Deliverable 7
ATDB and product specifications

EO AFRICA
Water Management

Document Id.:	P22S1956-45-v2.1	19/05/2025
Approved by:	Cristoforo Abbattista	
Checked by:	Giulio Ceriola	
Client Reference:	ESA Contract No 4000139810/22/I-DT	

Document History

List of reviews			
Version	**Authors**	**Date**	**Note**
0	Planetek Hellas, CIHEAMB, Planetek Italia	16/05/2024	1st issue
1.0	Planetek Hellas, CIHEAMB, Planetek Italia	10/09/2024	Revisions
2.0	Planetek Hellas, Planetek Italia	15/01/2025	2nd Issue
2.1	Planetek Hellas	06/05/2025	Revisions, addressed ESA comments on datasets, acquisitions, section 5.5

Table of Contents

Table of Figures

Table of Tables

1. Introduction

This document is the "ATBD and products specifications" and corresponds to the second release of the deliverable D7 according to ESA Contract No. 4000139810/22/I-DT and the Project Proposal P22S1956-02-v0.

1.1. Applicable documents

- Statement of Work and its applicable and reference documents
- "EO Africa Water Management" proposal "P22S1956-02-v0"
- "EO Africa Water Management" Negotiation Points P22S1956-03-v0.1
- "EO Africa Water Management" Minutes of the Preparatory Meeting P22S1956-06-v0
- Contract with ESA 4000139810/22/I-DT
- P22S1956-15-v1.1_D15_EO_AFRICA_EXPLORERS_PMP: Project Management Plan

1.2. Reference documents

- [RD01] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree". Advances in Neural Information Processing Systems 30 (NIPS 2017), pp. 3149-3157.

- [RD02] Buitinck, Lars, et al. "API design for machine learning software: experiences from the scikit-learn project." *arXiv preprint arXiv:1309.0238* (2013).

- [RD03] Taylor, C. Robert (1993). "Dynamic Programming and the Curses of Dimensionality". *Applications Of Dynamic Programming To Agricultural Decision Problems*. Westview Press. pp. 1–10. ISBN 0-8133-8641-1.

- [RD04] Hughes, G.F. (January 1968). "On the mean accuracy of statistical pattern recognizers". *IEEE Transactions on Information Theory*. 14 (1): 55–63. doi:10.1109/TIT.1968.1054102. S2CID 206729491

- [RD05] Trunk, G. V. (July 1979). "A Problem of Dimensionality: A Simple Example". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. PAMI-1 (3): 306–307. doi:10.1109/TPAMI.1979.4766926. PMID 21868861. S2CID 13086902

- [RD06] B. Chandrasekaran; A. K. Jain (1974). "Quantization Complexity and Independent Measurements". *IEEE Transactions on Computers*. 23 (8): 102–106. doi:10.1109/T-C.1974.223789. S2CID 35360973

- [RD07] Allen, R.G. "A Penman for all seasons," *J. Irrig. Drain. Eng.*, vol. 112, no. 4, pp. 348–368, 1986.

- [RD08] Allen, R.G., Pereira, L.S., Raes, D., & Smith, M. *Crop evapotranspiration—Guidelines for computing crop water requirements*, FAO Irrigation and Drainage Paper 56, FAO, Rome, 1998, pp. 300(9), D05109.

- [RD09] Allen, R.G., Tasumi, M., Morse, A., Trezza, R., Wright, J.L., Bastiaanssen, W., Kramber, W., Lorite, I., & Robison, C.W. "Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)—Applications," *J. Irrig. Drain. Eng.*, vol. 133, pp. 395–406, 2007.

- [RD10] Baldocchi, D.D. "Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: Past, present, and future," *Glob. Chang. Biol.*, vol. 9, no. 4, pp. 479–492, 2003.

- [RD11] Bastiaanssen, W.G.M., Noordman, E.J.M., Pelgrum, H., Davids, G., Thoreson, B.P., & Allen, R.G. "SEBAL model with remotely sensed data to improve water-resources management under actual field conditions," *J. Irrig. Drain. Eng.*, vol. 131, pp. 85–93, 2005.

- [RD12] Calera, A., Campos, I., Osann, A., D'Urso, G., & Menenti, M. "Remote Sensing for Crop Water Management: From ET Modelling to Services for the End Users," *Sensors*, vol. 17, no. 5, 2017. DOI: 10.3390/s17051104.

- [RD13] Chen, J.M., & Liu, J. "Evolution of evapotranspiration models using thermal and shortwave remote sensing data," *Remote Sensing of Environment*, vol. 237, pp. 111594, 2020.

- [RD14] Doorenbos, J., & Kassam, A.H. *Yield response to water*, Irrigation and Drainage Paper, vol. 33, pp. 257, 1979.

- [RD15] Doorenbos, J., & Pruitt, W.O. *Guidelines for predicting crop water requirements*, FAO Irrigation and Drainage Paper No. 24, 1977.

- [RD16] El-Shirbeny, M.A., Ali, A.M., Savin, I., Poddubskiy, A., & Dokukin, P. "Agricultural water monitoring for water management under pivot irrigation system using spatial techniques," *Earth Systems and Environment*, vol. 5, no. 2, pp. 341–351, 2021.

- [RD17] Ershadi, A., McCabe, M.F., Evans, J.P., Chaney, N.W., & Wood, E.F. "Multi-site evaluation of terrestrial evaporation models using FLUXNET data," *Agric. For. Meteorol.*, vol. 187, pp. 46–61, 2014.

- [RD18] Fuglie, K., Dhehibi, B., El Shahat, A.A.I., & Aw-Hassan, A. "Water, policy, and productivity in Egyptian agriculture," *American Journal of Agricultural Economics*, vol. 103, no. 4, pp. 1378–1397, 2021.

- [RD19] Gebremedhin, M.A., Lubczynski, M.W., Maathuis, B.H., & Teka, D. "Deriving potential evapotranspiration from satellite-based reference evapotranspiration, Upper Tekeze Basin, Northern Ethiopia," *Journal of Hydrology: Regional Studies*, vol. 41, pp. 101059, 2022.

- [RD20] Guo, D., Westra, S., & Maier, H.R. "An R package for modelling actual, potential and reference evapotranspiration," *Environmental Modelling & Software*, vol. 78, pp. 216–224, 2016.

- [RD21] Hargreaves, G.L., Hargreaves, G.H., & Riley, J.P. "Agricultural benefits for Senegal River basin," *J. Irrig. Drain. Eng.*, vol. 111, no. 2, pp. 113–124, 1985.

- [RD22] Hargreaves, G.H., & Samani, Z.A. "Reference crop evapotranspiration from temperature," *Appl. Eng. Agric.*, vol. 1, no. 2, pp. 96–99, 1985.

- [RD23] Irmak, S., & Irmak, A. "Reference and crop evapotranspiration in South Central Nebraska. II: Measurement and estimation of actual evapotranspiration for corn," *J. Irrig. Drain. Eng.*, vol. 134, no. 6, pp. 700–715, 2008.

- [RD24] Irmak, S., Howell, T.A., Martin, D.L., Payero, J.O., & Copeland, K.S. "Variability analyses of alfalfa-reference to grass-reference evapotranspiration ratios in growing and dormant seasons," *J. Irrig. Drain. Eng.*, vol. 134, no. 2, pp. 147–159, 2008.

- [RD25] Jackson, R.D., Idso, S.B., Reginato, R.J., & Pinter Jr, P.J. "Canopy temperature as a crop water stress indicator," *Water Resources Research*, vol. 17, no. 4, pp. 1133–1138, 1981.

- [RD26] Jensen, M.E., & Haise, H.R. "Estimating evapotranspiration from solar radiation," *J. Irrig. Drain. Div.*, vol. 89, no. 4, pp. 15–41, 1963.

- [RD27] Jensen, M.E., Robb, D.C.N., & Franzoy, C.E. "Scheduling irrigations using climate-crop-soil data," *J. Irrig. Drain. Div.*, vol. 96, no. 1, pp. 25–38, 1970.

- [RD28] Jensen, M.E., Wright, J.L., & Pratt, B.J. "Estimating soil moisture depletion from climate, crop and soil data," *Trans. ASAE*, vol. 14, no. 5, pp. 954–959, 1971.

- [RD29] Beyer, K.; Goldstein, J.; Ramakrishnan, R.; Shaft, U. (1999). "When is "Nearest Neighbor" Meaningful?". *Database Theory — ICDT'99*. LNCS. Vol. 1540. pp. 217–235. doi:10.1007/3-540-49257-7_15. ISBN 978-3-540-65452-0. S2CID 206634099.

- [RD30] Majozi, N.P., Mannaerts, C.M., Ramoelo, A., Mathieu, R., & Verhoef, W. "Uncertainty and sensitivity analysis of a remote-sensing-based Penman–Monteith model to meteorological and land surface input variables," *Remote Sensing*, vol. 13, no. 5, 882, 2021.

- [RD31] Makkink, G.F. "Examination of Penman's revised formula," *Netherlands J. Agric. Sci.*, vol. 5, no. 4, pp. 290–305, 1957.

- [RD32] Masri, M.I., et al. "Effect of water stress and fertilization on yield and quality of sugar beet under drip and sprinkler irrigation systems in sandy soil," *Int. J. Agric. Sci.*, vol. 5, no. 3, pp. 414–425, 2015.

- [RD33] McCabe, M.F., Ershadi, A., Jimenez, C., Miralles, D.G., Michel, D., & Wood, E.F. "The GEWEX LandFlux project: Evaluation of model evaporation using tower-based and globally-gridded forcing data," *Geosci. Model Dev. Discuss.*, vol. 8, pp. 6809–6866, 2015.

- [RD34] Michel, D., Jiménez, C., Miralles, D.G., Jung, M., Hirschi, M., Ershadi, A., Martens, B., McCabe, M., Fisher, J.B., & Mu, Q. "The WACMOS-ET project – Part 1: Tower-scale evaluation of four remote-sensing-based evapotranspiration algorithms," *Hydrol. Earth Syst. Sci.*, vol. 20, pp. 803–822, 2016.

- [RD35] Miralles, D.G., Brutsaert, W., Dolman, A.J., & Gash, J.H. "On the Use of the Term 'Evapotranspiration'," *Water Resources Research*, vol. 56, no. 11, 2020. DOI: 10.1029/2020WR028055.

- [RD36] Nichols, J., Eichinger, W., Cooper, D., Prueger, J., Hipps, L., Neale, C., & Bawazir, A. "Comparison of Evaporation Estimation Methods for a Riparian Area," College of Engineering, University of Iowa, Iowa City, IA, USA, p. 50, 2004.

- [RD37] Penman, H.L. "Vegetation and hydrology," *Soil Sci.*, vol. 96, no. 5, pp. 357, 1963.

- [RD38] Priestley, C.H.B., & Taylor, R.J. "On the assessment of surface heat flux and evaporation using large-scale parameters," *Mon. Weather Rev.*, vol. 100, no. 2, pp. 81–92, 1972.

- [RD39] Thornthwaite, C.W., & Holzman, B. "The role of evaporation in the hydrologic cycle," *Eos, Trans. Am. Geophys. Union*, vol. 20, no. 4, pp. 680–686, 1939.

- [RD40] Vinukollu, R.K., Wood, E.F., Ferguson, C.R., & Fisher, J.B. "Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches," *Remote Sens. Environ.*, vol. 115, pp. 801–823, 2011.

- [RD41] Wanniarachchi, S., & Sarukkalige, R. "A review on evapotranspiration estimation in agricultural water management: Past, present, and future," *Hydrology*, vol. 9, no. 7, pp. 123, 2022.

- [RD42] Walter, I.A., et al. "ASCE's standardized reference evapotranspiration equation," in *Watershed Management and Operations Management 2000*, pp. 1–11.

- [RD43] Williams, L.E., & Ayars, J.E. "Grapevine water use and the crop coefficient are linear functions of the shaded area measured beneath the canopy," *Agric. For. Meteorol.*, vol. 132, no. 3–4, pp. 201–211, 2005.

- **WEBSites references:**
    - https://earsc.org/
    - https://gdal.org/en/latest/
    - https://git.earthdata.nasa.gov/projects/LPDUR/repos/ecostress_swath2grid/browse
    - https://pypi.org/project/pyresample/

1.3. Acronyms

AARSE	African Association of Remote Sensing of the Environment
AI	Artificial Intelligence
ATDB	Algorithm Theoretical Baseline Documents
AWS	Amazon Web Services
cal/val	Calibration and Validation
CIHEAM-IAMB	International Center for Advanced Mediterranean Agronomic Studies - Mediterranean Agronomic Institue of Bari
CHIME	Copernicus Hyperspectral Imaging Mission
CSW	Catalogue Service for the Web
CWSI	Crop Water Stress Index
DAME	Data Intensive Technologies for Multi-mission Environments
DIAS	Data and Information Access Services
DSS	Decision Support System
EA	Early adopter
EARSC	European Association of Remote Sensing Companies
ECOSTRESS	ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station
EO	Earth Observing or Earth Observation
EO AFRICA	African Framework for Research Innovation, Communities and Applications in Earth Observation
EODC	Earth Observations Data Cube
ESI	Evaporative stress index

ET	Evapotranspiration
ETa	actual evapotranspiration
ETc	crop evapotranspiration
ETo	Potential Evapotranspiration
FAIR	Findability, Accessibility, Interoperability, and Reusability
FAO	Food and Agriculture Organization of the United Nations
GDP	Gross Domestic Product
GeoTIFF	Geographic Tagged Image File Format
GIS	Geographic Information System
GUI	Graphical User Interface
ICT	Information and Communication Technologies
IDE	Integrated development environment
ISS	International Space Station
ITT	Invitation To Tender
IWRM	Integrated Water Resources Management
JPL	Jet Propulsion Laboratory
Kc	Crop coefficient
KO	Kick-Off
Ky	Yield response factor
LAI	Leaf Area Index
LST	Land Surface Temperature
LSTM	Land Surface Temperature Monitoring
M2M	Machine to Machine
NARSS	National Authority for Remote Sensing & Space Sciences
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NoR	Network of Resources
OGC	Open Geospatial Consortium
OSAP	Sixth of October Agricultural Projects Company
PKH	Planetek Hellas
PKI	Planetek Italia
PRISMA	Hyperspectral Precursor of the Application Mission
R&D	Research and Development
SARE	remote Sensing Approach to estimate the Reference Evapotranspiration
SDG	Sustainable Development Goals
SEB	surface energy balance
SME	Small and Medium-sized Enterprise
SoW	Statement of Work
SRL	Societal Readiness Level
TDR	Time Domain Reflectometry
TIR	Thermal Infrared Radiation
TRL	Technology Readiness Level
UAV	Unmanned Aerial Vehicle
USGS	United States Geological Survey
VI	Vegetation Index
VIS/NIR	visible/near-infrared
VM	Virtual Machine
VNIR/SWIR	Visible and Near-Infrared / Short-wave infrared
Wb	Water productivity
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service
WP	Work Package
WUE	Water Use Efficiency
Ya	Actual harvested yield
Ym	Potential harvested yield

2. Background

In regions with arid climates and sparse vegetation, such as Egypt, inefficient water consumption poses a significant challenge to agricultural sustainability. Agriculture, being one of the most water-intensive sectors, suffers greatly from this issue, leading to considerable water loss.

The aim is to address this challenge by leveraging thermal Earth Observation data to accurately estimate crop water consumption, specifically through actual evapotranspiration (ETa). This data-driven approach offers valuable insights into the true water requirements of crops, enabling the optimization of irrigation practices for more effective and sustainable water resource utilization.

Accurate ETa estimation is crucial for effective water management in irrigation systems, water resource planning, allocation, and enhancing water use efficiency. By providing insights into crops' consumptive water use, decision-makers can better allocate available water resources to different land uses. This, in turn, can lead to improvements in irrigation scheduling, water allocation, and overall efficiency, thereby enhancing agricultural productivity while conserving precious water resources.

In-situ measurements of these variables are considered accurate but are complex, expensive, and labor-intensive and therefore cannot be afforded over large areas and in many developing contexts of the world. Conversely, the current unprecedented availability of free and open EO satellite data has been recognized as the most feasible mean to provide, in near-real-time, temporally and spatially continuous information for monitoring large cropped areas necessary to adopt both evidence-based governance instruments (e.g. water pricing policies) and managerial actions.

2.1. Scope and Limitations

The integration of space data into agricultural practices represents a transformative tool for supporting farmers and planners in optimizing irrigation water management. By providing small farmers with access to satellite data, the primary goal is to enhance sustainability and increase farmers' income through informed decision-making processes. This encompasses leveraging satellite data to anticipate and mitigate crop stress caused by water losses, particularly through evapotranspiration. Furthermore, the adoption of satellite data facilitates precision irrigation planning and scheduling, promoting optimized water use efficiency across agricultural landscapes.

In addition to aiding individual farmers, the utilization of satellite-derived information informs the development of more effective policies concerning regional water management, allocation, and distribution. This approach aims to ensure equitable and efficient utilization of water resources, contributing to broader agricultural sustainability initiatives.

However, it's important to recognize the inherent limitations in this approach. Many farmers may lack awareness of the full capabilities offered by satellite technology, leading to challenges in effectively harnessing its potential benefits. Additionally, farmers may encounter difficulties in identifying the most suitable solutions for their specific needs, necessitating targeted training initiatives to bridge this knowledge gap.

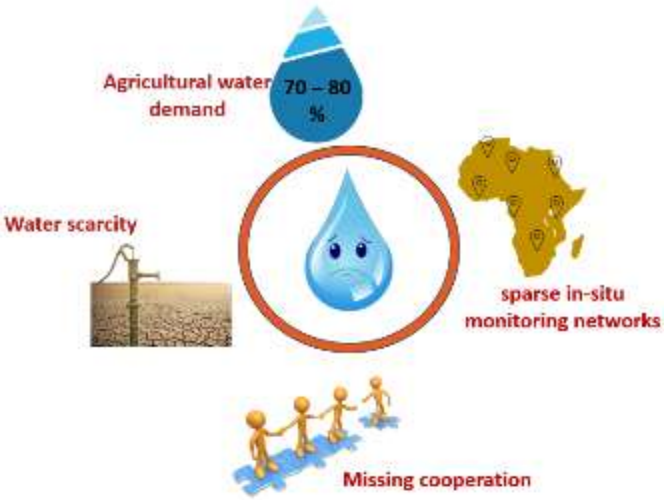


Figure 2-1: Challenges that have to be addressed due to the pressing issue of water demand.

Moreover, while satellite data offers invaluable insights, it's essential to acknowledge its limitations in providing real-time, on-the-ground information. Weather conditions, cloud cover, and spatial resolution can affect data accuracy and timeliness, requiring complementary ground-based monitoring and verification methods.

Despite these challenges, the integration of satellite data into agricultural practices holds immense promise for improving irrigation water management. Recognizing and addressing these limitations through education, training, and ongoing refinement of data analysis methods will be crucial for maximizing the benefits of satellite technology in agriculture.

2.2. Purpose of the Algorithm

The agricultural sector is already the largest user of water resources, accounting for roughly 70% of all freshwater withdrawals globally and requires even more water to sustain the continuous demand growth of food and biomass fostered by the population growth and changes in rainfall and temperature patterns caused by climate change. This is leading to an increase in competition and conflicts among users and sectors, especially in water-limited areas, for the use of the resources. The main purposes of this algorithm elaboration are :

- Acquire detailed knowledge of the spatial-temporal variation of soil and plant conditions.

- Provide the index NDVI acquired from Hyperspectral data, a really important factor with values [-1,1]. Negative values indicate water presence. On the other hand, with NDVI values close to +1, there’s a high possibility that it’s dense green leaves. NDVI close to zero indicates an absence of green leaves.

- Provide temporal and spatial scales that match rapidly evolving capabilities to vary cultural procedures, irrigation and agrochemical inputs[1].

