - For some indicators, historical data existed only for specific years or crops, making cross-year comparisons or multi-parcel aggregations unreliable.
  - This heterogeneity constrains the model's ability to compute accurate performance indicators uniformly across administrative units.

2.7.3 Summary of the first 30 months of the project / RIL Crop Management - Sub lab Agri-environmental monitoring for Policy Makers

2.7.3.1 Lessons Learned

- EO-based indices (NDVI, NDWI, NDRE) proved essential in bridging data gaps in farm records and in scaling insights from sensor-monitored fields to the regional level.
- However, aggregating field-level data into coherent LAU or regional metrics proved more difficult than expected. Ground-truth data was often fragmented spatially, with pilot parcels spread across administrative borders or located in isolated clusters. Furthermore, temporal inconsistencies (e.g., gaps in field data between years) also reduced the feasibility of creating continuous or representative datasets at a regional scale.
- This limitation highlighted the need for either (a) denser and more evenly distributed ground data or (b) improved data assimilation methods that weigh EO observations more heavily when parcel coverage is sparse.

- Despite these challenges, the modular architecture and semantic pipelines ensured that even fragmented datasets could be partially integrated for local-scale decision support.

2.7.3.2 Planning vs. Execution

- Originally planned parcels were substituted in 2024 due to crop rotation practices, but data continuity was ensured by selecting similar agronomic profiles.
- Sensor and EO data acquisition aligned with the plan. Parcel rotation required updated sensor deployment in 2024.

2.7.3.3 Actions for Final Stage

- Update SVM model with 2024 field data and expand EO-based phenology tracking (calculation for 2023 almost done, calculation for 2024 pending, integration in the dashboard pending).
- Finalize API tools to deliver regionalized indicators (e.g., irrigation use, pesticide intensity) with sensor-enhanced validation.
- Quality Control and ingest 2025 data, then publish to the dashboard.
- Ship Region+Crop UX and integrate crop-classification layers; do a quick user walkthrough and capture feedback.
- Finalize aggregation methods for field data scalability to regional level (L3 - 2nd Iteration)
- Integrate ESA Worldcover data to enhance the classification results.
- Pesticide data for one season have been integrated but a second-year deployment is pending: Install the pesticide sensor in a second tomato field to aid validation of farm-recorded pesticide use and integrate it in the dashboard through a REST API.
- Finalize dashboard UI and functionalities (continuous refinements)

2.7.3.4 Other Notes

- EO-driven models enabled regional upscaling of crop monitoring even in areas lacking sensor coverage.
- The multi-source data fusion pipeline *EO + IoT + farm logs* supports the Agro-environmental Policy Indicator Monitoring Tool, currently at TRL 5–6.

2.8 RIL Crop management – Sub lab Sustainability performance

2.8.1 Sensor Selection

The need of this RI Sub lab is to have weather data that can feed models and rules which are present in the various functionalities of the Horta's Decision Support System grano.net®. This Decision Support System (DSS) is an online tool that supports farmers in the sustainable management of their wheat crop, providing advice on the main crop operations to be performed in the field (i.e., sowing, fertilization, crop protection interventions, yield forecasts, etc.). Several kinds of inputs are used by models and algorithms in the DSS, such as weather data, soil data, crop operations data, and remote sensing data. Weather data are collected by means of in-situ weather stations, which have sensors to register air temperature, rain amount, relative humidity, and leaf wetness.

2.8.2 Data Acquisition Plan

2.8.2.1 General information

Horta will monitor 10 wheat fields in the north of Italy also in the 2024-2025 cropping season. Each field will be associated with a representative weather station, which will collect data on temperature, relative humidity, and leaf wetness. A total of 10 weather stations will then be used in the RI Sub Lab (see Figure 28).

Figure 28. Location of weather stations in the North of Italy monitored in the RIL Crop management (Sub Lab Sustainability performance)

2.8.2.2 Schedule

Weather stations considered in this RI Sub lab are already installed. The wheat cropping season in northern Italy starts in October and ends in July next year. The data relevant for the crop will then be collected in these months. The first season considered was the 2023-24 cropping season, the second one was the 2024-2025 cropping season.

2.8.2.3 Resources

The data flow for data collection from the weather station is automatized and is up and running. Resources are needed to maintain the data flow, and to fix eventual problems (sensor failures and anomalous data are automatically corrected).

2.8.2.4 Special data constraints

Additional data expected to be used in this RI Sub-lab are Sentinel-2 data, which are being retrieved from a commercial provider. In addition to this, this RI Sub lab has received data from VITO and DHI , which is expected to be further discussed and needs to be further evaluated.

2.8.3 Summary of the first 30 months of the project / RIL Crop management – Sub lab Sustainability performance

The retrieval of weather data from *in situ* sensors was automatized and included data quality control. This was an essential aspect, to check for proper sensor functioning and timely intervention in case of anomalies in the sensors functioning. Actual sensor data retrieval was performed as expected, with no major problems or deviations recorded. The data acquisition will continue in the second iteration in the same area identified in the first one.

2.9 RIL Crop management – Sub lab Early Pest Detection

2.9.1 Sensor Selection

The system for organizing agrophage signaling in public agricultural advisory centers working within the eDWIN Advisory Platform would be expanded to include algorithms for optimizing data collection. Various data sets will be used, based on analysis of which eDWIN will suggest to observers and those organizing the observation system the optimal observation dates for specific agrophages. The result of the introduced changes will be the carrying out of observations at times of increased risk of occurrence of given agrophages in a given area.
Data from the following sources including sensors will be analyzed:

- Meteorological data from the public network of agro-meteorological stations equipped with the following sensors:
    - Air temperature
    - Humidity
    - Wind speed and direction
    - Amount of precipitation
    - Value of leaf wetting
- Observation and phenological stations – these stations are equipped with PTZ cameras and cloud software allowing remote management and control of actions. The stations allow observation of fields for the presence of disease and pests on them and can accurately track the state of plant development.
- Agrophage mathematical models and their results are based on field data (entered by farmers and advisors in eDWIN system) and the meteorological station network.
- NDVI data (value) for the smallest possible area, reduced to the point and its trend line in time
- Structured field observations collected by advisors and farmers within the eDWIN system as e-service.

2.9.2 Data Acquisition Plan

2.9.2.1 General information

We will use about 100 meteorological stations from a network of nearly 600 stations installed all over Poland (see Figure 29). The stations used will be in the Wielkopolska region. The number of fields and plants to be used in the analysis will depend on the order of the regional agrophage observation coordinator. Usually during the growing season, there are about 100 – 120 fields (observation points).

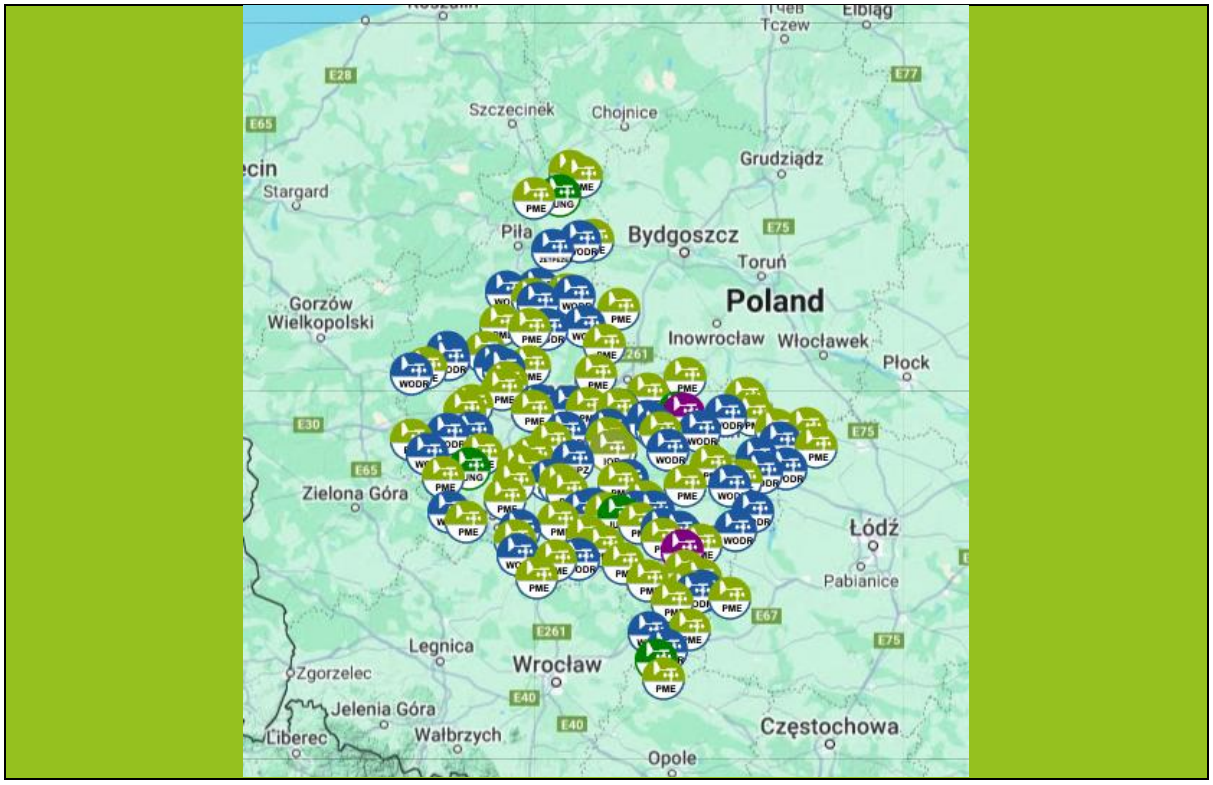

Figure 29. The meteorological stations used in RIL Crop management (Sub lab Early Pest Detection)

Eight phenological stations will also be used for automatic observations. These stations are in the Wielkopolska region as well. Some of the stations may be located in fields where, in addition, manual validation of observations by a qualified agricultural advisor will be carried out.

2.9.2.2 Schedule

The methodology for conducting field observations has been developed and has been operating effectively since 2020 in eDWIN, and much earlier outside of it (one of the elements of the eDWIN implementation was the optimization of the process of planning and execution of field observations and integration with the national agrofagi.com.pl system). Data will continue to be collected and used for the acquisition optimization and scale process for future years.
The methodology consists of:

1. Designating individuals for monitoring (the coordinator does this based on knowledge of the advisor's place of work)
2. Designation of the crops that a particular advisor will monitor (the coordinator does this based on knowledge of the crops in the area),
3. Indication of the exact points where the observations will be carried out (the advisor does this in a dedicated system and enters, e.g. coordinates and data on the crop)
4. Implementation of observations at the designated point through a dedicated tool
5. Notification of threats to farmers in areas where agrophages have been confirmed.

In parallel to this process, we will conduct an analysis of the possibility of optimizing the planning and execution of field observations based on the integration of observation data with data from meteorological stations, observation stations and NDVI.

2.9.2.3 Resources

Most of the resources are provided by the partner (PSNC), i.e. computing power, and data space.

