SpaceSUITE

Impuls AgriTech Call 2025

Digital Twins for Agrifood: from Satellites to Smart Irrigation

2-hour tutorial

Prof. Adriano Camps*

With contributions from Dr. Hong Wang, Prof. Mercé Vall·llossera, Prof. Carlos López, Zhongmin Ma, Gerard Portal, Miriam Pablos

**CommSensLab-UPC, Dept. of Signal Theory and Communications, Universitat Politècnica de Catalunya*

E-mail: adriano.jose.camps@upc.edu

Generalitat de Catalunya Departament d'Empresa i Treball

Cofinançat per la Unió Europea

Tutorial subvencionat pel Departament d'Empresa (**Programa Primer**) i amb el cofinançament del Fons Social Europeu Plus

SpaceSUITE

Motivation: Why Digital Twins for Agriculture?

Massive Water Use

Agriculture consumes approximately 70% of freshwater resources globally, making efficient use critical.

Climate Pressures

Increasing frequency of extreme droughts and rainfall due to climate change destabilizes water availability.

Inefficiencies in Traditional Irrigation

Conventional irrigation methods are reactive, inefficient, and often result in water wastage.

Predictive Capability of Digital Twins

DTs integrate data and models to forecast water needs and optimize irrigation schedules proactively.

What is a Digital Twin?

Dynamic Virtual Replica
A digital twin is a continuously updated virtual representation of a physical system, mirroring its behavior and state.

Origins in Industry
Initially developed for aerospace and Industry 4.0 to simulate engines, machines, and factories.

Real-Time Data Ingestion
Ingests data from sensors, satellites, and models in real time to maintain an accurate state.

Predictive Modeling & Feedback
Runs simulations to forecast future conditions and enables proactive adjustments in physical systems.

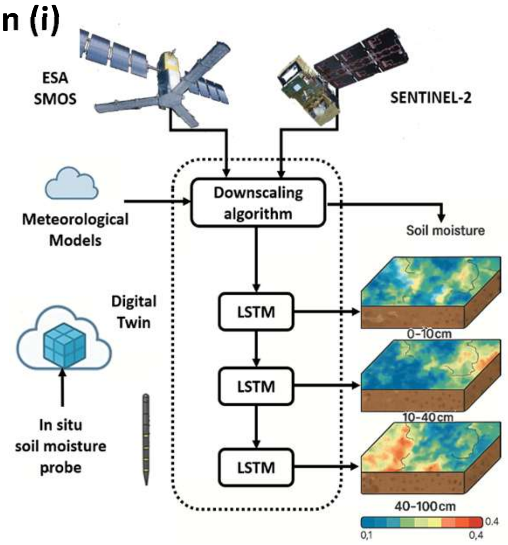
Digital Twins in Agriculture

- **Virtual Crop & Soil Model:** Digital Twins create a real-time, data-driven model of crops and soils, enabling dynamic monitoring and control.
- **Simulation Capabilities:** Supports 'what-if' analyses for irrigation scheduling, fertilizer application, and yield forecasting.
- **Risk Management:** DTs help assess and mitigate risks related to droughts, flooding, and crop diseases.
- **Precision Agriculture Support:** Enhances decision-making with fewer inputs and higher efficiency through integrated data insights.

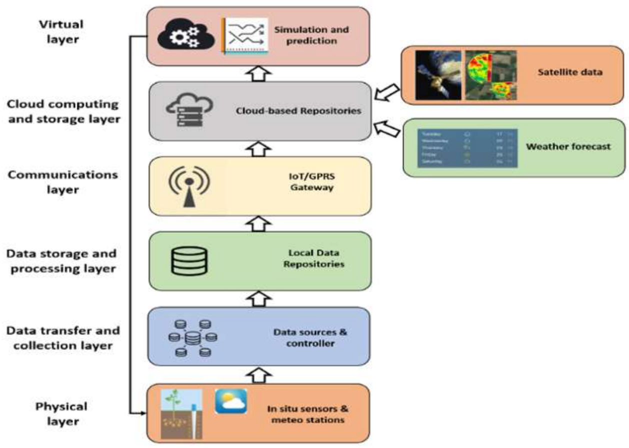
Plantation with green crops growing in agricultural farm [pexels.com]

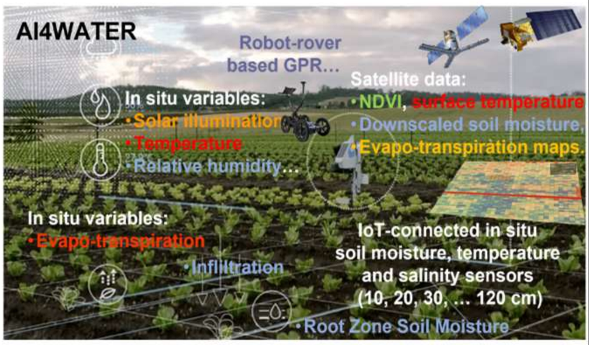
Case Study: AI4WATER Project Digital Twin (*i*)

- **Project Objective:** Developed a Digital Twin for irrigation optimization in Catalonia using multi-source data.
- **Multi-Scale Forecasting:** Predicted soil moisture from field to regional scale using machine learning models.
- **Data Integration:** Merged EO data (SMOS, SMAP), meteorological datasets (ERA5, APIs), and in situ probes.
- **LSTM-Based Prediction:** Used Long Short-Term Memory (LSTM) networks to forecast soil moisture 14 days ahead.



Case Study: AI4WATER Project Digital Twin (ii)





Tutorial Outline

- **Data Sources:** Explore remote sensing, meteorological, and in situ datasets essential for Digital Twins.
- **Data Integration & Management:** Understand how diverse data streams are ingested, stored, and aligned using FAIR principles.
- **Neural Network Training:** Dive into LSTM models trained to predict soil moisture using multi-source inputs.
- **Applications & Insights:** Analyze results from field deployments, including irrigation recommendations and water savings.
- **Conclusions & Future Outlook:** Review key findings and envision future developments like continental-scale DTs.

Traditional Management of Irrigation Water (i)

Plants need water to develop, and only a portion of the total biomass is converted into crops
$\Rightarrow$ You can't produce "more with less"

FIGURA 2
Repartición de la evapotranspiración en evaporación y transpiración durante el periodo de crecimiento de un cultivo anual

(1) $B = WP \cdot \Sigma Tr$

Where, B is the biomass produced cumulatively (kg per m²), Tr is the crop transpiration (either mm or m³ per unit surface), with the summation over the time period in which the biomass is produced, and WP is the water productivity parameter (either kg of biomass per m² and per mm, or kg of biomass per m³ of water transpired).

For most crops, only part of the biomass produced is partitioned to the harvested organs to give yield (Y), and the ratio of yield to biomass is known as harvest index (HI), hence:

(2) $Y = HI \cdot B$

The underlying processes culminating in B and in HI are largely distinct from each other. Therefore, separation of Y into B and HI makes it possible to consider effects of environmental conditions and stresses on B and HI separately.

WATER USE & PRODUCTIVITY

Total cumulative evapotranspiration (ET) of wheat crops typically ranges from 200 to 500 mm, although it can be less in non-irrigated semi-arid areas and reach 600-800 mm under heavy irrigation. The slope of the plot of grain yield vs. ET can be taken as the water productivity in terms of yield and consumptive use ($WP_{Y/ET}$). If the x-intercept of this relationship is taken as a measure of cumulative soil evaporation, then the slope can be interpreted as the water productivity in terms of transpiration ($WP_{Y/Tr}$). On this basis, $WP_{Y/Tr}$ is typically reported to be around 1.0-1.2 kg/m³ (10-12 kg/ha per mm) for grain production (French and Schultz, 1984). An international analysis has indicated the maximum achievable efficiency (for grain) in current wheat systems is likely to be around 2.2 kg/m³ (Sadras and Angus, 2006).

Adapted from "Experiencia en la Mejora de la Eficiencia en el Uso del Agua en Agricultura de Regadío: Tecnología, Información e Implicación de los usuarios", *Vicente Bodas, Albacete, 29/2-1/3, 2024*

SpaceSUITE

Traditional Management of Irrigation Water (ii)

- FAO56 model supported by Remote Sensing

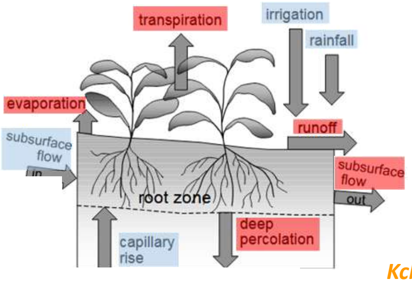
ET = Transpiration + Evaporation

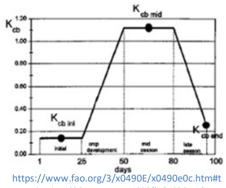
Satellite: Temporal Evolution — Agroclimatology SiAR network[1] / RuralCat[2]

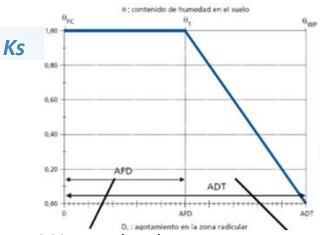
$$ET = \underbrace{Ks}_{\text{Water balance in the root zone}} \cdot \overbrace{Kcb}^{\text{Satellite}} \cdot \overbrace{ETo}^{\text{Agroclimatology}} + \underbrace{Ke}_{\text{Water balance in the surface (evaporation)}} \cdot \overbrace{ETo}^{\text{Agroclimatology}}$$

Water balance in the root zone

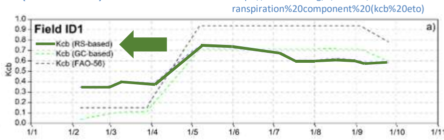
Water balance in the surface (evaporation)







Tarazona de La Mancha (Albacete)



https://www.fao.org/3/x0490E/x0490e0c.htm#transpiration%20component%20(kcb%20eto)

[1] https://servicio.mapa.gob.es/websiar/NecesidadesHidricas.aspx
[2] https://ruralcat.gencat.cat/web/guest/agrometeo.estacions

Adapted from "Evapotranspiración y balance de agua en suelo en cultivos leñosos", *Juan Manuel Sánchez & Jaime Campoy, Albacete, 29/2-1/3, 2024*

SpaceSUITE

Traditional Management of Irrigation Water (iii)

Difference between actual ET and potential ET is a measure of the degree of water stress of the plant.

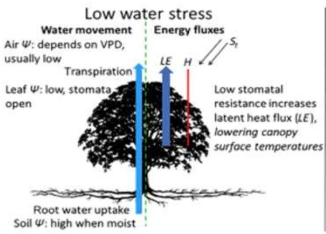
Hydric stress coefficient

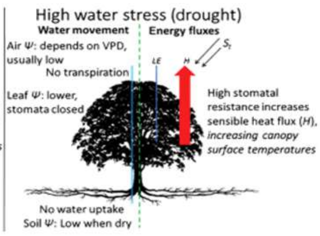
real ET

$$CWSI = 1 - \frac{ET}{ET_c}$$

$$0 < CWSI < 1$$

potential ET

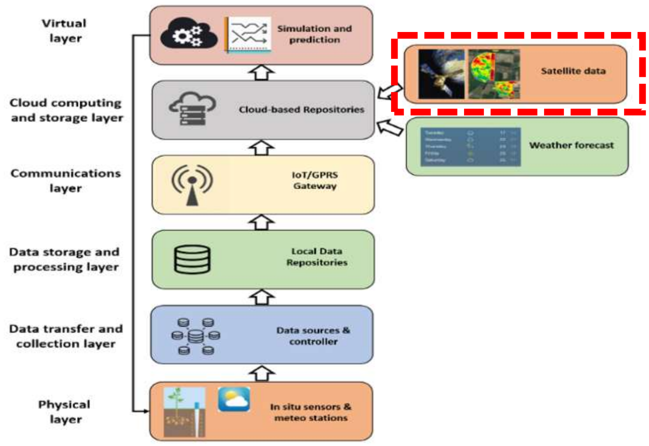




➢ AquaCrop - The FAO Crop Water Productivity Model [https://www.fao.org/aquacrop]

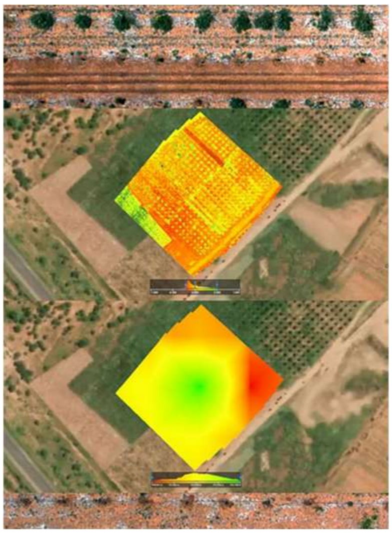
Adapted from "Evapotranspiración y balance de agua en suelo en cultivos leñosos", *Juan Manuel Sánchez & Jaime Campoy, Albacete, 29/2-1/3, 2024*

The Role of Remote Sensing (i)



The Role of Remote Sensing (ii)

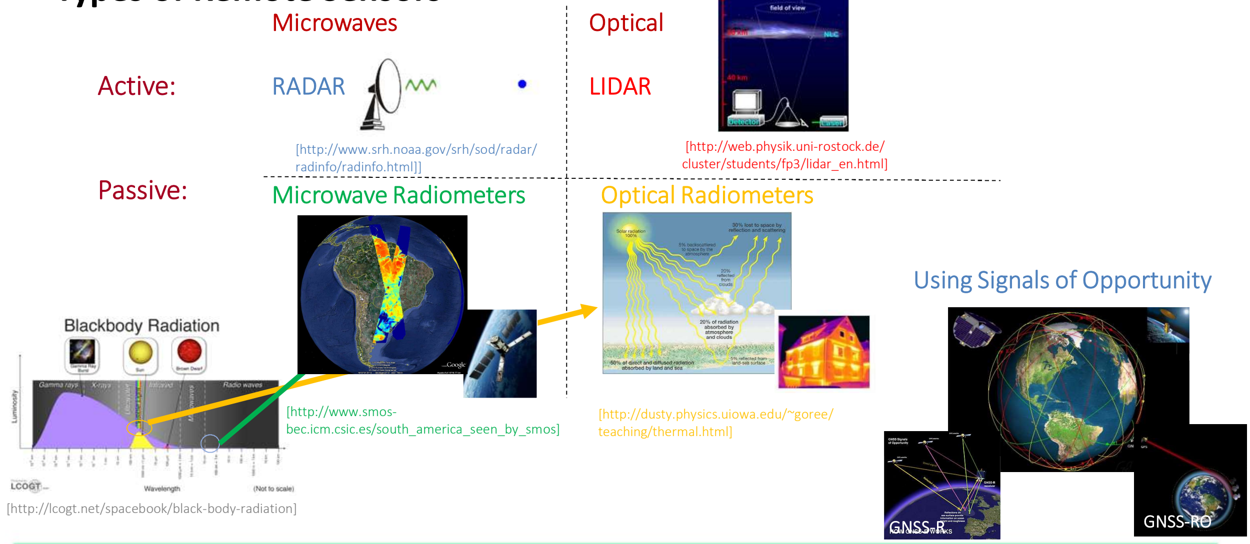
- **Extensive Spatial Coverage:** Satellites observe large areas frequently, providing consistent data for entire regions and enabling macro-scale monitoring.
- **Multiscale Monitoring:** Remote sensing captures vegetation, soil, and hydrological variables across spatial scales—from field plots to entire watersheds.
- **Non-Intrusive & Scalable:** Aerial observations are cost-effective, repeatable, and non-invasive, ideal for monitoring dynamic agricultural landscapes.
- **Critical Input for Digital Twins:** Satellite data serve as essential layers in DTs, complementing in situ and meteorological inputs.