- Provide Evapotranspiration using Earth Observation (EO) data, eliminating the need for scientists or farmers to conduct on-site measurements. This ensures that evapotranspiration data is provided in a timely manner for the relevant period. The algorithm providing ETa represents the reference data used by farmers for scheduling irrigation, providing crucial insights into actual evapotranspiration of the current cultivated crops and allowing to schedule future cultivations based on past results.

- Acquire the Stress Coefficient Ks which in desert area reflects the regions with water deficit. Ks considers factors like soil salinity and poor land fertility, particularly relevant in areas where water availability is a limiting factor. It adjusts ET estimates based on the magnitude of water deficit, providing insights into crop stress levels.

2.3. Audience/Users

The developed tool endeavors to function as a powerful asset, empowering stakeholders involved in the complex realm of water management. From individual farmers to policymakers working at regional or national levels, a wide spectrum of stakeholders stands poised to gain valuable insights from the capabilities offered by remote sensing tools.

Small farmers

For farmers, the algorithm serves as a powerful tool for precision agriculture, offering insights into the condition of their crops, the soil and the selection of cultivated goods each season. By accurately assessing crop health, farmers can make informed decisions on irrigation schedules, fertilizer application, and pest management, ultimately optimizing resource utilization and enhancing crop yield. The algorithm's results such as Ks factor and Evapotranspiration are an indication of early signs of stress or knowing the state of actively growing plants based on Eta's results and may allow farmers to take proactive measures, preventing potential losses and choose regions with more appropriate water conditions. Farmers could leverage the algorithm by selecting certain indexes that could indicate crop health and decide what should be cultivated and during which time period. By harnessing remote sensing tools like satellite imagery and ground-based sensors, these farmers gain access to near real-time data on critical parameters such as crop evapotranspiration, crop development, and crop water stress indicators like NDVI and CWSI, respectively. Drawing upon this data empowers them to make well-informed decisions concerning the timing and quantity of irrigation, thus amplifying both agricultural productivity and water efficiency.

Agricultural organizations

Agricultural organizations can leverage the algorithm to provide targeted support and advisory services to farmers, improving overall agricultural productivity in desert regions. The adoption of such technology not only benefits individual farmers but also contributes to the broader goal of ensuring food security and promoting the

[1] A. Calera, I. Campos, A. Osann, G. D'Urso, and M. Menenti, "Remote Sensing for Crop Water Management: From ET Modelling to Services for the End Users," Sensors, vol. 17, no. 5, Art. no. 5, May 2017, doi: 10.3390/s17051104.

sustainable use of resources in challenging agricultural landscapes. Organizations could exploit the information derived from NDVI for specific cultivations in certain time periods and advise farmers what to cultivate based on past data derived from EO data processing.

Policymakers

In another scenario, policymakers grapple with the challenge of formulating effective water management policies amidst fluctuating environmental conditions and competing socioeconomic demands. By integrating remote sensing tools into their decision-making processes, policymakers gain access to comprehensive data sets that illuminate water usage patterns, agricultural dynamics, and environmental stressors at various spatial and temporal scales. Leveraging the output of the developed solution, they can assess the impact of existing policies, identify areas of vulnerability, and anticipate future water resource trends with greater accuracy.

For instance, policymakers may use the developed solution data to evaluate the efficacy of water conservation initiatives, such as irrigation efficiency programs or land-use regulations. By analyzing indicators like land cover changes, vegetation health, water requriement and availability over time, they can gauge the effectiveness of different policy interventions and adjust strategies accordingly.

3. Algorithm Overview

3.1. Objectives of the Algorithm

The developed solution allows to map ETa and thus the actual water consumption of the cultivated crops. The main cultivated crops present in the test site are : wheat, maize and peanuts, representing 40% of the total cultivated area.

This algorithm utilizes Earth Observation (EO) data (hyperspectral and multispectral) and is designed to enhance agricultural practices in desert regions with sparse vegetation. By leveraging hyperspectral and multispectral information, the algorithm provides a comprehensive understanding of the water consumption patterns of cultivated crops.

The multifaceted objectives of the algorithm include crop growth monitoring, enabling farmers to track the development of their crops over time. Additionally, it serves as a valuable tool for stressed crop monitoring, allowing for the early detection of signs of stress or disease. Furthermore, the algorithm contributes to yield prediction, providing farmers and organizations with insights into potential harvest outcomes. The parameters of water productivity and water use efficiency are also key focuses, empowering stakeholders to optimize resource allocation and foster sustainable agricultural practices in challenging desert environments. In essence, this algorithm goes beyond conventional monitoring by providing a holistic approach to precision agriculture, supporting water conservation, and bolstering the resilience of crop cultivation in arid landscapes.

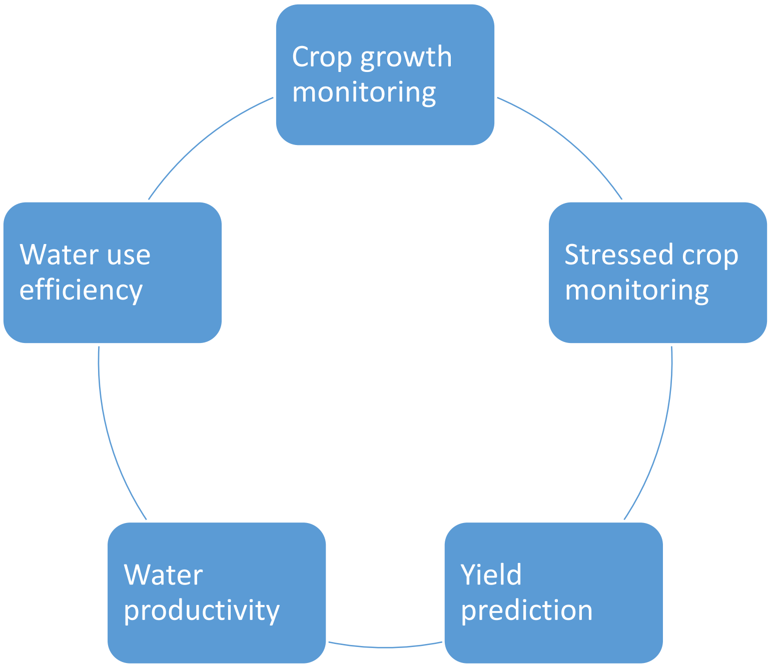


Figure 3-1: Objectives of the algorithm aiming crop growth monitoring

4. Literature Review

4.1. Existing Evapotranspiration Calculation Methods

Evapotranspiration plays a crucial role in the hydrological cycle, influencing water resource management and irrigation scheduling. It encompasses the evaporation of water from various surfaces like water bodies, land, and moist vegetation, as well as the transpiration process by plants (Wanniarachchi and R. Sarukkalige, 2022; Chen and Liu 2020). Distinguishing between evaporation and transpiration is challenging, so they are often collectively referred to as evapotranspiration (Stoy et al., 2019; Miralles et al., 2020). There is a continuous focus on accurately estimating evapotranspiration, especially in arid and semiarid regions with irrigation.

Evapotranspiration can be categorized primarily into potential, reference, and actual types. Potential evapotranspiration (ET_p) occurs when both soil and plant surfaces are wet, and it relies on surface attributes such as roughness and atmospheric conditions. However, since the simultaneous wet conditions of soil and plant surfaces is not always the case, the practical utility of ET_p is limited, mainly serving as a benchmark for the maximum evapotranspiration rate (Guo et al., 2026; Gebremedhin et al., 2022).

To mitigate some of the uncertainties surrounding ET_p, Doorenbos and Pruitt 1977, proposed adopting the concept of reference evapotranspiration (ET_0), which could function as a consistent climatic indicator for evapotranspiration. ET_0 represents the rate at which soil moisture, if readily available, would evaporate under specific atmospheric conditions and on a particular type of reference surface. Typically, the leaf surfaces of a well-watered reference crop are not saturated, resulting in some minimum surface resistance. When estimating actual evapotranspiration (ET), factors such as crop cover (e.g., leaf area) and growth stage (e.g., maturation) are accounted for by relating them to ET_0 through a crop coefficient (Kc). In instances of non-standard conditions or potential water stress, the term evapotranspiration refers to actual evapotranspiration (ET_a).

Evapotranspiration can be directly measured or estimated based on field measurements, but it can also be estimated using remote sensing and developed models.

4.1.1. Most common/used direct measurements methods

Direct field measurements of ETa through various instruments offer the advantage of obtaining precise values for a specific site. However, these instruments are costly, require time-consuming data collection and maintenance, and often need calibration for long-term estimation based on meteorological variables.

The most accurate ETa measurements are achieved using lysimeters (Williams and Ayars 2005), Bowen Ratio Energy Balance Systems (Irmak and S. Irmak 2008), and the eddy covariance technique (Baldocchi 2003).

4.1.1.1. Lysimeters

Lysimeter is a field equipment measures ET by controlling small environmental units for water balance monitoring. These devices, categorized into water balance lysimeters and weighing lysimeters, are useful for calibrating ET estimation equations and models. They help mitigate measurement errors from other systems, like wind and rain gauges, but their high cost and maintenance requirements make them impractical for continuous monitoring.

Figure 4-1: Lysimeters in vadose zone, (a) instrumentation and (b) installed lysimeter

4.1.1.2. The Bowen ratio

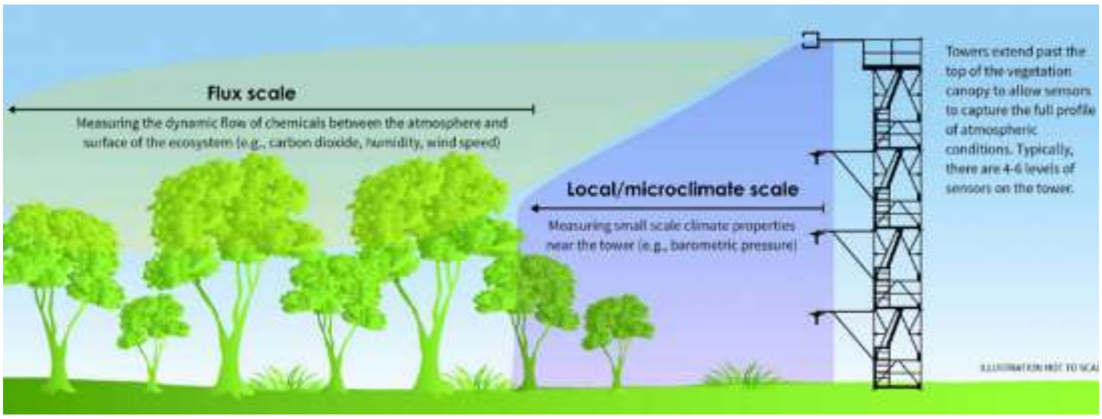
The Bowen ratio energy balance method is a reliable technique for micrometeorological conditions involving the measurement of air-temperature and water vapor gradients over the land surface. It estimates latent heat flux for short periods, but its accuracy can be affected by soil water availability and measurement errors in net radiation and soil heat flux. Overestimation may occur under certain conditions, such as positive or negative sensible heat flux.

Figure 4-2: The Bowen ratio stations in Logan, UT, USA

4.1.1.3. Eddy covariance

Eddy covariance is a direct measurement method that estimates water vapor flux based on principles established in the 1950s. It relies on covariance between vertical wind velocitys and specific humidity, utilizing turbulent wind motion at the surface. Advances in measurement equipment have facilitated its increased usage, allowing for the measurement of instantaneous eddies' vertical speed and specific humidity fluctuations.

Figure 4-3: Schematic of flux tower measurements as part of the eddy covariance method https://www.neonscience.org/data-collection/flux-tower-measurements)



4.1.2. Most common/used ET estimation methods

When the above-mentioned techniques are unavailable, ETa can still be indirectly estimated using information such as weather and radiation data. Numerous models were developed including both empirical and physically based analytical methods, such as the Penman, Penman-Monteith, Thornthwaite, and Priestley-Taylor methods. While empirical methods like Stanghellini and Hargreaves-Samani are computationally efficient, they may lack accuracy over large areas with diverse land surface characteristics. Physically based analytical methods offer better estimations but require extensive data inputs.

The different methods, advantages and disadvantages are presented in the following table.

Table 4-1: Advantages and disadvantages of notable ET estimation models

Type	Method	Advantages	Disadvantages
	Penman	Easy to apply	Underestimates ET under high movement conditions of atmospheric air masses
Combined	FAO Penman-Monteith	Provides satisfactory results	considers many components, which may result in complex calculations
	ASCE-EWRI Standardized PM	Provides ET_0 for both grass and alfalfa (hourly and daily)	Using a fixed ratio of surface resistance for the entire day may induce some errors in estimating ET_0
Temperature -based	Thornthwaite	Reliable for long-terms	Underestimates the ET during the summer and is not precise for short terms
	Blaney-Criddle	Easy to use and the data is usually available	The crop coefficient depends a lot on the climate
	Blaney-Criddle (FAO)	The given crop coefficient depends less on the climate	In high elevations, coasts, and small islands, there is no relation between temperature and solar radiation
	Hargreaves	Requires a minimum of climatological data	Underestimates ET_P on the coasts and under high movements of air masses.

	Hargreaves and Samani	Requires only maximum and minimum temperature data	Needs to be evaluated in many locations for its acceptance
	Linacre	Precise on annual basis	Precision decreases on daily base
Radiation-based	Makkink	Good for humid and cold climates	It is not reliable in arid regions
	Priestly Taylor	Reliable in humid areas	Not adequate for arid zones
	Stephens-Stewart	Reliable on the western side of the USA (where it was developed)	Should be evaluated in other locations
	Jensen-Haise	Reliable under calm atmospheric conditions	Underestimates ET under conditions of high movement of atmospheric air masses.

The FAO Penman-Monteith (FAO PM) method is widely regarded as a universal standard for reference evapotranspiration (ET_0) estimation, offering good results compared to lysimeter measurements. However, its use may be limited by the availability of input variables, which can sometimes be estimated. Various studies have evaluated different ET_0 models globally, with the FAO PM method consistently showing adaptability and accuracy.

Despite its effectiveness, implementing the FAO PM method requires calculating canopy surface resistance and access to weather data records. Nonetheless, it remains a preferred choice for ET_0 estimation due to its reliability and widespread adoption.

In the project framework, the FAO PM method is used for ET_0 estimation, followed by the incorporation of crop and water stress coefficients to assess actual evapotranspiration (ET_a) and validate the model.

4.1.3. Most common/used remote sensing-based models

Many remote sensing models for ET estimation exist, each classified based on different criteria. Adopting the classification into temperature-based and conductance-based models. The listed bellow methods are thoroughly elucidated, with explicit references to their respective authors, in the state-of-the-art deliverable (D3).

4.1.3.1. Temperature-based ET Models:

Simple Surface Energy Balance Model: This model utilizes radiometric temperature to estimate sensible heat flux and subsequently ET. It was developed using an airplane-mounted thermal sensor to obtain radiative temperature data for various surface types, allowing for the estimation of soil moisture and ET using a basic surface energy budget model.

Regional ET Model Using Crop Surface Temperature: Developed by Soer in 1980, this model maps actual evapotranspiration using surface temperatures and energy balance equations. It was initially tested for grassland in the Netherlands using measurements from a thermal sensor mounted over a crop canopy, along with meteorological and soil variables measurements.

Surface Energy Balance Algorithm for Land (SEBAL) Model: Originally developed for regions including Egypt, Spain, and Niger, SEBAL has been widely used for ET estimation. It was calibrated and validated using data from agrometeorological stations and flux sites in semi-arid regions, demonstrating high simulation accuracy with minimal ground-based data requirements.

Surface Energy Budget System (SEBS): SEBS estimates atmospheric turbulent fluxes and surface evaporative fraction using remote sensing-derived data, meteorological parameters, and radiation data. It consists of modules for deriving energy balance terms, stability parameters, and roughness length for heat transfer, enabling ET estimation for different surface conditions.

Mapping Evapotranspiration at High Resolution with Internalized Calibration (METRIC) Model: Built upon the SEBAL process, METRIC incorporates substantial refinements for improved accuracy. It accounts for temperature variations over topographical variations and uses standardized equations for energy balance calibration, enabling high-resolution ET mapping without specific crop classification.

Two-Source Energy Balance (TSEB) Model: Developed to simulate heat fluxes from vegetation and soil separately, TSEB incorporates an equation to calculate the soil resistance using wind speed near the soil surface. It allows for separate estimation of sensible heat for vegetation and soil components, enabling more detailed ET modelling.

Atmosphere-Land Exchange Inverse (ALEXI): ALEXI enhances ET estimation by incorporating a new component based on time-integrated planetary boundary layer (PBL) thermal energy. This addition improves sensible heat flux estimation, particularly in the morning when the PBL is warming up.

Disaggregated ALEXI (DisALEXI): Building upon ALEXI, DisALEXI disaggregates large-pixel sensible heat flux estimates into smaller pixels using TSEB. By utilizing both high and low-resolution remote sensing images, it provides high-resolution distributions of sensible and latent heat fluxes for regional ET mapping.

Various Contextural ET Models: These models utilize spatial contextual information from remote sensing data for ET estimation. Notable examples include TEFM, TIM, S-SEBI, and others, each employing different approaches to incorporate thermal and short-wave data for improved ET estimation accuracy.

4.1.3.2. Conductance-based ET Models:

Conductance-based ET models offer valuable alternatives to temperature-based models, especially in dense vegetated areas where sensible heat flux is small. These models utilize vegetation structural information from short-wave remote sensing to quantify leaf stomatal conductance, a key factor controlling transpiration. Unlike temperature-based models reliant solely on surface temperature measurements, conductance-based models can map spatially and temporally continuous ET, even in cloud-affected regions. Moreover, they are crucial for land modelling in Earth system models, as they can forecast future ET and hydrology by simulating required inputs. Additionally, conductance-based models integrate carbon cycle information, aiding in both ET and photosynthesis estimation.

Penman-Monteith Model: Initially used for regional ET estimation, this model employs the Penman-Monteith equation to estimate ET0 based on meteorological conditions and adjusts it using factors like NDVI and Ks for actual ET estimation.

Canopy Photosynthesis Models: These models focus on the close coupling between carbon and water cycles in plant canopies. Variants include leaf-level models, big-leaf canopy models, two-big-leaf models, and two-leaf models, each offering insights into leaf photosynthesis and its relation to stomatal conductance.

ET Models Coupled with Plant Photosynthesis: Upscaling photosynthesis from leaf to canopy impacts ET model design. Modifications, like transitioning from big-leaf to two-leaf models, aim to improve the coupling between carbon and water cycles. Widely used variants include the big-leaf Penman-Monteith model, the two-big-leaf model, and the two-leaf model.

4.1.3.3. The proposed model

The proposed model, known as the Stand-alone Remote Sensing Approach to Estimate Reference Evapotranspiration (SARE), offers a straightforward and efficient means of determining ETo using satellite data. At the core of the model are two key dynamic satellite data parameters: NDVI and Brightness Temperature (BT), which fluctuate in response to alterations in both land surfaces and atmospheric conditions.

The SARE model revolves around five fundamental fractions: vegetation fractions (Vf), location fractions (Lf), elevation fractions (Ef), seasonal fractions (Sf), and thermal fractions (Tf). It integrates these fractions using three distinct types of data input (El-Shirbeny et al. 2022). Firstly, Spatial Variation Layers (SVL) provide location-specific variability that remains consistent over time, encapsulating both Ef and Lf data. Secondly, Temporal Variation Layers (TVL) account for time-based changes, aligning with the seasonal variations in the northern hemisphere, and include Sf data. Lastly, Spatio-Temporal Variation Layers (STVL) capture the complex surface conditions of the Earth through the interactions of chemical, physical, and biophysical elements, as reflected in the electromagnetic radiation received by space-borne sensors. STVL data encompass both Tf and Vf components.

The equations and concepts behind these parameters are elaborated in detail in the state-of-the-art deliverable (D3).