Data is being collected on server infrastructure using continuous aggregation services. Most repositories have an extended retention policy, using disk resource scaling and maintaining the entire history of measurements while regularly creating data backups. Complex data acquisition services are used from various sources simultaneously - FTP servers, API interfaces, and reading directly from the database. The data has registers describing point data sources and a series of sensor sets. Access to the data is done by selecting the appropriate measuring point in the interface intended for reading. Data from various sources are unified and described with as much metadata as possible.
There were first algorithms of data validation introduced, with further potential to extend scope data quality supervision. Technologies used: REST API, Kafka streams, Logstash, time series DB, virtualization, containerization, microservices.

2.9.2.4 Special data constraints

To the data we already have, the methodology for acquiring NDVI data at specific points where agrophage occurrence data will be collected (crop fields) must be developed and positioned. Also, for historical field observations, for which we have, e.g., meteorological data and mathematically modeled threat data, it will be necessary to acquire NDVI data.

2.9.3 Summary of the first 30 months of the project / RIL Crop management – Sub lab Early Pest Detection

As part of our Research and Innovation Lab (RIL) within the ScaleAgData project, we have designed a set of tools aimed at increasing the effectiveness and flexibility of pest and disease monitoring in agriculture. The following sections describe the key solutions implemented:

Monitoring Dashboard
One of the core components of the developed system is the Monitoring Dashboard, which allows users to define their own reference point—either by selecting a previously defined monitoring location or freely choosing a spot on the map.

Upon selecting a reference point, the system automatically presents a set of information relevant to the advisor, including:

- Forecasts from disease models for the selected location,
- Pest and disease alerts submitted by other advisors nearby,
- Records of plant protection treatments conducted by farmers in the surrounding area,
- The Normalized Difference Vegetation Index (NDVI) for nearby fields helps to assess crop conditions.

This comprehensive view allows advisors to quickly and effectively evaluate the situation in a specific region without navigating multiple modules.

Dashboard Notifications
Each created dashboard not only provides real-time data visualization but also serves as the basis for weekly automated email notifications. For every dashboard, the system generates a PDF summary of the same key data presented in the interface.

Each notification includes:

- Results from disease forecasting models,
- The latest nearby threat reports,
- Information on protective treatments by farmers,
- NDVI index values for adjacent fields.

The summary is automatically compiled into a PDF file and attached to the email, enabling users to stay informed without logging into the system. A direct link to the dashboard is also provided in the message for those wishing to conduct a more detailed analysis.

Ad Hoc Reports

A new feature introduced in the mobile application is the ability to submit ad hoc reports. These reports are not bound to predefined monitoring points and can be submitted from any location.

This is especially useful for advisors monitoring areas where crop protection is strong and threats rarely occur. In such cases, traditional reporting becomes ineffective. Ad hoc reporting allows advisors to flexibly respond to actual field observations and gather valuable data from diverse locations.

Weekly Regional Reports

The system also includes a feature for generating weekly regional reports that can be compiled by regional coordinators. These reports summarize the phytosanitary status of a given region for a specific week.

The system supports the entire report creation process by:

- Automatically calculating key statistics based on submitted reports,
- Generating a draft version of the report that can be reviewed and edited before publishing.

This solution significantly reduces manual effort and saves time in producing weekly summaries.

Notifications and API Access for Weekly Reports

Given the high informational value of the weekly regional reports, an automatic notification system has also been introduced. Every user involved in the monitoring process receives an email whenever a new weekly report is published.

In addition, the report data is made available via API, which allows integration with external systems. For example:

- Displaying current threat levels on public information screens in agricultural advisory offices,
- Embedding data on websites that inform the public about regional phytosanitary conditions.

Testing phase

Currently, the system is in the testing phase, involving selected agricultural advisors from ODR who actively use the newly developed solutions in their daily work. To ensure that the evaluation is structured and comparable, advisors have been provided with detailed testing scenarios describing the sequence of actions to be performed in the system. These scenarios cover the key functionalities introduced in the first 30 months of the project, including the Monitoring Dashboard, automated notifications, weekly reports, and ad hoc reporting.

In addition to carrying out the defined tasks, advisors are required to complete a feedback survey. The survey collects information on usability, clarity of the interface, efficiency of the newly introduced mechanisms, and the overall impact of the tools on the monitoring process. Particular attention is given to the added value of integrating observation data with meteorological data, NDVI, and disease forecasting models.

The feedback collected will be systematically analyzed by the development team. Based on these insights, developers will be able to refine existing functionalities, address reported issues, and prioritize future improvements. This iterative process ensures that the solutions created within the

project not only meet technical requirements but also respond to the real needs of agricultural advisors in the field.

By engaging end-users directly in the testing and validation phase, the project creates a feedback loop between practitioners and developers, strengthening the relevance and long-term applicability of the developed mechanisms in pest and disease monitoring.

Feedback from Advisors – Testing Phase

The initial feedback collected from advisors testing the new functionalities provides valuable insights into both strengths and areas requiring improvement:

Positive remarks:

- In general, the system is perceived as working correctly.
- Users appreciate the simplification of the interface and availability of ad hoc reports.

Identified issues and suggestions:

- Advisors find it useful to add new observation points from other locations to broaden the scope of visible threats. Current observation fields are often well-protected, which may limit the representativeness of the threat picture.
- A data inconsistency was observed: regardless of the selected crop (wheat, rapeseed), the system displayed monitoring data for peas.
- Lack of incoming email notifications was reported in some cases.
- Suggestions include further interface simplification and ensuring proper alignment between selected crops and displayed monitoring data.

Next steps:

The development team will analyze the feedback in detail, with priority given to fixing data consistency issues and improving the flexibility of selecting observation points. Further iterations of testing will continue to gather advisor experiences and ensure the tools meet their needs in real-world monitoring scenarios.

3 Sensor Development

3.1 Nanoparticle Gas-Sensing Array

The development of nanoparticle gas sensors in the RIL Crop management (Sub lab Agri-environmental monitoring for Policy Makers) has the potential to provide valuable information through the detection of pesticides in crops and the estimation of pest applications in the fields.

3.1.1 Chemical sensors

Chemical sensors convert a chemical input signal into a measurable electrical output signal and are widely used in biomedicine and environmental sciences. Also, they are often used for the detection of potentially toxic substances. They consist of two functional parts: an identification medium that selectively reacts with the measured chemical substance and a transducer medium that converts the reaction of the first medium into an electrical signal. The performance of the chemical sensors is judged by their response time, signal-to-noise ratio, selectivity, and detection limits, among others. All these factors depend on the two functional parts already mentioned, i.e., the identification and the transducer medium. Therefore, the construction of efficient chemical sensors is directly linked to the development of materials that have good recognition and mutagenic properties. Nanoparticles can perform both recognition and mutation roles and are widely used for the development of chemical sensors.

3.1.2 Electrochemical sensors

Electrochemical sensors are a class of chemical sensors whose output signal is expressed by a change in their capacitance or resistance. This is achieved as their sensing medium is a dielectric material (membrane) between the mutating medium, which is conductive, effectively creating electrodes that are separated by a dielectric surface.

In the case of electrochemical resistance sensors, a predetermined voltage is applied to the electrodes. The sensing membrane reacts to the chemical signal by changing its conductivity resulting in a change in the measured resistance in the circuit. In the case where resistive sensors incorporate nanoparticles, their conductivity is achieved by coating the electrodes with two-dimensional or three-dimensional films of metallic nanoparticles. The sensor film can be polymer, ions, surfactants, or biomolecules and surrounds the nanoparticles and electrodes. The sensing membrane responds to the chemical signal by changing its morphology, which results in a change in the distance between the nanoparticles and thus their resistance/conductivity. Thus, the sensor responds to the chemical stimulus by changing its resistance in a constant potential difference circuit.

3.1.3 Description of the sensing array to be deployed

The sensing array that will be used by the RIL Crop management (Sub-lab Agri-environmental monitoring for Policy Makers) will have as a goal to discern between four widely used pesticides and humidity, which is a gas that is commonly present in real conditions of the fields. The sensing array consists of gold electrodes on which two-dimensional conductive films of platinum nanoparticles have been deposited. On top of the nanoparticle layer, there are four polymer layers that are sensitive to volatile organic compounds and will be used as gas-sensitive layers. The transducing layer absorbs gases that are present in the short vicinity whereas each one of the four polymer layers is susceptible to different gases. The absorption of the gases leads to the swelling of the four polymer layers. As a result, the nanoparticle layer is deformed, and the mean value of the inter-nanoparticle distance is increased. This leads to an increase in the resistance of the sensor due to its exposure to pesticides.

The sensing array is integrated as an autonomous data source within the IoT station. Data collection on pesticide detection is initially locally processed and then transferred to the cloud data repository - repository for integration and further processing. Figure 30 provides an illustration of the integration of the sensing array with the Gaiasense IoT gateway. More information for the Gaiasense IoT station is available in section 2.7 RIL Crop management.

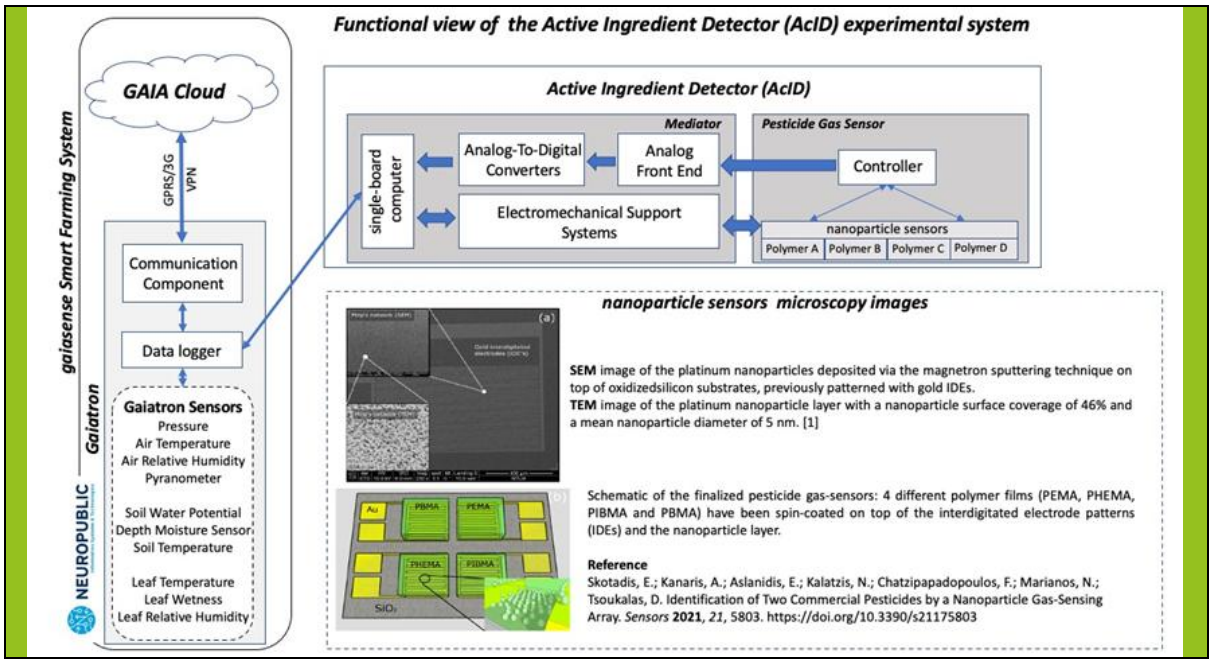

Figure 30. Integration of the nanoparticles sensing array with the IoT gateway.