NDVI crop mapping. Credits: poco_bw [istockphoto.com]

SpaceSUITE

Types of Remote Sensors



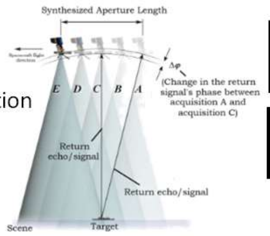
SpaceSUITE

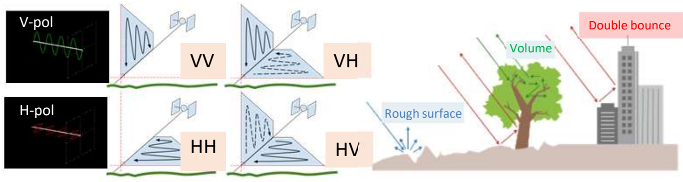
Key satellites for Agricultural Monitoring

- **Sentinel-1 (SAR):** Synthetic Aperture Radar penetrates clouds to monitor soil moisture and detect flooding events with high temporal resolution.
- **Sentinel-2 (Optical):** Captures multispectral imagery for calculating vegetation indices like NDVI and LAI, aiding in crop monitoring.
- **Sentinel-3 (Thermal):** Provides surface temperature and evapotranspiration data crucial for modeling crop water use.
- **SMOS & SMAP (Radiometry):** Dedicated soil moisture missions offering global coverage of topsoil conditions used in drought and irrigation management.
- **Sentinel-6 (Altimetry):** Monitors hydrological systems like rivers and lakes through satellite altimetry.

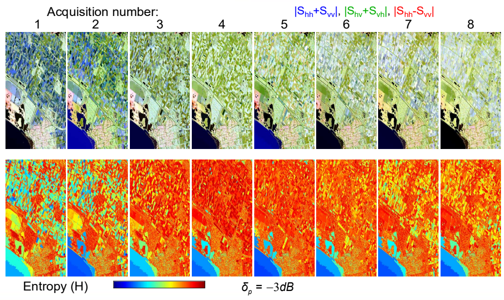
Synthetic Aperture Radar - SAR (i): Sentinel-1

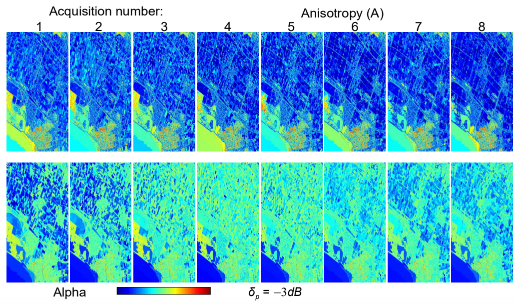
- **SAR Polarimetry** (PolSAR)
- **Multi-temporal** data:
  ⇒ object & change detection





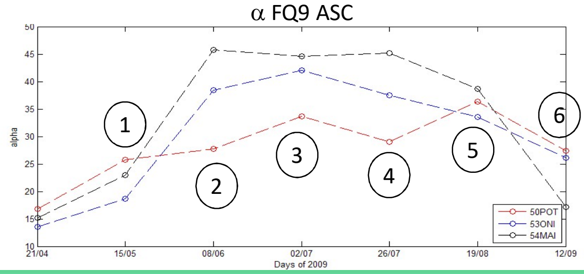
[https://hyp3-docs.asf.alaska.edu/hyp3-docs/guides/introduction_to_sar/]

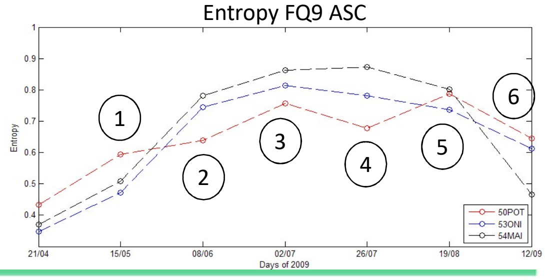




Synthetic Aperture Radar - SAR (ii): Sentinel-1

SAR Polarimetry (PolSAR) & Multi-temporal data for crop monitoring





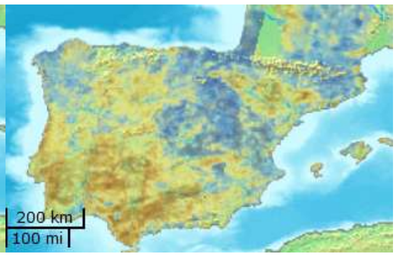
Synthetic Aperture Radar - SAR (iii): Sentinel-1

Daily 1 km **Soil Moisture Index** (SSM)

[https://land.copernicus.eu/global/products/]

- It is not Soil Moisture, it is an “index” from [0, 1]
- Measures reflectivity changes
- SAR is affected by speckle noise $\Rightarrow$ average from 10 m to 1000 m

Single pass: 19/6/2023



6 combined passes: 14-19/6/2023

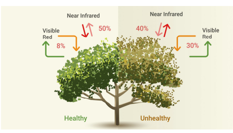
Impuls Agritech · SpaceSUITE · UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH · UPCxels · agrixels · Financiado por la Unión Europea NextGenerationEU · Plan de Recuperación, Transformación y Resiliencia

Optical Sensors: Sentinel-2 and Sentinel-3

[https://physicsopenlab.org/wp-content/uploads/2017/01/veg.gif]

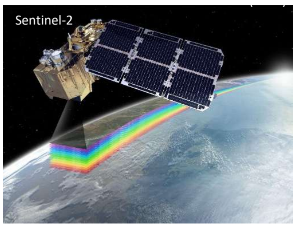
NDVI (Normalized Difference Vegetation Index):

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}}$$



[https://bikeshbade.com.np/media/uploads/2020/05/07/ndvichart_eD4HXBw.png]

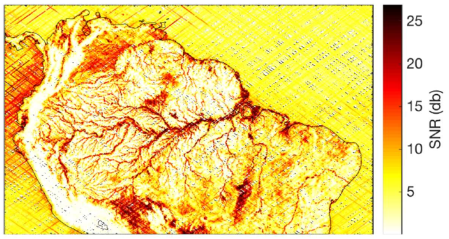
- It does not measure the soil moisture
- It does measure the plant “health”
- It is sensitive to the outer layer
- Many applications for laptop and cell phone



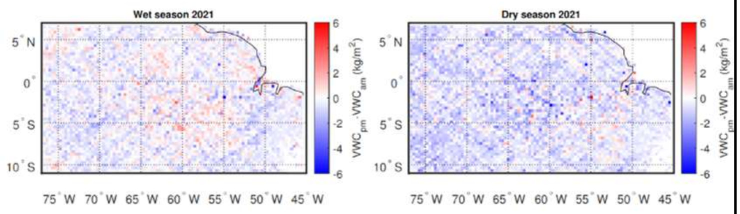
[credits ESA]

GNSS-Reflectometry: NASA CyGNSS (present), and ESA HydroGNSS & Atlantic Constellation (future)

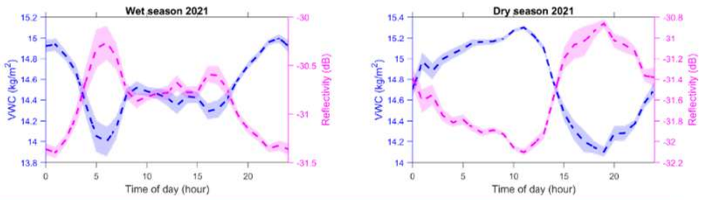
- Reflectometry using satellite navigation signals (**GNSS-R**) allows to estimate **Surface Soil Moisture** and **vegetation opacity** (linked to vegetation water content).
- Average difference between afternoon – morning VWC computed in 2 h Windows centered at 5 h and 17 h, from 12/2000 to 6/2021 (wet season) and 7/2021-11/2021 (dry season), at 36 km resolution.



Vegetation water content (VWC) seasonal cycle

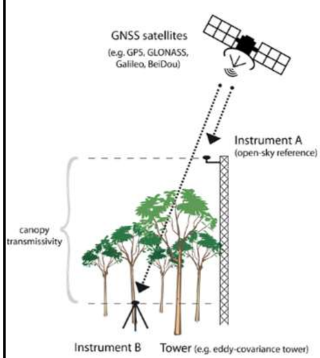


Vegetation water content (VWC) diurnal cycle



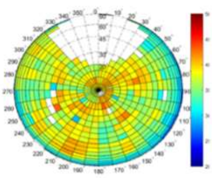
GNSS-Transmissometry (in situ)

- **Transmissometry** using navigation signals (GNSS-T) allows to estimate the opacity of vegetation, which is linked to water content, locally and continuously over time

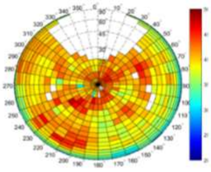


2015/10/22

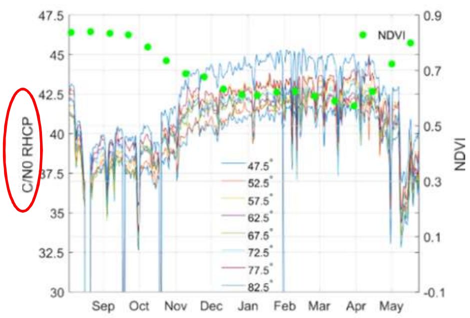
2016/03/31



2015/10/22

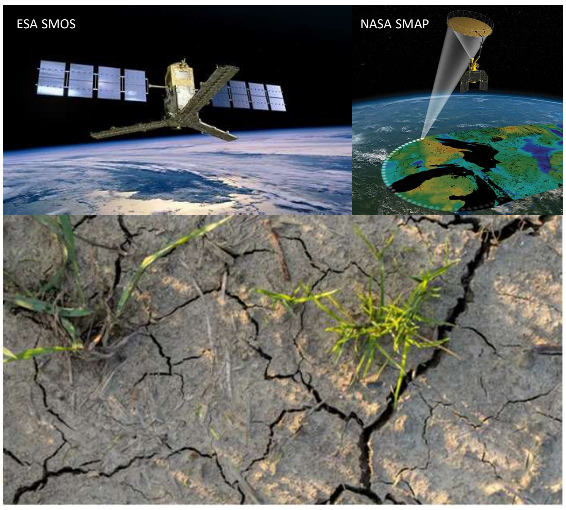


2016/03/31



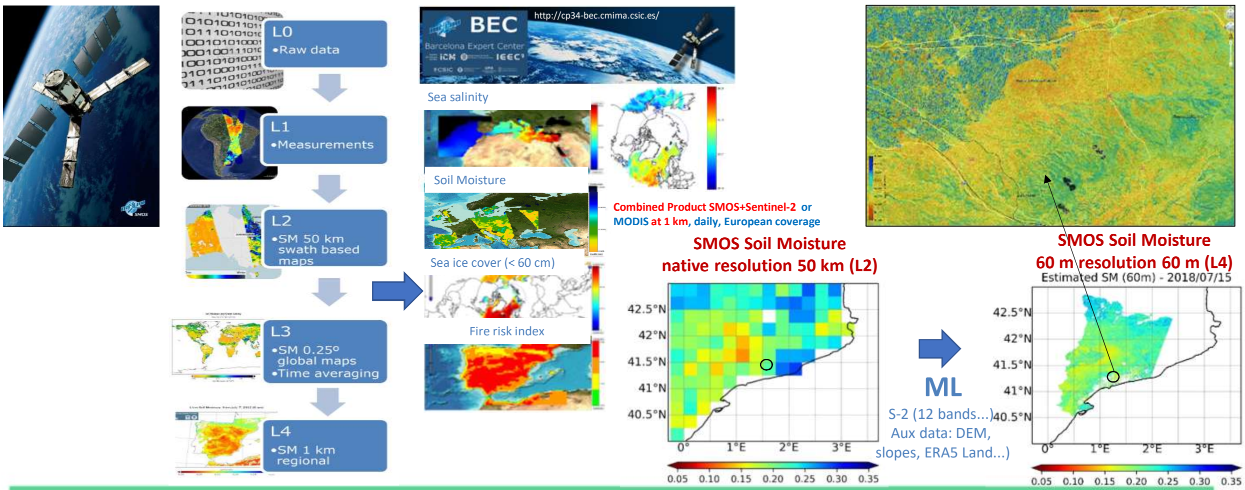
Soil Moisture Missions: SMOS & SMAP (i)

- **SMOS – ESA (2009–):** Provides global surface soil moisture at ~40 km resolution. Uses passive L-band radiometry to observe topsoil dynamics.
- **SMAP – NASA (2015–):** Delivers higher-resolution (~9 km) data on soil moisture. Complements SMOS with improved spatial detail.
- **Applications:** Support drought monitoring, irrigation planning, and early warning systems. Widely adopted in agrifood and hydrology sectors.
- **Limitation:** Only captures the surface layer (0–5 cm), requiring models and in situ data to estimate deeper moisture.



Soil Moisture Missions: SMOS & SMAP (ii)

- L-band Microwave Radiometers: Very sensitive to surface soil moisture and vegetation water content



Credits: Miriam Pablos (UPC) & Gerard Portal (UPC)

Copernicus ERA-5 Land Reanalysis

Long-Term Reanalysis
Provides consistent climate and hydrological data from 1950 to present, using model assimilation of observations.