4.2. Comparison and Rationale for the Chosen Approach

Intercomparison of remote sensing-based ET estimation models reveals a range of advantages and limitations. Two-source models theoretically promise more accurate ET estimation over sparse vegetation by separately considering soil and vegetation for energy balance closure. Studies indicate superior accuracy with Two-Source Energy Balance (TSEB) compared to Surface Energy Balance Algorithm for Land (SEBS) in sparsely vegetated grasslands, while single-source models perform better in semiarid rangeland.

Models like Surface Energy Balance Algorithm for Land (SEBAL) and Mapping Evapotranspiration at High Resolution with Internalized Calibration (METRIC) share similarities in theoretical frameworks and data requirements. METRIC offers advantages over SEBAL in mountainous areas and provides relatively more accurate ET estimates.

Non-iterative single-source models, such as Surface Energy Balance Index (SEBI), Simplified Surface Energy Balance Index (S-SEBI), and Surface Energy Balance System (SSEB), calculate sensible heat (H) using evaporative fraction (EF). SSEB, verified for different climatic zones, tends to underestimate H in dry open areas.

Iterative models like Two-Source Energy Balance Priestley–Taylor (TSEB-PT), Two-Source Time Integrated Model (TSTIM), and Atmosphere-Land Exchange Inverse (ALEXI) show promising performance across various climatic zones. Among them, TSEB-PT has gained more attention in recent literature.

Reflectance-based models estimate potential crop transpiration using crop coefficients (Kc) and actual ET values obtained from remote sensing methods. They have been shown to reduce seasonal water requirements by 18% when used in irrigation schedules.

Triangular and Trapezoidal methods have their own advantages and limitations. While the Triangular method requires a large number of pixels with wide ranges of soil moisture and vegetation fraction, the Trapezoidal method requires ground-measured data but fewer pixels. Both methods do not immediately detect water stress.

Based on all the above, the Stand-alone Remote Sensing Approach to Estimate Reference Evapotranspiration (SARE) is proposed as a solution. It was used to estimate ET_0 from 2005 to 2020 with LST and NDVI from MODIS. The results were validated against the FAO-Penman–Monteith applied over 35 ground meteorological stations distributed across Arab countries and covering all climate classes based on the most recent Köppen–Geiger climate classification (El-Shirbeny et al. 2022). The statistical analysis showed good results with average Root Mean Squared Error (RMSE) ranging from 6.9 to 17.3 (mm/month), a correlation coefficient (r) and an index of agreement (d) with more than 0.9. Being calibrated to the local climatological conditions, SARE will be selected in the framework of the present project.

4.3. Previous Studies and Research

The previous studies, research, and foundational theories of the aforementioned models have been extensively examined in State-of-the-art review document (D3) and the accompanying review article within the ongoing project framework titled "Advancements in Remote Sensing for Evapotranspiration Estimation: A Comprehensive Review of Temperature-Based Models". This comprehensive review delineates the methodologies employed, elucidates the chronological progression of modifications proposed and implemented by scholars, and accentuates both the merits and limitations of the models.

Furthermore, the review confronts the enduring challenge of accurately gauging evapotranspiration across various scales. It furnishes a retrospective comparative analysis spanning a 15-year interval, empowering practitioners to discern the most suitable model for particular circumstances.

Moreover, it deliberates on the strides made in satellite missions, notably the Copernicus Space Component (CSC) and Landsat Next, and their pivotal role in augmenting ET estimation models.

5. Algorithm Architecture

5.1. High-Level Structure

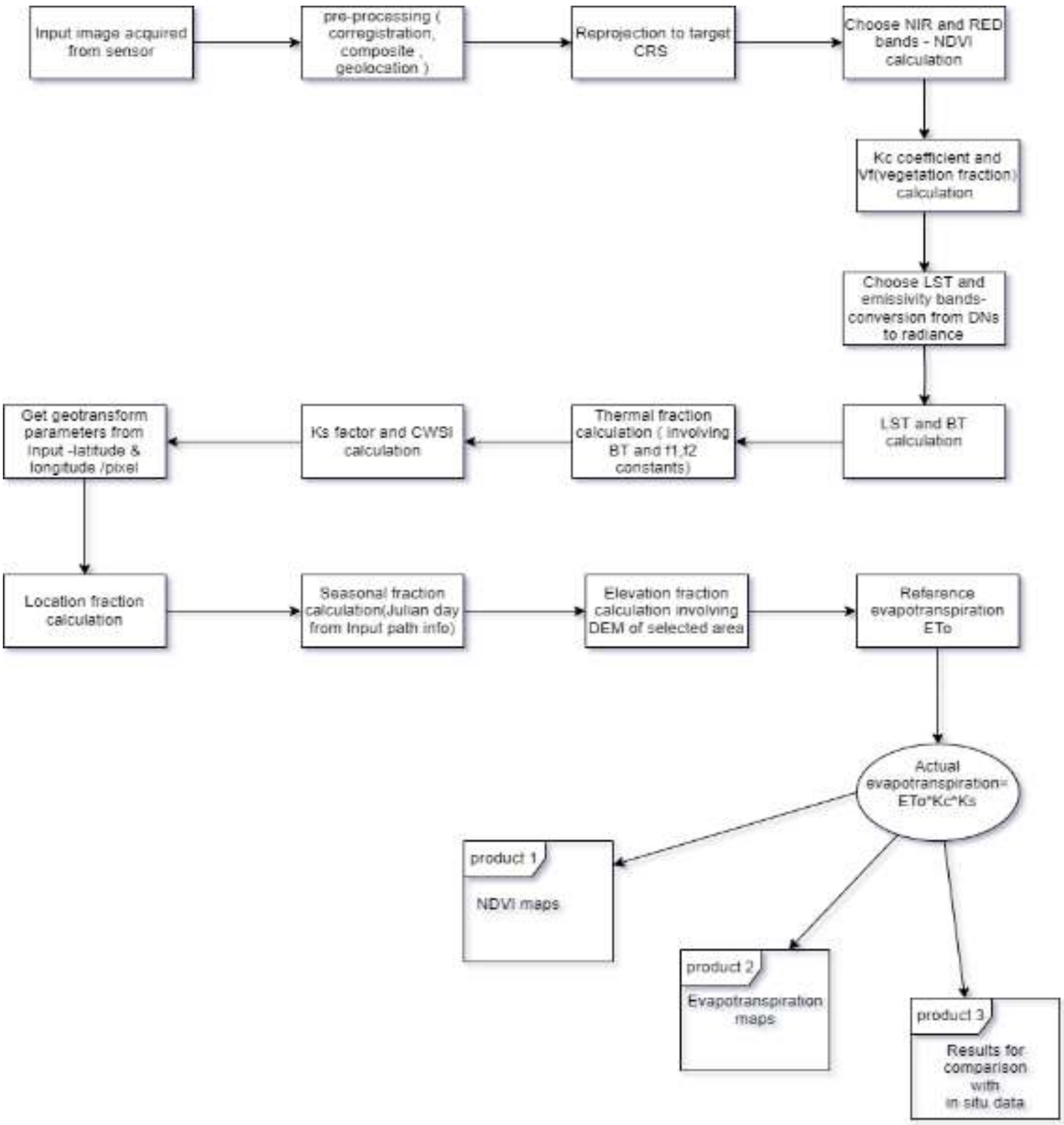


Figure 5-1: Workflow of the algorithm from Input image up to required Product outputs

In the image above the steps from image acquisition up to taking the result products are presented. Once the input image is ordered and acquired it needs a pre-processing depending the sensor which is described in the section below. Then leveraging the optical data of input images NIR and VNIR bands are used for calculating NDVI and the Vegetation fraction which highly affected by NDVI. From thermal data the thermal bands are acquired, and LST & BT are extracted in order to calculate the KS and CWSI. Leveraging the geotransform attributes of the raster input the latitude and longitude are acquired so the location fraction can be acquired. The step then is to multiply the needed factors and get reference evapotranspiration. Finally, the actual Evapotranspiration is extracted along with the products such as Eta, Ks and NDVI maps.

5.2. Data Inputs

The algorithm relies on various satellite data inputs, encompassing both hyperspectral and multispectral imagery, to facilitate precise and detailed analysis of field measurements. The main optical input comes from the PRISMA satellite, operated by the Italian Space Agency. PRISMA is known for its hyperspectral capabilities, capturing a wide range of spectral bands (400–2500 nm) with a spatial resolution of 30 meters. A total of 13 PRISMA images were included, each selected to closely match a Landsat acquisition within a ±2-day window to ensure good alignment in time.

To complement the hyperspectral data, Landsat imagery—specifically from Landsat 8 and 9, operated by NASA and USGS—was used for its multispectral and thermal data. Landsat provides regular global coverage at a 30-meter resolution and plays a key role in the dataset thanks to its reliable revisit schedule and its provision of thermal data used to derive Land Surface Temperature (LST) and Brightness Temperature. A total of 52 Landsat images were initially collected across the time period of interest, spanning summer (May-September) 2023, winter (February-April) 2024, and summer (May-October) 2024. After filtering out cloudy scenes, 36 high-quality Landsat images remained, with roughly one usable image per week from May to September for the summer seasons and from February to April for the winter measurements.

In addition, data from Sentinel-2 (part of ESA's Copernicus Program) were included to provide additional optical support, especially when PRISMA was not available or when comparing across datasets. Like Landsat, 52 Sentinel-2 images were acquired and filtered down to 36 usable ones, based on cloud cover and temporal matching with Landsat acquisitions.

Each satellite input was carefully chosen to align with specific field measurement dates, ensuring that data from different sensors were collected on the same or very similar days. This coordination across missions—PRISMA, Landsat, and Sentinel—was essential for building a consistent and comparable dataset.

To further enhance the dataset, ECOSTRESS data were used. ECOSTRESS, installed on the International Space Station, provides thermal data at 70-meter resolution. These observations offer valuable insights into land surface temperature patterns and vegetation stress. All standard ECOSTRESS products are delivered in HDF5 format, a widely used format in Earth observation. These files contain structured scientific data, including groups and datasets, which were converted to GeoTIFF for easier handling and accurate georeferencing. As mentioned in section 12, the limited availability of ECOSTRESS was complemented by Landsat data for the thermal component.

	29-May	6-Jun	10-Jun	16-Jun	22-Jun	14-Jul	22-Jul	30-Jul	15-Aug	31-Aug	16-Sep	25-Sep
Prisma	X	X	√	√	X	√	√	X	X	X	√	X
Sentinel-2	√	√	√	√	√	√	√	√	X*	√	√	√
Ecostress	X	X	X	√	X	X	X	√	√	X	X	√
EnMap	X	X	X	√	√	X	X	X	√	√	X	X
LandSat 8/9	√	√	√	√	X*	√	√	√	X*	√	√	√

Dates with in situ data

√ Success X Fail X*Cloudy

Figure 5-2: EO data available during the first cycle showing the coincidence with in situ data

	6-Jan	22-Jan	7-Feb	23-Feb	10-Mar	26-Mar	3-Apr	11-Apr	27-Apr
Prisma	√	X	√	X	X	√	X	X	X
Sentinel-2	√	X	√	X*	√	X*	√	√	√
Ecostress	√	√	X	√	√	√	√	√	X
EnMap	√	X	X	X	X	X	X	X	X
LandSat 8/9	√	√	√	X*	√	√	√	√	√

Figure 5-3: EO data available during the second cycle showing the coincidence with in situ data

	13/5/2024	29/5/2024	6/6/2024	14/6/2024	30/6/2024	16/7/2024	24/7/2024	1/8/2024	9/8/2024	17/8/2024	3/9/2024	18/9/2024
Prisma	√	X	X	X	X	X	√	X	√	√	X	X
Sentinel-2	√	√	√	√	√	√	√	√	√	√	X	[illegible]
Ecostress	X	X	X	X	√	X	√	X	√	√	X	X
EnMap	√	X	√	√	X	√	√	√	√	X	X	X
LandSat 8/9	√	√	√	X	√	√	√	√	√	√	[illegible]	X

Figure 5-4: EO data available during the third cycle showing the coincidence with in situ data

To fill in occasional gaps—either due to unavailable dates or excessive cloud cover—data from EMIT (Earth Surface Mineral Dust Source Investigation) were also explored. EMIT is a hyperspectral imaging mission developed by NASA's Jet Propulsion Laboratory (JPL), designed to study surface mineral dust composition. It operates in the Visible and Near-Infrared (VNIR) spectral range, covering wavelengths from 380 to 2500 nanometers, with 285 spectral bands. EMIT's data are particularly useful for identifying surface mineral types and understanding their roles in atmospheric processes. The mission provides Level 2 (L2) products in GeoTIFF format, which include processed reflectance and mineral classification data ready for geospatial analysis. Although EMIT data are publicly available for selected regions and time windows, only two images were found that matched the study area and timeframe. Unfortunately, one image only partially covered the Area of Interest (AOI), and the other was largely cloud-obstructed, making both unsuitable for the analysis.

In total, more than 72 satellite images from various missions were collected and used to generate maps and extract relevant information. These acquisitions cover three key time windows: summer 2023, winter 2024, and summer 2024.

The following tables report the datasets procured during the project, including PRISMA new acquisitions, Landsat data (in replacement of ECOSTRESS for the thermal component) and Sentinel-2 data as complementary dataset.

Area of Interest	El Salheya El Gedida - Sharqia governorate (northeast of Egypt)
Area extent	13.800 hectares
Location	30° 22' 35" and 30° 31' 19" N 31° 55' 24" and 32° 02' 38" E

	PRISMA new acquisitions dataset	
	Acquisition mode	**Acquisition Date**
1	PRS_L2D_STD	10 June 2023
2	PRS_L2D_STD	16 June 2023
3	PRS_L2D_STD	15 July 2023
4	PRS_L2D_STD	21 July 2023
5	PRS_L2D_STD	17 September 2023
6	PRS_L2D_STD	17 September 2023
7	PRS_L2D_STD	30 December 2023
8	PRS_L2D_STD	05 January 2024
9	PRS_L2D_STD	09 February 2024
10	PRS_L2D_STD	12 May 2024
11	PRS_L2D_STD	26 July 2024
12	PRS_L2D_STD	07 August 2024
13	PRS_L2D_STD	18 August 2024

	Landsat8 dataset	
	Acquisition mode	**Acquisition Date**
1	Landsat 8 L2SP	04 June 2023
2	Landsat 8 L2SP	04 June 2023
3	Landsat 8 L2SP	12 June 2023
4	Landsat 8 L2SP	28 June 2023
5	Landsat 8 L2SP	14 July 2023
6	Landsat 8 L2SP	22 July 2023
7	Landsat 8 L2SP	07 August 2023
8	Landsat 8 L2SP	23 August 2023
9	Landsat 8 L2SP	31 August 2023
10	Landsat 8 L2SP	08 September 2023
11	Landsat 8 L2SP	16 September 2023
12	Landsat 8 L2SP	26 October 2023
13	Landsat 8 L2SP	03 November 2023
14	Landsat 8 L2SP	27 November 2023
15	Landsat 8 L2SP	13 December 2023
16	Landsat 8 L2SP	06 January 2024
17	Landsat 8 L2SP	14 January 2024
18	Landsat 8 L2SP	30 January 2024
19	Landsat 8 L2SP	15 February 2024
20	Landsat 8 L2SP	02 March 2024
21	Landsat 8 L2SP	05 May 2024
22	Landsat 8 L2SP	13 May 2024
23	Landsat 8 L2SP	21 May 2024
24	Landsat 8 L2SP	29 May 2024
25	Landsat 8 L2SP	06 June 2024
26	Landsat 8 L2SP	14 June 2024
27	Landsat 8 L2SP	22 June 2024
28	Landsat 8 L2SP	30 June 2024
29	Landsat 8 L2SP	16 July 2024
30	Landsat 8 L2SP	24 July 2024

31	Landsat 8 L2SP	01 August 2024
32	Landsat 8 L2SP	09 August 2024
33	Landsat 8 L2SP	17 August 2024
34	Landsat 8 L2SP	25 August 2024
35	Landsat 8 L2SP	02 September 2024
36	Landsat 8 L2SP	10 September 2024

	Sentinel-2 dataset	
	Acquisition mode	**Acquisition Date**
1	Sentinel-2 L2A	02 June 2023
2	Sentinel-2 L2A	07 June 2023
3	Sentinel-2 L2A	12 June 2023
4	Sentinel-2 L2A	17 June 2023
5	Sentinel-2 L2A	22 June 2023
6	Sentinel-2 L2A	02 July 2023
7	Sentinel-2 L2A	07 July 2023
8	Sentinel-2 L2A	12 July 2023
9	Sentinel-2 L2A	17 July 2023
10	Sentinel-2 L2A	22 July 2023
11	Sentinel-2 L2A	27 July 2023
12	Sentinel-2 L2A	01 August 2023
13	Sentinel-2 L2A	06 August 2023
14	Sentinel-2 L2A	21 August 2023
15	Sentinel-2 L2A	26 August 2023
16	Sentinel-2 L2A	31 August 2023
17	Sentinel-2 L2A	25 September 2023
18	Sentinel-2 L2A	20 October 2023
19	Sentinel-2 L2A	25 October 2023
20	Sentinel-2 L2A	04 November 2023
21	Sentinel-2 L2A	09 November 2023
22	Sentinel-2 L2A	14 November 2023
23	Sentinel-2 L2A	29 November 2023
24	Sentinel-2 L2A	04 December 2023
25	Sentinel-2 L2A	14 December 2023
26	Sentinel-2 L2A	19 December 2023
27	Sentinel-2 L2A	24 December 2023
28	Sentinel-2 L2A	03 January 2024
29	Sentinel-2 L2A	08 January 2024
30	Sentinel-2 L2A	13 January 2024
31	Sentinel-2 L2A	18 January 2024
32	Sentinel-2 L2A	28 January 2024
33	Sentinel-2 L2A	07 February 2024
34	Sentinel-2 L2A	13 March 2024
35	Sentinel-2 L2A	02 April 2024
36	Sentinel-2 L2A	12 April 2024
37	Sentinel-2 L2A	17 April 2024
38	Sentinel-2 L2A	22 April 2024
39	Sentinel-2 L2A	27 April 2024
40	Sentinel-2 L2A	02 May 2024
41	Sentinel-2 L2A	07 May 2024
42	Sentinel-2 L2A	12 May 2024
43	Sentinel-2 L2A	17 May 2024
44	Sentinel-2 L2A	22 May 2024
45	Sentinel-2 L2A	27 May 2024
46	Sentinel-2 L2A	01 June 2024

47	Sentinel-2 L2A	06 June 2024
48	Sentinel-2 L2A	11 June 2024
49	Sentinel-2 L2A	16 June 2024
50	Sentinel-2 L2A	21 June 2024
51	Sentinel-2 L2A	26 June 2024
52	Sentinel-2 L2A	01 July 2024
53	Sentinel-2 L2A	11 July 2024
54	Sentinel-2 L2A	16 July 2024
55	Sentinel-2 L2A	21 July 2024
56	Sentinel-2 L2A	26 July 2024
57	Sentinel-2 L2A	31 July 2024
58	Sentinel-2 L2A	05 August 2024
59	Sentinel-2 L2A	10 August 2024
60	Sentinel-2 L2A	15 August 2024
61	Sentinel-2 L2A	20 August 2024
62	Sentinel-2 L2A	25 August 2024
63	Sentinel-2 L2A	30 August 2024
64	Sentinel-2 L2A	09 September 2024
65	Sentinel-2 L2A	19 September 2024
66	Sentinel-2 L2A	24 September 2024
67	Sentinel-2 L2A	29 September 2024

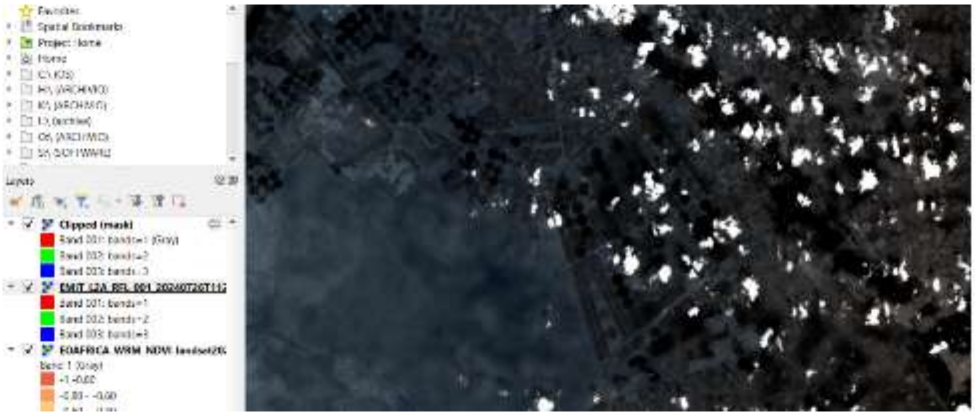


Figure 5-5 : EMIT sensor image for the 26th of July 2024 (season 3)

5.3. Meteorological Data

The data collection process involves gathering meteorological data and relies on the automatic meteorological station that is located at the edge of the agricultural farm at 30°31'12.0"N 31°57'36.0"E and altitude of 5 m a.s.l. The station's hourly measurements are averaged over a daily mean and are reported every two months. The meteorological data includes the average, the maximum and the minimum air temperatures, and the dew/frost point temperatures at a height of 2 meters. The data includes the relative humidity average as well, the precipitation, the wind speed and the solar radiation.

