EO MAJI
EO AFRICA EXPLORERS

STATE OF THE ART REVIEW

V1

Date: 27/04/2023

Contract No.

4000139395/22/I-DT

Submitted by

In Cooperation with:

DOCUMENT RELEASE SHEET

Author:	Héctor Nieto (AECSIC), Vicente Burchard (AECSIC), M. Pilar Martín (AECSIC)	Sign Date:
In collaboration with:	Radoslaw Guzinski (DHI) Michael Munk (DHI) Darren Ghent (UoL)	
Approval	Zoltan Szantoi	Sign Date:
Distribution:	ESA and partners	

CHANGE RECORD

Version	*Date*	*Page(s)*	*Change record*	*Release*
0	27/04/2023	12	First submission to ESA	1
1				
2				
3				

1 Introduction

1.1 Project objective

This project aims to implement a prototype for irrigation mapping and crop yield estimation using inputs from the scientific ECOSTRESS and PRISMA missions. The final aim is to develop workflows, in collaboration with the African Early Adopters and EO partner(s), that support African irrigation and food security management, as well as transfering these R&D learnings and results to African end-users and stakeholders. More specifically the project objectives in this project can overall be listed as:

- Exploration of the capabilities for future operational Copernicus missions (LSTM+CHIME) to estimate ET and crop water stress.
- Investigate the potential of PRISMA hyperspectral observations and thermal-based crop stress metrics to improve crop yield/biomass estimations to support agricultural monitoring
- Complement the ET retrievals with crop yield, in order to acquire a better understanding of water use efficiency (WUE) of cultivated landscapes.
- Direct involvement of Africal Early Adopters, in order to secure the usefulness and applicability of the prototype.
- Publish the findings in a freely available code repository and as scientifically peer-reviewed papers, as well as to promote the codes through other outreach activities such as development of digital notebooks.

All activities are to be carried out within the duration of the project lifetime from 1 December 2022 to 30 November 2024.

1.2 Scope of Document

This document presents the State of the Art for developing the suggested products in EO MAJI (EO Africa Explorers" ESA AO/1-11038/21/I-DT), with the goal evaluating the existing EO algorithm/s for irrigation delimitation, irrigation accounting and crop yield mapping. This review also aims at justifying the selected technique that will be developed and validated during the project

1.3 Reference documents

REF-1	Statement of Work: ESA-EOP-SD-SOW-0250 – EO AFRICA EXPLORERS
REF-2	EO MAJI proposal dated 18/02/2022
REF-3	Clarification request from ESA dated 06/06/2022
REF-4	Response to clarification dated 22/06/2022
REF-5	Contract No. 4000139395/22/I-DT

2 Methods

2.1 Starting point

2.1.1 Sharpening and sensor fusion

For land surface temperature (LST) data sharpening and sensor fusion a new physics-based approach to downscaling has been developed which combines high resolution optical data with coarser resolution thermal data. The approach utilises the information from the high resolution optical data to provide a sub-pixel variability in emissivity while crucially weighting the information from the coarser resolution thermal data based on the respective uncertainties in a covariance matrix.

A downscaled LST product is derived by taking LST from the Sea and Land Surface Temperature (SLSTR) product at 1 km resolution and using Sentinel-2 derived Land Surface Emissivity (LSE) data to iteratively update the LST in an Optimal Estimation (OE) scheme. In this methodology the OE scheme will assimilate the medium resolution SLSTR LST and the Sentinel-2 LSE and attempt to minimise the difference between the SLSTR calculated Bottom-Of-Atmosphere (BOA) brightness temperatures (BTs) and simulated BOA BTs generated from the SLSTR LST and the Sentinel-2 LSE. This would be an operation performed at the resolution of the Sentinel-2 data.

The processing first obtains the normalised difference vegetation index (NDVI) from the Sentinel-2 data and maps the SLSTR LST data onto the Sentinel-2 NDVI pixel grid. The NDVI data is used with the NDVI threshold method[1] to estimate the LSE.

The re-gridded SLSTR LST combined with the original coarse resolution LSE associated with the pixel in the Split-Window processing (whether this be explicit or through biome estimation) is used with Planck's Law to estimate the BOA BT values. In order to transfer the traceable uncertainties from the Split-Window algorithm, the random component of the total uncertainty from the SLSTR LST algorithm is applied as a noise on these BOA BT values using a Gaussian random distribution.

The SLSTR LST is then used with the Planck function again, but this time with the LSE derived from the Sentinel-2 NDVI threshold method. Two sets of BOA BTs are evaluated and the differences used to iteratively update the estimate for LSE used in the retrieval until an empirically determined threshold is reached from where there is no longer any significant improvement to be made by further iteration.

This methodology produces a full uncertainty breakdown including both the input total uncertainty of the SLSTR LST retrieval as well as the uncertainties due to the downscaling. The use of the SLSTR LST means that any additional uncertainties are a result of the retrieval process and the surface parameters, as the atmospheric uncertainty should have been fully captured in the SLSTR LST retrieval.