High Temporal Resolution
Hourly data enables detailed analysis of short-term dynamics like evapotranspiration and rainfall events.

High Spatial Resolution
ERA5-Land offers data at 0.1° (~9 km) resolution, suitable for regional to local-scale modeling.

Water Balance Closure
Delivers key variables (rainfall, temperature, ET) needed to model water balance and soil moisture changes.

Accessing Data via Sentinel Hub

Open Platform
Sentinel Hub provides access to various EO products through user-friendly APIs and web interfaces.

Index Computation
APIs support generation of vegetation indices like NDVI, NDWI, and LAI on-demand.

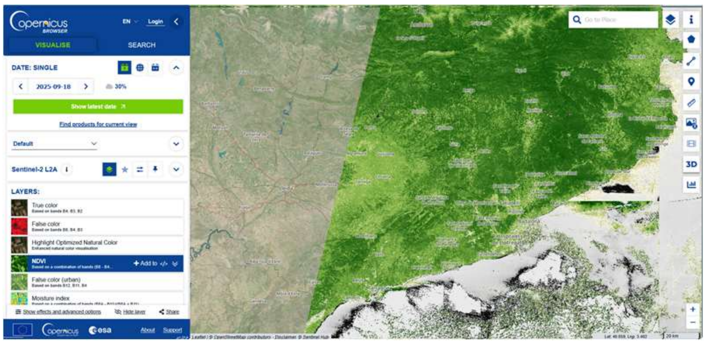
Custom Queries
Users can define areas of interest and time ranges to extract tailored EO data products.

Automation-Ready
Simple HTTP/JSON calls enable integration into automated pipelines and Digital Twins.

[https://browser.dataspace.copernicus.eu/]

Vegetation Indices for Irrigation Insight

- **NDVI (Normalized Difference Vegetation Index):** Quantifies vegetation greenness and photosynthetic activity—widely used for crop health monitoring.
- **EVI (Enhanced Vegetation Index):** Improves sensitivity in high-biomass regions and corrects for atmospheric distortions compared to NDVI.
- **LAI (Leaf Area Index):** Measures canopy density, estimating total leaf surface area—strongly linked to biomass and transpiration.
- **Irrigation Relevance:** Vegetation indices help detect crop stress early, enabling optimized irrigation before visible symptoms emerge.



Sample NDVI Time Series

Impuls Agritech SpaceSUITE UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH UPCxels agrixels Financiado por la Unión Europea NextGenerationEU Plan de Recuperación, Transformación y Resiliencia

NDVI Time Series & Crop Calendar Alignment

Field-Level NDVI Monitoring
Satellite-derived NDVI values are tracked over time to observe crop greenness dynamics and stress responses.

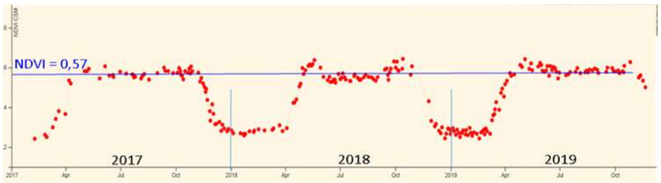
Seasonal Evolution
NDVI curves show clear phases: emergence, growth, peak, senescence—mirroring crop development.

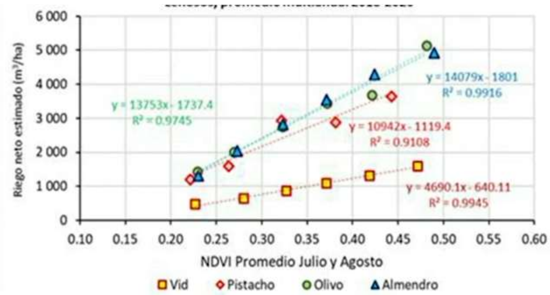
Crop Calendar Comparison
Aligning NDVI with phenological stages helps identify anomalies and plan irrigation around sensitive periods.

Operational Use
These insights guide proactive decisions—e.g., intensifying monitoring during flowering or reducing irrigation post-peak.

NDVI Time Series & Crop Calendar Alignment

Estimated net irrigation ratio vs average NDVI July-August Multiannual Woody Average 2018-2020

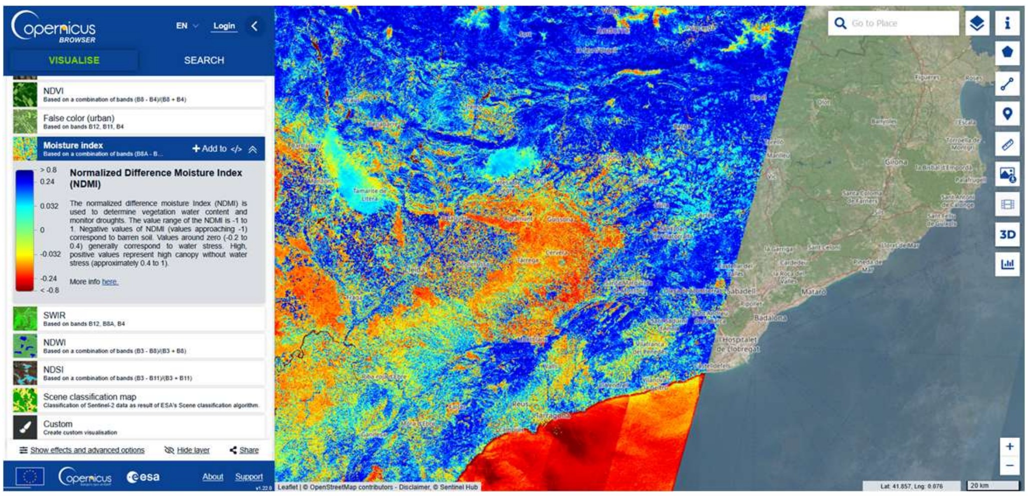




Woody crop	Consumption (m^3/ha)
Olive trees and pistachio (support irrigation)	1500
Almond tree (support irrigation)	1800
Nuts, olive groves and others (0.35 ≤ NDVI < 0.45)	3000
Nuts, olive groves and other intensive irrigation (0.4 ≤ NDVI < 0.55)	4000
Nuts, olive groves and other super-intensive irrigation (NDVI ≥ 0.55)	5000

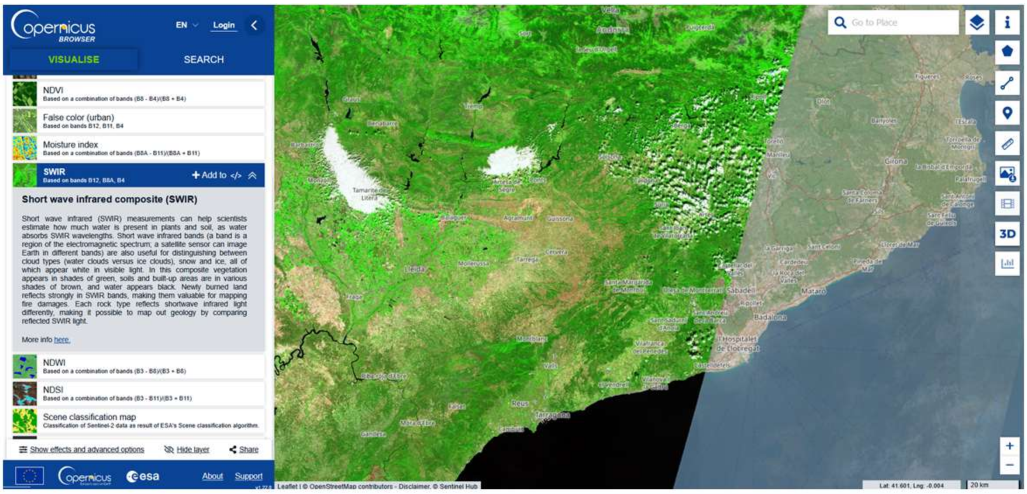
Impuls Agritech SpaceSUITE UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH UPCxels agrixels Financiado por la Unión Europea NextGenerationEU Plan de Recuperación, Transformación y Resiliencia

Other Vegetation Indices (i)



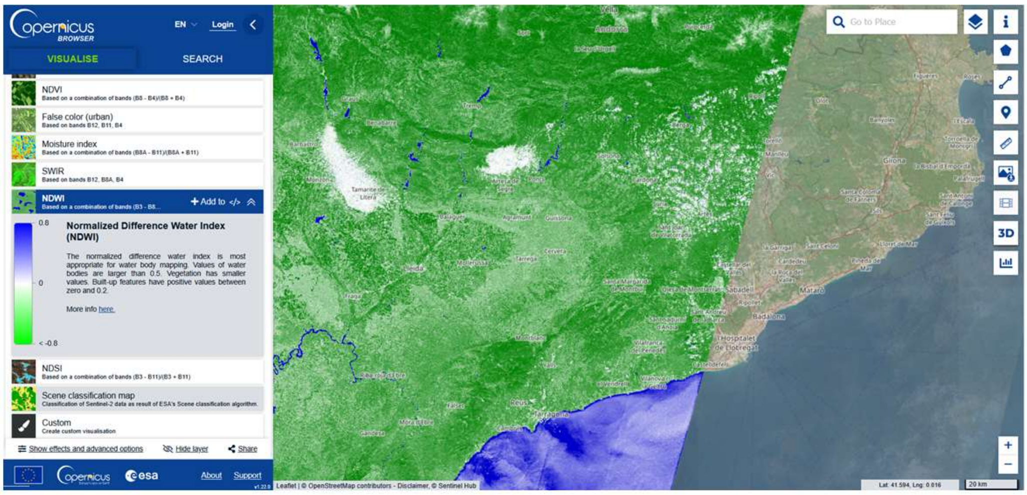
Impuls Agritech SpaceSUITE

Other Vegetation Indices (ii)



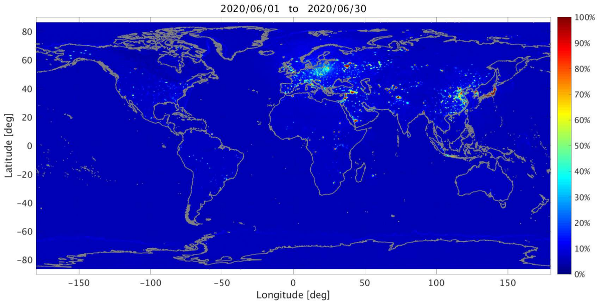
Impuls Agritech SpaceSUITE

Other Vegetation Indices (iii)



Remote Sensing: Data Quality Issues

- **Cloud Obstruction in Optical Data:** Clouds block surface visibility in optical sensors like Sentinel-2, leading to missing or noisy NDVI observations.
- **Radio Frequency Interference (RFI):** L-band radiometry from SMOS/SMAP is impacted by terrestrial radio noise, especially in densely populated areas.
- **Radar Signal Distortion:** SAR-based soil moisture can be distorted by vegetation cover, surface roughness, or dielectric variability.
- **Temporal Gaps:** Satellite revisit times and data dropouts introduce gaps that must be interpolated or modeled.



Evolution of RFI @ L-band over time

Fusion with Ground Data

Downscaling Satellite Data
Satellite soil moisture is coarse (~9–40 km); ground probes help refine estimates to field-scale (e.g., 60 m).

Data Assimilation Techniques
Combining EO with in situ sensors using models improves depth penetration and accuracy of root-zone moisture estimates.

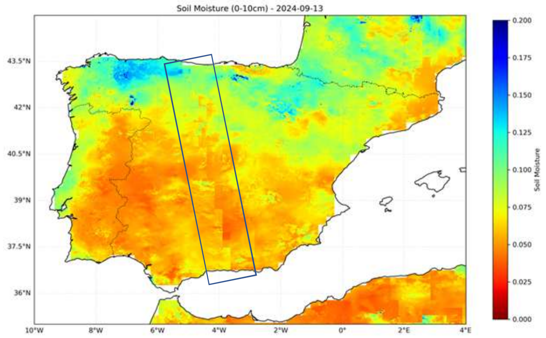
Multi-Sensor Synergy
Merging optical, SAR, and radiometric data creates more robust and reliable soil moisture profiles.

Operational Relevance
Ground-calibrated satellite data support actionable insights in irrigation management and drought monitoring.

Impuls Agritech SpaceSUITE UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH UPCxels agrixels Financiado por la Unión Europea NextGenerationEU Plan de Recuperación, Transformación y Resiliencia

Case Study: Soil Moisture Maps vs. Probe readings

- **Satellite-Derived Soil Moisture:** Maps from SMOS/SMAP show spatial patterns of wet and dry zones at coarse resolution.
- **In Situ Probe Observations:** Point-based measurements offer high precision soil moisture readings at specific depths.
- **Visual Correlation:** Comparison reveals alignment between satellite zones and probe values—validating EO data utility.
- **Training Data for ML:** Combined data serve as training and validation sets for LSTM models in Digital Twins.



Impuls Agritech SpaceSUITE UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH UPCxels agrixels Financiado por la Unión Europea NextGenerationEU Plan de Recuperación, Transformación y Resiliencia

Open Tools for Remote Sensing Data & Digital Twins

- **Google Earth Engine (GEE):** Cloud-based platform for big data EO analysis with access to petabytes of imagery and environmental data.

  [https://earthengine.google.com/]

- **Sentinel Hub:** Offers APIs and GUI tools to access and visualize Sentinel data and compute vegetation indices.

  [https://browser.dataspace.copernicus.eu/]

- **Barcelona Expert Center:** Provides tailored SMOS/SMAP soil moisture products and research-grade EO data services.

  [https://bec.icm.csic.es/]

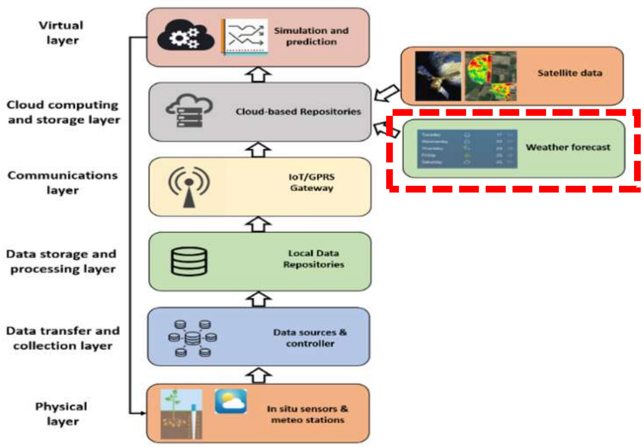
- **Democratizing EO Access:** These platforms enable scalable, open-source EO applications in agriculture, water, and climate.