An example of meteorological data is presented in the following table.

Table 5-1: Daily average meteorological data

Weather data of Egypt-Ismailia-Ksaseen								
Latitude: 30.52 | Longitude: 31.96 | Altitude: 5								
Parameter(s): **T2M:** Temperature Average at 2 Meters (°C) **TMIN:** Temperature at 2 Meters Minimum (°C) **TMAX:** Temperature at 2 Meters Maximum (°C) **TDEW:** Dew/Frost Point at 2 Meters (°C) **RH2M:** Relative Humidity Average at 2 Meters (%) **RAIN:** Precipitation (mm) **WIND:** Wind Speed at 2 Meters (m/s) **SRAD:** Solar Radiation (MJ/m^2/day)								
DATE	**T2M**	**TMIN**	**TMAX**	**TDEW**	**RH2M**	**RAIN**	**WIND**	**SRAD**
25/08/2023	30.5	21.9	40.7	16.3	51.2	0	2.6	25.8
26/08/2023	31.6	22.9	41.3	16.6	49.9	0	3	25.7
27/08/2023	32.1	24.3	40.6	17	46.6	0	2.7	25.2
28/08/2023	31.2	24.3	39.6	15.6	44	0	2.8	25.3
29/08/2023	30.2	23	39.1	16.7	49.9	0	2.7	25.4
30/08/2023	30.8	22.1	41	16	49.3	0	2	25.3
…	-	-	-	-	-	-	-	-

5.4. Processing Steps

Initially, the image is opened, and a mask is applied to exclude any no-data values, commonly represented as 0 or -9999. Subsequently, the bands are read as arrays, enabling efficient manipulation and analysis of the pixel data. For Landsat images, an additional step may be necessary to convert the Digital Numbers (DN) to Top of Atmosphere (TOA) radiance, from which the Brightness Temperature (BT) and subsequently the Land Surface Temperature (LST) can be derived. For facilitating this though we used Band10 Level 1 Landsat data for BT and Band10 Level 2 Landsat data for LST. Furthermore, geotransform parameters of the input file are obtained, providing essential spatial referencing information. Leveraging the dimensions of the raster data, arrays are created to hold latitude and longitude values for each pixel, facilitating geospatial analysis and interpretation. These coordinates are then stored in .npy files for future reference. Additionally, the elevation Digital Elevation Model (DEM) file is read as an array, and adjustments are made to ensure compatibility with the desired range. Subsequently, the Elevation Fraction is calculated, with its value being proportional to the square of the DEM data.

5.5. Pre-processing

The PRISMA data come as Level 2 data which are geolocated and have the atmospheric correction radiance. PRISMA data come as HE5 files. In order to extract their corrected geolocation, they need to be opened first with PRISMA Toolbox which is a software totally devoted to the interaction and manipulation of PRISMA satellite mission hyperspectral products. PRISMA Toolbox v1.0 allows to import, view and convert L1, L2B, L2C, L2D products in a very simple and immediate way (in MS Windows based PC). Thanks to its features the user can interact in a very simple and quick manner with all the spectral bands and the metadata of the HDF products without taking care of the HDF format. The desired information (bands) are extracted one per one. Because PRISMA comprises a high-spectral resolution VNIR-SWIR imaging spectrometer, it works in numerous, narrow and contiguous bands arranged from the visible to the near infrared (VNIR, Visible and Near InfraRed) and up to the infrared short-wave (SWIR, Short Wave InfraRed). This means that the bands Red and NIR that are needed for the calculations don't correspond only to one band in PRISMA. For that, the average band was visualized and extracted in each case as a GeoTiff.

The ECOSTRESS data, as referred to previously come as HDF5 files. As they come, they are not geolocated, and their format seems to be difficult to visualize and process directly. A script was used to convert ECOSTRESS swath data to projected GeoTiffs. Moreover, their resolution was changed from 70m to 30m either with the Software QGIS or with a Python script developed for upscaling/downscaling according to the input data needs. When executing this script, a user will submit a desired output projection and input directory containing ECOSTRESS swath data products as command line arguments. The script begins by opening any of the ECOSTRESS products listed below that are contained in the input directory. Next, it uses the latitude and longitude arrays from the ECO1BGEO product to resample the swath dataset to a grid using nearest neighbor resampling (Pyresample/kdtree[2]) The script exports the gridded array as a GeoTIFF (GDAL).

Reading ECOSTRESS HDF5 Input Files:

```
f = h5py.File(ecoList[0])            # Read in ECOSTRESS HDF5 file

ecoName = ecoList[0].split('.h5')[0]  # Extract original filename

print(ecoName)
```

Loading Latitude and Longitude Arrays from the Corresponding ECO1BGEO File: These geolocation arrays are essential for defining the swath.

```
g = h5py.File(geo[0])

geo_objs = []

g.visit(geo_objs.append)

latSD = [str(obj) for obj in geo_objs if isinstance(g[obj], h5py.Dataset) and '/latitude' in obj]

lonSD = [str(obj) for obj in geo_objs if isinstance(g[obj], h5py.Dataset) and '/longitude' in obj]

lat = g[latSD[0]][()].astype(float)

lon = g[lonSD[0]][()].astype(float)

dims = lat.shape
```

[2] https://git.earthdata.nasa.gov/projects/LPDUR/repos/ecostress_swath2grid/browse

print(dims)

Swath-to-Grid Conversion Using Nearest Neighbor Resampling (via pyresample):

swathDef = geom.SwathDefinition(lons=lon, lats=lat)

(The swathDef is later used for nearest-neighbor resampling with pyresample.kd_tree.)

By default, the script will loop through and perform the steps for each science dataset (SDS) in the HDF5 file. There is an optional argument that allows you to select a subset of SDS layers within a given product. In this project's case the only layers needed were Land Surface Temperature and Emissivity.

To begin the data processing pipeline, the input image is opened in either JP2 or TIFF format, depending on the specific data source. Subsequently, an image composite is created by selecting and combining the desired bands relevant to the optical or thermal characteristics of the scene. For optical data from sensors like PRISMA, Sentinel, or ENMAP, the standard sequence comprises green, red, and near-infrared (NIR) bands, resulting in a three-band image composite. Conversely, for thermal data such as Landsat and ECOSTRESS, the composite consists of additional bands reflecting thermal properties. For Landsat, the sequence encompasses green, red, NIR, land surface temperature (LST), and emissivity, yielding a five-band composite. In contrast, ECOSTRESS data is composed of LST and emissivity bands, forming a three-band composite. Using QGIS, the resolution of these composites is adjusted to ensure consistency, typically converting them to a 30-meter resolution. This process facilitates co-registration, ensuring that images from different sources are spatially aligned for accurate analysis and interpretation.

Before the algorithm calculations begin, a crucial pre-processing step involving Coordinate Reference Systems (CRS) is undertaken to ensure consistency and accuracy in subsequent calculations. Input data often arrive with varying CRS specifications, typically denoted as EPSG 4326 or 32636. These differing CRS assignments may arise from the diverse sources of the data, leading to potential spatial discrepancies. To address this, all data are uniformly transformed into a common CRS, namely EPSG 4326, at the onset of the algorithm. This standardization serves multiple purposes, foremost among them being the facilitation of seamless integration and comparison of spatial data layers. By aligning all datasets to a common CRS, spatial analyses and calculations can be executed accurately and effectively across different datasets

5.6. Post-processing

In situ measurements obtained from pivot-field measurements are characterized by specific latitude and longitude coordinates that serve as essential reference points for our analysis. In our data processing methodology, we compile a comprehensive list of latitude and longitude coordinates corresponding to the in-situ measurements and subsequently endeavor to identify the corresponding pixel within the satellite imagery output. By pinpointing the exact pixel associated with each latitude and longitude coordinate, we extract the pixel value as a representative measure of the environmental conditions observed at that specific location. This extracted pixel value serves as a pivotal anchor point. We implemented a calibration technique to ensure the accuracy and reliability of the pixel values corresponding to specific latitude and longitude coordinates. To achieve this, we applied a rounding approach to identify the nearest pixel to the target latitude and longitude coordinates. Once the nearest pixel was identified, we expanded our analysis to include the surrounding pixels in a 3x3 grid pattern, encompassing a total of 9 pixels. By considering the values of the adjacent pixels, we calculated the average value of this pixel neighborhood. This calibration process serves as a robust quality control measure, aimed at identifying and potentially mitigating anomalous pixel values, such as excessively high or irregular readings. By averaging the values of neighboring pixels, we mitigate the impact of potential outliers or inaccuracies, thereby enhancing the overall reliability and integrity of the dataset. This approach ensures that our analysis is based on robust and representative data, minimizing the influence of potential artifacts and discrepancies in the pixel values associated with specific latitude and longitude coordinates.

5.7. Output Format and Interpretation

The generated outputs include maps of evapotranspiration, Crop Water Stress Index (CWSI), and Normalized Difference Vegetation Index (NDVI), which are provided in TIFF format. These TIFF files are compatible with various Geographic Information System (GIS) software like QGIS and can also be uploaded to the EO Africa platform for further visualization and timeseries analysis. Each TIFF file retains the initial image name and includes pertinent information within its file name. To ensure spatial consistency, these images are reprojected to match the initial spatial extent, including the Coordinate Reference System (CRS), of the initial image. This reprojection process is facilitated using the Rasterio library, particularly the **src.transform** function, along with the GDAL datatype to specify the desired datatype for the new output TIFF files.

Along with these outputs, three distinct QML files were created, one for each key output—NDVI, ETa (Actual Evapotranspiration), and CWSI (Crop Water Stress Index). These QML files were designed to apply custom styling and color gradients that correspond to typical value ranges for each index, aiding in the visual interpretation of the data. For NDVI and CWSI, the QML files follow their standard ranges: NDVI typically spans from -1 to +1, while CWSI values range from 0 to 1. For ETa, the range was dynamically set based on the minimum and maximum values found in each specific dataset. These QML files were created to enhance visualization and ease the understanding of the spatial patterns within the cultivated fields in the Area of Interest (AOI). The QML files were uploaded to the project’s platform, where they can be applied for immediate use in visualizing these indices across different timescales. By employing these pre-defined ranges and color schemes, the QML files offer consistent, standardized visuals that make it easier for users to interpret key parameters related to crop health, water usage, and stress levels. The inclusion of these QML styles, along with the TIFF files, allows for seamless integration with GIS tools like QGIS, ensuring that all data products can be visualized clearly and meaningfully within the platform’s interface.

6. Mathematical Formulas and Models

6.1. Equations Used in Evapotranspiration Calculation

For the calculation Evapotranspiration several factors needed to be calculated concerning the vegetation or the thermal behavior of the surface.

Among these essential factors are the Normalized Difference Vegetation Index (NDVI), a fundamental indicator of vegetation vigor and density, which serves as the basis for calculating the Crop Coefficient (Kc) and Vegetation Fraction. NDVI values typically range from -1 to 1, with higher values indicating healthier vegetation cover.

$$NDVI = (NIR - Red) / (NIR + Red) \quad \textbf{(1)}$$

$$Vf = 1.013 - NDVI^5 \quad \textbf{(2)}$$

$$Kc = 0.9 / (NDVI\_dv * (NDVI - NDVI\_mv)) + 0.3 \quad \textbf{(3)}$$

Land Surface Temperature (LST) and Brightness Temperature (BT) are key thermal parameters reflecting surface and atmospheric conditions, respectively, with expected ranges influenced by local climate measured in ºC. LST and BT were needed for the Thermal factor's calculation but also for the Ks index. The KS factor represents the thermal component of water stress, influenced by factors such as canopy temperature and atmospheric conditions and is highly connected to the CWSI which quantifies plant stress levels.

$$CWSI = (LST - air\_temp\_2m) / (LST\_max - LST\_min) \quad \textbf{(4)}$$

$$Ks = (1 - CWSI) \quad \textbf{(5)}$$

The Julian Day serves as a temporal reference for calculating the Seasonal Fraction, capturing the influence of seasonal variations on evapotranspiration patterns.

$$Lf = 1 - 0.0063 * \sqrt{lat^2} \quad \textbf{(6)}$$

$$f1 = 1.13 * Lf \quad \textbf{(7)}$$

$$f2 = 4.88 * Lf * Sf^2 \quad \textbf{(8)}$$

$$Tf = 0.7 * BT * (f1 / f2) \quad \textbf{(9)}$$

$$A = Year / 100 \quad \textbf{(10)}$$

$$B = 2 - A + (A / 4) \quad \textbf{(11)}$$

$$Jday = (365.25 * (Year + 4716)) + (30.6001 * (Month + 1)) + Day + B - 1524.5 \quad \textbf{(12)}$$

$$Sf = 1 / (1 + 0.033 * cos(2 * \pi * Jday / 365)) \quad (13)$$

A Digital Elevation Model (DEM) is a representation of terrain elevation values at regularly spaced intervals. Using elevation data from a DEM to create a terrain model. In this process, elevation values from the DEM are used to construct a 3D representation of the terrain surface.

$$Ef = 1 - 0.00011 * \sqrt{DEM^2} \quad (14)$$

6.2. Parameters and Variables

Using the SARE model described in the proposal using the factors described in previous chapter Vf, Lf, Ef, Sf and Tf , which are respectively the vegetation, location, elevation, seasonal and thermal fractions which play a key role in the reference evapotranspiration. Consequently, the Kc and Ks factors which are an exact product of the thermal and optical data define the actual evapotranspiration of the specific place and at an exact date.

$$ETo = Vf * Lf * Ef * Sf * Tf \quad (15)$$

$$ETa = ETo * Kc * Ks \quad (16)$$

7. Implementation Details

7.1. Programming Language and Tools

In the implementation of the algorithm, Python3 serves as the primary programming language, chosen for its versatility, readability, and extensive ecosystem of libraries. To facilitate the execution and experimentation of the algorithm, Jupyter notebooks serve as the preferred environment. These interactive notebooks provide a user-friendly interface for running code, visualizing results, and documenting insights in a seamless manner. Moreover, the algorithm and associated notebooks are encapsulated within a Docker image, a lightweight containerization solution. This Docker image is constructed within a Linux virtual machine running Ubuntu 22.04 LTS. The underlying hardware configuration features an Intel 11th Gen Core i7 processor.

Furthermore, while the algorithm is primarily designed to run within the specified Docker environment, efforts have been made to ensure compatibility with alternative setups. The codebase is thoroughly tested and verified to run seamlessly within virtual machines using tools such as Visual Studio Code or Conda environments. However, it's important to note that these alternative environments do not include the Docker image, necessitating additional configuration and setup steps for deployment. By accommodating diverse development and execution environments, the algorithm maximizes accessibility and flexibility.

7.2. Code Structure

The development and validation of the algorithm was done in the sequence developing > testing > validating. The code developed was open-source and deposited in a dedicated GitHub repository and properly documented to ensure its reusability as well as uptake by others. In particular, the code has been written in Python as it is a widely used programming language (especially in the EO community) benefiting also from many libraries available that can extend the analysis capabilities.

7.3. Resources

In the image below, the available resources for code implementation were provided. We utilized Oracle VirtualBox to create a virtual machine environment, incorporating Linux Ubuntu 22.04 as the operating system. Linux offered a rich array of command-line utilities, programming libraries, and development frameworks. Additionally, the virtual environment facilitated experimentation, testing, and deployment of code in a controlled setting.

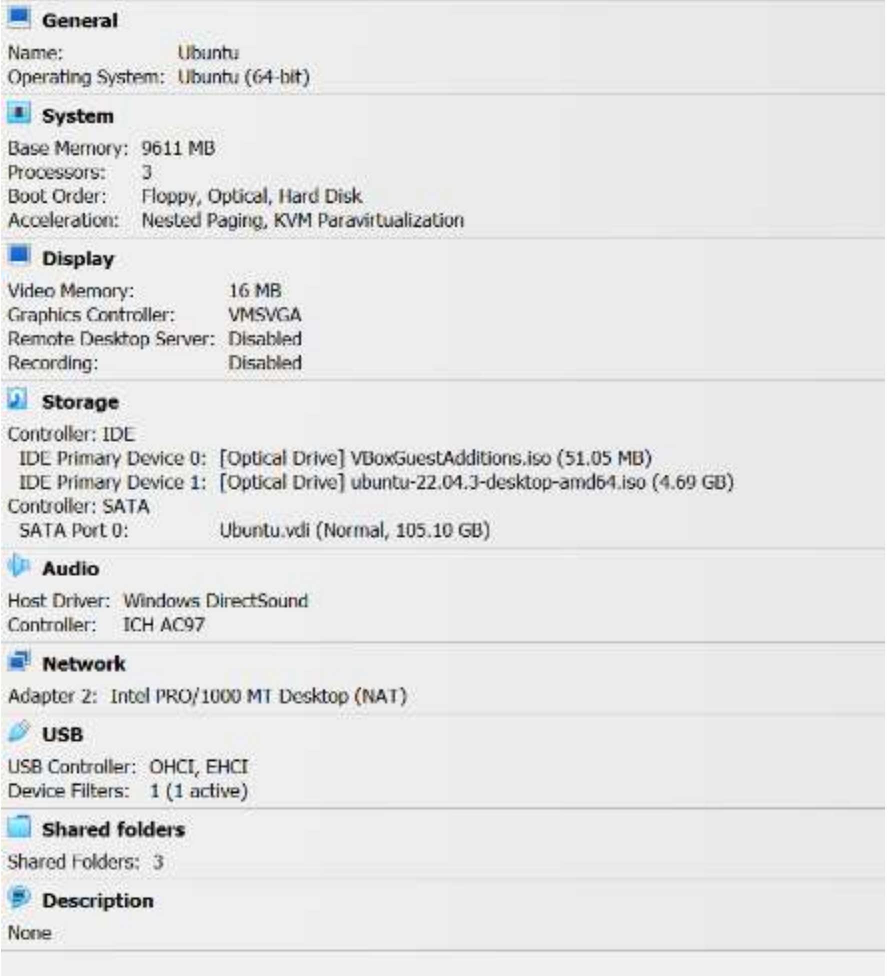


Figure 7-1: VM implementation set for the installation and running the algorithm

7.4. Dependencies and Libraries

Key Libraries include the rasterio and GDAL libraries, essential for handling geospatial data and raster manipulation tasks with efficiency and precision. Additionally, the matplotlib library is employed for visualization purposes, enabling the generation of maps, plots and graphical representations of the algorithm's output. Alongside these specialized libraries, fundamental Python libraries such as math, numpy, sys, skimage, and datetime are utilized for various data manipulation, numerical computation, and date-time calculations. More analytically the version if each library used is :

- rasterio==1.3.6
- pandas==1.3.4
- geopandas==0.12.2
- osgeo==0.0.1
- matplotlib==3.4.3
- numpy>=1.21.1
- pandas==1.3.4
- geopandas==0.12.2
- geopy==2.3.0
- h5py==3.8.0
- scikit-image==0.22.0

8. Results

The analysis of different sensor combinations has been performed: Landsat (Thermal) – Landsat (VNIR), PRISMA (VNIR) – Landsat (Thermal), Sentinel 2 (VNIR) – Landsat (Thermal) and PRISMA (VNIR) -ECOSTRESS (Thermal). In each of these combinations, NDVI, Ks and the crop coefficients (Kc) were computed: a comprehensive comparison of these results was conducted to evaluate the algorithm's performance.