3.1.4 Fabrication of the Sensors

As already mentioned, the array of sensors is based on the combination of polymer films and platinum nanoparticles. Their fabrication is achieved through the following steps:

- On top of a 10 cm diameter oxidized silicon substrate, gold interdigitated electrodes are produced through the e-gun technique and lift-off techniques.
- The wafers are broken down into wafers that contain 4 individual sensors which form a sensor array.
- The platinum nanoparticles, which have a diameter of 5 nm, are deposited on individual wafers.
- The nanoparticles are produced and deposited on the electrodes, through the DC sputtering technique.
- Four different polymers are spin coated on top of the nanoparticle substrates and a polymer layer was obtained with a thickness of 500 nm.

Each sensor array/chip consists of a total of four individual sensors with the same polymer film. For each of the four different polymeric films, a total of two sensor arrays (eight individual sensors) were fabricated and characterized (Figure 31) [1].

The arrays of sensors that will be developed with films from the following four different polymers are:

- Poly (ethyl methacrylate) (PEMA)
- Poly (2-hydroxyethyl methacrylate) (PHEMA)
- Poly (isobutyl methacrylate) (PIBMA)
- Poly (butyl methacrylate) (PBMA)

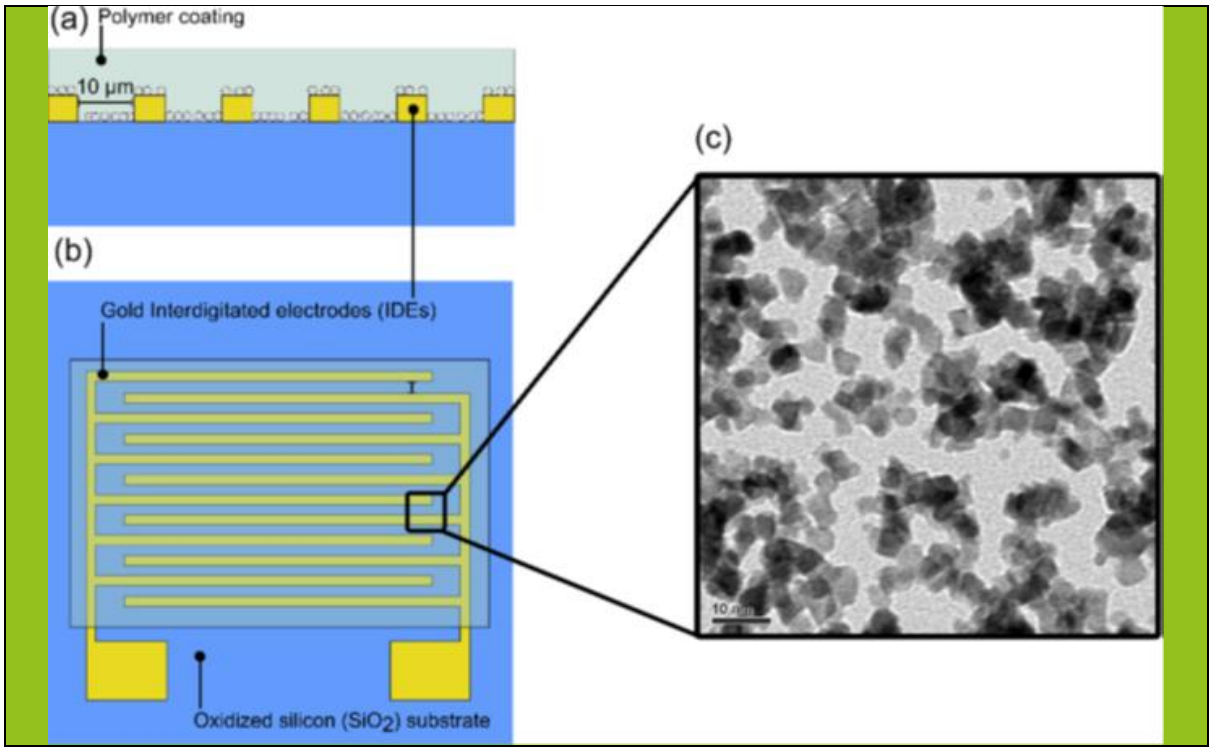

Figure 31. Sensor schematic and microscopy images (a) cross -section

of the sensing device (b) top-down view of the sensor (c) transmission electron microscopy image, the Pt nanoparticle distribution and size can be seen (source: Skotadis et al., 2020)

3.1.5 Current status

As of 31 January 2024, the sensor array was fully developed and ready for installation. Its deployment was scheduled to coincide with the period of pesticide application in tomato cultivation, which typically occurs during the summer months. This strategic timing ensured that the sensor array is operational during a critical phase for data collection on pesticide detection and application monitoring. The pesticide sensor array was installed at the 2nd pilot site in Farsala, Greece (see Figure 32, specific field location is subject to laws about personal data and thus is confidential info). However, during commissioning in August 2025 we encountered station-level issues that prevented stable data generation. Corrective maintenance is underway. Upon completion of the fix, the sensor will be redeployed, and a set of controlled spray tests will be executed to capture a validated dataset. The spraying campaign has been rescheduled in September 2025 to align with the revised maintenance timeline. The technical design remains unchanged and only the spraying test schedule is affected.

Figure 32: General area which includes the pilot field where the pesticide sensor was installed in Farsala, Greece.

3.2 Hyperspectral Fabry-Pérot interferometer camera

3.2.1 Overview

For developing a hyperspectral sensor for this project, the main objective was to be compatible with airborne measurements using drones that, in turn, require fast data acquisition. The acquisition speed requirement demands powerful light-gathering optics. To enable this, a new type of Fabry-Pérot interferometer (FPI) filter, with a large clear aperture, has been developed. The hyperspectral imager designed and built in this study is shown in Figure 33 below.

Figure 33: Hyperspectral imager developed in this project.

In addition to an FPI as the adjustable bandpass filter, the main components of the imager are:

- Industrial camera module with an InGaAs sensor
- Control electronics and low-pressure housing for the FPI
- Off-the-shelf commercial camera lens along with static long-pass and short-pass filters

The main specifications of the imager are presented in Table 4.

Table 4. Key specifications of the hyperspectral instrument

Parameter	Value
Parameter	Value
Pixel resolution	1280 x 1024
Pixel size	5 µm
Field of view	10.5° x 8.4°
Wavelength range	650–1200 nm
Spectral bandwidth	30–35 nm
FPI aperture diameter	43 mm
Focal length	35 mm
f/N	1.8
Mass	1.7 kg

3.2.2 Fabry-Pérot interferometer

To increase the ability of the instrument to collect signals, a large FPI filter prototype has been developed. The filter is based on two 43-mm clear aperture metal mirrors separated by a small gap. By adjusting this distance between the mirrors, the wavelength of the transmitted electromagnetic radiation can be tuned. The size of the gap is altered by varying the control voltage applied to the three piezoelectric elements attached to the mirror substrates. The structure of the FPI developed for the imager is schematically illustrated in Figure 34.

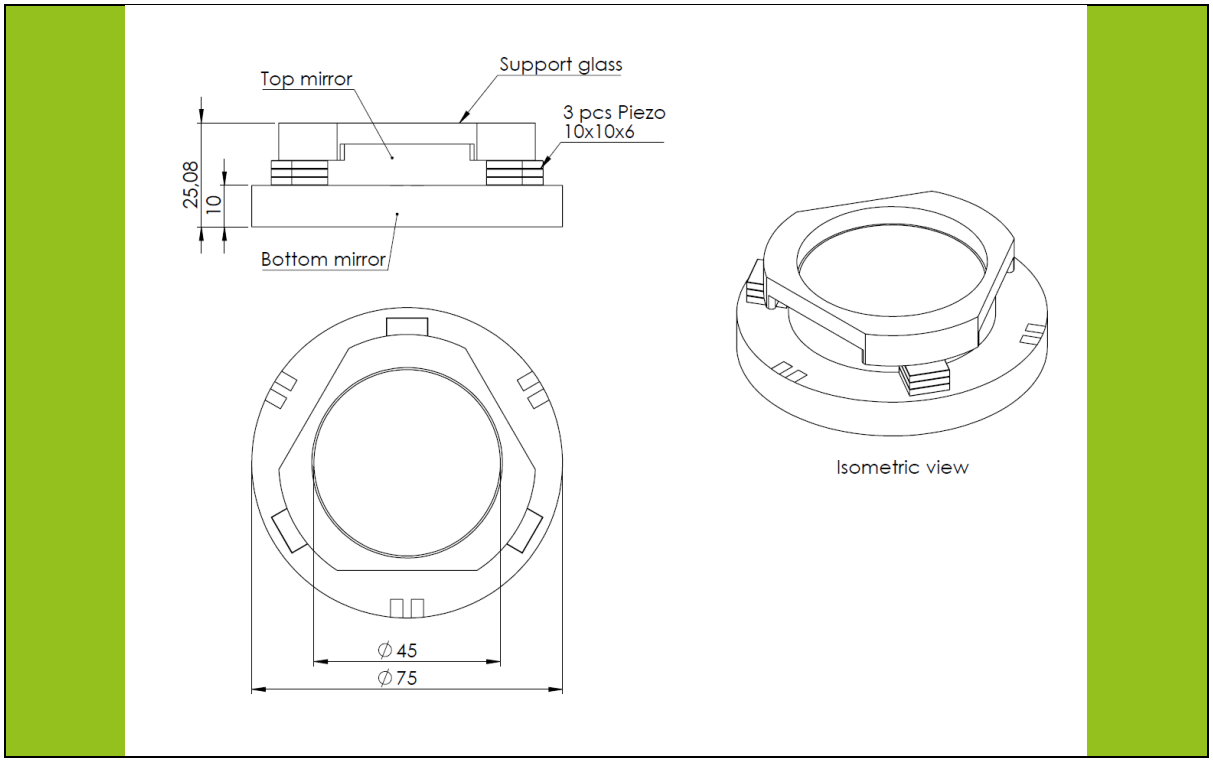

Figure 34. Simplified schematic of the FPI filter.

The wavelength λ transmitted by the FPI depends on the gap according to the equation $\lambda = 2d / n$, where d is the distance between the mirrors and n is a positive integer indicating the order of the constructive interference.

Atmospheric pressure resists and dampens the movement of the mirror substrates required for adjusting the distance between the mirrors. This damping effect is especially significant when reducing the gap, as the air pressure between the mirrors rises.