Summary: Remote Sensing in Digital Twins for Agriculture

- **Provides Large-Scale Observations:** Remote sensing delivers consistent, regional-scale data on soil, vegetation, and hydrology.
- **Complemented by Other Data:** Satellite data must be integrated with meteorological inputs and in situ sensors to cover limitations.
- **Supports Dynamic Monitoring:** Enables frequent updates of crop condition, soil moisture, and evapotranspiration in Digital Twins.
- **Foundation for Predictive Analytics:** Feeds machine learning models and irrigation simulations within the DT framework.

[Picture from Unsplash]

The Role of Meteorological Data (i)



The Role of Meteorological Data (ii)

- **Primary Driver of Soil–Plant System:** Weather variables control water input (precipitation) and output (evapotranspiration) in the soil–crop–atmosphere continuum.
- **Evapotranspiration Modeling:** Temperature, radiation, wind, and humidity determine crop water demand and reference evapotranspiration (ET_0).
- **Forecast Integration:** Weather forecasts support predictive irrigation scheduling by anticipating near-future water needs.
- **Critical for Water Balance:** Accurate rainfall and ET estimates are essential for calculating the soil water budget.

[Picture from Unsplash]

Impuls Agritech SpaceSUITE UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH UPCxels agrixels Financiado por la Unión Europea NextGenerationEU Plan de Recuperación, Transformación y Resiliencia

Meteorological Data Sources

Copernicus ERA5-Land
Historical reanalysis dataset with hourly weather variables since 1950. Provides climate context and model training input.

Open-Meteo API
Real-time and forecast weather data service. Delivers 7–14 day predictions via simple API calls.

Complementary Coverage
ERA5 offers consistency over decades; Open-Meteo ensures up-to-date forecasts for proactive decision-making.

Supports Multiple Applications
Both datasets feed irrigation scheduling, soil moisture modeling, and anomaly detection.

ECMWF, NOAA, national and regional meteorological agencies → May require specific APIs for each

Key Meteorological Variables

Precipitation

Primary water input to soil; drives infiltration, runoff, and recharge. Highly variable and forecast-sensitive.

Temperature

Affects soil evaporation, plant growth rates, and evapotranspiration. Critical for stress modeling.

Solar Radiation

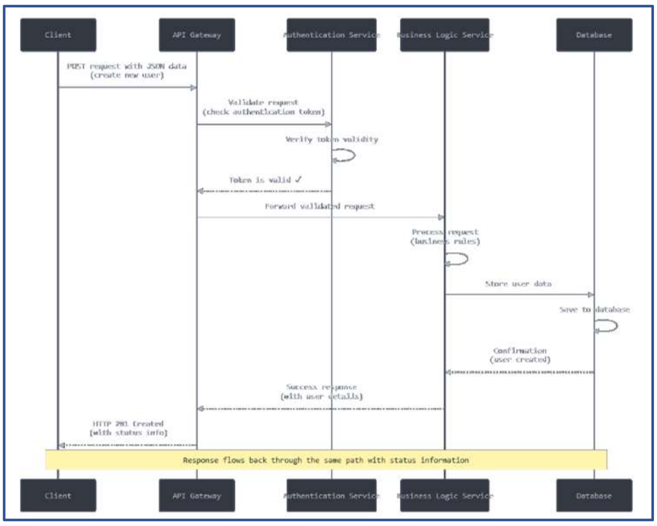
Drives photosynthesis and crop energy balance; key input to ET_0 estimation.

Wind Speed & Humidity

Influence vapor pressure deficit and crop transpiration rate; modulate daily ET values.

Example API Call for Automated Weather Data Retrieval:

1. Client sends a POST request with JSON data to create a new user
2. API Gateway validates the request with the authentication service
3. Authentication Service confirms the token is valid
4. Business Logic Service processes the validated request
5. Database stores the data and returns confirmation
6. The response flows back through the same path with status information



Integration with EO

Aligning Meteorological Forecasts with Satellite Observations

Temporal Harmonization

Forecasts (hourly/daily) must align with infrequent satellite overpasses—this requires resampling, aggregation, and temporal interpolation techniques to ensure consistency in model inputs.

Forecast-Based Forward Simulation

Meteorological forecasts extend EO data utility by enabling forward simulations—LSTM models can ingest both to predict 14-day soil moisture trends for irrigation scheduling.

Data Alignment Strategies

Techniques like time-window matching, lag correction, and feature engineering bridge the gaps between asynchronous EO and forecast datasets, crucial for machine learning pipelines.

Operational Synchronization

For real-time DT operation, automated synchronization scripts ingest EO and meteo inputs with timestamps normalized to server time, ensuring temporal integrity across data streams.

Uncertainty Issues

- **Forecast Horizon Limits:** Forecast accuracy declines sharply beyond 5–7 days, especially for precipitation. Digital Twins must account for this decay to avoid false irrigation triggers.
- **Spatial Representativity Gaps:** Forecast grid cells often span multiple microclimates—resulting in generalized outputs that may misrepresent farm-level variability critical for water decisions.
- **Bias Correction Needs:** Systematic over-/under-estimation in model outputs necessitates bias correction via ground station anchoring or dynamic reweighting in the ML pipeline.
- **Impact on Decision Confidence:** Uncertainty directly affects trust in DT recommendations. Farmers are less likely to follow automated advice unless risk levels and reliability are clearly communicated.

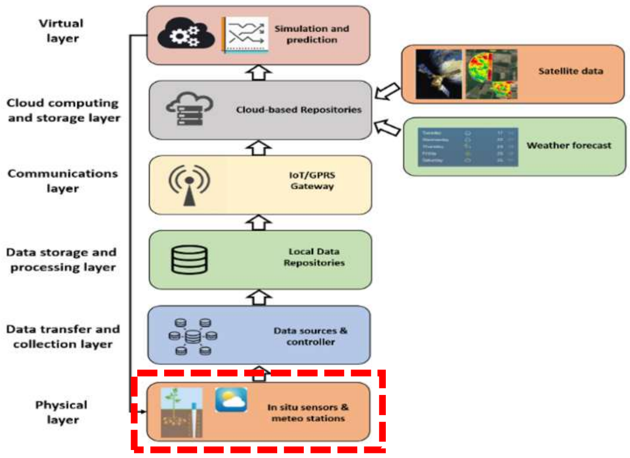
[Picture from Unsplash]

Summary: Meteorological Data in Digital Twins for Agriculture

- **Key Role of Weather:** Meteorological conditions—especially rainfall, radiation, temperature, and wind—drive evapotranspiration and crop water needs, forming the core of DT logic.
- **Data Source Synergy:** Combining historical reanalysis (e.g., ERA5-Land) with real-time forecasts enables backward-looking validation and forward-looking irrigation predictions.
- **Forecast Integration Challenges:** Temporal misalignment, spatial generalization, and model bias require harmonization steps before forecasts can be used effectively in ML-driven Digital Twins.
- **Predictive Capability Boost:** By ingesting forecasts, DTs evolve from descriptive monitoring tools into predictive decision-support systems, enabling proactive irrigation planning.

[Picture from Unsplash]

The Role of Ground Observations (i)



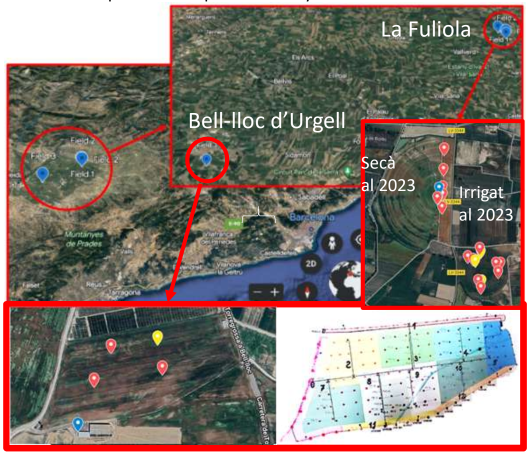
SpaceSUITE

The Role of Ground Observations (ii)

- **High-Resolution, Site-Specific Data:** Probes measure soil moisture, temperature, and salinity at various depths—vital for local model calibration.
- **Temporal Continuity:** In situ sensors provide continuous readings, capturing short-term dynamics not resolved by satellites.
- **Validation & Downscaling:** Used to validate EO data and scale it down to field resolution, bridging remote sensing gaps.
- **Critical for Irrigation Management:** Ground data guide real-time irrigation decisions and alert for thresholds like field capacity or wilting point.

AI4WATER in situ sensors:

- Blue: meteorological stations
- Red: soil moisture/temperature/electric conductivity probes
- Yellow: probes not operational any more



SpaceSUITE

Types of Ground Data

Soil Moisture Probes
Multi-depth sensors (e.g., Sentek Drill & Drop) measure volumetric water content through the root zone.

Weather Stations
Collect temperature, humidity, wind, and rainfall data locally for ET and crop stress modeling.

Plant-Based Sensors
Devices such as dendrometers and leaf-turgor sensors directly monitor plant water status.

Data Logging & Transmission
In situ systems connect via telemetry (LoRa, GSM) to send data to cloud platforms like IrriMAX Live.

SpaceSUITE

Soil Probes (Sentek/Irrimax)

Vertical Profiling
Probes measure volumetric water content at depths such as 0–10, 10–40, and 40–100 cm—critical for understanding root-zone water availability across growth stages.

Precision Irrigation Support
Data from different soil layers inform irrigation decisions tailored to crop root depth and soil retention characteristics, minimizing over- or under-watering.

Real-Time Monitoring
High-frequency updates (15–30 minutes) provid continuous visibility into infiltration dynamics, enabling fine-tuned responses to rainfall or irrigation events.

Training Data for ML Models
These measurements serve as training targets in LSTM-based models, anchoring simulations and improving accuracy in soil moisture forecasting.

20 x IRRIMAX probes: SM, T, EC every 10 cm from 5 to 115 cm

SpaceSUITE

ATMOS-41 Station

Comprehensive Sensing Suite
ATMOS-41 measures precipitation, solar radiation, temperature, humidity, wind speed/direction, barometric pressure—covering all key ET0 variables in one device.

Compact and Solar-Powered
Fully integrated and self-powered via solar panel, the station simplifies deployment in remote fields without requiring grid electricity or complex infrastructure.

Critical for ET0 Calculation
Supplies direct inputs for Penman-Monteith ET0 model—enabling real-time estimates of crop water demand critical for Digital Twin irrigation logic.

Robust Field Performance
Designed for harsh agricultural environments, ATMOS-41 ensures consistent data transmission even under wind, rain, and temperature extremes.

2 x ZENTRA ATMOS 41 meteorological stations: rays, ETo, atmospheric pressure, vapor pressure, precipitation, solar radiation, air temperature…

Impuls Agritech SpaceSUITE UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH UPCxels agrixels Financiado por la Unión Europea NextGenerationEU Plan de Recuperación, Transformación y Resiliencia

Example measurements

Layered Moisture Response

Topsoil (0–10 cm) shows immediate spikes after rain or irrigation, while deeper layers (40–100 cm) respond gradually, reflecting percolation time and soil structure.

Infiltration Dynamics

Time series reveal the rate at which water penetrates through the profile, enabling analysis of soil texture, compaction, and retention across depths.

Event-Based Analysis

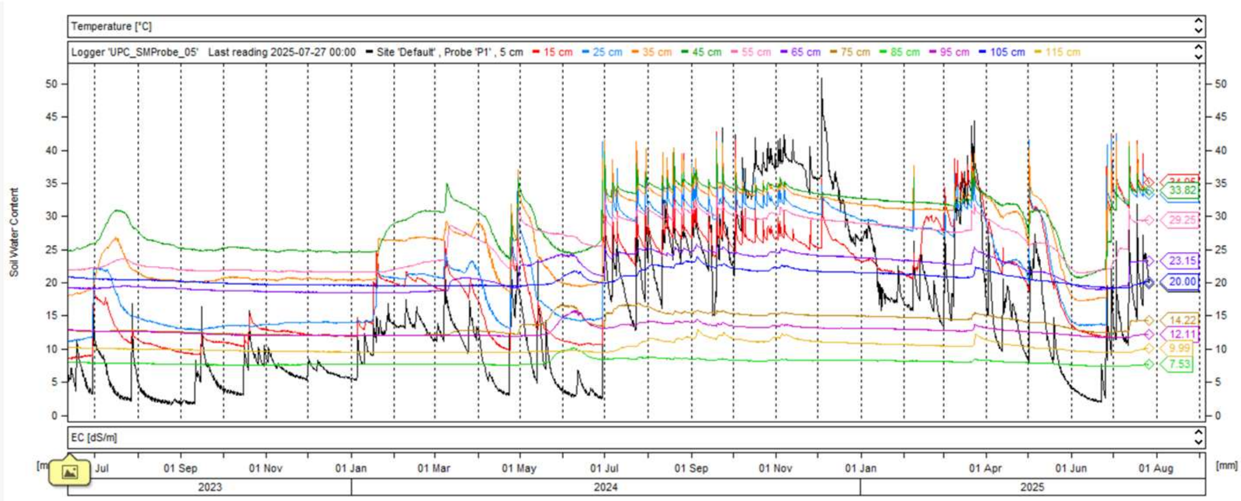
By aligning rainfall timestamps with probe data, we assess how much water was retained vs. lost—vital for evaluating irrigation efficiency.

Model Validation Role

These measurements validate whether LSTM forecasts realistically mimic the timing and amplitude of real-world soil moisture dynamics.