In the maps and graphs below some of the results are presented showing the experimentations with different sets of data, the fluctuations of soil characteristics that affect the Evapotranspiration results and the importance of vegetation indexes such Kc, Ks and NDVI. Comparisons with in-situ data for selected fields have been conducted to improve algorithm calibration.

The procedure the team followed is the same for the results presented. As Eta intermediate, the actual evapotranspiration is presented with the use of Ks from in situ, instead of the EO calculated, for better accuracy and understanding of the results.

8.1. Season 1

8.1.1. Product maps

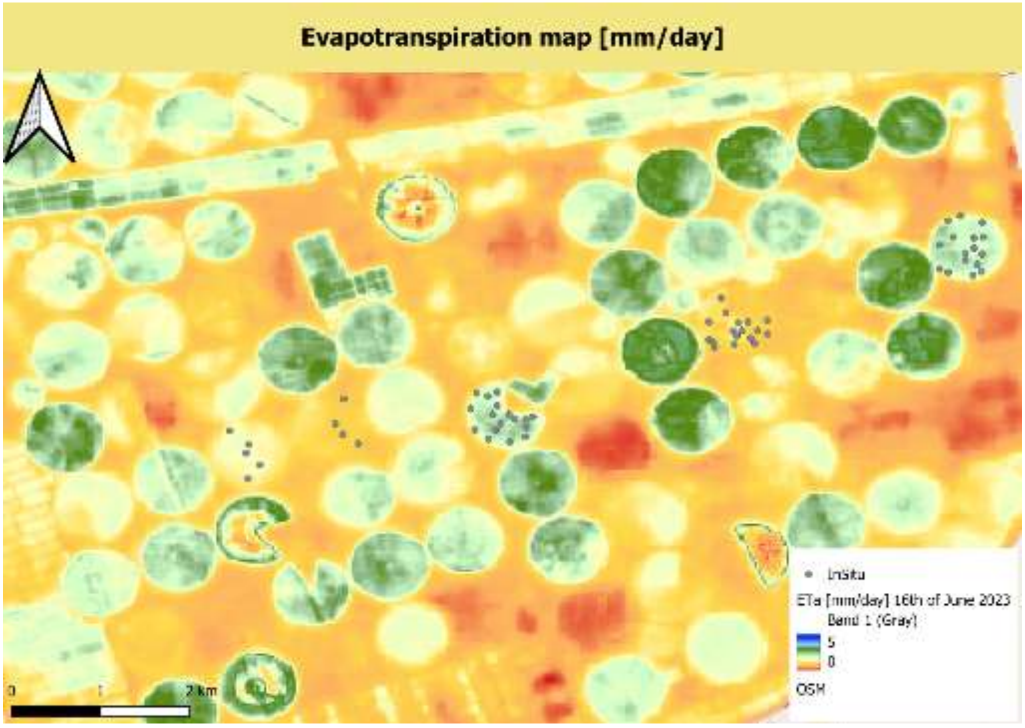


Figure 8-1 : ETa map corresponding to the 16th of June 2023 with in situ data over the EO output

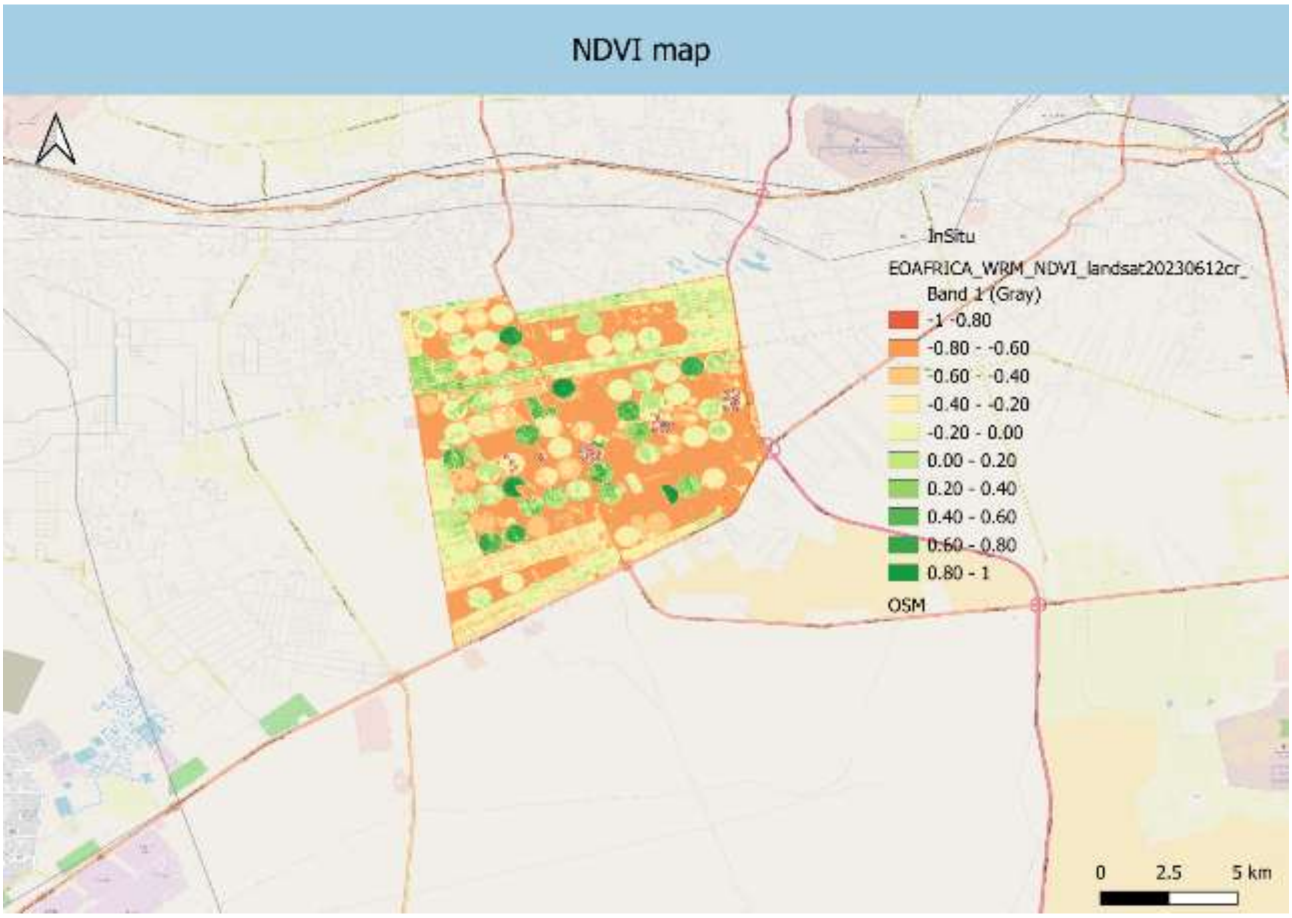


Figure 8-2: NDVI map corresponding to the 16th of June 2023 with in situ data over the EO output

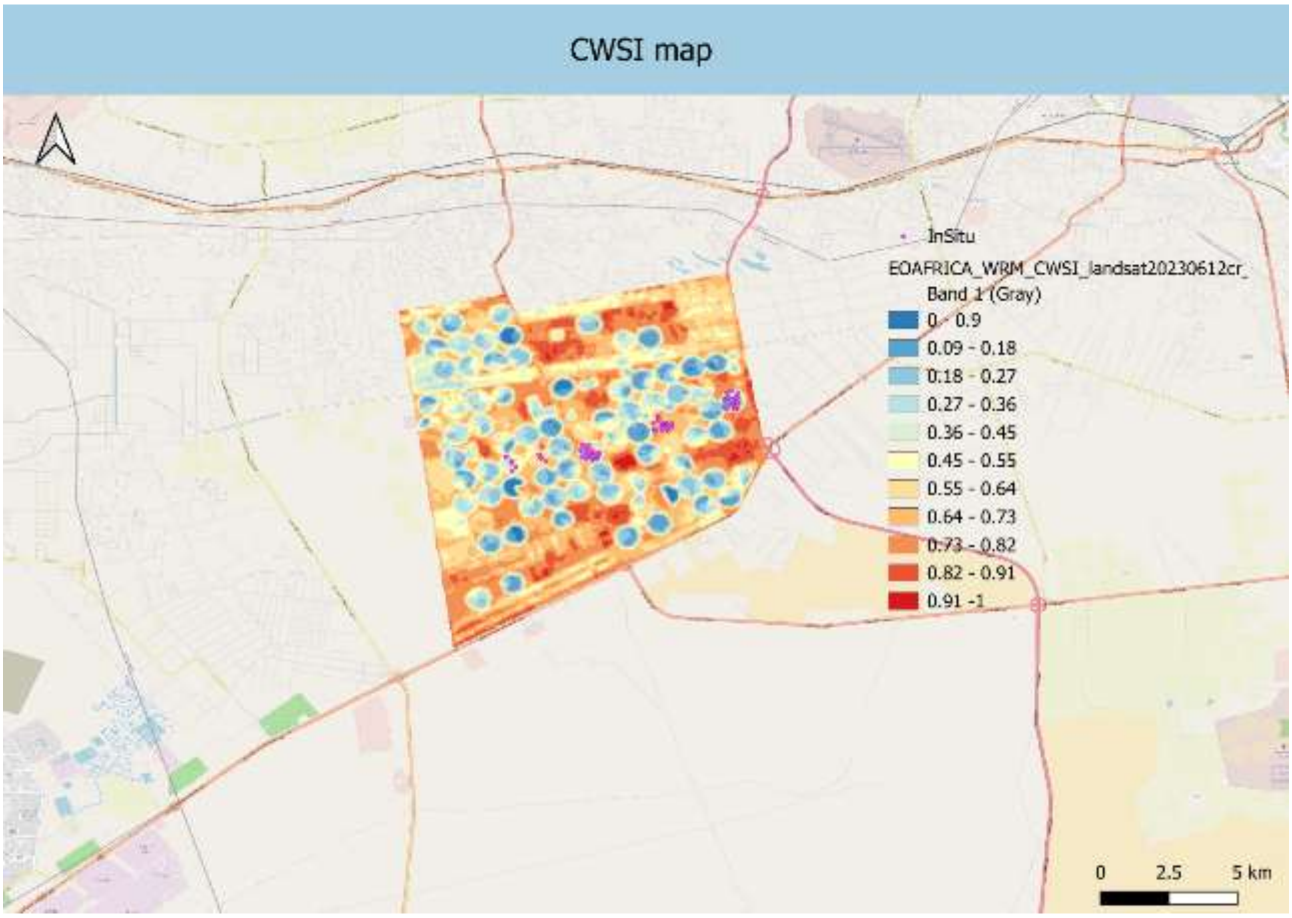
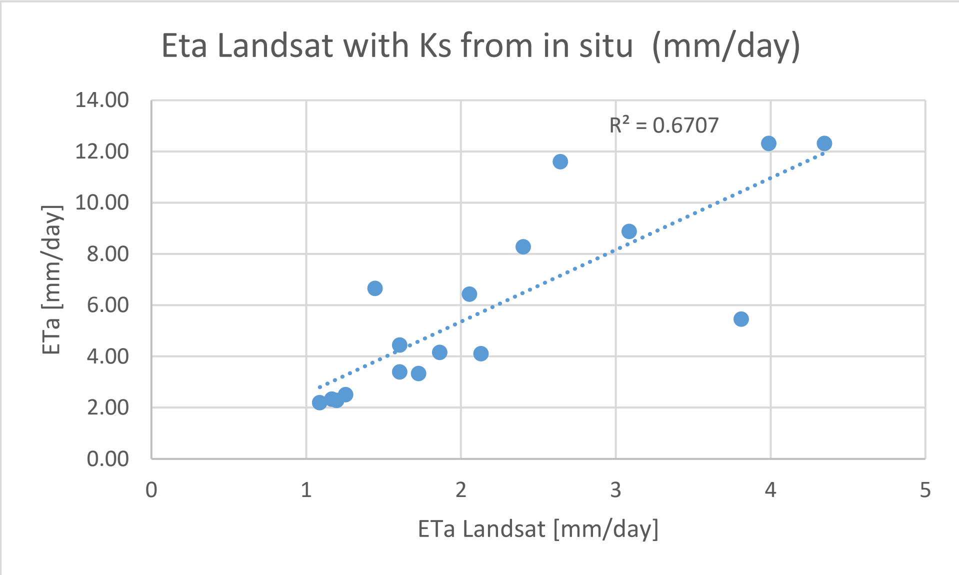


Figure 8-3: CWSI map corresponding to the 16th of June 2023 with in situ data over the EO output

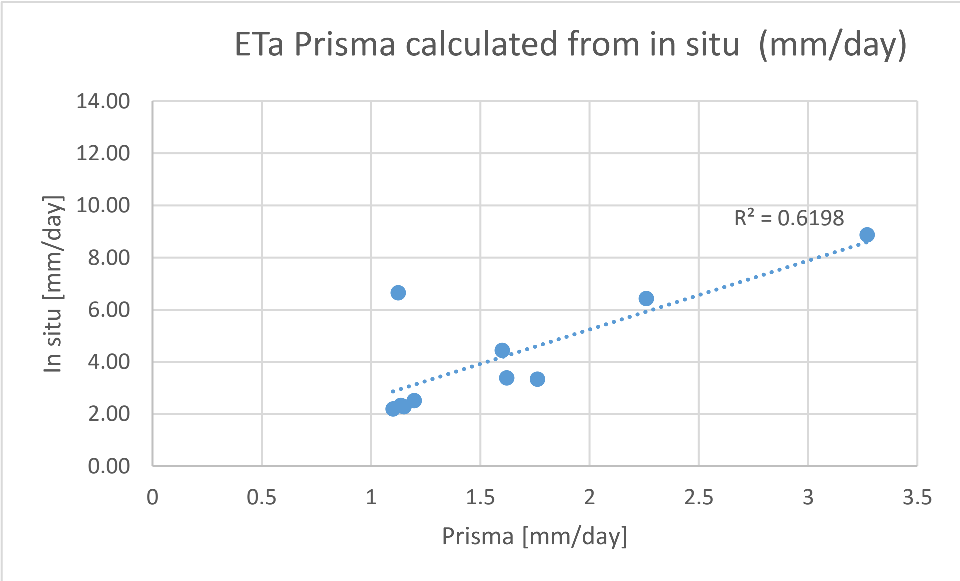
8.1.2. Pivot 1

This study involves a total of five pivots, but for now, we'll focus on Pivot 1 as a test pivot. We'll dive into the results from Pivot 1 in detail, using it as a starting point to explore key metrics like evapotranspiration (ETa) and crop performance. Pivot 1 was chosen because of its representative data, giving us a solid foundation to work with.

The other four pivots will come into play in a separate Validation Report. They'll be used to confirm and verify the findings we get from Pivot 1, making sure the results hold true across different areas. By comparing data from these other pivots, we can ensure the patterns and insights from Pivot 1 are reliable and consistent, giving us a clearer picture of the overall system. This approach lets us test our methods thoroughly.



re 8-4: Evapotranspiration acquired from Landsat data for all season 1 regarding Pivot 1 compared to in situ



e 8-5: Evapotranspiration acquired from Prisma data for all season 1 regarding Pivot 1 compared to in situ

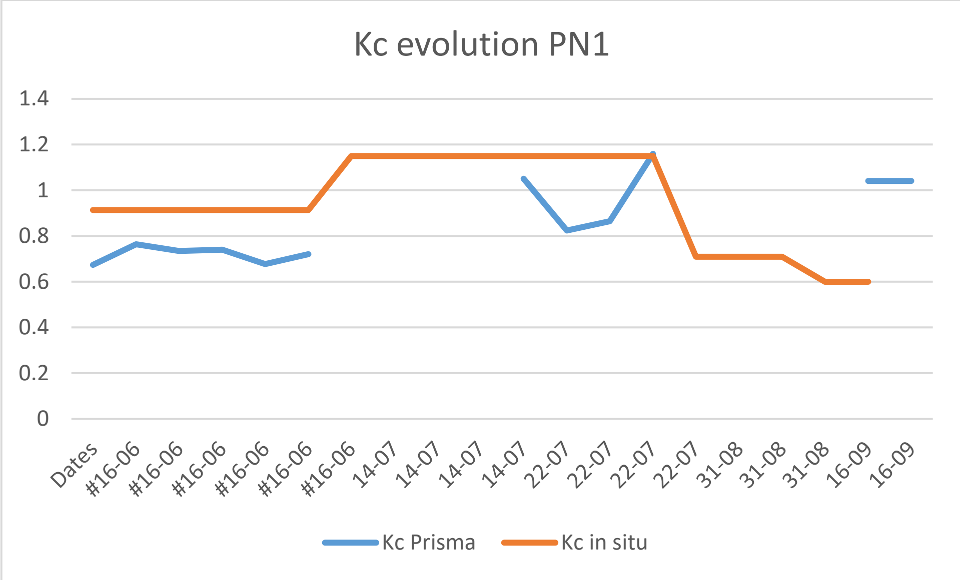


Figure 8-6: Kc vegetation index derived using NDVI from Prisma data during season 1

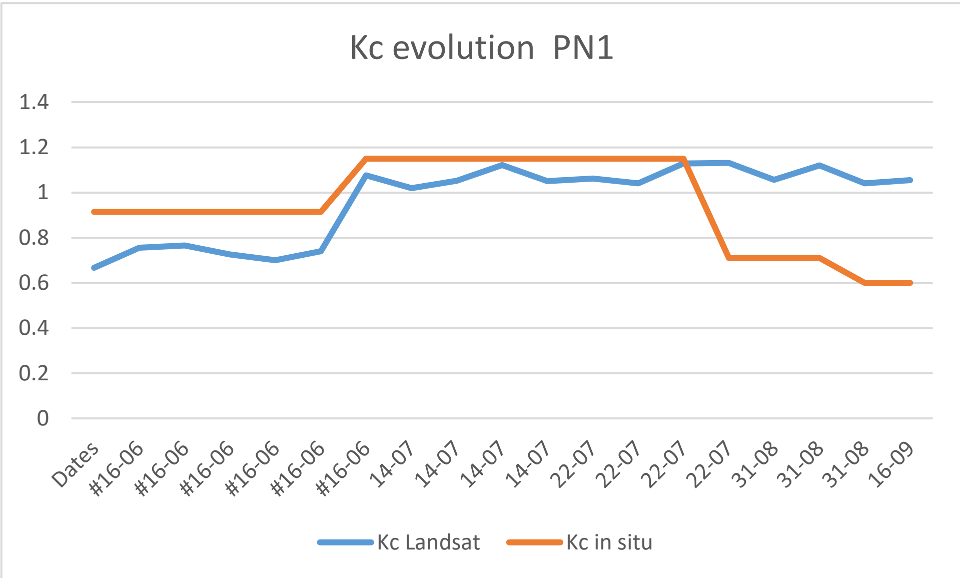


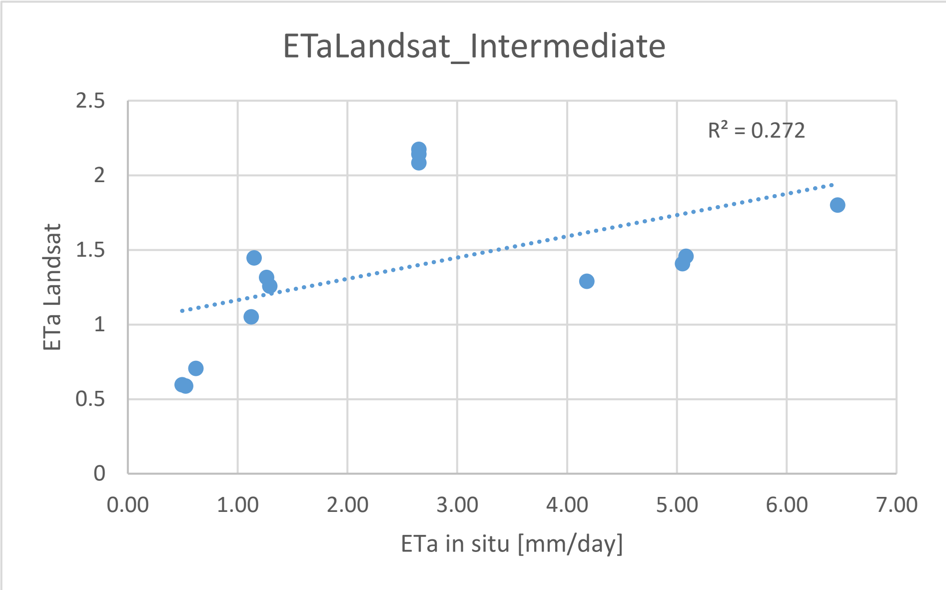
Figure 8-7: Kc vegetation index derived using NDVI from Landsat data during season 1