3.2.3 Electronics

The distance of the mirrors of the FPI is continuously controlled by measuring the capacitance of the electrodes etched on the surface of the mirror substrates. The capacitance measurement feedback loop is also used to maintain the mirror surfaces parallel to each other to keep the gap constant throughout the whole mirror area.

The FPI controller electronics are implemented as a ring-shaped printed circuit board (PCB) around the FPI. The controller is installed in the same low-pressure chamber with the FPI to make the disturbance-sensitive connections between the controller and the FPI as short as possible. This solution also reduces the number of signals required to be fed through the low-pressure chamber.

In addition to controlling the FPI, the electronics are also designed to trigger image acquisition once the FPI has settled on the desired gap, and to actuate the mirrors to the next gap once the exposure of each frame is ready.

The key components of the controller implementation are capacitance to digital converter (CDC), microcontroller unit (MCU), piezo drivers with integrated high-voltage boost-converter. The implementation also contains a digital-to-analogue converter (DAC) to generate control signals for piezo drivers, an analogue-to-digital converter (ADC) to convert piezo voltages to

digital format for the MCU and environmental sensors to measure the temperature and pressure inside the FPI housing. The block diagram of the implementation is shown in Figure 35.

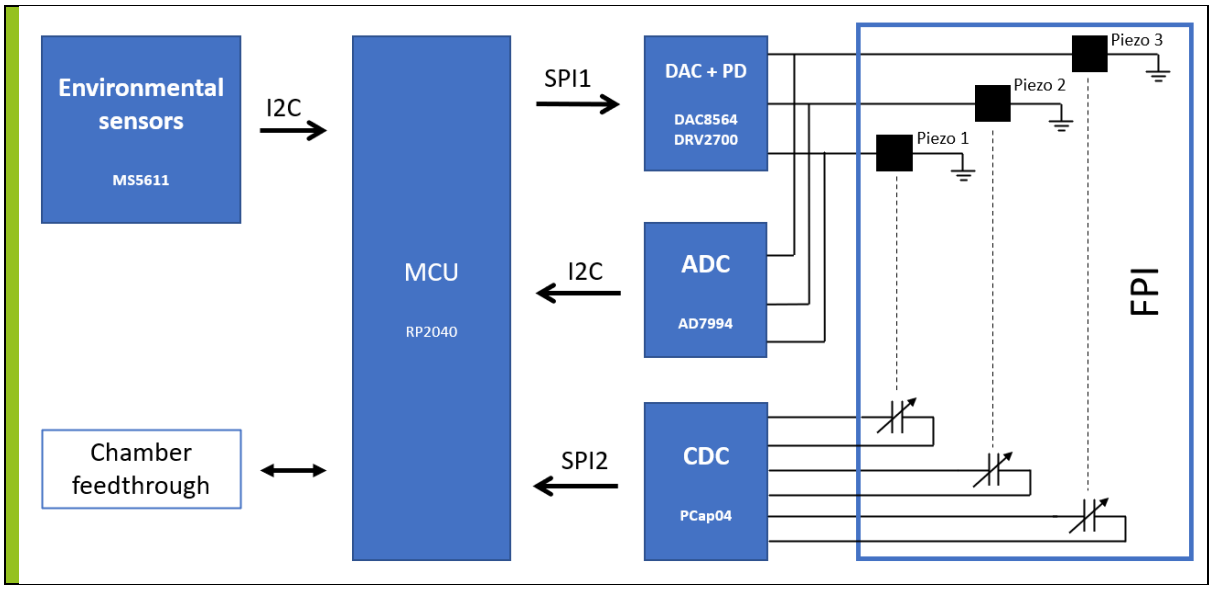

Figure 35. Block diagram of the implemented controller. Bus arrows present directions of data transfer in normal operation conditions.

The PCB layout was implemented in a ring-shaped PCB with a 100 mm outer diameter and 66 mm inner diameter. The shape of the PCB posed a challenge for isolating the switching mode voltage controller noise from the sensitive capacitance measurement. To minimize the disturbances, the capacitance measurement circuit and switching mode voltage converters are laid on the opposite sides of the PCB, and ground fills were added on the upper and lower copper layers with capacitance measurement tracks. The layout of the FPI control PCB is shown in Figure 36.

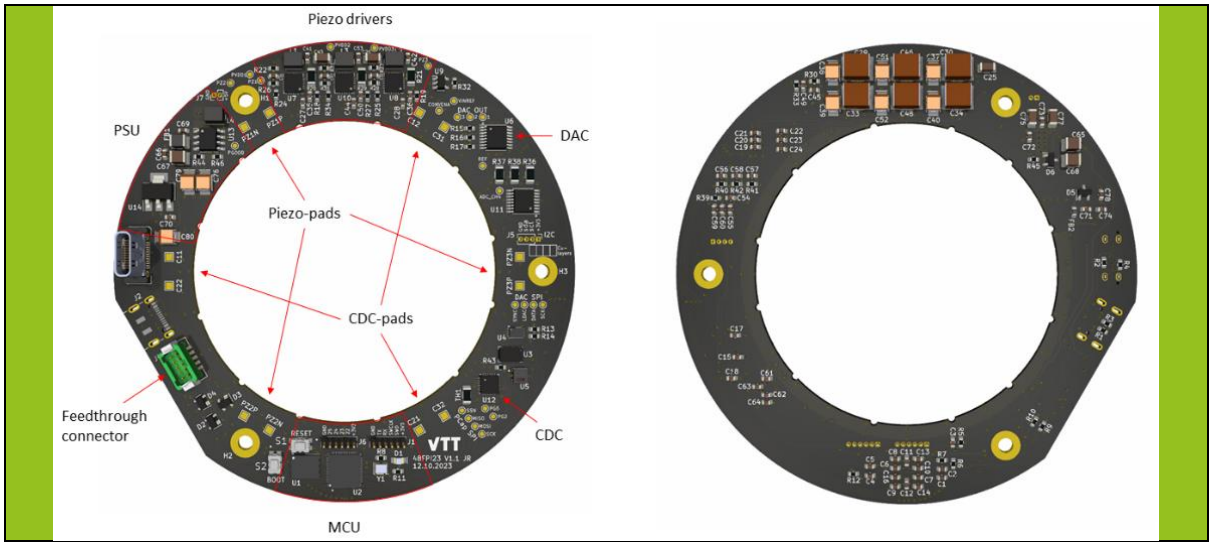

Figure 36. Layout overview of the implemented FPI-controller.

The firmware of the controller is written in C language. The firmware utilizes both cores of the MCU, using one for execution of the gap control loop and tasks synchronized with it. The other core is used for handling serial communication with the master device and executing other tasks related to received serial commands.

The gap controller is a digital PI controller. The error signal for the controller is filtered with a moving median filter of adjustable length for both proportional and integral terms. The execution frequency of the control loop is approximately 8 kHz. The controller employs variable controlling parameters for the settled and dynamic states.

3.2.4 Optics and mechanics

To reduce the impact of air pressure on the actuation speed of the FPI, the filter is enclosed in low-pressure housing. A short-pass and a long-pass filter installed on the optical axis on both sides of the housing act as the windows of the chamber. A ten-pin connector is used to feed through the supply power and electrical signals of the FPI, and a one-way valve is used to create and upkeep the low-pressure conditions inside the chamber. The mechanical design schematic is shown in Figure 37.

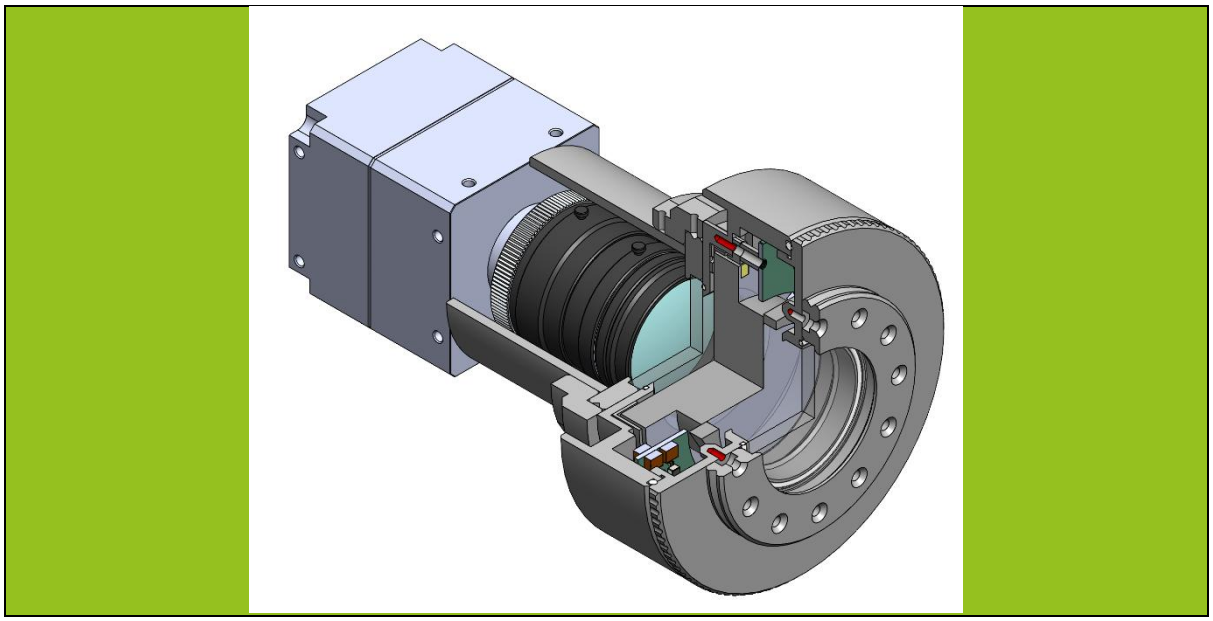

Figure 37. Mechanical design of the hyperspectral imager.

For lightweight structures, aluminum is the preferred choice due to its good mechanical properties and low density. EN AW6082 aluminum alloy was selected as the material as it is commonly available, has good machinability, and is easy to anodize. Each seam interface in the construction increases the rate at which the housing is filled with leaking air. Typically, vacuum sealed flanges are installed with multiple screws around the pressure side of the seal. In this case, the main lid was implemented with a single non-standard thread to reduce the diameter of the housing. Altogether, the following interfaces have been identified as the potential points for air leaks:

- the main lid of the housing,
- two 50-mm filters (short-pass and long-pass), functioning as the windows of the chamber,
- custom-sealed electrical feedthrough,
- manual one-way valve for depressurizing the chamber.

To improve the air tightness of the instrument, all the O-ring interfaces are treated with vacuum grease.

The FPI control electronics are enclosed inside a vacuum chamber together with the FPI, and the camera lens is installed inside a lens tube to obtain weatherproof construction. The FPI housing is mounted at the end of the lens tube using three stop screws.

The electrical feedthrough connector is sealed using epoxy to reduce air leak rate. The seal was achieved by pumping vacuum at one end of the connector while flooding the other end with epoxy. To reduce any possible outgassing, the epoxy was first prepared in a dedicated vacuum chamber before treating the connector. Installation of the connector into the housing using epoxy makes the installation permanent. The components of the imager are presented in an exploded view in Figure 38.