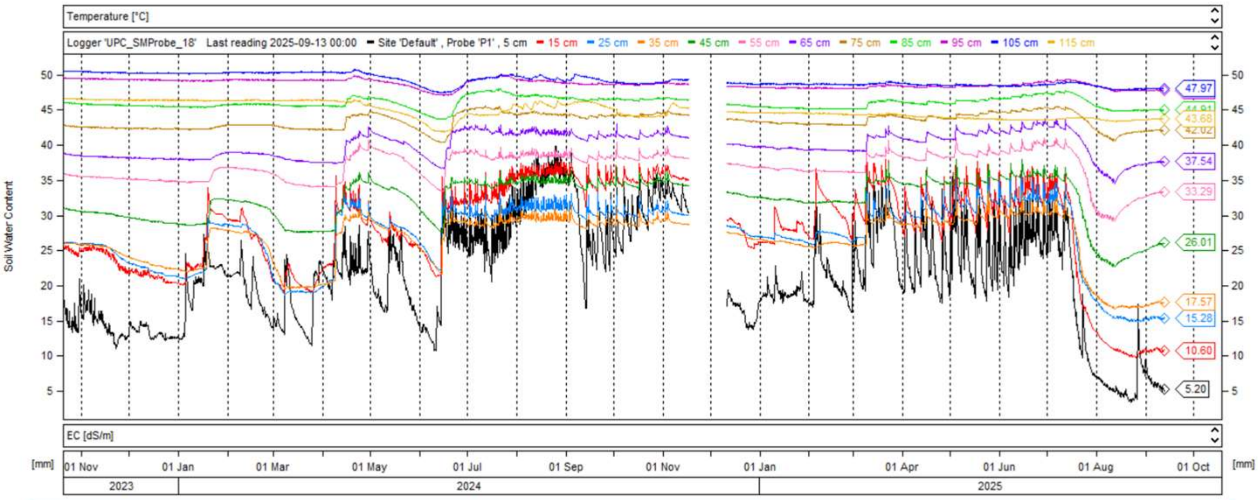
Impuls Agritech SpaceSUITE UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH UPCxels agrixels Financiado por la Unión Europea NextGenerationEU Plan de Recuperación, Transformación y Resiliencia

Example of moisture profiles (rainfed field 2023, La Fuliola)



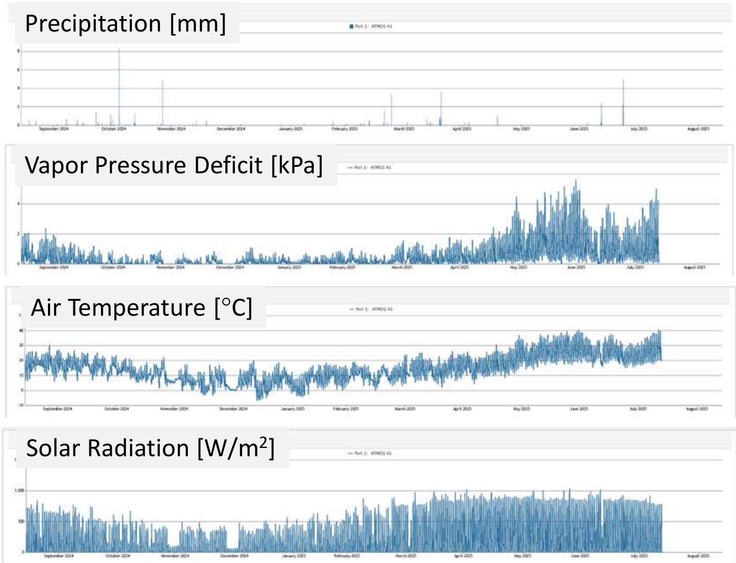
Impuls Agritech SpaceSUITE — UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH — UPCxels — agrixels — Financiado por la Unión Europea NextGenerationEU — Plan de Recuperación, Transformación y Resiliencia

Example of moisture profiles from a probe (irrigated field, Bell-lloc d'Urgell)



Impuls Agritech SpaceSUITE — UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH — UPCxels — agrixels — Financiado por la Unión Europea NextGenerationEU — Plan de Recuperación, Transformación y Resiliencia

Example of meteorological station data – ATMOS 41



Impuls Agritech · SpaceSUITE · Universitat Politècnica de Catalunya BarcelonaTech · UPCxels · agrixels · Financiado por la Unión Europea NextGenerationEU · Plan de Recuperación, Transformación y Resiliencia

Data Integration

- **Training Neural Networks:** In situ soil moisture readings serve as targets for LSTM models, enabling supervised learning based on multi-source inputs (EO, meteo, probes).
- **Bias Correction Anchor:** Sensor data correct systematic errors in satellite- or forecast-derived estimates, ensuring Digital Twins remain accurate and trustworthy.
- **Temporal Synchronization:** Combining asynchronous data sources requires careful timestamp alignment and interpolation to ensure meaningful integration into training datasets.
- **Fusion at Feature Level:** Inputs from EO indices (NDVI, LAI), weather (ET0, rainfall), and probes are fused into feature vectors that capture both spatial and temporal variability.

[Picture from Unsplash]

Impuls Agritech · SpaceSUITE · Universitat Politècnica de Catalunya BarcelonaTech · UPCxels · agrixels · Financiado por la Unión Europea NextGenerationEU · Plan de Recuperación, Transformación y Resiliencia

Challenges

- **Sensor Calibration and Drift:** In situ sensors require frequent recalibration due to electronic drift and environmental exposure. Dust, temperature cycles, and corrosion alter readings over time, demanding periodic field validation.
- **Maintenance Burden:** Physical maintenance is resource-intensive—sensors must be checked for biofouling, cabling faults, and battery depletion, particularly in remote agricultural plots.
- **Spatial Representativeness:** Point sensors measure small soil volumes, which may not represent heterogeneous field-scale conditions such as texture variation or irrigation gradients.
- **Data Continuity Risks:** Power failures, vandalism, or connectivity gaps can create missing data periods, compromising time-series integrity crucial for model calibration.

… a ~2400 € probe smashed by a tractor (top) or animal (bottom)

Impuls Agritech · SpaceSUITE · Universitat Politècnica de Catalunya BarcelonaTech · UPCxels · agrixels · Financiado por la Unión Europea NextGenerationEU · Plan de Recuperación, Transformación y Resiliencia

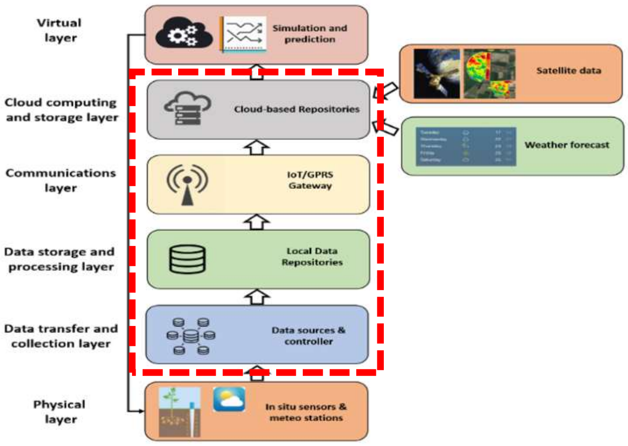
Summary: Ground Observations in Digital Twins for Agriculture

- **Ground Truth Role:** In situ networks anchor digital twins to measurable physical conditions, providing the calibration foundation for all satellite and model-based soil and climate products.
- **Temporal and Spatial Fineness:** High-frequency, localized data allow for resolving sub-daily processes and microclimate effects, far exceeding the revisit frequency of satellites.
- **Validation and Benchmarking:** Ground observations form the benchmark for model validation, ensuring AI-driven predictions and Earth Observation analyses remain tied to empirical data.
- **Integration with Digital Twin Ecosystem:** By linking in situ, EO, and model data streams, systems achieve dynamic calibration—updating the digital twin with real-time field conditions.

[Picture from Unsplash]

Impuls Agritech · SpaceSUITE · Universitat Politècnica de Catalunya BarcelonaTech · UPCxels · agrixels · Financiado por la Unión Europea NextGenerationEU · Plan de Recuperación, Transformación y Resiliencia

Data Pipeline (i)



Impuls Agritech SpaceSUITE UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH UPCxels agrixels Financiado por la Unión Europea NextGenerationEU Plan de Recuperación, Transformación y Resiliencia

Data Pipeline (ii)

Acquisition

Data are collected from in situ sensors, satellites, and meteorological models. Each source operates at different temporal and spatial scales requiring harmonization at ingestion.

Ingestion

Automated ETL (Extract–Transform–Load) processes convert heterogeneous raw data into standardized formats such as NetCDF or GeoTIFF, tagging with metadata for traceability.

Storage

Centralized cloud repositories or distributed data lakes store structured and unstructured data, ensuring redundancy, scalability, and access control.

Dissemination

Processed data products are distributed via APIs, dashboards, and user portals, enabling seamless integration into AI pipelines and Digital Twin visualization layers.

Impuls Agritech SpaceSUITE UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH UPCxels agrixels Financiado por la Unión Europea NextGenerationEU Plan de Recuperación, Transformación y Resiliencia

FAIR Principles in Meteorological Data

Findable

Data must have rich metadata and identifiers so users can locate datasets easily via search and indexing.

Accessible

Data should be retrievable using standard protocols (e.g., APIs), with clear usage licenses.

Interoperable

Datasets must use standardized formats and vocabularies to enable seamless integration across platforms.

Reusable

Clear documentation and provenance ensure that data can be applied in new contexts and long-term studies.

Impuls Agritech SpaceSUITE UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH UPCxels agrixels Financiado por la Unión Europea NextGenerationEU Plan de Recuperación, Transformación y Resiliencia

Data Platforms

Google Earth Engine (GEE)
A planetary-scale platform enabling petabyte-scale analysis of satellite and geospatial data, ideal for machine learning integration and large-scale environmental monitoring.

Barcelona Expert Center (BEC)
Provides specialized ocean and land data products, calibration datasets, and algorithmic validation services through ESA's EO data ecosystem.

AWS and Cloud-Native Storage
Amazon Web Services supports elastic data storage and serverless processing pipelines using services such as S3, Lambda, and SageMaker for scalable EO computation.

Interoperability Layer
Modern platforms increasingly expose OGC-compliant APIs to interconnect analysis environments, enabling cross-platform Digital Twin workflows.

Impuls Agritech SpaceSUITE UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH UPCxels agrixels Financiado por la Unión Europea NextGenerationEU Plan de Recuperación, Transformación y Resiliencia

Data Fusion Challenges

- **Spatial Scale Discrepancy:** Field sensors measure at point or plot scale (~1 m), while satellite data often represent aggregated pixels of 30–1000 m, requiring upscaling or downscaling strategies.
- **Temporal Resolution Mismatch:** Satellites typically provide daily or multi-day revisits, while in situ networks capture sub-hourly variability. Synchronizing these time steps is essential for accurate fusion.
- **Noise and Bias Propagation:** Merging heterogeneous data sources introduces biases and uncertainties that propagate through models unless properly corrected with statistical harmonization techniques.
- **Computational Demands:** Large-scale data fusion across multiple sensors and time steps requires advanced cloud computing resources and optimized algorithms for efficient processing.

[Picture from Unsplash]

AI4WATER Server Architecture

Data Sources
Combines data streams from satellites, in situ sensors, and meteorological networks, forming a continuous inflow into the AI4WATER ecosystem.

Server Core
A centralized processing hub ingests, harmonizes, and stores multi-source data. Edge computing nodes support low-latency processing at sensor locations.

Machine Learning Layer
Deep learning models (e.g., LSTMs) process temporal patterns and spatial correlations, generating soil moisture and irrigation forecasts.

Visualization and Dashboard
Interactive web-based dashboards present predictions, uncertainty metrics, and sensor network status in real time for decision support.

Metadata Standards

STAC Framework
The SpatioTemporal Asset Catalog (STAC) provides a universal metadata schema to describe geospatial datasets, supporting discoverability and interoperability.

Core Elements
Includes asset ID, bounding box, datetime, license, and data links—allowing automated cataloging and API-based search.

Soil Moisture Dataset Example
A STAC Item might define a GeoTIFF asset for soil moisture, with metadata linking to sensor type, temporal coverage, and processing history.

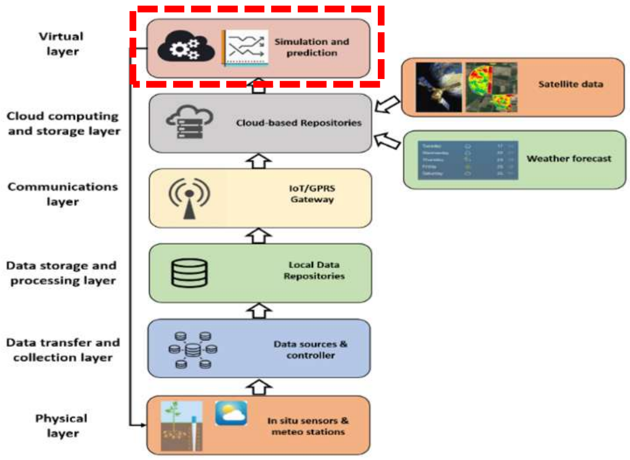
Integration with Digital Twins
STAC metadata enables dynamic data ingestion into digital twin architectures, supporting version control and provenance tracking.

Summary: Data Pipeline

- **Centralized Knowledge Base:** Data servers act as the persistent memory of the Digital Twin, integrating real-time and historical data streams across multiple domains.
- **FAIR and STAC Compliance:** By adhering to FAIR and STAC standards, data remain discoverable, interoperable, and reusable within global digital ecosystems.
- **Fusion and Computation Hub:** Servers host AI models and data fusion processes, transforming heterogeneous inputs into harmonized knowledge representations.
- **Enabling Predictive Intelligence:** Efficient management and access frameworks empower the twin's neural components to forecast, diagnose, and optimize environmental conditions.

[Picture from Unsplash]

Virtual Layer - the “core” of the Digital Twin (i)



Virtual Layer - the “core” of the Digital Twin (ii): Why Machine Learning ?

- Capture Non-Linear Interactions in Agro-Environmental Systems
- No need to “implement” FAO56 models... ML “learns” them from the data

Complex Interactions
Soil moisture dynamics depend on highly non-linear feedbacks between climate, soil properties, vegetation, and management practices that traditional models struggle to capture.

Limitations of Physical Models
Process-based hydrological models are data-hungry and rigid, requiring extensive calibration and often failing to generalize across scales or conditions.

Data Availability
Remote sensing, IoT sensors, and meteorological networks now provide vast, multi-modal datasets ideal for AI-driven inference.

Predictive Flexibility
Machine learning can learn complex dependencies directly from data, enabling adaptive forecasting even under variable climatic regimes.

Impuls Agritech SpaceSUITE UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH UPCxels agrixels Financiado por la Unión Europea NextGenerationEU Plan de Recuperación, Transformación y Resiliencia

Time Series Forecasting Methods

Classical Models
ARIMA, VAR, and Kalman filters model linear temporal dependencies and are effective for short-term forecasting with stationary datasets.