8.2. Season2

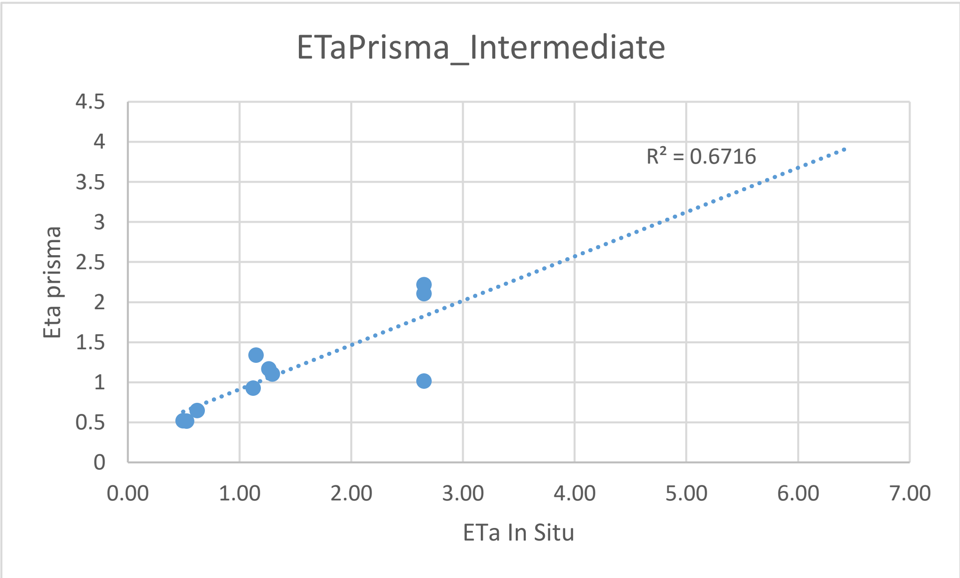
This study involves a total of five pivots, but the description is focused on Pivot 1 as a test pivot. We'll dive into the results from Pivot 1 in detail, using it as a starting point to explore key metrics like evapotranspiration (ETa) and crop performance. Pivot 1 was chosen because of its representative data, giving us a solid foundation to work with.

The other four pivots will come into play in a separate validation report. They'll be used to confirm and verify the findings we get from Pivot 1, making sure the results hold true across different areas. By comparing data from these other pivots, we can ensure the patterns and insights from Pivot 1 are reliable and consistent, giving us a clearer picture of the overall system. This approach lets us test our methods thoroughly.

8.2.1. Pivot 1



re 8-8: Evapotranspiration acquired from Landsat data for all season 2 regarding Pivot 1 compared to in situ



e 8-9: Evapotranspiration acquired from Prisma data for all season 2 regarding Pivot 1 compared to in situ

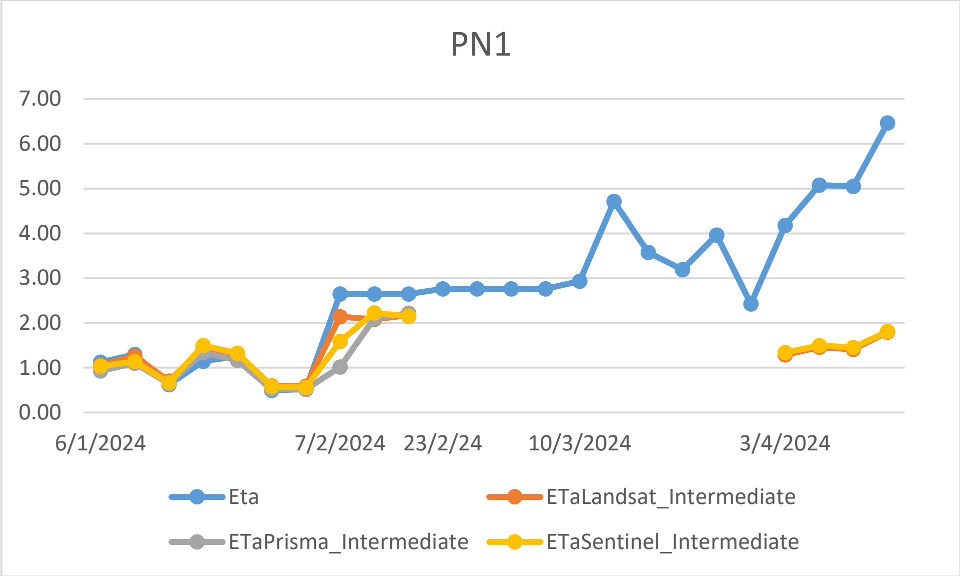


Figure 8-10: Chronological evolution of Eta for Pivot 1 from all selected sensors and in situ data

9. Machine Learning approach

In this project, the goal was to enhance the estimation of evapotranspiration (ET) by integrating in situ measurements with Earth Observation (EO) data. The focus was on comparing in situ data with EO-based estimates for specific days when field measurements were available. However, the limited availability of these dates, especially across different seasons, posed a challenge for comprehensive analysis. To address this limitation, an alternative approach was developed to leverage the in situ data while introducing a new method for ET computation, beyond the commonly used SARE model. This approach aimed to improve temporal coverage and enhance the accuracy of ET estimates, providing an understanding of water dynamics across the study area.

9.1. Introduction

Machine learning (ML) involves developing algorithms that enable computers to learn from data and make predictions. In the conducted tests, two popular ML algorithms, Light Gradient Boosting Machine (LGBM) regressor *(Guolin Ke et al, 2017*) and Random Forest (RF) regressor *(Buitinck, Lars, et al, 2013),* were trained to assess their performance in predicting parameters related to evapotranspiration. The parameters that the algorithms were trained to predict are listed and described in Table 9-1Table 9-1: List of parameters and their description to be predicted by using ML approach below:

Table 9-1: List of parameters and their description to be predicted by using ML approach

Parameter	Acronym	Description
Crop Stress Coefficient	Ks	A factor that reduces the crop coefficient to account for water stress conditions, indicating reduced water availability or drought.
Actual Evapotranspiration	ETa	The actual rate of water loss from soil and plants, reflecting real conditions including water availability and crop type.

9.2. Experimental design

The experiment was designed to assess the relationship between satellite-derived features and ground truth data related to agricultural water use, specifically focusing on crop coefficient (Kc), crop stress coefficient (Ks), potential evapotranspiration (ET_0), and actual evapotranspiration (ETa). The study was conducted across several agricultural pivots, which are circular fields irrigated using center-pivot systems.

For each pivot, ground truth data were meticulously collected. This data included measurements of Kc, Ks, ET_0, and ETa, obtained through field sensors and validated through manual observations. These ground measurements provided a reliable reference to compare with the satellite-derived estimates.

To complement the ground truth data, images were acquired from two satellites: Landsat 8, which provides multispectral data with a 30-meter resolution, and PRISMA, a hyperspectral satellite offering a much finer spectral resolution. The images from Landsat 8 and PRISMA were first aligned using geometric and radiometric correction techniques to ensure that they accurately corresponded to the same spatial locations across the different datasets.

For each pixel entirely contained within the boundaries of a pivot, relevant features from the aligned satellite images were extracted. These features included thermal bands to estimate surface temperature and various hyperspectral indices that can be sensitive to specific crop conditions. The extracted satellite features for each pixel were then aggregated and associated with the corresponding ground truth data for the entire pivot. This integration allowed for a comprehensive dataset that linked remotely sensed information with on-the-ground measurements, facilitating a detailed analysis of how well satellite data can predict or reflect actual crop conditions and water use.

This experimental design enabled us to evaluate the accuracy and reliability of satellite-derived estimates for Kc, Ks, ET_0, and ETa at the pivot scale, providing valuable insights into the potential of remote sensing technologies in precision agriculture and water resource management.

9.3. Dataset building

The dataset construction began with the extraction of features from satellite images provided by both the PRISMA and Landsat-8 satellites, corresponding to pixels that intersect cultivation pivots for 5 different dates (Table 9-2). Specifically, one image per satellite per date was used, resulting in a total of five PRISMA images and five Landsat-8 images. All the information related to the data satellite missions are provided in 5.2. These features included spatial, spectral, textural, and ground data, which are essential for accurate analysis. Invalid or corrupted data points were carefully removed.

ıle 9-2: Dates of data collected for both in situ and satellite measurements and related experimental crop cycle.

PRISMA acquisition time closest to in-situ date	Landsat-8 acquisition time closest to in-situ date	Date	Crop cycle
10 June 2023	12 June 2023	June 6th 2023	1
16 June 2023	28 June 2023	June 16th 2023	1
15 July 2023	14 July 2023	July 14th 2023	1
21 July 2023	22 July 2023	July 22nd 2023	1
26 July 2024	16 July 2024	July 16th 2024	3

The final step in the dataset construction was the addition of several hyperspectral indices related to vegetation health and moisture content for both plants and soil. These indices are listed and described in Table 9-3 below:

Table 9-3: Hyperspectral indices related to vegetation health and moisture content.

Name of Index	Acronym	Formula	Satellite	Reference	Note
Normalized Difference Water Index	NDWI-1640	$\frac{860nm - 1649nm}{860nm + 1649nm}$	PRISMA	IDB - Index DataBase	
Normalized Difference Water Index	NDWI-2130	$\frac{860nm - 2130nm}{860nm + 2130nm}$	PRISMA	IDB - Index DataBase	
Normalized Difference Vegetation Index	NDVI	$\frac{860nm - 670nm}{860nm + 670nm}$	PRISMA	IDB - Index DataBase	
Water Index	WI	$\frac{900nm}{970nm}$	PRISMA	IDB - Index DataBase	
Moisture Stress Index	MSI	$\frac{1649nm}{820nm}$	PRISMA	IDB - Index DataBase	
Simple Ratio Water Index	SRWI	$\frac{860nm}{1240nm}$	PRISMA	IDB - Index DataBase	Calculated and added to the dataset at test 5
Crop Stress Water Index	CSWI	$\frac{T - Tmin}{Tmax - Tmin}$	Landsat8	Deliverable D6.1	Calculated and added to the dataset at test 5
Normalized Difference Moisture Index	NDMI	$\frac{820nm - 1600nm}{820nm + 1600nm}$	PRISMA	IDB - Index DataBase	
Land Surface Temperature	LST	*(BT / (1 + (0.00115 * BT / 1.4388) * Ln(ε)))*	Landsat8	https://giscrack.com/how-to-calculate-land-surface-temperature-with-landsat-8-images/	Calculated and added to the dataset at test 5

Starting from the original dataset, several tests were conducted. Below, are described the main key tests that illustrate the steps taken to improve data handling and enable the development of algorithms capable of making predictions with real-world data.

The main differences between the tests involve the addition or removal of specific data elements. All datasets were normalized based on their mean and standard deviation, except for normalized vegetation indices, which were normalized individually. To evaluate the performance of the model, R-squared (R^2) was used as a key metric. R^2, also known as the coefficient of determination, measures the proportion of the variance in the dependent variable that is predictable from the independent variables. It ranges from 0 to 1, where a value closer to 1 indicates that the model explains most of the variability in the data, while a value closer to 0 suggests that the model fails to capture the underlying patterns. R^2 is particularly useful for understanding how well the model fits the data, with higher values reflecting a better fit

The main key tests and their respective challenges are described below.

9.3.1. Test n°1

In the first test, the entire original dataset was used to train the RF and LGBM algorithms. Initially, the features were averaged for each pivot, and then the dataset was split into training and testing sets by randomly selecting features from the averaged data. The original dataset included data related to the pivot's location, such as latitude and longitude. After the first training run, both algorithms performed well, with R2 value shown in Table 9-4. However, some issues arose when analyzing the table of important variables. For both algorithms, latitude and longitude emerged as the most significant features. While geographic coordinates can be valuable for estimating evapotranspiration in global models, their dominance in a model trained on localized data suggests overfitting. This overreliance on location data may lead to a model that is poorly calibrated and less generalizable to other regions.

Table 9-4: R^2 for each predicted parameter by using dataset built for test 1.

Model	Accuracy R^2	
	Ks	**ETa**
LGBM	0,79	0,91
RF	0,81	0,90

9.3.2. Test n°2

In the second test, latitude and longitude were removed from the dataset, and a new training process was conducted. While the R^2—particularly for ETa—decreased slightly, the overall performance remained good. Additionally, after predicting Kc, this variable was added back into the training dataset to improve accuracy. However, this introduced several challenges related to using predicted variables to boost the model's performance, as prediction errors could propagate through the model. Initially, this approach appeared promising, but critical issues soon emerged. Specifically, during the dataset splitting process, features were randomly assigned to the training and testing sets. This random selection allowed features from the same pivot, which shared the same ground truth measurements, to appear in both sets. This overlap resulted in data leakage, undermining the validity of the results. Moreover, incorporating predicted variables like Kc—an essential component of the ETa formula—into subsequent re-training steps increased the overall error, as inaccuracies compounded throughout the model. Due to these concerns, this approach will not be repeated in future tests. The results of this test are shown in Table 9-5 below:

able 9-5: R^2 for each predicted parameter by using dataset built for test 2. Eta(2a) represents the rediction of variable Eta after the addiction of predicted variable Kc into the train and test dataset.

Model	Accuracy R^2		
	Ks	**ETa**	**Eta(2a)**
LGBM	0,41	0,44	0,44
RF	0,25	0,54	0,69

9.3.3. Test n°3

In the third test, latitude, longitude, and all Landsat data (except for band 10, due to its relationship with surface temperature) were excluded. The main issue, however, was that the same features were used for both training and test validation for each pivot. This led to overfitting, making the results unsuitable for real-world applications.

The results of this test are reported in Table 9-6 below:

Table 9-6: R^2 for each predicted parameter by using dataset built for test 3.

Model	Accuracy R^2	
	Ks	**ETa**
LGBM	0,35	0,40
RF	0,21	0,61

9.3.4. Test n°4

In the fourth test, the dataset was split by assigning certain pivots for training and others for testing. All features were used for the training pivots, while for the testing pivots, the features were averaged to evaluate the algorithms. Although the results were not particularly strong, this approach provided a valuable foundation for developing models that could be applied to real-world scenarios. Additionally, the prediction of the ET_0 parameter was included in this test, further expanding the model's capabilities. Results are shown in Table 9-7:

Table 9-7: R^2 for each predicted parameter by using dataset built for test 4.

Model	Accuracy R^2	
	Ks	**ETa**
LGBM	0,07	0,48
RF	0,09	0,52

9.3.5. Test n°5

In the fifth test, new indices were introduced, specifically Land Surface Temperature (LST), Crop Water Stress Index (CSWI), and Simple Ratio Water Index (SRWI), which are related to surface temperature and plant water content. The dataset was split in the same manner as in the fourth test, with certain pivots designated for training and others for testing. Incorporating these new indices led to an improvement in results to an acceptable level, especially considering the limited amount of data available. The details are shown in Table 9-8 below:

Table 9-8: R^2 for each predicted parameter by using dataset built for test 5.

Model	Accuracy R^2	
	Ks	**ETa**
LGBM	0,15	0,73
RF	0,05	0,54

As reported in Table 8, the accuracy, in terms of R-squared, is considered acceptable, especially given that the features used for training and testing were kept distinct for each pivot. However, the results for the crop stress coefficient (Ks) were notably poor. This discrepancy may be attributed to timing differences between in-ground data collection and satellite acquisition. This issue is critical, particularly in arid and hot regions, where the timing of data collection can significantly affect the results. For example, if ground data is collected early in the morning or shortly after irrigation, it may differ substantially from satellite data, which is often collected during the hottest part of the day, potentially long after irrigation has occurred

Following in Table 9-9 are reported the most ten important variables for Test n°5:

Table 9-9: Most important variables for each parameter predicted in test 5.

Importance x predicted parameter				
	LGBM		RF	
	Ks	**ETa**	**Ks**	**ETa**
Band / hyper-spectral index	NDMI	Landsat_B10	VNIR_4	LST
	SWIR_41	SWIR_40	VNIR_3	Landsat_B10
	VNIR_40	SWIR_102	NDMI	NDMI
	SWIR_97	NDMI	SWIR_86	SWIR_49
	SWIR_40	WI	Landsat_B10	SWIR_50
	SWIR_89	SWIR_42	SWIR_42	SRWI
	Landsat_B10	SWIR_95	SWIR_136	VNIR_1
	SWIR_35	SWIR_101	SRWI	VNIR_2
	SWIR_98	CSWI	LST	VNIR_3
	SWIR_91	SWIR_97	VNIR_2	VNIR_4

9.3.6. Test n°6

Considering the results obtained in Test n°5, further adjustments were made to improve the model's accuracy. Specifically, additional tests followed Test n°5, including the following steps:

- Eliminating all indices that were not ranked among the top 10 most important variables;
- Eliminating all indices that were not ranked among the top 20 most important variables

A dimensionality reduction of the dataset was performed as a final step, selecting only the top twenty most important variables from Test n°5 to re-train the model and evaluate its capability to predict Eta and Ks. The top twenty variables were selected for each parameter and model, and a new model was trained using this reduced dataset.

For most models, there was a decrease in accuracy, except for the LGBM model trained to predict Eta. The results are not shown for models where accuracy decreased, as the best models for these parameters are already presented in Test n°5. However, in the case of the LGBM model for Eta, the reduction in dimensionality improved the model's accuracy, with the R^2 value increasing from 0.73 to 0.81, as shown in Table 9-10: R^2 for each predicted parameter by using dataset built for test 5 below:

Table 9-10: R^2 for each predicted parameter by using dataset built for test 5.

Model	Accuracy R^2
	ETa
LGBM	0.81

The variables used for this model are reported in Table 9-11 below:

Table 9-11: Variables composing the dataset of Test n°6 for LGBM model.