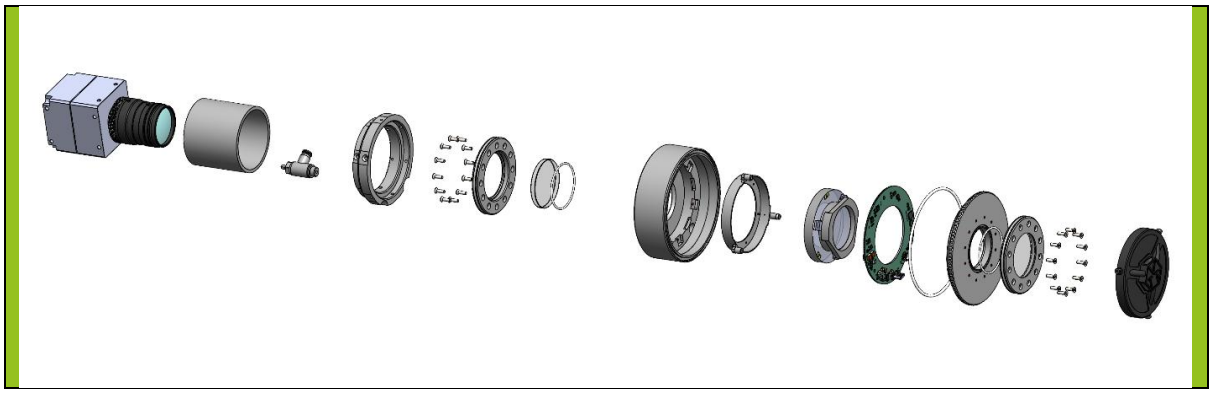

Figure 38. Exploded view showing the parts of the instrument.

3.2.5 Current status

As of summer 2025, the instrument has been updated based on the feedback received from the RIL lab partners. The wavelength range has been extended from 650–950 to 650–1200 nm by replacing the old FPI filter with a new unit. The new FPI filter can achieve much smaller air gaps between the mirrors, which allows employing the first order of interference transmission wavelength bands. The hyperspectral imager was recalibrated and characterized after the modifications.

In addition to the new FPI filter, the measurement system has been upgraded with a global navigation satellite system (GNSS) receiver. The receiver was requested by the RIL lab partners to make it easier to pre-process the hyperspectral image data captured from a moving drone. The measurement software was updated to facilitate the changes in the hardware.

The updated camera was delivered to ILVO in June 2025 (and to be later sent to AUTH) for the second iteration of RIL Soil field measurements.

4 Sensor Spatial Planning

Ground sensing and laboratory sampling are important for RILs, it serves as the actual, measured ground truth for each one of the fields of interest. In most RILs, in-situ measurements are required for full data acquisition either as standalone solution or complementary to satellite or airborne data. Ground sensors or sampling permits the evaluation of the validity of the earth observation data and the related developed models such as in the RIL Soil Health.

The way the sensors are spatially arranged may affect the quality and adequacy of the collected data. Terrain and cost restrictions as well as the number of available custom-made sensors may prohibit a dense sensor placement over the whole area of interest that could provide an accurate representation of the field ground truth. Accurate monitoring of fields such as SOC and humidity among others often requires recovering not only central tendencies but also spatial extremes. Conventional uniform or grid sampling prioritizes even spatial coverage but can under-sample rare, extreme-value regions. The proposed method aims to (i) bias sampling toward extremes in a controlled, non-biased way, (ii) enforce a loose mean-preserving constraint to avoid systematic bias, and (iii) include geographic dispersion so that extreme-value emphasis does not collapse all sensors onto the same small area. Hence, the following strategy was developed in ScaleAgData and may be followed by RILs to select the best possible positioning of the available sensors and/or sampling locations ensuring adequate coverage with minimum error on the re-constructed field.

The following methodology was followed by ICCS for designing sparse sensor networks using satellite products with an emphasis on accurate recovery of spatial extremes. First, each field is statistically analyzed. Through a constrained convex optimization over histogram bin weights, it pushes the distribution toward extremes while preserving the mean value. The field is then sub-sampled using the new weighted distribution. Then uses the biased sample pool into *k* sensor locations using a Gaussian Mixture Model (GMM) clustering with spatial weighting to avoid extreme clustering. The final *optimal* sensor locations are defined after a stochastic run of several GMM instances to ensure unbiased results.

Finally, a reconstruction-based evaluation pipeline compares the *optimized* sensor placement to a regular grid baseline using spatial interpolation and a set of global and extreme-aware metrics. As an example, Figure 39 below shows two sensor arrangement designs, the reconstruction of an NDVI field, which was used as a proxy for Soil Organic Carbon (SOC) measurement, in parcel “Caritas” by RIL Soil health (ILVO) and the related reconstruction spatial errors. The first column shows the reconstruction of the ground truth (3rd column) using extreme-aware optimal placement while the second column the reconstruction using a regular grid of sensors. The *optimal* design captured the extreme values, suffering from mild error mainly in average-valued regions. On the contrary, regularly spaced grid

design missed completely the extreme values.

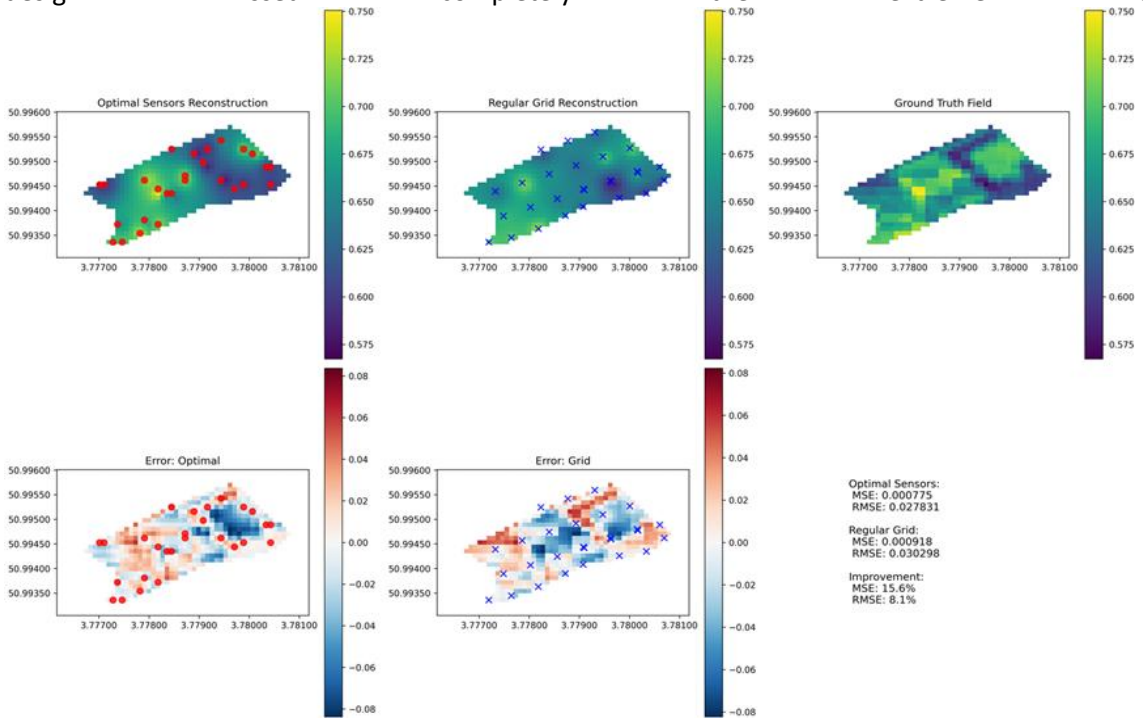

Figure 39. Field reconstruction from 2 different sensing designs and the related reconstruction error maps.

Furthermore, the optimal design's reconstruction shows distributional similarity with the ground truth while the gridded design distribution is more centered, as shown in Figure 40. It is also shown that the extremes are better aligned in the reconstruction of the *optimal* design with smaller systematic error (see point vs pixel error scatter plots in the same figure). The basic error metrics (see Figure 41), computed over valid pixels, show a better overall performance for the *optimal* design, with MSE and RMSE being improved by 15.6% and 8.1% respectively compared to the regular grid design. The improved performance is further supported by imbalance metrics with emphasis to the extreme values, such as Squared Error Relevance Area (SERA), which was improved by 110.81%. The performance comparison between the *optimal* design and the regular grid design in terms of extreme values. The reconstructed *optimal* design showed 5.5%, 94.76% and 57.1% improvement in terms of minimum values error, maximum values error and range error respectively.

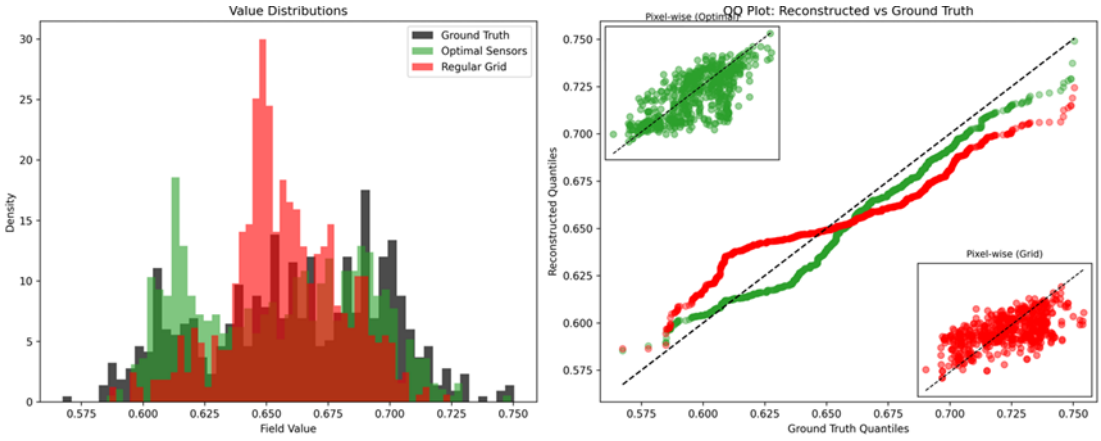

Figure 40: Distributional similarities between two different sensors placement designs and the ground truth, and point-wise view errors.