Ensemble Methods
Tree-based algorithms like Random Forest, XGBoost, and CatBoost capture non-linearities and feature interactions without explicit time modeling.

Deep Learning Approaches
Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Transformers can learn long-term dependencies and complex patterns.

Hybrid Techniques
Combining statistical and deep learning methods improves robustness and interpretability in soil moisture and hydrological forecasting.

Why LSTM ?

Capturing Long-Term Dependencies in Environmental Data

Memory Retention
LSTM networks store information across long sequences, overcoming vanishing gradient issues that affect standard RNNs.

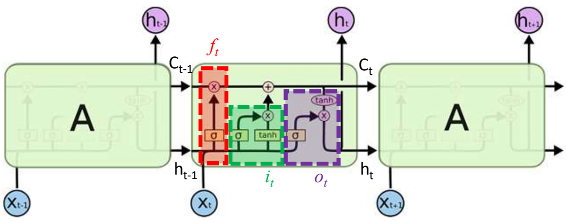
Temporal Context Awareness
Their gating mechanisms—input, forget, and output—enable selective information flow, maintaining relevant dependencies over time.

Multivariate Capability
LSTMs efficiently process multi-dimensional time series including EO, meteorological, and in situ data simultaneously.

Proven Robustness
Widely used across hydrology and agriculture, LSTMs outperform linear models for soil moisture forecasting and drought prediction.

LSTM Architecture [1]

- **Input Layer:** Multivariate inputs (e.g., EO indices, temperature, precipitation) are normalized and encoded as sequential vectors for processing.
- **Hidden LSTM Cells:** Each LSTM cell contains three gates—input, forget, and output—that regulate the flow and retention of information over time steps.
- **Output Layer:** The final dense layer transforms hidden states into continuous outputs, such as predicted soil moisture values for each depth or pixel.
- **Architecture Advantages:** This architecture enables long-term context retention, gradient stability, and scalability across multiple variables and spatial units.



Mathematical formulation

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad \text{forget gate}$$
$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad \text{input gate}$$
$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad \text{candidate state}$$
$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad \text{cell update}$$
$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad \text{output gate}$$
$$h_t = o_t \odot \tanh(C_t) \quad \text{hidden state}$$

Key properties

- The additive update in C_t allows gradients to flow without vanishing.
- Multiplicative gates (f_t, i_t, o_t) act as differentiable logic controls.
- Each gate has independent parameters (W_*, b_*) learned by back-propagation through time.

[1] Hochreiter, S. and Schmidhuber, J. , Long Short-Term Memory. Neural Computation, 9(8), 1735-1780,1997.

Training Setup

Integrating Multi-Source Data for Model Calibration

Input Variables
Earth Observation indices (NDVI, NDWI), meteorological variables (precipitation, evapotranspiration), and in situ soil moisture are combined into unified time-series inputs.

Output Targets
The model predicts soil moisture values at multiple depths and locations, capturing vertical soil dynamics and regional heterogeneity.

Normalization and Scaling
Data are standardized to remove unit biases and to ensure stable training convergence across spatially diverse datasets.

Training Infrastructure
Training is performed on GPU-enabled cloud environments, enabling efficient hyperparameter optimization and fast model iteration.

Training Curves

Evaluating Model Convergence and Generalization

- **Loss Evolution:** Training and validation loss curves are monitored across epochs to ensure convergence and avoid overfitting.
- **Early Stopping:** Training halts automatically when validation loss plateaus, preventing excessive optimization on training data.
- **Learning Rate Scheduling:** Dynamic learning rates stabilize training and promote faster convergence, adapting to gradient changes.
- **Interpretation:** A stable separation between curves indicates good generalization; divergence signals underfitting or overfitting.

Preventing Overfitting

Ensuring Robust and Generalizable Models

- **Cross-Validation:** Data are partitioned into training and validation folds to assess model consistency and detect overfitting tendencies early.
- **Dropout Layers:** Randomly deactivating neurons during training reduces co-adaptation and enhances model generalization.
- **Regularization:** L1/L2 penalties constrain model weights, discouraging excessive complexity and improving interpretability.
- **Data Augmentation:** Temporal shuffling and noise injection broaden the input distribution, mimicking real-world variability.

Validation Metrics

Quantifying Model Accuracy and Performance

Root Mean Square Error (RMSE)
Measures the average magnitude of prediction error. Example: RMSE ≈ 0.0263 at 60 m resolution indicates high model precision.

Mean Absolute Error (MAE)
Captures average deviation from observed values, providing an interpretable measure of forecast reliability.

Mean Squared Error (MSE)
Emphasizes large deviations by squaring errors, making it sensitive to outliers and robust for optimization tuning.

Benchmarking
Comparing these metrics across regions and seasons validates generalizability and model robustness under different climatic conditions.

Training and validation @ 60 m spatial resolution (i)

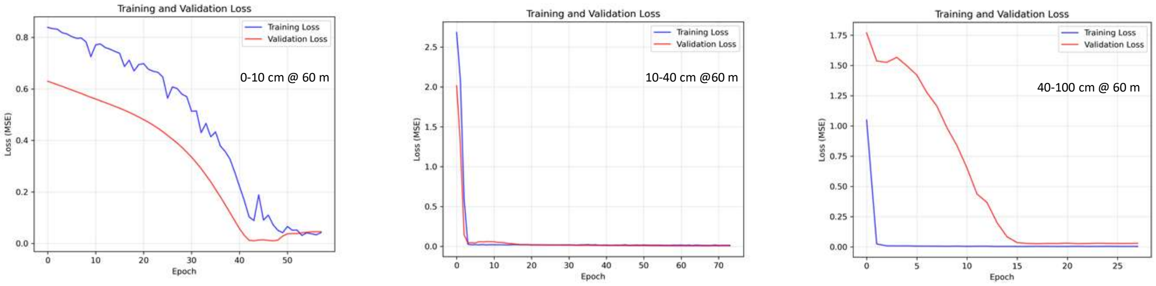
Parameter settings

Layers	Input variables	Learning rate	Batch size	epochs	lookback	Forecast length	Number of layers	dropout	hidden size	Train/test split
0 – 10 cm @ 60m	'0_10cm soil moisture', 'temperature_2m', 'relative_humidity_2m', 'precipitation', 'cloud_cover', 'et0_fao_evapotranspiration', 'wind_speed_10m', 'wind_gusts_10m', 'vapor_pressure_deficit'	0.0001	128	100 (early stop at 58)	7	14	1	0.1	64	81%/19%
10 – 40 cm @ 60m	'0_10cm soil moisture', 'soil_moisture_10_to_40cm', 'temperature_2m', 'relative_humidity_2m', 'precipitation',	0.0001	128	100 (early stop at 20)	1	14	1	0.1	64	81%/19%
40 – 100 cm @ 60m	'0_10cm soil moisture', 'soil_moisture_10_to_40cm', 'soil_moisture_40_to_100cm', 'relative_humidity_2m', 'precipitation'	0.0001	128	100 (early stop at 28)	7	14	1	0.1	64	81%/19%

- Learning Rate: The step size that controls how much the model's weights are updated during each optimization step.
- Batch Size: The number of samples processed together before the model's weights are updated.
- Epochs: The number of times the entire training dataset passes through the model during training.
- Lookback: The number of past time steps used as input to predict the next value in a time series.
- Forecast Length: The number of future time steps the model predicts.
- Number of Layers: The number of stacked LSTM layers in the neural network.
- Dropout: A regularization technique that randomly deactivates a fraction of neurons during training to prevent overfitting.
- Hidden Size: The number of neurons in each LSTM layer, representing the dimensionality of the hidden state.
- Train/Test Split: The proportion of the dataset allocated to training versus validation or testing.

Training and validation @ 60 m spatial resolution (ii)

Training performance



Evaluation metrics

Layers	MSE	RMSE	MAE
0 -10 cm @ 60 m	0.0009	0.0296	0.0233
10 - 40 cm @ 60 m	0.0012	0.0356	0.0296
40 – 100 cm @ 60m	0.0018	0.0430	0.0357

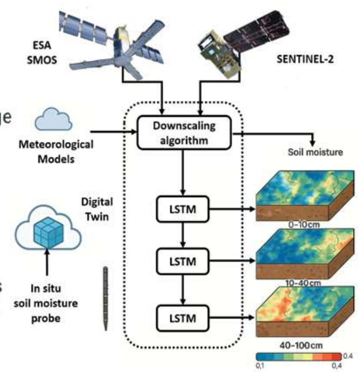
Multi-layer Predictions

Layered Output Structure
Predictions are generated for three depth intervals—0–10 cm, 10–40 cm, and 40–100 cm—representing distinct soil horizons with unique hydrological behavior.

Vertical Continuity Modeling
The LSTM captures inter-layer dependencies, linking surface moisture pulses to deeper storage and percolation effects.

Improved Irrigation Insights
Layer-specific forecasts inform irrigation scheduling, ensuring water reaches the target root zone while avoiding wasteful over-irrigation.

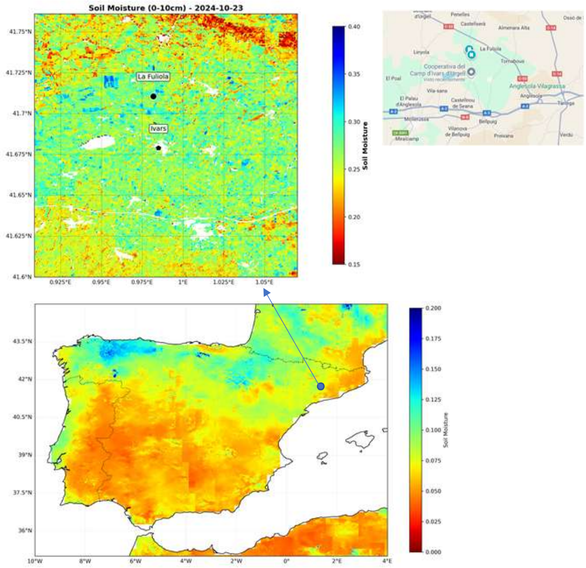
Beyond Surface Data
Subsurface predictions extend the digital twin's situational awareness, enabling long-term soil water balance assessment.



SpaceSUITE

Dataset Resolution

- **Regional-Scale Data (1 km):** Ideal for large-scale monitoring and long-term trend analysis; enables computational efficiency for continental or national DTs.
- **Field-Scale Data (60 m):** Captures fine-grained variability in soil and crop conditions; supports precision agriculture and localized decision-making.
- **Resolution Trade-offs:** Higher resolution improves spatial detail but increases data volume, noise, and training time—requiring careful model design.
- **Multi-Resolution Fusion:** Combining coarse and fine data layers allows simultaneous regional insight and site-specific precision.

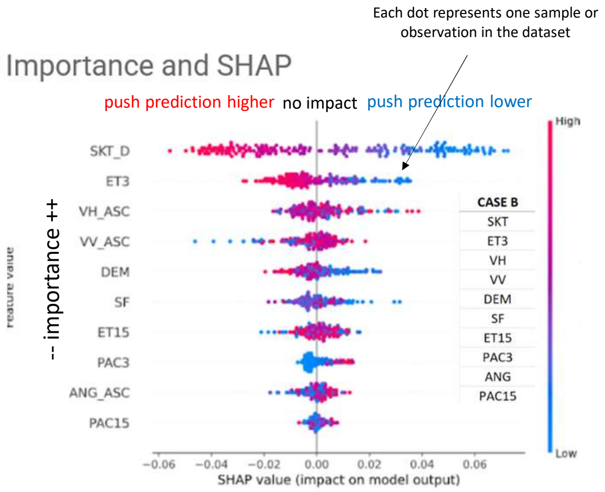


SpaceSUITE

Interpretability

Explaining Model Decisions Using Feature Importance and SHAP

- **Feature Attribution:** SHAP (SHapley Additive exPlanations) quantifies each feature's contribution to a prediction, revealing how inputs like precipitation or EO indices drive results.
- **Variable Importance:** Global importance rankings identify key predictors—often precipitation, temperature, and NDVI—for model transparency and insight.
- **Local Explanations:** Case-level SHAP values visualize the relative influence of each feature on individual forecasts, supporting interpretability in real-time.
- **Decision Support:** Transparent models enhance trust among stakeholders, aligning AI outputs with agronomic expertise and physical reasoning.

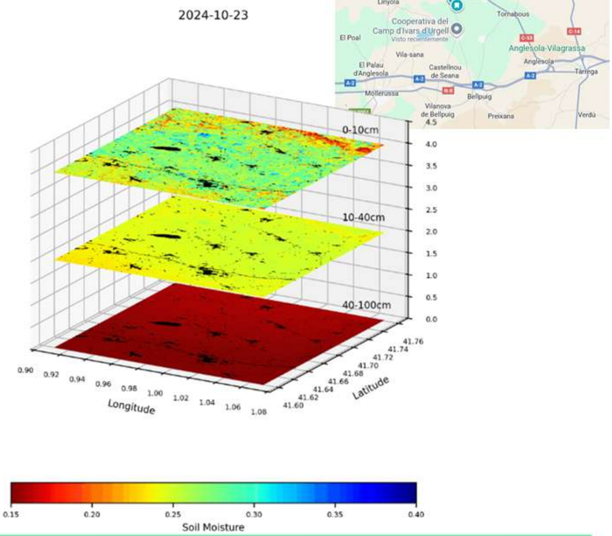


Sample SHAP values of NN to downscale SMOS soil moisture maps (0-10 cm)

Example Predictions

- **Forecast vs Observation:** The LSTM model reproduces the temporal evolution of soil moisture closely, aligning with in situ probe measurements across varying conditions.
- **Pattern Recognition:** Model captures seasonal trends and rainfall-driven spikes, demonstrating ability to generalize under both dry and wet regimes.
- **Uncertainty Representation:** Prediction intervals reflect sensor noise and model uncertainty, offering confidence estimates for decision-making.
- **Operational Relevance:** Time-series predictions feed directly into irrigation scheduling tools and Digital Twin dashboards for actionable insights.

Soil moisture profiles forecast up to 14 days from present



Scalability

Expanding AI4WATER Models Across Regions

From Pilot to Regional Scale
Models initially trained at La Fuliola are generalized and deployed across Mediterranean basins with climate and soil parameter adaptation.

Transfer Learning
Pre-trained weights from pilot models accelerate adaptation to new regions, minimizing the need for full retraining.

Infrastructure Scaling
Cloud-native deployment enables elastic scaling of data ingestion, training, and inference workloads across distributed clusters.