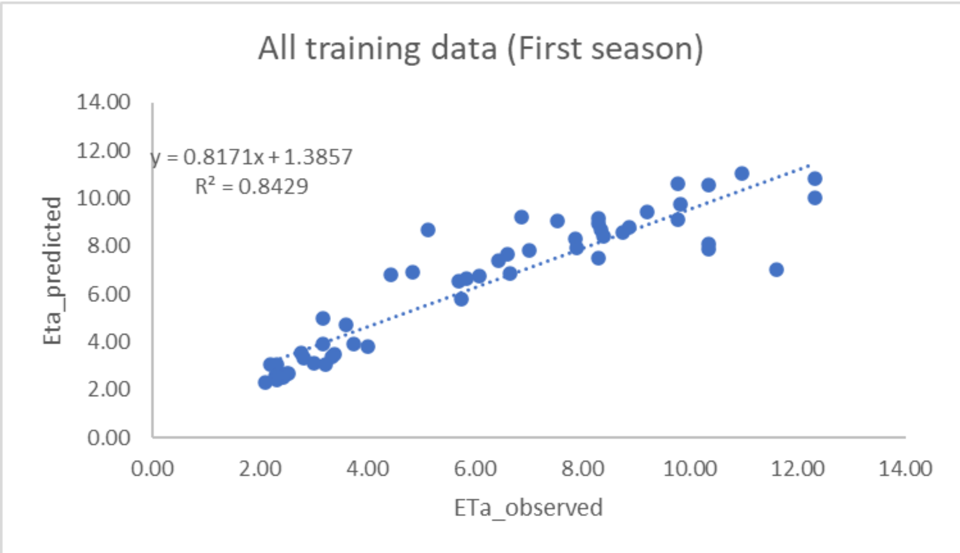
	List of 20 variables used for the training of the model
Band / hyper-spectral index	Landsat B10, NDMI, VNIR 1, SWIR 162, wi, SWIR 87, SWIR 42, SWIR 89, SWIR 102, SWIR 92, SWIR 41, SWIR 97, SWIR 40, SWIR 101, SWIR 98, SWIR 171, SWIR 62, SWIR 99, SWIR 95, CSWI

9.4. Results of ML approach

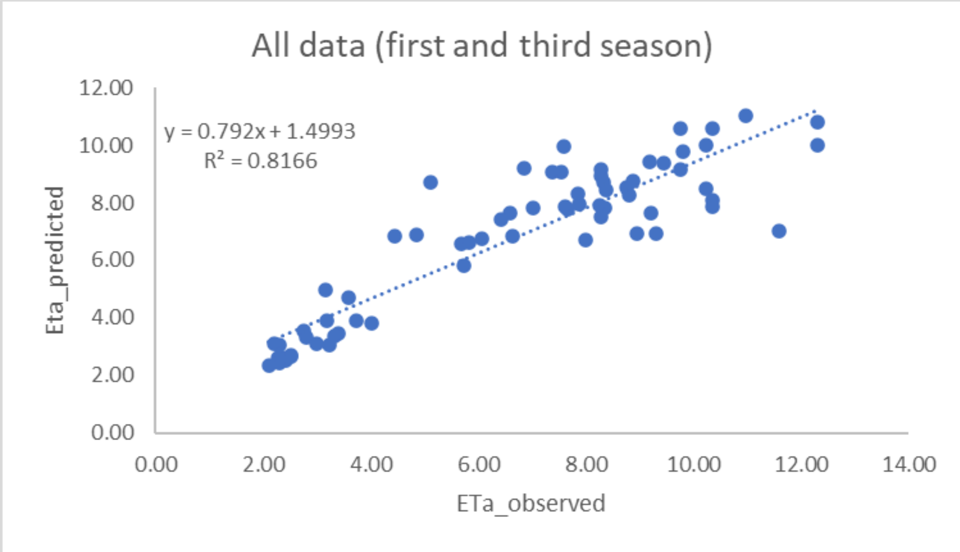
As explained in paragraph 9.3 and in the tests conducted, the aim was to identify the best variables for constructing an optimal dataset to estimate Eta and Ks. As previously noted, the results for predicting the Ks parameter were unsatisfactory, while good results were achieved for estimating Eta. Therefore, the following section will present and discuss the results of the best-trained model for Eta, which is the LGBM model from Test n°6. Additionally, the Eta maps generated for each date in the dataset will be shown below.

As reported in paragraph 9.3.6, the best-performing model was the one trained with only 20 variables rather than the complete set of 236. This phenomenon is commonly known in machine learning as the *Curses of Dimensionality (Taylor, C. Robert, 1993),* where the accuracy of a classifier or regressor initially improves as the number of variables used for training increases, but after reaching a certain point, further dimensionality starts to degrade the model's performance instead of enhancing it *(Hughes, G.F, 1968; Trunk, G. V, 1979; B. Chandrasekaran; A. K. Jain, 1974).*

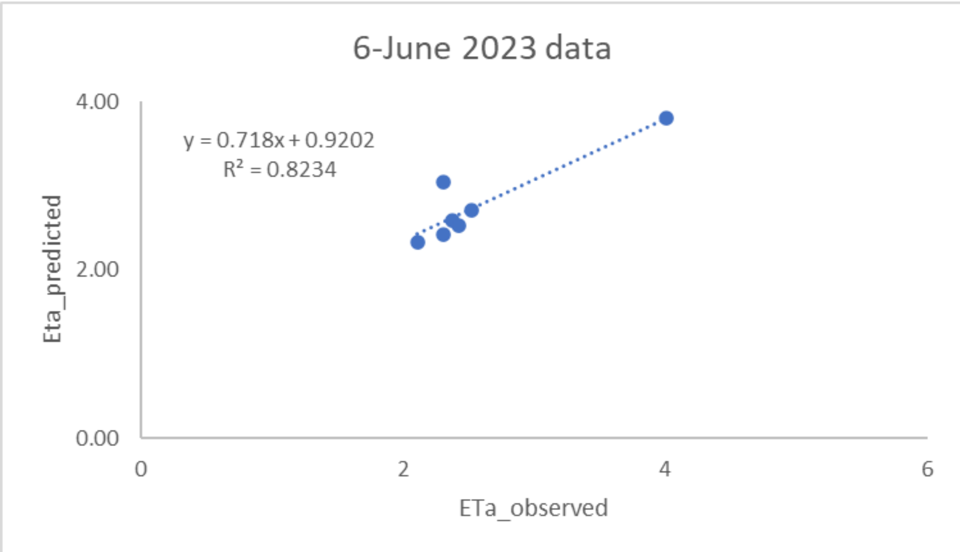
The correlation between the predicted values and those measured in the ground is shown below for all available data and each date analyzed. The correlation between observed and predicted values is quite high for each date and also for all data evaluated together. The only exception is the date of the third cycle harvest (16 July 2024) shown in Figure 9-7, because the PRISMA images are not from the same day as the ground measurements.



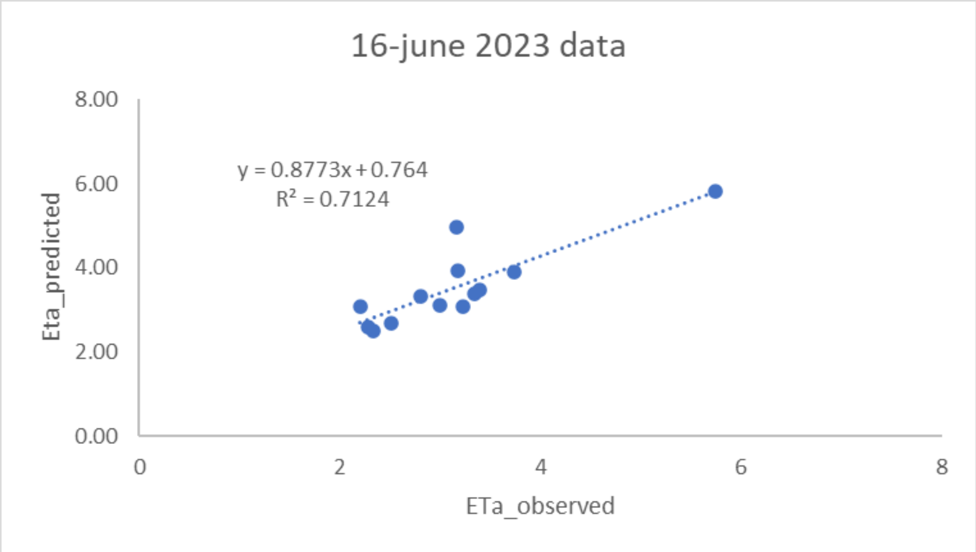
ıre 9-1: scatterplot showing the correlation between predicted and observed values for all data from the first crop cycle used in the training phase.



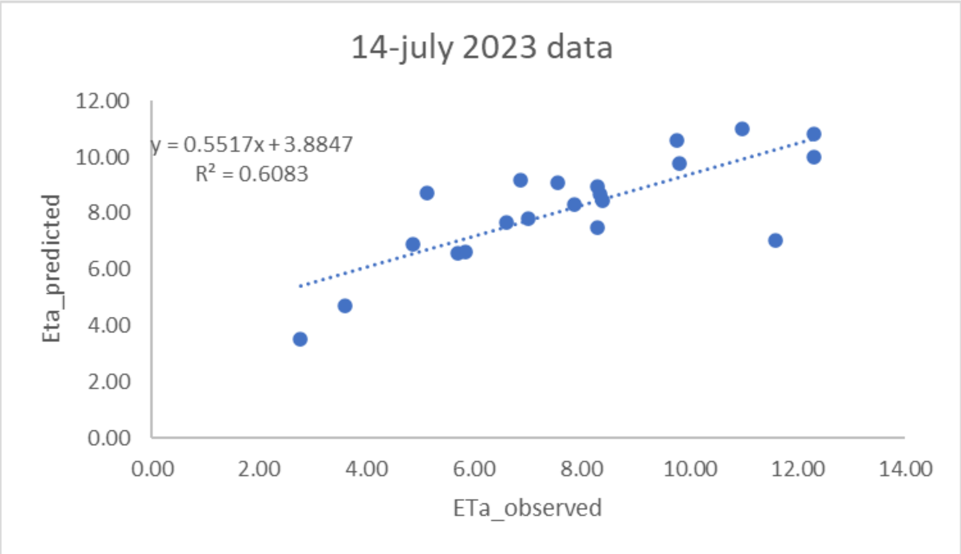
ıre 9-2: scatterplot showing the correlation between predicted and observed values for all available data from the first and third crop cycle.



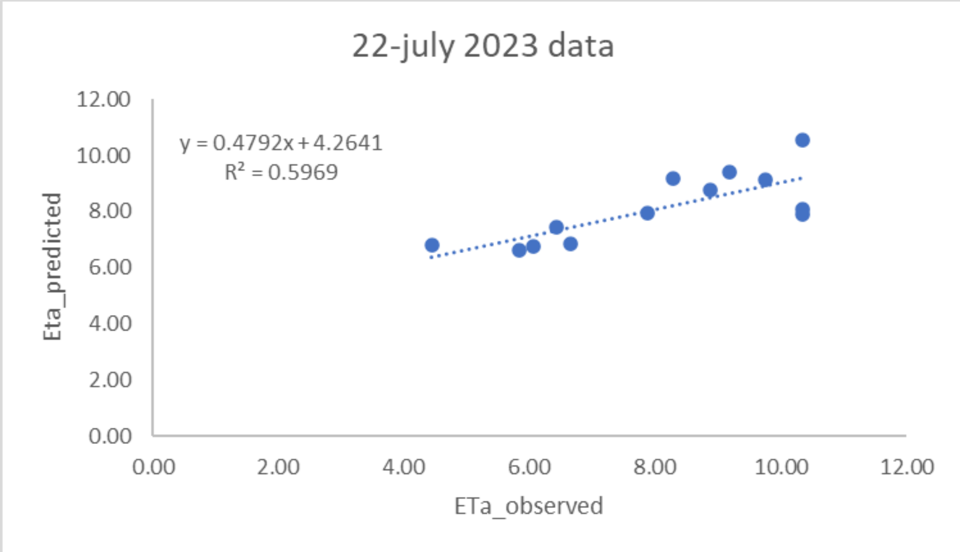
e 9-3: scatterplot showing the correlation between predicted and observed values for the date of June 6th 2023.



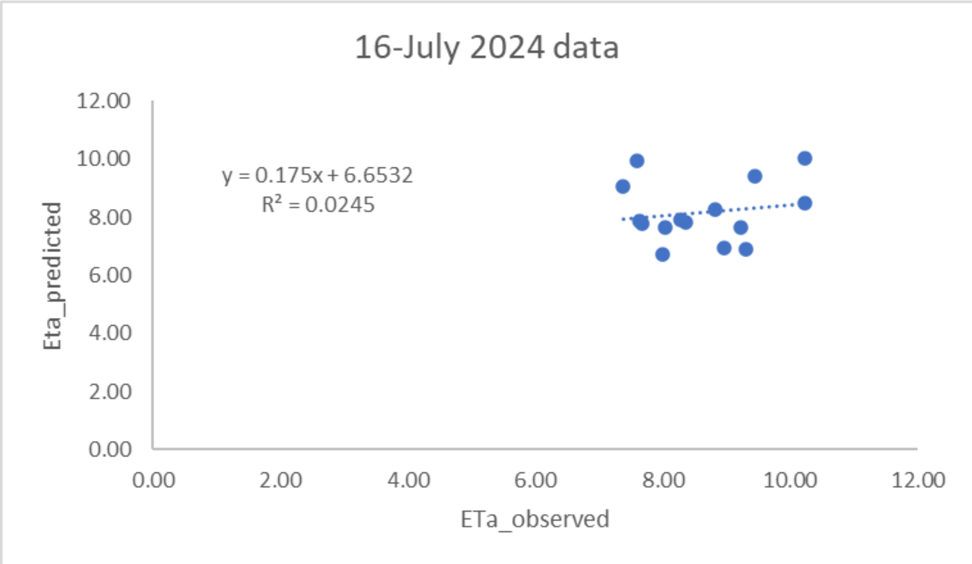
e 9-4: scatterplot showing the correlation between predicted and observed values for the date of June 16th 2023.



e 9-5: scatterplot showing the correlation between predicted and observed values for the date of July 14th 2023;



'e 9-6: scatterplot showing the correlation between predicted and observed values for the date of July 22nd 2023.



'e 9-7: scatterplot showing the correlation between predicted and observed values for the date of July 16th 2024.

9.4.1. Eta maps

An Eta map was generated for all the selected dates, where the correlation between predicted and observed values was shown in paragraph 9.4. It is reported below. As the summer season progresses, it is possible to observe how Eta becomes higher.

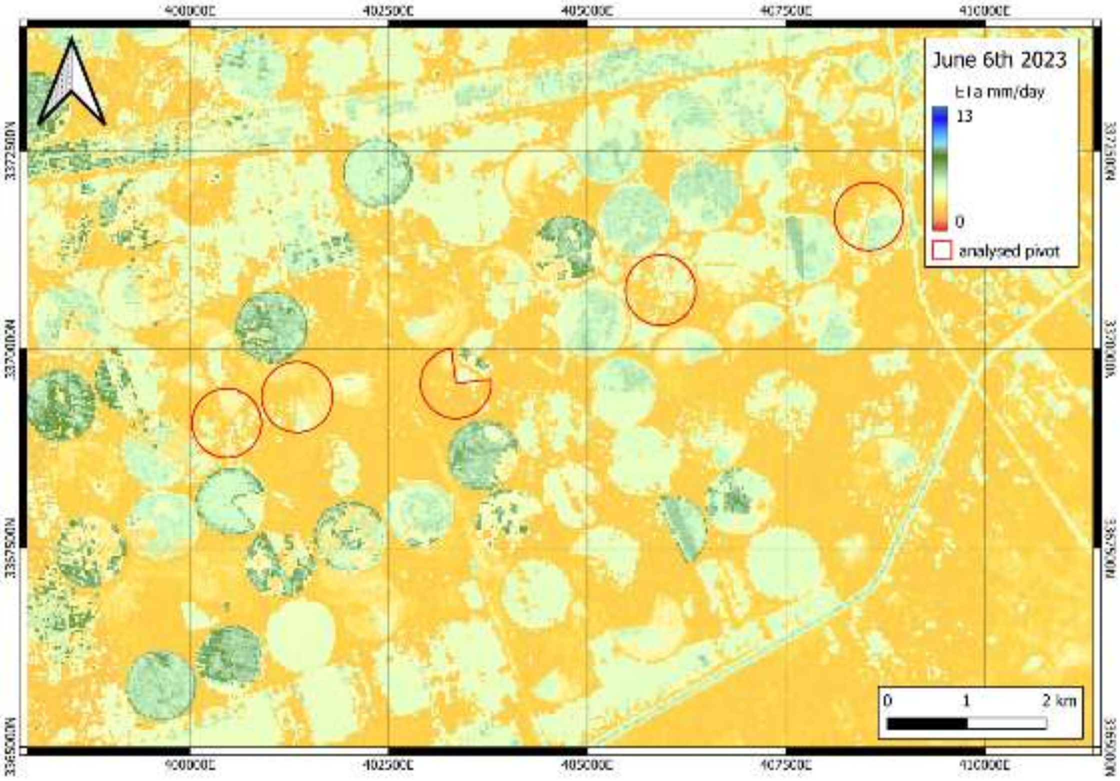


Figure 9-8: ETa map of June 6th, 2023. Inside the map are reported the study areas.

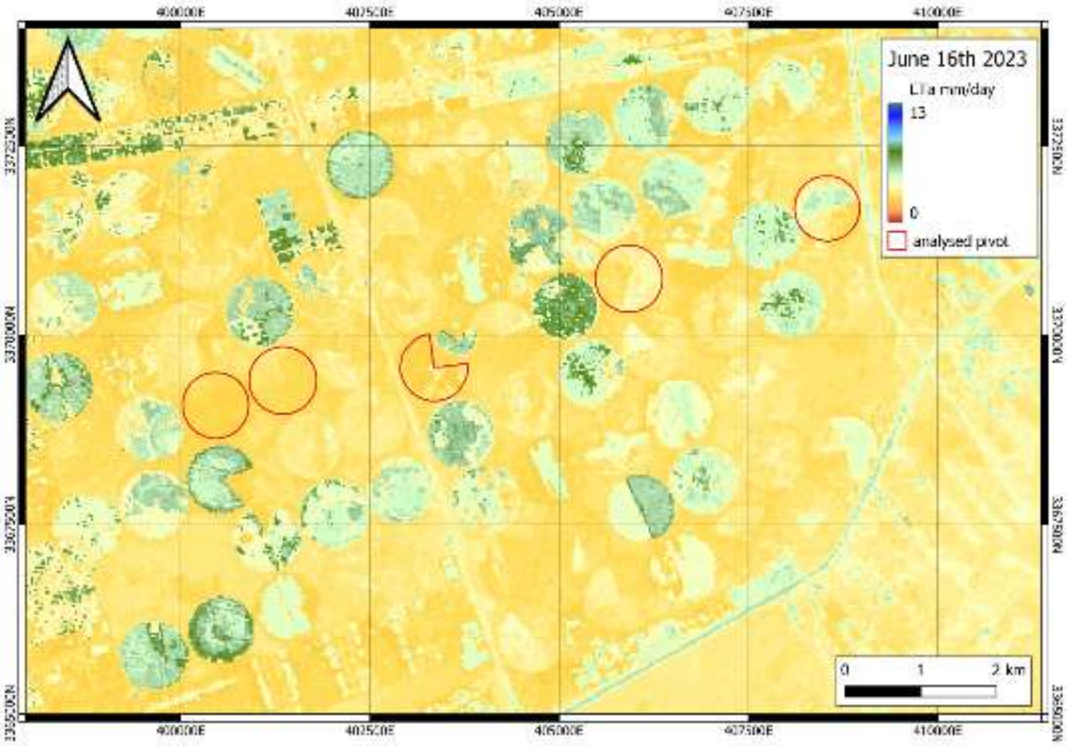


Figure 9-9: ETa map of June 16th, 2023. Inside the map are reported the study areas.