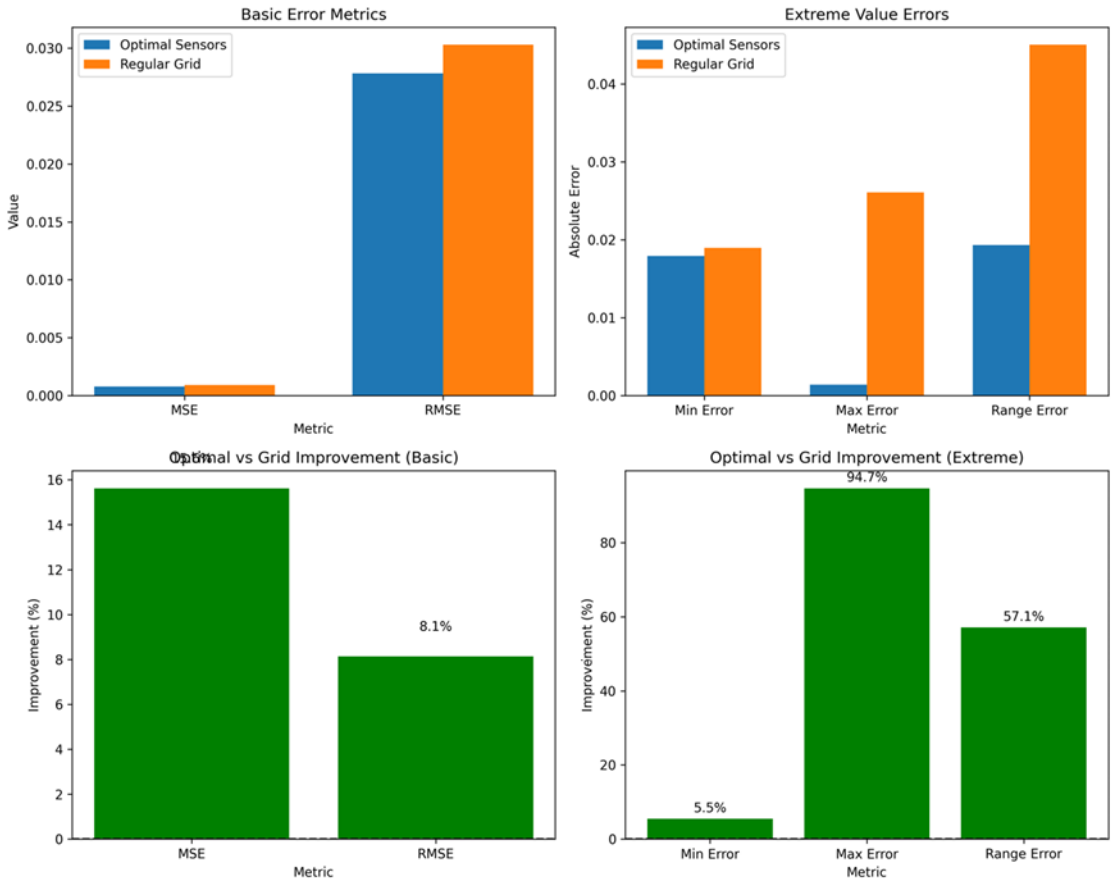

Figure 41: Regular error metrics and extreme values performance for two different sensor placement design strategies.

As a subsequent step, the methodology will be applied during the second iteration for both the RIL on Soil Health and the RIL on Crop Management (NP). For the RIL on Soil Health, NDVI will serve as the input, guiding the calculation of the placement of SOC sampling positions. For the RIL on Crop Management, Soil Moisture EO data provided by DHI will be used as input to determine the placement of 20 IoT soil moisture sensors. A formal validation process will then be carried out thereafter.

5 Edge processing enabling technologies, real time processing and privacy

Edge computing is a paradigm that enables data processing and analysis near the source of the data, rather than relying on centralized cloud servers. It can reduce the latency, bandwidth, and energy consumption of data transmission, as well as enhance the privacy and security of the data. It can also allow us to tailor more resilient applications in case of an unreliable network, since computations and decisions taking are implemented on-site. Edge computing can also enable more complex and intelligent applications that require real-time or near-real-time feedback.

In the case of data for agriculture, which is the topic of this project, edge computing may be of special interest for data quality assessment and improvement, data compression, early alerting systems, or on-the-field interpretation of complex sensors or even for federated learning.

5.1 YaraSense Platform

In the frame of T3.2, we are building a hardware and software Platform for Integrated Sensing and Edge Computing tailored to the needs of the agriculture sector that we call YaraSense (see Table 5).

Table 5. Specifications of the YaraSense Platform.

ID	Description
1	The YaraSense hardware can be installed permanently in the middle of a field.
2	The YaraSense hardware is autonomous for its energy supply.
3	The YaraSense hardware can be connected to the cloud using some long-range communication technologies (LoRaWAN, 4G, satellite communications, …).
4	The YaraSense hardware accepts various types of sensors, that can be plugged and unplugged easily on the field.
5	The YaraSense firmware easily accepts new sensor drivers to be installed when it is deployed on the field. (This could be a firmware update)
6	It is possible to connect wired sensors to the YaraSense hardware, using standard communication protocols (I2C, SDI12, SPI, RS485, …).
7	It is possible to connect BTLE sensors, especially using a mesh network.
8	The YaraSense firmware runs an acquisition and processing pipeline (made of processing steps).
9	The YaraSense firmware pipeline can be modified by sending new processing steps or configurations via the long-range communications.
10	The YaraSense enclosure must be weather and dust proof, stable enough to stay on the field in harsh weather conditions and visible enough.

The design of this YaraSense platform is based on the Edge Spot IoT box, made by EGM. It is a flexible and expandable IoT device capable of connecting to various sensors and actuators, transmitting data across diverse networks. Its versatility stems from the presence of three mikroBUS™ slots, facilitating the integration of extension cards. With over 1000 available options, and the ease of designing custom cards tailored to specific data collection needs, this device offers extensive adaptability. Furthermore, its power management system is highly versatile, enabling energy harvesting from solar panels and efficient battery load management. In scenarios with limited energy resources, the device can enter an ultra-low power mode, consuming mere tens of microamps. See D3.1 for more details.

Figure 42. The Edge Spot hardware.

The Edge Spot hardware (Figure 42) is stable and has already been used for real-life deployments. Its software is essentially a Software Development Kit (SDK) and an application template that can be both mixed to develop applications for specific uses. During this project, the SDK will be improved to add new functionalities linked to the needs of the project (edge computing). These developments will lead to a complete, deployable platform to fully leverage the potential of the far-edge computing in this project. These edge computing capabilities will be demonstrated using multi-sensor integration, like an electronic nose to assess various gas emissions that could lead to estimates of organic content or nutrients in the soil.

The global architecture for the YaraSense system has been developed, as shown in (Figure 43). The YaraSense platform is designed to support three types of serial communication protocols for wired sensors: I2C, RS-485 and SDI-12. The YaraSense box is intended to serve as a centralized hub for collecting data from various sensors, processing it locally through edge processing pipelines, and can transmit some processed data to the cloud for further analysis or integration.

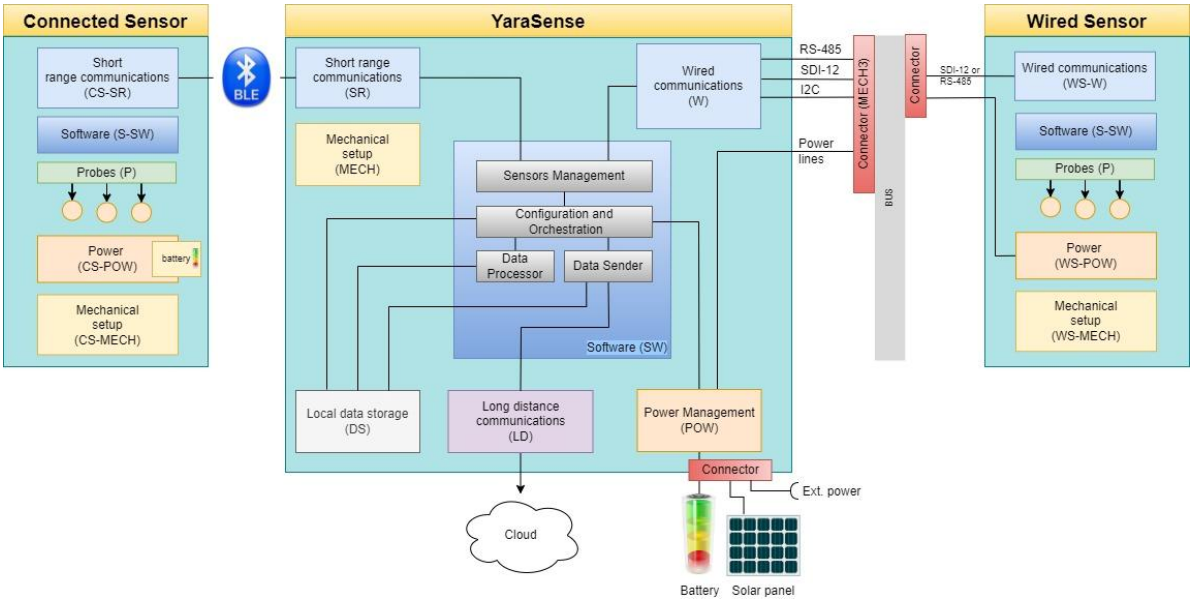

Figure 43. YaraSense Architecture

5.2 YaraSense's sensors

To design, evaluate, and improve the edge computing capabilities of YaraSense, we integrated the platform with real-world sensors of varying complexity. As a baseline, we connected a SDI-12 soil probe (SWAP Soil Redox probe), which provides relatively simple scalar measurements. In parallel, we experimented with more advanced and prototype sensors such as an electronic nose (eNose) and a spectrometer with infrared temperature sensing. These sensors produce arrays of high-dimensional data, making them well-suited for on-site processing with embedded AI techniques. This setup allows us to assess the value of performing computation directly at the edge, reducing data transmission requirements and enabling faster, local insights.

5.2.1 Soil probes

Soil probes are essential tools in modern agriculture, providing critical data about soil conditions that help farmers make informed decisions. These sensors measure various parameters such as moisture, temperature, pH, and nutrient levels, which are vital for optimizing crop growth and yield. One specific type of soil probe is the Redox probe, which measures the soil's oxidation-reduction potential. This information can provide valuable insights into soil health and microbial activity, both of which influence nutrient availability and plant growth.

To explore the relevance of soil oxidation-reduction potential in agriculture, SWAP Soil Redox probes were selected for testing. These probes require a 12V power supply and communicate using the SDI-12 protocol. SDI-12 is a widely used communication standard for environmental sensors. It operates on a single data communication line, allowing multiple sensors to connect to a single data logger. The protocol is designed for low-power applications and is ideal for remote or battery-powered systems, making it well-suited for agricultural deployments.

An early prototype of the YaraSense system that was designed to integrate and evaluate these probes. The EdgeSpot device supplies the power to the probes and operates using a battery supported by a solar panel. The SDI-12 driver is incorporated into the EdgeSpot SDK to facilitate sensor data acquisition.

To ensure data transmission, a LoRaE5 module is connected to the EdgeSpot device. The module leverages the LoRaWAN protocol, which is particularly well-suited for this application due to its low-power consumption, long-range communication capabilities, and ability to handle small periodic data transmissions. This makes it an ideal choice for agricultural environments where sensors are often deployed in remote locations with limited access to power and network infrastructure. This enables data to be transferred efficiently to the EGM data platform, Stellio.

The system was deployed in a vineyard. The probes were placed in two separate rows of vines, each receiving a different fertilizer treatment. The goal of this test was to determine whether differences in treatment could be detected through variations in the soil's oxidation-reduction potential.