[1] Sobrino, J. A.; Jimenez-Munoz; J. C.; Soria, G.; Romaguera, M.; Guanter, L.; Moreno, J.; Martinez, P. Land surface emissivity retrieval from different VNIR and TIR sensors. *IEEE Transactions on Geoscience and Remote Sensing*, **2008**, 46(2), 316-327. https://doi.org/10.1109/TGRS.2007.904834

2.1.2 ET

The ET algorithm in this project will take advantage of results and methods developed in ESA's Sentinels for Evapotranspiration (Sen-ET[2,3]) and ET4FAO[4] projects. It exploits the synergies between high spatial resolution and lower revisit time shortwave sensors with the high revisit but coarser thermal infrared sensors. The developed algorithm have shown to be robust in a wide range of landscapes and conditions[2,3,4,5,6] including delivering products at scales suitable for agronomic monitoring and applications.

The algorithm will make use of the two-source energy balance (TSEB) model as it provides a good balance between being physically-based while having relatively few inputs and parameters. It also has the practical advantage of already having an established open-source Python code implementation (pyTSEB, https://github.com/hectornieto/pyTSEB) to run TSEB independently or to easily integrate it within other pipelines.

The main inputs required by TSEB is land surface temperature (LST) derived from thermal infrared (TIR) sensors. In this case ECOSTRESS mission will provide highly valuable TIR data acquisitions with a frequent revisit time and 70m spatial resolution. In order to compute continous daily ET estimates, whenever ECOSTRESS data is not available, the sharpened LST with Sentinel-3 + Sentinel-2 will be used instead. In addition, for cloudy oversvations, gap-filling techniques will be applied as the ones proposed in [4,7].

2.2 Crop Yield

Earth Observation (EO) methods have long been established as a promising approach to support agricultural activities and food security at multiple spatio-temporal scales. Indeed, satellite remote sensing is increasingly being used to support government agencies ability to monitor agricultural water use, yield and irrigation delimitation at large spatial scales, allowing to support policy making, account for natural resources and mitigate the effects of climate change on agricultural production and food security, especially in data-scarce regions such as the African continent. Over the years, different methods have been developed to estimate crop yield based on EO imagery.

These techniques can be largely grouped into 1) regression-based empirical models, 2) light-use efficiency (LUE) models, 3) data assimilation methods that calibrate/force crop growth models using remote sensing data and 4) hybrid models that use crop growth models to train remote sensing-based regression models.

[2] Guzinski, R.; Nieto, H. Evaluating the feasibility of using Sentinel-2 and Sentinel-3 satellites for high-resolution evapotranspiration estimations. *Remote sensing of Environment*, **2019**, 221, pp.157-172. https://doi.org/10.1016/j.rse.2018.11.019

[3] Guzinski, R.; Nieto, H.; Sandholt, I.; Karamitilios, G. Modelling high-resolution actual evapotranspiration through Sentinel-2 and Sentinel-3 data fusion. *Remote Sensing*, **2020**, 12(9), p.1433. https://doi.org/10.3390/rs12091433

[4] Guzinski, R.; Nieto, H.; Sánchez, J.M.; López-Urrea, R.; Boujnah, D.M.; Boulet, G. Utility of Copernicus-based inputs for actual evapotranspiration modeling in support of sustainable water use in agriculture. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **2021**, 14, pp.11466-11484. https://doi.org/10.1109/JSTARS.2021.3122573

[5] Bellvert, J.; Jofre-Ĉekalović, C.; Pelechá, A.; Mata, M.; Nieto, H. Feasibility of using the two-source energy balance model (TSEB) with Sentinel-2 and Sentinel-3 images to analyze the spatio-temporal variability of vine water status in a vineyard. *Remote Sensing*, **2020**, 12(14), p.2299. https://doi.org/10.3390/rs12142299

[6] Aguirre-García, S.D.; Aranda-Barranco, S.; Nieto, H.; Serrano-Ortiz, P.; Sánchez-Cañete, E.P.; Guerrero-Rascado, J.L. Modelling actual evapotranspiration using a two source energy balance model with Sentinel imagery in herbaceous-free and herbaceous-cover Mediterranean olive orchards. *Agricultural and Forest Meteorology*, **2021**, 311, p.108692. https://doi.org/10.1016/j.agrformet.2021.108692

[7] Delogu, E.; Olioso, A.; Alliès, A.; Demarty, J.; Boulet, G. Evaluation of multiple methods for the production of continuous evapotranspiration estimates from TIR remote sensing. *Remote Sensing*, **2021**, 13(6), p.1086. https://doi.org/10.3390/rs13061086