Interoperable Framework
OGC-compliant APIs allow integration of regional Digital Twins, supporting cross-boundary water management initiatives.

Impuls Agritech SpaceSUITE

Training and validation @ 1 km spatial resolution (i)

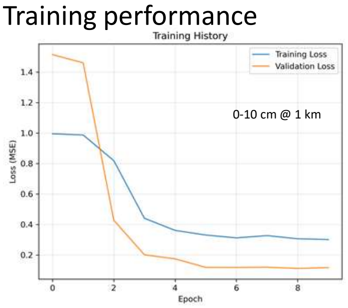
Parameter settings

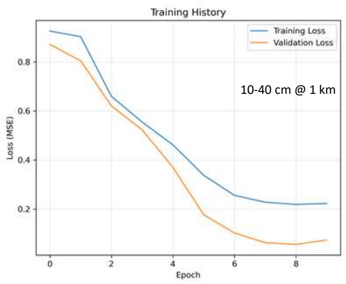
Layers	Input variables	Learning rate	Batch size	epochs	lookback	Forecast length	Number of layers	dropout	hidden size	Train/test split
0_10cm_1km	'0_10cm soil moisture', 'temperature_2m', 'relative_humidity_2m', 'precipitation', 'evapotranspiration', 'vapor_pressure_deficit'	0.0001	64	10	1	14	3	0.3	128	70%/30%
10_40cm_1km	'0_10cm soil moisture', 'soil_moisture_10_to_40cm', 'temperature_2m', 'relative_humidity_2m', 'precipitation', 'evapotranspiration', 'vapor_pressure_deficit'	0.0001	64	10	1	14	3	0.3	128	70%/30%
40_100cm_1km	'0_10cm soil moisture', 'soil_moisture_10_to_40cm', 'soil_moisture_40_to_100cm', 'relative_humidity_2m', 'precipitation', 'evapotranspiration', 'vapor_pressure_deficit'	0.0001	64	10	1	14	3	0.3	128	70%/30%

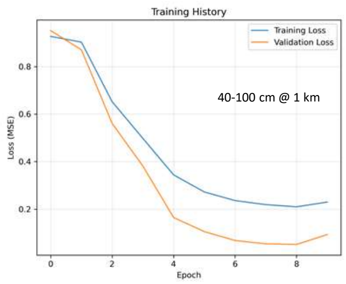
Impuls Agritech SpaceSUITE

Training and validation @ 1 km spatial resolution (ii)

Training performance







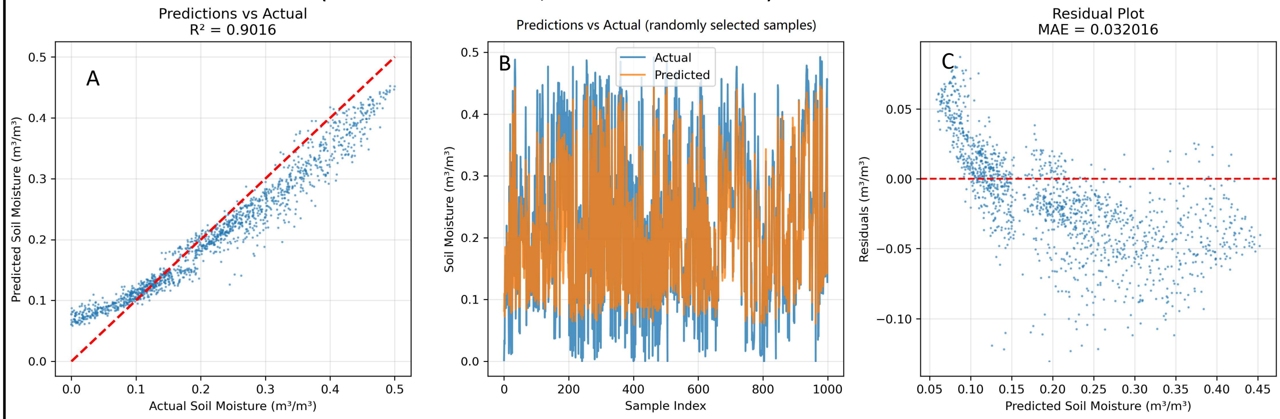
Evaluation metrics

Layers	MSE	RMSE	MAE
0-10 cm @ 1 km	0.0016	0.0402	0.0320
10-40 cm @ 1 km	0.0002	0.0143	0.1003
40-100 cm @ 1 km	0.0003	0.0172	0.0128

Impuls Agritech · SpaceSUITE · Universitat Politècnica de Catalunya BarcelonaTech · UPCxels · agrixels · Financiado por la Unión Europea NextGenerationEU · Plan de Recuperación, Transformación y Resiliencia

Training and validation(1 km spatial resolution)

Predictions vs. Actual (0 – 10 cm @ 1 km, validation dataset)

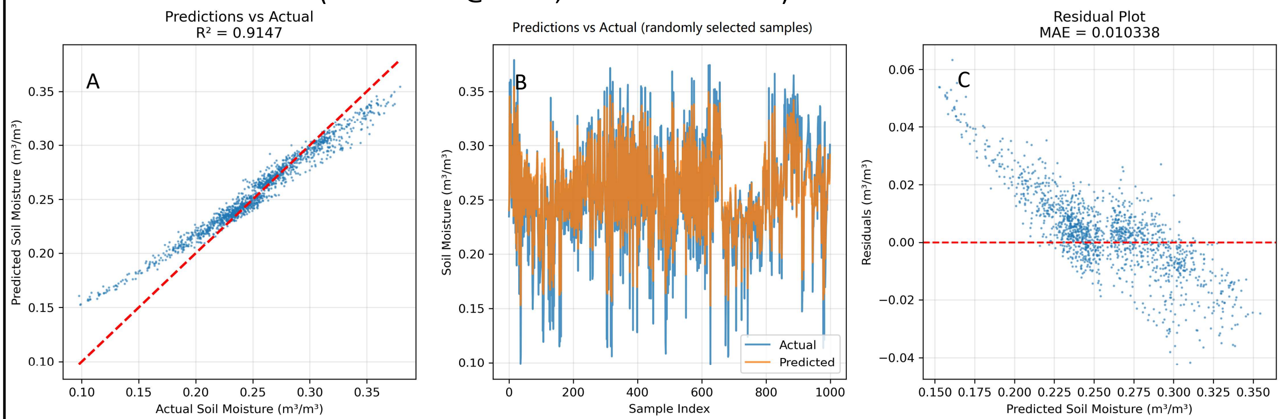


A. Predictions vs Actual. R^2 **is the coefficient of determination**; the red dashed line is the perfect prediction line y=x (1:1 line)
B. Predictions vs Actual. **Randomly selected from the 1000 samples** from validation dataset to show the predictions with their corresponding actual values.
C. Residual plot. **Residuals** are calculated as predictions minus their corresponding actual values; the red dashed line is y=0; **MAE** is the mean absolute value of residuals.

Impuls Agritech · SpaceSUITE · Universitat Politècnica de Catalunya BarcelonaTech · UPCxels · agrixels · Financiado por la Unión Europea NextGenerationEU · Plan de Recuperación, Transformación y Resiliencia

Training and validation(1 km spatial resolution)

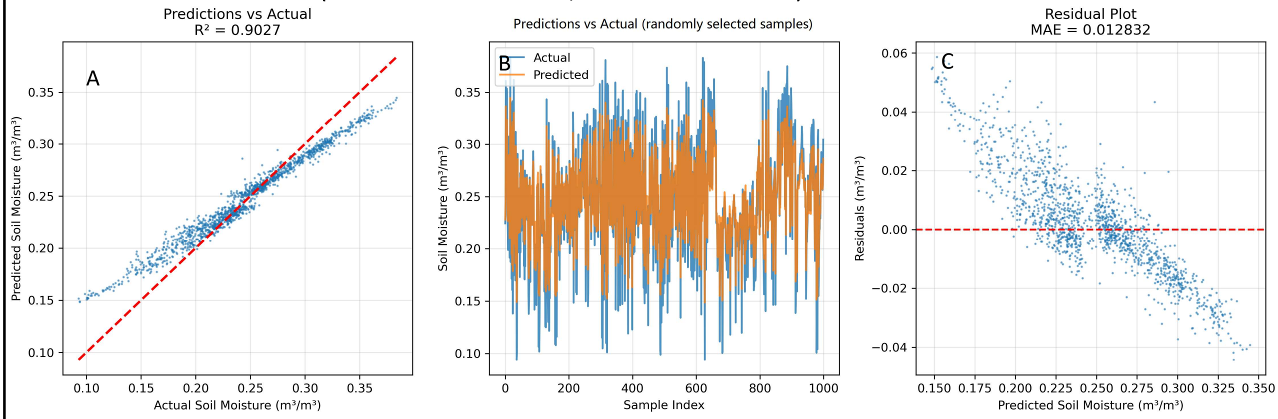
Predictions vs. Actual (10 – 40 cm @ 1 km, validation dataset)



A. Predictions vs Actual. R^2 **is the coefficient of determination**; the red dashed line is the perfect prediction line y=x (1:1 line)
B. Predictions vs Actual. **Randomly selected from the 1000 samples** from validation dataset to show the predictions with their corresponding actual values.
C. Residual plot. **Residuals** are calculated as predictions minus their corresponding actual values; the red dashed line is y=0; **MAE** is the mean absolute value of residuals.

Training and validation(1 km spatial resolution)

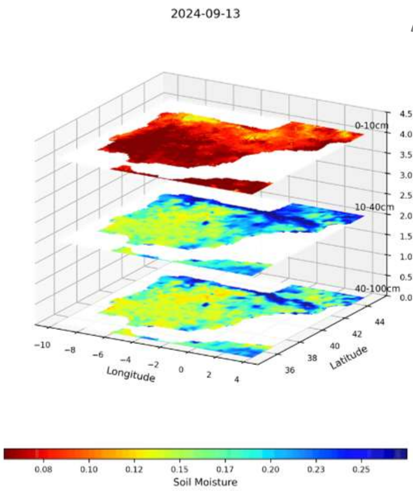
Predictions vs. Actual (40 – 100 cm @ 1 km, validation dataset)

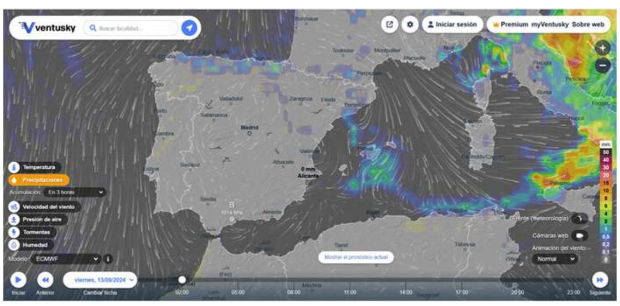


A. Predictions vs Actual. R^2 **is the coefficient of determination**; the red dashed line is the perfect prediction line y=x (1:1 line)
B. Predictions vs Actual. **Randomly selected from the 1000 samples** from validation dataset to show the predictions with their corresponding actual values.
C. Residual plot. **Residuals** are calculated as predictions minus their corresponding actual values; the red dashed line is y=0; **MAE** is the mean absolute value of residuals.

Forecasting Soil Moisture - 14-day SM Predictions for Multi-layer profiles

Actual Rain Rate from Meteo Data





Soil moisture profiles forecast up to 14 days from present

Impuls Agritech SpaceSUITE UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH UPCxels agrixels Financiado por la Unión Europea NextGenerationEU Plan de Recuperación, Transformación y Resiliencia

Summary: Virtual Layer - the "core" of the Digital Twin

LSTM as the Brain

Long Short-Term Memory networks capture temporal dependencies across climatic, soil, and crop variables, forming the cognitive engine of the digital twin.

Multi-Source Learning

Training integrates EO, meteorological, and in situ data streams, enhancing robustness and adaptability across scales.

Predictive and Prescriptive Power

Beyond forecasting, neural models support scenario analysis and decision optimization for water and crop management.

Continuous Calibration

Ongoing ingestion of new observations keeps models dynamically tuned, ensuring sustained performance and relevance.

Impuls Agritech SpaceSUITE UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH UPCxels agrixels Financiado por la Unión Europea NextGenerationEU Plan de Recuperación, Transformación y Resiliencia

Multi-layer Visualization

3D Representation of Soil Moisture Profiles

Depth Profile Mapping

Visualization combines predictions from 0–10, 10–40, and 40–100 cm layers, creating a full 3D soil moisture distribution.

Temporal Evolution

Dynamic animations show moisture movement over time, allowing users to track infiltration and drying cycles.

Data Fusion Display

Combines in situ, EO, and model data to generate geospatially consistent volumetric maps of soil water content.

Decision Support Interface

Interactive dashboards provide cross-sections, time sliders, and statistical summaries for expert users.

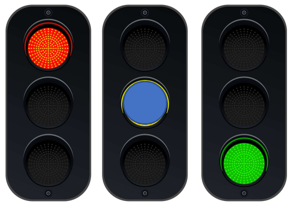
SpaceSUITE

Future Activities:

1. Addition of farmers' irrigation: when, where, how much ?

2. Improved GUI

- **Dashboards for Farmers & Policymakers:** Interactive dashboards provide real-time insights into crop conditions, soil moisture, and irrigation needs. Farmers can access localized data, while policymakers can monitor regional water use efficiency and sustainability metrics.
- **Mobile-Friendly Visualization:** Responsive, mobile-first interfaces ensure accessibility in the field. Farmers can visualize forecasts, receive irrigation alerts, and input on-site data from smartphones or tablets, enabling a seamless human–DT interaction.
- **Data-Driven Decision Support:** The UI integrates AI analytics to recommend optimal irrigation timing, detect anomalies, and visualize risk zones using intuitive color-coded maps and scenario simulations.



Easier interpretation (at a given day from d until d+14 days)

- Red: drier than threshold for more than X days
- Blue: wetter than threshold for more than X days
- Green: SM OK, lower than upper threshold & higher than lower threshold

SpaceSUITE

3. Irrigation Scheduling (i)

Forecast-Based Recommendations
The system suggests optimal irrigation timing and volume using 14-day soil moisture forecasts, reducing waste while preventing crop stress.

Root-Zone Focus
Multi-layer predictions target the active root zone, aligning irrigation depth with crop water demand.

Adaptive Control
Integration with IoT-enabled irrigation systems allows automatic scheduling adjustments in response to updated forecasts.