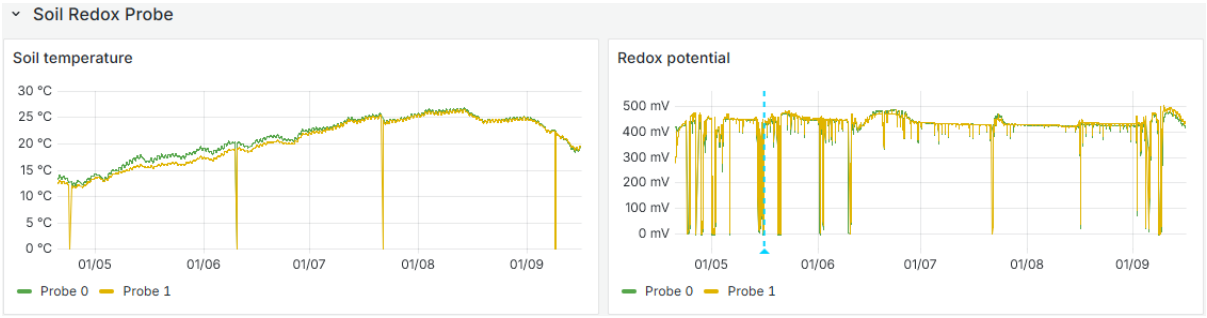

Figure 44 Temperature and redox values collected by the 2 SWAP probes with the early prototype.

The result of this early deployment (Figure 44) showed that the SDI-12 protocol drivers' integration and the data transmission protocols were properly integrated in the YaraSense embedded application. This validated the basis of the software used for after for the more advanced prototype.

5.2.2 Spectrometer & IR temperature surface sensors

The Water RIL Lab requested a spectrometer sensor to interpolate satellite data and provide more detailed ground-level insights. To meet this need, we selected the SparkFun Triad Spectroscopy Sensor – AS7265x. This sensor covers a wide range of wavelengths, making it suitable for detecting and analyzing spectral data that can complement satellite imagery. It communicates via the I2C protocol, ensuring straightforward integration with the existing YaraSense system.

In addition, an infrared (IR) temperature surface sensor, the MLX90614 module, has been added to the system. This sensor, which can also be controlled using the I2C protocol, measures the surface temperature of plants and soil without physical contact. The sensors are powered by the 5V supply from the EdgeSpot device. Both sensors are placed next to each other in a camera box, allowing their data to be correlated for deeper insight:

Figure 45. Pixelo head sensor unit

This unit, which we call the *Pixelo head (see* Figure 45*)*, can be connected to our YaraSense box. Two prototypes were developed and deployed by the RIL Lab Water. The YaraSense reads data from both sensors every 30 minutes without draining the battery, thanks to the solar panel. It uses the spectrometer data to calculate the NDVI (Normalized Difference Vegetation Index), which is sensitive to both vegetation vigor and biomass. The NDVI is computed as follows:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Where:

- NIR = reflected light in the near-infrared spectrum
- RED = reflected light in the red spectrum

NDVI values range from -1 to 1. Values close to 1 indicated dense, healthy vegetation whereas values near -1 illustrate the absence of vegetation.

Currently, YaraSense transmits both the raw sensor data and the computed NDVI values, ensuring that the measurements are reliable and relevant.

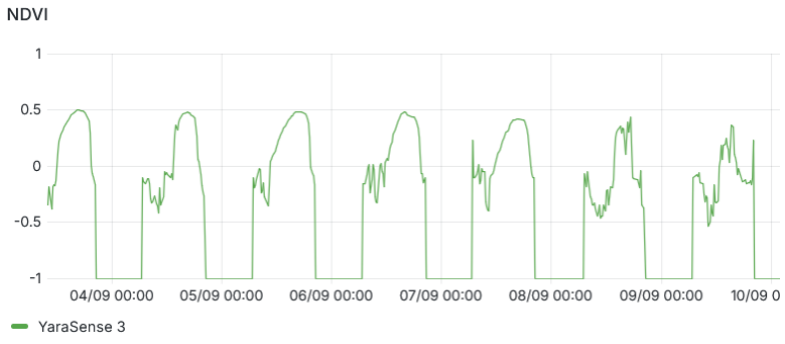

Figure 46.Timeseries of the NDVI index as measured by the sensor. One can see the daily pattern which will be compared and correlated to EO measurements.

A YaraSense box was deployed on the rooftop of the EGM office, with the sensor pointing toward the horizon. Over the course of the week, we observed the following trend (Figure 46):

- At night, the NDVI is not defined and takes the default value of -1, as expected in the absence of light

- During the day, NDVI values are stable at the same hours of the day, which matches our expectations since vegetation indices would not fluctuate significantly under consistent daylight conditions

We plan to extend the software so that NDVI computation and transmission are handled directly through the scripting engine (see 5.4). This will allow the device to send only the relevant NDVI values, demonstrating a real-world application of edge processing while ensuring both efficiency and reliable vegetation monitoring.

5.2.3 eNose

EGM is working on the development of an electronic nose, named eNose, designed to assess various gas emissions, which could provide valuable estimates of organic content or nutrients in the soil. The eNose consists of multiple low-cost gas sensors, each targeting different compounds (Table 6):

Table 6. eNose gas sensors.

Gas	Sensors
Dioxygen	SEN0219
Ammonia	TGS826-A00
Nitrogen dioxide	SEN0574
Dihydrogen	MQ-8
VOC (Volatile Organic Compounds)	TGS2602
Methane	TGS2611
Butane, propane	TGS2610
Air contaminants	TGS2620

The data from these sensors are collected by measuring their internal resistance, using a proper circuitry and an ADC. To complement these measurements, a BME280 sensor was integrated to monitor environmental conditions, including temperature, humidity, and pressure. This latter sensor communicates using the I2C protocol.

At this point, the eNose was built as an independent device, especially to perform lab tests. It sends data directly using UART and LoRaWAN connectivity (Figure 47).

The plan is to make this sensor as one of the SDI-12 sensors in the YaraSense system: the power being provided by the main box through a power bus and communicating with it with the SDI-12 protocol. At this stage, this sensor is not yet an SDI-12 client, but the corresponding driver will be incorporated into the EdgeSpot SDK to transmit the data to the YaraSense SDI-12 server. It will be assessed on the field for the monitoring of soil respiration.

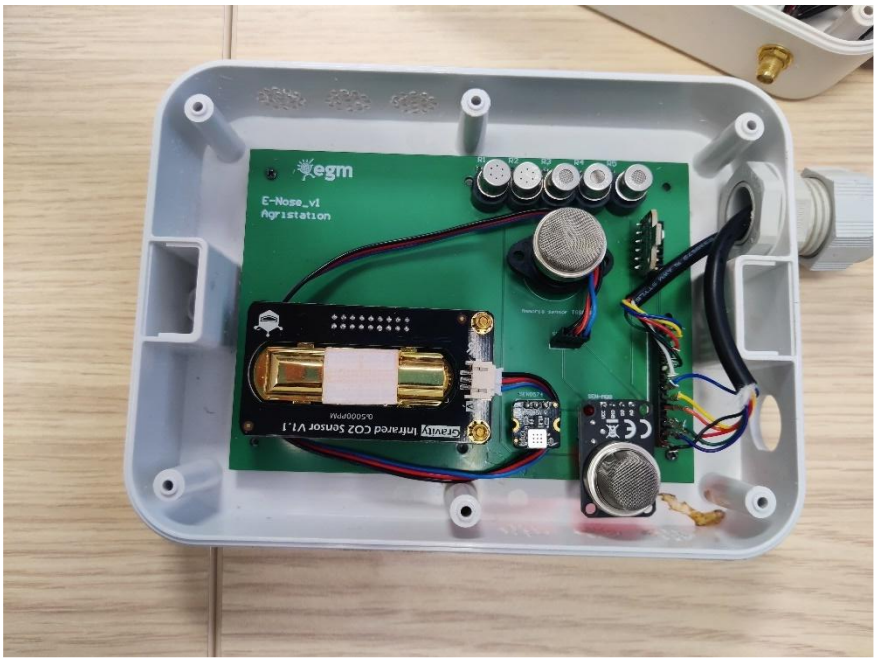

Figure 47. Board containing gas sensors of the eNose

5.3 Firmware Update Over-The-Air

Firmware Update Over-The-Air (FUOTA) is a method for remotely updating the firmware of devices, such as sensors or embedded systems, without the need for physical access. In the context of edge processing, FUOTA is particularly useful because it allows updates to be applied to devices that perform data processing locally, at the "edge" of the network. By updating the firmware remotely, it ensures that edge devices can maintain their capabilities without requiring manual intervention. This is crucial in environments where devices are deployed in remote or hard-to-reach locations, such as agricultural fields or industrial sites.

5.3.1 Bootloader Implementation

The first stage of development focused on the bootloader, a lightweight program responsible for managing firmware updates and application startup. Its role is to enable the microcontroller unit (MCU) to install and validate new firmware images without requiring external hardware tools.

The YaraSense bootloader currently supports the following capabilities:

- Firmware acquisition via serial interface – allowing initial testing and controlled deployment.
- Partitioned flash memory layout with three dedicated sections:
    1. Bootloader
    2. Active user application
    3. Candidate (new) application
- Firmware validation mechanism – before execution, the bootloader verifies the integrity and authenticity of the new image. If validation fails, the MCU continues running the previous firmware.

5.3.2 Firmware Distribution via MQTT

Once the bootloader foundation was stable, development extended to wireless firmware distribution using the MQTT protocol. MQTT is a lightweight publish-subscribe protocol designed for IoT

communications, offering minimal overhead and robustness in environments with low bandwidth, high latency, or intermittent connectivity.

The first implementation was successfully demonstrated using MQTT over a 4G network. The process includes:

- Cryptographic signing and fragmentation. A custom host-side script generates a signature for the firmware image and splits it into fragments of 3200 bytes.
- Fragment reception and reassembly. The MCU subscribes to update messages, reassembles the fragments into a complete firmware package, and verifies the provider's authenticity using signature validation.
- Fail-safe mechanism. In case of signature mismatch or incomplete download, the MCU reverts to the previously validated firmware, ensuring device continuity.
- Seamless integration with the bootloader. Enabling safe execution of the newly validated firmware image.

5.3.3 Preliminary Results

Initial tests were carried out with firmware packages of approximately 55 kB in size. The following performance metrics were observed:

- Update time: 2–3 minutes, depending on network quality.
- Transmission: 18 fragments of 3200 bytes each.
- Reliability: Bootloader fallback mechanism confirmed successful operation in test scenarios involving corrupted or incomplete updates.

5.4 Scripting engine for far-edge computing

We explored the use of WebAssembly (Wasm) as a scripting engine for far-edge IoT devices. The objective was to enable remote deployment of small, portable modules that can extend device functionality without requiring full firmware updates.