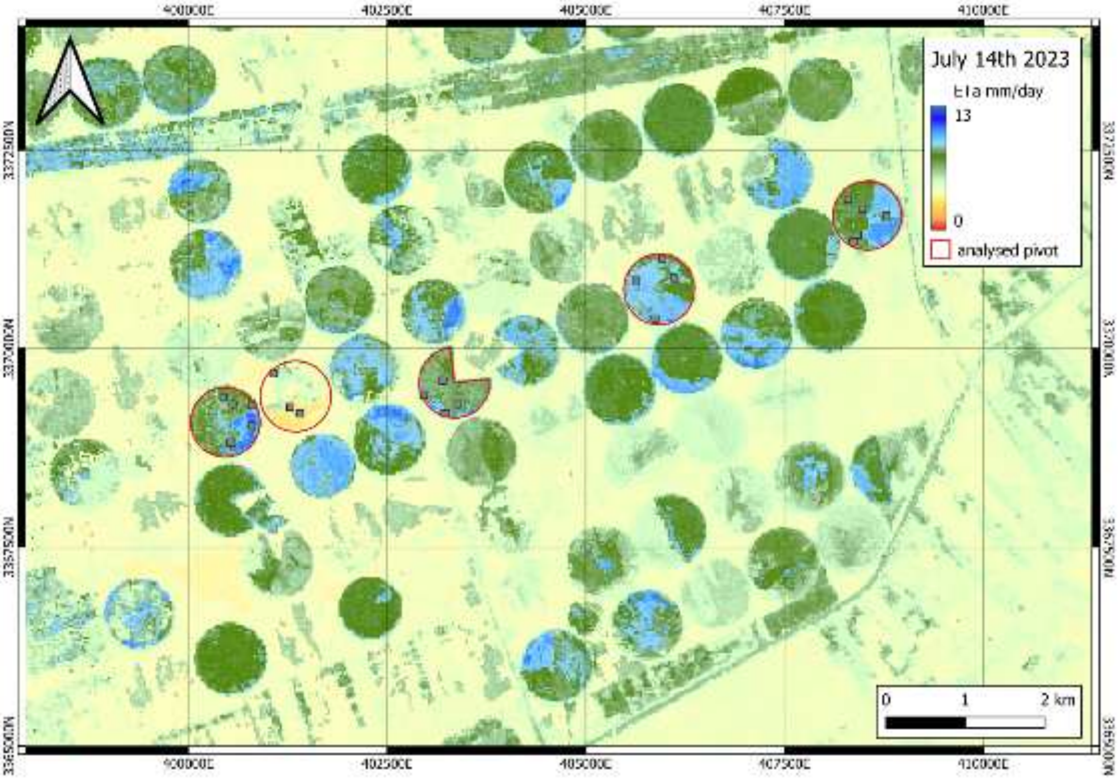


Figure 9-10: ETa map of July 14th, 2023. Inside the map are reported the study areas.

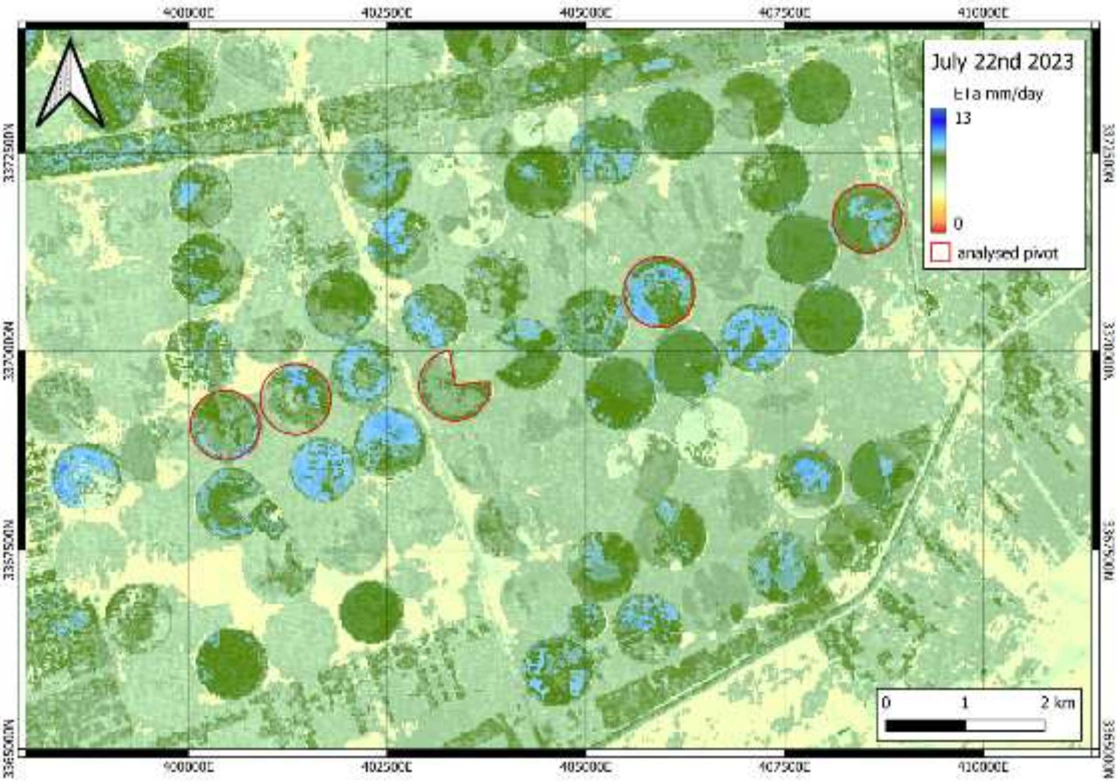


Figure 9-11: ETa map of July 22nd, 2023. Inside the map are reported the study areas.

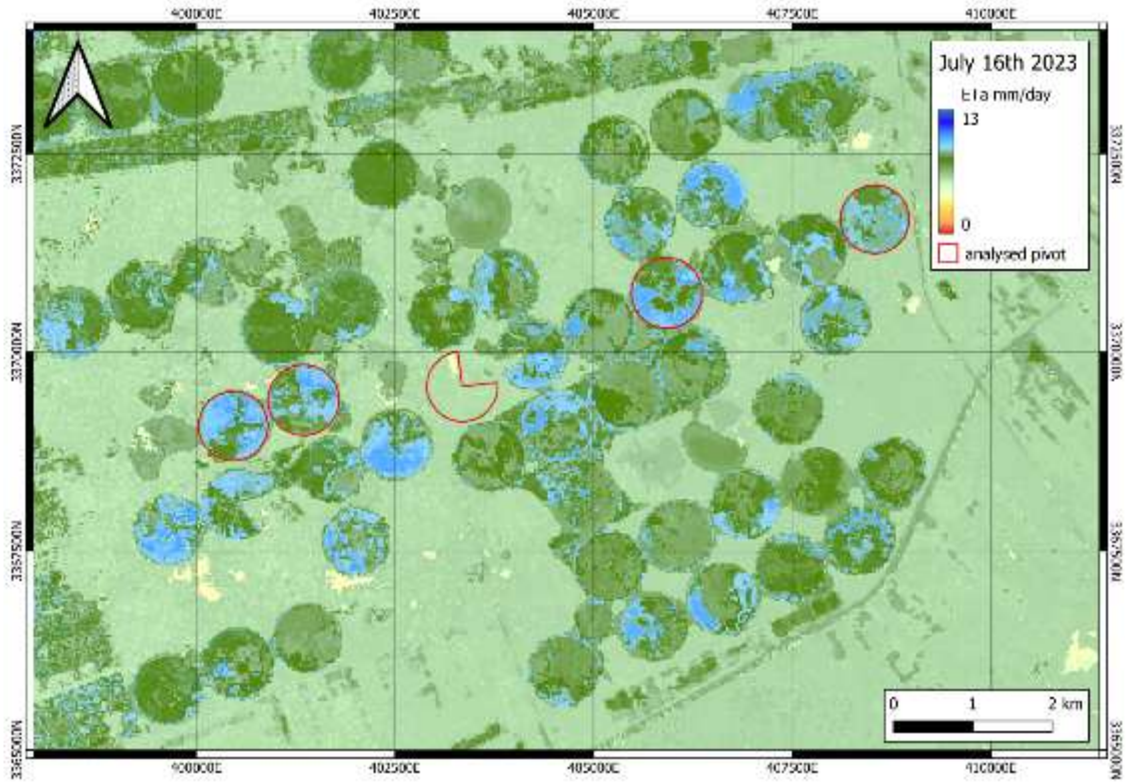


Figure 9-12: ETa map of July 16th, 2024. Inside the map are reported the study areas.

10. Products Specifications

The definition of the products specifications is a key element that drives the process of algorithm development. Technical specifications of the products developed on the demonstration site are described in this section using a set of tables whose content is adapted from the “Document Requirements Definition for Earth Observation Product Specifications” of the European Association of Remote Sensing Companies (EARSC)[3].

The following products will be demonstrated during the project:

- Product 1: Mapping Actual evapotranspiration (ETa) for yield prediction, water efficiency, and water productivity assessment
- Product 2: Mapping Crop Water Stress Index (CWSI) for preventive actions against water stress
- Product 3: Mapping vegetation indices for monitoring of growth development stages

[3] EASRC Document: EARSC/guideline/2013/001, accessible for example at: https://mafiadoc.com/queue/earsc-product-specification-drd-guideline-v10-v2_59d59b5cp723dd4bf42ce79a.html

Product 1: Mapping Actual Evapotranspiration (ETa maps)
Content
Mapping of the actual crop evapotranspiration as a product of ET0 estimated using SARE, EO-derived crop coefficient (Kc) and water stress coefficient (Ks) to predict yield, estimate water productivity, and water use efficiency.
Geographic coverage
• Demonstration site is in El Salheya under the Sharqia governorate in the north of Egypt
Input data sources
• Satellite data: hyperspectral (PRISMA) and thermal EO data (Landsat 8&9 and ECOSTRESS) • In situ data: Meteorological data (max and min air temperature, humidity, wind speed, radiation, precipitation); Crop type, planting date agricultural practices, harvest date; Plant density, leaf area index (LAI), plant height; Water application (time and amount); Soil moisture content; Water holding capacity, wilting point, total available water; Irrigation system efficiency/uniformity; Crop yield.
Methodology
The solution is composed of three main sub-models: EO-based ETa, ground-based ETa, and EO-based crop development monitoring. The ground-based ETa serves for the validation of the EO-based ETa whose output is used for near real time crop development monitoring
Spatial resolution and coverage
• 30m (PRISMA) , 30m (Landsat) and 70m (ECOSTRESS) resolutions over the geographic
Coordinate Reference System
• WGS84 UTM36 (or according to the User needs)
Accuracy assessment approach
A statistical analysis performed for the model's validation foresees the application of a series of statistical criteria including the mean error (ME), the mean relative error (MRE), the root mean square error (RMSE) and the correlation coefficient (r).
Frequency
• 15 days as baseline (according to EO data availability)
Availability
• 24 hours after data importing
Delivery/Output format
• GIS-ready GeoTIFF output • Delivery by WMS or downloading and publication on a web portal
Data type
• Geospatial raster data
Raster coding
• Cell by cell
Metadata
• INSPIRE standard compliant

Table 10-1: Product 2 CWSI maps

Product 2: Mapping Crop Water Stress Index (CWSI maps)
Content
Map of Crop Water Stress Index the water stress map is fundamental to preventive irrigation management actions (adjustment of scheduling: volumes and frequency) against water stress the irrigation scheduling.
Geographic coverage
• Demonstration site is in El Salheya under the Sharqia governorate in the north of Egypt
Input data sources
• Satellite data: hyperspectral (PRISMA) and thermal EO data (Landsat 8&9 and ECOSTRESS) • In situ data: Meteorological data (max and min air temperature, humidity, wind speed, radiation, precipitation); Crop type, planting date agricultural practices, harvest date; Plant density, leaf area index (LAI), plant height; Water application (time and amount); Soil moisture content; Water holding capacity, wilting point, total available water; Irrigation system efficiency/uniformity; Crop yield.
Methodology
The solution is composed of three main sub-models: EO-based ETa, ground-based Eta, and EO-based crop development monitoring. The ground-based ETa serves for the validation of the EO-based ETa whose output is used for near real time crop development monitoring
Spatial resolution and coverage
• 30m (PRISMA), 30m (Landsat) and 70m (ECOSTRESS) resolutions over the geographic
Coordinate Reference System
• WGS84 UTM36 (or according to the User needs)
Accuracy assessment approach
A statistical analysis performed for the model's validation foresees the application of a series of statistical criteria including the mean error (ME), the mean relative error (MRE), the root mean square error (RMSE) and the correlation coefficient (r).
Frequency
• 15 days as baseline (according to EO data availability)
Availability
• 24 hours after data importing
Delivery/Output format
• GIS-ready GeoTIFF output • Delivery by WMS or downloading and publication on a web portal
Data type
• Geospatial raster data
Raster coding
• Cell by cell
Metadata
• INSPIRE standard compliant

Table 10-2: Product 3 Vegetation indices maps

Product 3: Mapping Vegetation indices (Vegetation indices maps)
Content
Mapping vegetation indices for monitoring crop development stages as a potential indicator for crop health including agronomic and engineering parameters (Distribution efficiency and uniformity).
Geographic coverage
• Demonstration site is in El Salheya under the Sharqia governorate in the north of Egypt
Input data sources
• Satellite data: hyperspectral (PRISMA) and thermal EO data (Landsat 8&9 and ECOSTRESS) • In situ data: Meteorological data (max and min air temperature, humidity, wind speed, radiation, precipitation); Crop type, planting date agricultural practices, harvest date; Plant density, leaf area index (LAI), plant height; Water application (time and amount); Soil moisture content; Water holding capacity, wilting point, total available water; Irrigation system efficiency/uniformity; Crop yield.
Methodology
The solution is composed of three main sub-models: EO-based ETa, ground-based ETa, and EO-based crop development monitoring. The ground-based ETa serves for the validation of the EO-based ETa which output, is used for near real time crop development monitoring
Spatial resolution and coverage
• 30m (PRISMA) , 30m (Landsat) and 70m (ECOSTRESS) resolutions over the geographic
Coordinate Reference System
• WGS84 UTM36 (or according to the User needs)
Accuracy assessment approach
A statistical analysis performed for the model's validation foresees the application of a series of statistical criteria including the mean error (ME), the mean relative error (MRE), the root mean square error (RMSE)and the correlation coefficient (r).
Frequency
• 15 days as baseline (according to EO data availability)
Availability
• 24 hours after data importing
Delivery/Output format
• GIS-ready GeoTIFF output • Delivery by WMS or downloading and publication on a web portal
Data type
• Geospatial raster data
Raster coding
• To be defined

11. Use Cases and Applications

The developed solution offers versatile applications across water management and agricultural sectors. It can be effectively utilized by water managers, farmers, and governmental organizations to streamline various processes including irrigation scheduling, water accounting and allocation, and water rights administration.

11.1. Real-world Scenarios

11.1.1. Crop development

The tool can be employed to monitor crop development by utilizing satellite imagery and ground-based sensors to track various indicators of plant growth and health over time. One commonly used indicator is the Normalized Difference Vegetation Index (NDVI). NDVI provides valuable information about vegetation density and vigor, serving as a proxy for crop development. By analyzing changes in NDVI values throughout the growing season, farmers and agronomists can assess the progression of crop growth, identify areas of stress or underperformance, and make informed decisions regarding management practices such as fertilization, irrigation, and pest control.

11.1.2. Crop water stress

Remote sensing can be instrumental in monitoring crop water stress by capturing key indicators of plant hydration status and physiological responses to water availability. One widely used metric for assessing crop water stress is the Crop Water Stress Index (CWSI), which can be derived from thermal infrared imagery.

High CWSI values indicate greater levels of water stress within the crop canopy, signaling the need for irrigation or other water management interventions. However, even when information about CWSI is obtained in near real-time, it can serve as a valuable resource for informing and improving future irrigation practices.

11.1.3. Crop yield estimation

The output generated by the developed solution functions as an input for the yield estimation, following the prescribed equation.

$$\left(1-\frac{Y_a}{Y_m}\right)=K_y\left(1-\frac{ET_a}{ET_c}\right)$$

- With Ym the maximum potential harvested yield (available from literature/libraries and/or local data sets), and Ky the crop-specific yield response factor (available from literature/libraries and/or local sets).
- Ya (actual or harvested yield) will be validated using the in-field harvested yield available for the case study.

The yield obtained serves as the input for estimating the water productivity, facilitating the assessment of efficiency in water usage within the agricultural system.

11.1.4. Water productivity (Wp)

The developed solution provides a comprehensive toolkit for evaluating the relationship between water input and crop productivity for evaluating the efficiency of water usage in agriculture through indicators such as Water productivity (Wp).

$$WUE = \frac{\text{Yield}}{Applied\ water}$$

This tool empowers farmers and water managers to tailor irrigation strategies according to specific crop requirements, thus maximizing agricultural output while minimizing water consumption. Moreover, the application of such tools enables the identification of areas with suboptimal water use practices, facilitating targeted interventions to enhance Wp and alleviate water scarcity concerns.

12. Challenges and Future Improvements

12.1. Known Challenges

Navigating the complexities of agricultural monitoring in arid regions like Egypt presented several formidable challenges throughout the project. Due to a lack of precise scheduling, instances occurred where irrigation events did not align with sensor imaging sessions (season 1), creating disparities in data collection timelines. This fact resulted in strange results where in situ values seemed a lot higher than the EO ones and pretty inconsistent with each other. This anomaly is attributed to irrigation events that precede the measurements. Following irrigation, the upper soil undergoes a drying phase due to wind and solar radiation. Consequently, the evaporation portion of evapotranspiration decreases during this period. Using the Time Domain Reflectometry (TDR) at the field level allows to access soil content at greater root depths, leading to an increase in field-based ETa measurements.

Regarding season 2 and 3, field visits typically yield only a small number of data points, usually around 4 to 7 measurements per visit, making it tough to draw broad comparisons. Additionally, aligning Earth Observation (EO) data with in-situ collection is tricky due to the limited number of monitoring dates, which are often dictated by the specific pivot being considered. The needs for monitoring also shift depending on the crop's growth stage—early growth stages may only require occasional observation, while more advanced stages, especially during peak irrigation, call for more frequent satellite passes to capture critical data. Inaccuracies in traditional in-situ data collection methods can lead to poor correlations between EO data and what's measured on the ground. Using more advanced techniques like eddy covariance flux towers could significantly improve the reliability of this data. Another issue is the mismatch between soil moisture levels derived from EO thermal data, such as the Crop Water Stress Index (CWSI), and those measured directly in the field using tools like TDR probes. Factors like soil water retention, sunlight, or wind may affect this relationship, leading to discrepancies. When it comes to assessing leaf temperature from satellite imagery to gauge crop water stress, the 30-meter spatial resolution can introduce errors, especially at the outer edges of pivots where bare soil may distort land surface temperatures. This makes it harder to accurately pinpoint "hot" and "cold" spots within the crop.

Furthermore, the scarcity of ECOSTRESS data was a challenge, as the limited availability prevented the pre-ordering of images to synchronize with in-situ observations. Consequently, the reliance on Landsat data for thermal insights became imperative. Compounding these issues, some ECOSTRESS data were captured in the early morning hours, typically around 2 or 4 AM. In the desert climate of Egypt during the night, characterized by cooler temperatures and minimal atmospheric interference, such early captures introduced complexities in thermal measurements. Capturing LST relies on reflectance, making those dates difficult to work with because of this irrelevance.

Under the same conditions, the ML approach shows promising results compared to the SARE approach, encouraging further investigation and exploration of this methodology. To enhance its potential, it is crucial to increase the quantity of data, allowing the construction of a more robust dataset. Expanding the volume of input data for training will enable the development of a more resilient and accurate ML model, ultimately improving performance and reliability. Based on the ML tests conducted and described in paragraph 9.3, as well as the type of data collected and the results discussed in paragraph 9.4, there is significant potential for machine learning to achieve better outcomes in the future. Achieving these improvements requires a substantial increase in data collection, particularly through in-situ sampling conducted on dates that align with satellite acquisitions over the area of interest. A well-defined data sampling scheme is also crucial for improving consistency and accuracy. For example, synchronizing the timing of in-situ sampling with satellite acquisitions would be particularly important for critical variables like temperature, which can exhibit significant variations throughout the day in arid and hot environments. The ML results also suggest that this approach could be effectively scaled over time, as demonstrated in the results reported in paragraph 9.4.1 and illustrated in Figure 9-12, where the estimated ETA values align with data from previous years.