Resource Efficiency
AI-driven scheduling reduces water use by up to 20–30% without compromising yield or quality.

3. Irrigation Scheduling (ii) – Water savings

- **Reduced Consumption:** Field trials indicate 20–35% reduction in irrigation water use while maintaining or improving yields.
- **Dynamic Optimization:** Forecast-based scheduling aligns irrigation with actual soil moisture dynamics, avoiding unnecessary water application.
- **Energy and Cost Benefits:** Lower water usage translates to decreased pumping energy and operational expenses for farmers.
- **Sustainability Impact:** Water savings contribute to broader sustainability goals under the EU Green Deal and UN SDG 6 (Clean Water).

[Picture from Unsplash]

4. Crop Yield Improvements

- **Optimized Soil Moisture Levels:** Maintaining optimal root-zone moisture reduces crop stress and improves photosynthetic efficiency, leading to higher biomass accumulation.
- **Data-Informed Irrigation:** Predictive irrigation ensures water availability aligns with critical growth stages, maximizing yield potential.
- **Measured Gains:** Field validation shows 5–15% yield increases in cereals and horticultural crops when guided by AI4WATER recommendations.
- **Resilience Enhancement:** Improved soil-water balance enhances drought tolerance and stabilizes yields under variable climatic conditions.

[Picture from Unsplash]

5. Coupling with other Digital Twins – ex. UCo, other countries

SpaceSUITE

Conclusions:

- **Improved Water Management:** Digital Twin forecasts optimized irrigation scheduling, resulting in up to 20% water savings across pilot sites while maintaining or improving crop yields.
- **Operational Efficiency:** Integration of remote sensing, IoT, and AI reduced manual interventions and improved reaction time to drought stress events by 35%, enhancing field-level responsiveness.
- **Resilience & Sustainability:** DT-driven decision support increased system resilience under variable climate conditions and improved compliance with EU water directives and sustainability goals.

Contributing to UN SDG 2 – Zero Hunger

Digital Twins enhance food security by optimizing production efficiency, reducing waste, and supporting sustainable resource management, particularly in water-scarce regions.

Supporting UN SDG 6 – Clean Water

Through precise irrigation forecasting and monitoring, DTs promote water-use efficiency and ensure compliance with integrated water resource management principles.

Advancing the EU Green Deal (Farm-to-Fork)

DT-enabled transparency and traceability across the agrifood chain contribute to the Green Deal's sustainability targets, reducing emissions, chemical inputs, and environmental impact.

SpaceSUITE

Acknowledgements:

- AI4WATER - **Gemelos digitales para agricultura de regadío** (TED2021-131877B-I00)
- GENESIS - **Gnss ENvironmEntal and Societal mISsions** (PID2021-126436OB-C21)
- AGRÍXELS - **Estándar de información agraria orientado a servicios basados en Inteligencia Artificial para casos de uso en agroalimentación** (TSI-100121-2024-19)
- SpaceSUITE - **SPACE downstream Skills development and User uptake through Innovative curricula in Training and Education (**PI-ALL-101140269-SpaceSUITE)

Generalitat de Catalunya Departament d'Empresa i Treball
Cofinançat per la Unió Europea

Tutorial subvencionat pel Departament d'Empresa (**Programa Primer)**
i amb el cofinançament del Fons Social Europeu Plus

To know more about the AI4WATER project: [https://youtu.be/4E_Jd-oXzsw?si=qAdTzdskPQUh48PZ]

Thanks for your attention.

Questions?

Backup slides

STAC Framework

- **Purpose and Structure:** STAC defines a unified JSON-based schema for cataloging spatiotemporal assets (e.g., satellite images, drone captures). Core components: Item (single asset), Catalog (collection), and API (queryable interface).
- **Data Flow Model:** STAC enables discoverability and automation across cloud repositories: ingestion → indexing → search → access. It links metadata, geolocation, and time attributes for scalable Earth Observation pipelines.
- **Mathematical Abstraction:** A STAC Item I can be represented as I = {A, S, T, M}, where A is asset link, S spatial footprint, T timestamp, and M metadata vector. Query: Q = f(S, T, M) → returns subset of Items satisfying filter constraints.

LSTM Architecture

Mathematical formulation

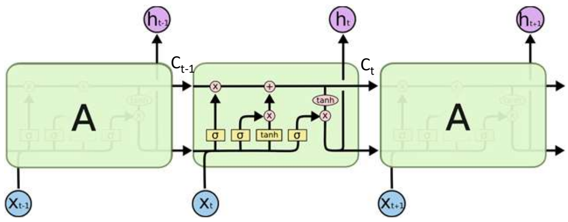
$$
\begin{aligned}
f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) && \text{forget gate}\\
i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) && \text{input gate}\\
\tilde{C}_t &= \tanh(W_c[h_{t-1}, x_t] + b_c) && \text{candidate state}\\
C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t && \text{cell update}\\
o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) && \text{output gate}\\
h_t &= o_t \odot \tanh(C_t) && \text{hidden state}
\end{aligned}
$$

Key properties

- The additive update in C_t allows gradients to flow without vanishing.
- Multiplicative gates (f_t, i_t, o_t) act as differentiable logic controls.
- Each gate has independent parameters (W_*, b_*) learned by back-propagation through time.

Classical cell schematic

- Horizontal line = cell state C_t flowing through time.
- Vertical paths = gates receiving $[h_{t-1}, x_t]$.
- ⊙ nodes = element-wise multiplication; + nodes = additive update.
- Output h_t exits through the top of the block.



- An LSTM cell is a gated dynamical system combining multiplicative and additive pathways: f_t decides what to forget, i_t what to add, o_t what to expose. The memory cell C_t provides constant error flow stabilizing long-term gradient propagation
- The linear recurrence of C_t C_t C_t maintains long-range memory, while gating non-linearities regulate information flow.
- In Digital Twins, LSTMs enable multi-scale temporal forecasting—learning coupled dynamics between climate drivers and observed responses such as evapotranspiration or biomass evolution.

ARIMA (AutoRegressive Integrated Moving Average) model

- ARIMA forms the statistical backbone of time-series forecasting.
- By decomposing a signal into autoregressive and moving-average components after differencing, it captures linear temporal structure.
- In Digital Twin frameworks, ARIMA provides a physically interpretable baseline for process forecasting, anomaly detection, and uncertainty quantification before higher-order ML models are applied.

Mathematical Formulation

A stationary ARMA(p,q) process:

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q},$$

where $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$.

ARIMA(p,d,q) extends this by differencing d times:

$$\nabla^d y_t = (1-B)^d y_t, \quad \Phi(B)\nabla^d y_t = \Theta(B)\epsilon_t,$$

with B the back-shift operator and Φ, Θ polynomials in B.

Interpretation

- **AR(p)** captures persistence (temporal correlation).
- **I(d)** removes trend or seasonality via differencing.
- **MA(q)** models shock propagation through past errors.
- Parameter estimation by maximum likelihood; model order by AIC/BIC.

Applications

- Forecasting climate or hydrological variables driving Digital Twins.
- Baseline for comparing deep models (LSTM, Transformer).
- Hybrid ARIMA-ML pipelines fuse physical interpretability with nonlinear learning.

VAR (Vector Autoregression)

- VAR generalizes the AR model to vector time series, capturing inter-variable feedback loops.
- It enables analysis of how disturbances in one component propagate through the system.
- In Digital Twin applications, VAR provides a transparent statistical framework to diagnose causal relations and simulate multi-sensor co-evolution before embedding into non-linear neural models.

Mathematical Formulation

A VAR(p) model for an n-dimensional time series $y_t = [y_{1t}, \ldots, y_{nt}]'$ is

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \ldots + A_p y_{t-p} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \Sigma)$$

where

- $A_i \in \mathbb{R}^{n \times n}$ are coefficient matrices capturing cross-variable dependencies,
- Σ is the covariance of innovations.

The compact lag-operator form:

$$A(B) y_t = \epsilon_t, \qquad A(B) = I - A_1 B - A_2 B^2 - \ldots - A_p B^p.$$

Stability & Impulse Response

System stability requires det |A(z)| ≠ 0 for |z|≤1.

Impulse-response functions (IRFs) describe the propagation of shocks through time:

$$y_{t+h} = \Phi_h \epsilon_t, \quad \Phi_h = A_1 \Phi_{h-1} + \ldots + A_p \Phi_{h-p}.$$

Applications

- Captures coupled dynamics among multiple climate or sensor variables in Digital Twins (soil moisture, temperature, rainfall).
- Supports forecasting, granger causality, and scenario simulation.
- Forms a linear baseline for multivariate sequence models such as RNNs and Transformers.

Kalman filter

- Kalman filter is a recursive Bayesian estimator operating on linear Gaussian state-space models.
- It alternates between predicting the system evolution and correcting that prediction with new measurements.
- In Digital Twin systems, Kalman filtering ensures dynamic consistency between model forecasts and sensor observations, forming the core of real-time data assimilation pipelines.

Mathematical Formulation

State-space model:

$$\begin{aligned} x_t &= Ax_{t-1} + Bu_t + w_t, \quad && w_t \sim \mathcal{N}(0, Q) \\ y_t &= Hx_t + v_t, && v_t \sim \mathcal{N}(0, R) \end{aligned}$$

Prediction Step

$$\hat{x}_{t|t-1} = A\hat{x}_{t-1|t-1} + Bu_t$$

$$P_{t|t-1} = AP_{t-1|t-1}A^T + Q$$

Update Step

$$K_t = P_{t|t-1}H^T(HP_{t|t-1}H^T + R)^{-1}$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(y_t - H\hat{x}_{t|t-1})$$

$$P_{t|t} = (I - K_tH)P_{t|t-1}$$

Interpretation

- x_t: hidden system state; y_t: noisy observation.
- The filter provides the **optimal linear unbiased estimator** minimizing state error covariance.
- Recursive algorithm combining model prediction and measurement correction.

Applications

- Data assimilation for Digital Twins (soil moisture, temperature, hydrological states).
- Sensor fusion from heterogeneous inputs.
- Smoothing noisy signals and reconstructing latent dynamics.

Recurrent neural Networks

- RNNs introduce recurrence to neural computation, enabling context retention across time steps.
- The network's hidden state acts as a compact sufficient statistic of all past inputs.
- In Digital Twins, simple RNNs provide fast, interpretable baselines for temporal forecasting and serve as precursors to gated and attention-based architectures.

Mathematical Formulation

For input sequence $x_t \in \mathbb{R}^m$ and hidden state $h_t \in \mathbb{R}^n$:

$$\begin{aligned} h_t &= \phi(W_{hx}x_t + W_{hh}h_{t-1} + b_h) \\ y_t &= W_{yh}h_t + b_y \end{aligned}$$

where

- $\phi(\cdot)$ = non-linear activation (tanh / ReLU),
- W_{hx}, W_{hh}, W_{yh} = learned weights,
- h_t carries **temporal context** forward.

Training – Backpropagation Through Time (BPTT)

Gradients are computed recursively:

$$\frac{\partial L}{\partial W_{hh}} = \sum_t \frac{\partial L_t}{\partial h_t} \frac{\partial h_t}{\partial W_{hh}},$$

but repeated Jacobian multiplication can cause **vanishing/exploding gradients**, motivating gated variants (LSTM, GRU).

Interpretation

- RNNs are discrete-time nonlinear dynamical systems.
- Each hidden layer acts as a **memory** that integrates historical input.
- Suitable for short-term dependencies and sequence labeling.

Applications

- Forecasting meteorological sequences (e.g., precipitation ↔ humidity).
- Temporal fusion of satellite + IoT data in Digital Twins.
- Embedding dynamical priors in hybrid physical-ML simulators.

Gated Recurrent Unit (GRU)

- GRUs simplify LSTM design by collapsing forget and input gates into a single update mechanism.
- Its reduced parameterization allows faster convergence with comparable temporal modeling capacity.
- In Digital Twin contexts, GRUs offer an excellent trade-off between accuracy and efficiency for near-real-time forecasting across multi-sensor data streams.

Mathematical Formulation

$$
\begin{aligned}
z_t &= \sigma(W_z[h_{t-1}, x_t] + b_z) && \text{update gate} \\
r_t &= \sigma(W_r[h_{t-1}, x_t] + b_r) && \text{reset gate} \\
\tilde{h}_t &= \tanh(W_h[r_t \odot h_{t-1}, x_t] + b_h) && \text{candidate state} \\
h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t && \text{new hidden state}
\end{aligned}
$$

Interpretation

- **Update gate** (z_t) controls how much new information replaces memory.
- **Reset gate** (r_t) determines how strongly the past influences the candidate state.
- Unlike LSTM, GRU merges cell & hidden states → fewer parameters, faster training.
- Retains long-term context through additive updates, avoiding vanishing gradients.

Applications

- Lightweight alternative to LSTM in real-time Digital Twin forecasting.
- Performs well for medium-length dependencies (e.g., soil moisture trends).
- Integrated in IoT pipelines where computational efficiency is crucial.

Impuls Agritech SpaceSUITE UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH UPCxels agrixels Financiado por la Unión Europea NextGenerationEU Plan de Recuperación, Transformación y Resiliencia

Transformers

- Transformers replace recurrence with attention — enabling direct modeling of global temporal and spatial dependencies.
- By weighing all positions simultaneously, they excel at long-sequence prediction and multi-modal fusion.
- In Digital Twin applications, Transformers can assimilate satellite, sensor, and climate data streams for large-scale forecasting with unmatched interpretability via attention maps.

Core Mathematical Formulation

For an input sequence of embeddings $X = [x_1, x_2, ..., x_T]$,
we compute the **scaled dot-product attention:**

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V,$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V$$

The **multi-head attention** mechanism:

$$\text{MHA}(Q, K, V) = [\text{head}_1, ..., \text{head}_h] W_O, \quad \text{where head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

Each layer includes **residual connections** and **layer normalization:**

$$Y = \text{LayerNorm}(X + \text{MHA}(X, X, X))$$

$$Z = \text{LayerNorm}(Y + \text{FFN}(Y))$$

Architectural Interpretation

- Encoder-decoder structure with stacked self-attention layers.
- Captures global dependencies via attention weights instead of recurrence.
- Position encoding injects order information into embeddings.
- Parallelizable → efficient for long sequences and high-dimensional inputs.

Applications

- Spatiotemporal forecasting in Digital Twins (e.g., climate sequences, satellite time series).
- Fusion of heterogeneous sensor data.
- Foundation for advanced models (Vision Transformers, Temporal Transformers).