Wasm is an open standard (W3C, 2019) that defines a portable binary instruction format acting as a universal assembly language. Initially designed for web browsers, it is increasingly adopted outside the browser, including embedded microcontrollers. Wasm provides:

- Near-native execution speed with a compact runtime.
- Portability, thanks to its architecture- and OS-independent bytecode.
- Security, via a sandbox model with no direct hardware access unless explicitly exposed by APIs.
- Flexibility, allowing small, incremental updates of device logic.
- Language agnosticism, supporting many languages (C, C++, Rust, AssemblyScript, etc.) compiled to Wasm.

These properties make Wasm highly relevant for far-edge computing: new processing functions can be distributed as small modules, executed securely on the device, and updated over-the-air with minimal bandwidth and energy consumption.

5.4.1 Desktop Prototype

An initial proof-of-concept was developed on Linux, integrating:

- An acquisition loop simulation with MQTT communication.
- The Wasm Micro Runtime (WAMR) from the Bytecode Alliance. The compiled runtime footprint was under 60 kB.
- A CMake/Clang toolchain for compiling C code into Wasm modules.

This PoC, published as EdgeVM on GitHub, validated the feasibility of Wasm-enabled IoT devices and established a simulation framework for module testing.

5.4.2 Embedded Implementation (EdgeSpot, STM32L4, FreeRTOS)

We then ported Wasm runtimes to the EdgeSpot IoT platform, which runs only a real-time operating system (FreeRTOS). Integration steps included:

- Writing a custom API layer exposing only the Wasm interpreter and essential device functions.
- Embedding runtime code into the EdgeSpot build flow.
- Running simple Wasm modules directly on the MCU.

In parallel, a lightweight StreamLite library (based only on <math.h>) was created for simple stream-processing tasks in Wasm modules.

5.4.3 Runtimes Evaluated

Two Wasm runtimes were tested in depth:

Wasm Micro Runtime (WAMR)

- Developed by: Bytecode Alliance (active community, regular contributions).
- Strengths: ARM-compatible, feature-rich, supports multithreading in OS environments, strong documentation.
- Challenges: Designed primarily for OS-based environments; integration with bare-metal/RTOS systems required significant adaptation. CMake build system not designed for easy embedding into existing MCU projects.
- PoC: Linux simulation (EdgeVM) and preliminary EdgeSpot integration.

Wasm3

- Developed by an independent author (less active maintenance but still functional).
- Strengths: Written in plain C, no external dependencies, extremely small footprint, easy to embed by including source files directly in the MCU build. Especially suited to constrained MCUs (Arduino, ESP32).
- Challenges: Limited support and documentation. Stability issues when running under FreeRTOS.
- PoC: EdgeSpot PoC repository demonstrated simple Wasm execution.

5.4.4 Integration Architecture

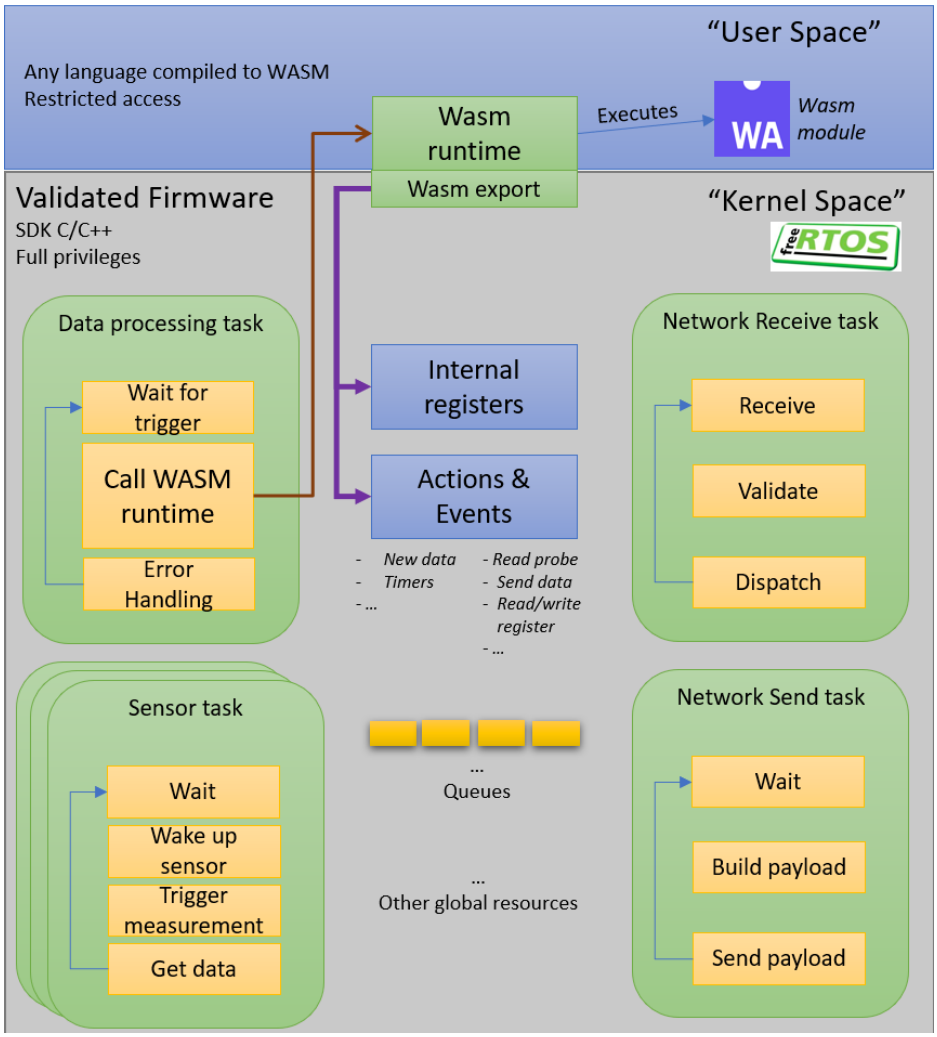

Figure 48. Embedded software architecture for the integration of the WASM engine for far-edge computing in the YaraSense.

The architecture, depicted in Figure 48., separates validated firmware (with full privileges) from Wasm modules (restricted). Key tasks are:

- Data processing: Wait for triggers, call the Wasm runtime, handle errors.
- Sensor management: Wake sensors, trigger measurements, and acquire data.
- Networking: Build and send data payloads.
- Internal registers & events: Enable controlled communication between modules and system resources.

This architecture ensures that only validated firmware controls the hardware directly, while Wasm modules can extend functionality securely through predefined APIs.

5.4.5 Evaluation Results

Advantages

- Wasm modules are architecture- and OS-independent.
- Integration requires exposing only a small, well-defined API.
- Modules can be compiled on a remote machine and deployed directly to MCUs.
- Execution is sandboxed, preventing uncontrolled hardware/system access.
- Modules are very small, enabling efficient over-the-air transfer and update.
- wasm3 in particular has minimal memory footprint and is easy to integrate.

- Multi-language source support broadens the developer ecosystem.

Limitations

- Instability with FreeRTOS: memory leaks, segmentation faults, and unexplained crashes observed.
- API behavior may vary depending on call order and module implementation.
- Portability requires case-by-case integration and extensive testing.
- Frequent segmentation faults with little diagnostic output; difficult to identify whether the fault originated from the runtime, the RTOS, or the Wasm module.
- System crashes can leave the MCU unrecoverable without a reset.
- Remote module update pipeline not yet fully validated.
- Resource constraints: limited RAM, no filesystem, and unreliable buffer transfers under FreeRTOS.
- High discipline required for Wasm module development; debugging is complex due to loss of conventional MCU debug tools.

Conclusions and next steps

Wasm has clear potential for extending device functionality at the far edge in a secure, portable, and lightweight manner. Our PoCs showed that Wasm runtimes are small enough to fit into MCU environments and that simple modules can be executed successfully.
However, significant robustness and integration challenges remain, particularly under FreeRTOS. At present, Wasm runtimes must be treated as experimental on bare-metal RTOS devices.

In the upcoming work, we plan to evaluate *MicroPython* as an alternative scripting engine, with the same goal of enabling dynamic, high-level edge logic but with potentially greater stability and reliability for production use.

6 Sensor Data Catalogue

The information of all sensor data acquired in the RILs will be collected into a common catalogue during the project. This catalogue is an Excel file that will be updated along with the data acquisition campaigns.

As ScaleAgData follows the principle *'as open as possible'*, ScaleAgData supports the goals of the Open Science Policy under Horizon Europe and thus, appropriate open science actions will be implemented as an integral part of its proposed methodology as described in deliverable D1.2 Open Science and Data management Plan.

6.1 Data catalogue Structure

The Excel file for cataloging the data set contains the sheets Definitions & Instructions, Acronyms & Abbreviations, and six sheets named by the RILs. The Definitions & Instructions sheet contains some additional information for filling up the table.

The sensor data set information will be given in the sheets dedicated to each RIL. A simple table format has been selected for inputting the data. The table contains two parts: A) The **Data Sets**, and B) the **Sensors**, defined here as:

Data Set

- A collection of (sensor) data acquired at different times and/or locations. The information recorded should be adequate for 1) identifying all the data sets acquired and used in ScaleAgdata project, and for 2) accessing (within terms defined by applicable licenses) those data sets that will be available during/after the project (e.g. FAIR data).

  Note: A data set may consist of data collected from a set of sensors for a certain time period (e.g. one growing season), as well as for several periods. A data set may also include data from several sensors of different types (e.g., weather data).

Sensors

- Sensor(s) acquired to collect the corresponding data set/data sets. The documentation included may consist of one or several sensors per data set.

The data catalogue tables shall be filled up to the level of detail relevant for the above purposes 1) and 2) given in the data set definition above.

For example, in the case of weather station data, containing air temperature, humidity, wind speed and direction, air pressure etc., it is not relevant to list all the separate sensors in the table, if the corresponding data set contains records of all separate sensor data (including location and acquisition time). In this case filling up the table columns for the data set will be enough. However, for tracking the end-to-end performance of a certain data process, detailed sensor information should be included in the sensor metadata.

Sensor Data Catalogue dated 25.9.2025: 

7 Conclusion

This deliverable summarizes the planning related to the sensor selection, and data acquisition activities, as well as the development of new sensor technology and edge processing methodologies in ScaleAgdata. This document is thus describing the practical implementation of local data collection design in the development of a generic agricultural data integration concept.

The work performed builds upon co-design development approach and follows directly from the workshops arranged in WP2 and their outcomes, namely the vision scenarios, and user stories, and the derived data flows, use cases, RIL backlogs, and the rolling plan described in the deliverable D2.1.

At the time of finalizing this second issue of the deliverable most of the data acquisition campaigns had already reached their second (or third) round and the results from the earlier campaigns could be considered for 2025 and 2026 acquisitions. The feedback from the partners in the first project iteration has facilitated significant progress in the technology development activities: the Neuropublic pesticide sensor array is in the piloting phase, the VTT hyperspectral camera is being used on its second field measurement campaigns in RIL Soil, and the edge processing technology by EGM has greatly advanced in several hardware and software development issues.

The experiences and the information obtained along with the first 30 months of the project will guide the RIL activities and technology developments still ahead in the project.