2.2.1 Empirical models

There is a large body of literature that have demonstrated a strong relation between crop yield and vegetation indices (VIs) from satellite imagery [8] [9] [10] . Indeed, these methods generally establish an empirical relationship between VIs and observed crop yield. Subsequently, these calibrated empirical models are applied to predict and/or forecast crop yield for upcoming seasons. Most of the developed models used aggregated VIs over a period of time, rather than instantaneous satellite overpasses, as the relation of crop yield and spectral characteristics varies with crop growth, while temporal aggregation also limits the effects of other factors (e.g. clouds, soils) that affect the vegetation spectral response[11]. Indeed, there is a general consensus that the relationship between spectral VIs and crop yield is largely dependent on the seasonal timing of observation and aggregation. The extensive literature on the NDVI-yield relationship suggest mid-to-late season NDVI better represents yield estimates than maximum NDVI or other seasonal integration[3]. For example, Rasmussen (1992)[12] demonstrated that early season NDVI had no significant relationship with millet yield (r2 > 0.1), while NDVI values 30 days after the peak maximum NDVI explained 90% of the variance of observed yields in Burkina Faso. Bognár et al. (2011)[1] found that they could could successfully forecast county-level winter wheat yield in Hungary up to 50 days before harvest and county-level corn up to 70 days harvest using near-infrared (NIR)-based spectral indices from the National Oceanic and Atmospheric Administration's (NOAA) Advanced Very High Resolution Radiometer (AVHRR). The study by Funk and Budde (2009)[3] also stressed the importance of temporal smoothing to reduce atmospheric contamination and applying masks to remove non-agricultural vegetation signals, including removing pre-season vegetation signals for increased robustness, in developing scale-invariant empirical models in in East Africa (i.e., Kenya and Zimbabwe).

While most empirical models for crop yield have been based using VIs from the shortwave optical region (i.e. 0.4-2.5 µm), certain studies have also explore the use of thermal infrared (TIR: 8-14 µm) remote sensing to better incorporate the effects of seasonal droughts on crop yield predictions. Johnson (2014)[13] showed strong empirical relations combining both MODIS-based NDVI and land surface temperature (LST) with corn and soybean yield over the midwest USA. Similarly, Anderson et al. (2016)[14] showed that the evaporative stress index (ESI) from MODIS showed higher correlation with observed yield of major crops in Brazil compared to traditionally indices such as leaf area index (LAI), including showing an earlier response (10 – 25 days) to extreme events. Gómez-Candón et al. (2021)[15] also suggested the high utility of surface energy balancing (SEB) modeling of ET, which combines both LAI and LST inputs, as en important indicator for crop yield. Indeed, Franch et al. (2019)[2] also incorporated the evaporative fraction, estimated through a SEB model, as a predictive variable in their empirical model to estimate winter wheat yield for USA and Ukraine at the county and *oblast* level, respectively.

2.2.2 Light-use efficiency (LUE) models

The Monteith Light-Use Efficiency (LUE) concept [16] has been widely applied to model vegetation carbon uptake, which can be related to crop yield with the so called harvest index (i.e. the ratio of yield to aboveground biomass). This approach suggests that a proportional relationship exists between carbon uptake or Gross Primary Production (GPP) and incoming solar radiation at the canopy level. GPP is function of incoming photosynthetically active radiation (PAR), the fraction of absorbed PAR (fAPAR) and a light use efficiency (LUE)

[8] P. Bognár et al., "Yield Forecasting for Wheat and Corn in Hungary by Satellite Remote Sensing," *International Journal of Remote Sensing* 32, no. 17 (2011): 4759–67, https://doi.org/10.1080/01431161.2010.493566.

[9] B. Franch et al., "Remote Sensing Based Yield Monitoring: Application to Winter Wheat in United States and Ukraine," *International Journal of Applied Earth Observation and Geoinformation* 76, no. November 2018 (2019): 112–27, https://doi.org/10.1016/j.jag.2018.11.012.

[10] Chris Funk and Michael E. Budde, "Phenologically-Tuned MODIS NDVI-Based Production Anomaly Estimates for Zimbabwe," *Remote Sensing of Environment* 113, no. 1 (January 15, 2009): 115–25, https://doi.org/10.1016/j.rse.2008.08.015.

[11] L. Karthikeyan, Ila Chawla, and Ashok K. Mishra, "A Review of Remote Sensing Applications in Agriculture for Food Security: Crop Growth and Yield, Irrigation, and Crop Losses," *Journal of Hydrology* 586 (July 1, 2020): 124905, https://doi.org/10.1016/j.jhydrol.2020.124905.

[12] M. S. RASMUSSEN, "Assessment of Millet Yields and Production in Northern Burkina Faso Using Integrated NDVI from the AVHRR.," *International Journal of Remote Sensing* 13, no. 18 (December 1, 1992): 3431–42, https://doi.org/10.1080/01431169208904132.

[13] David M. Johnson, "An Assessment of Pre- and within-Season Remotely Sensed Variables for Forecasting Corn and Soybean Yields in the United States," *Remote Sensing of Environment* 141 (February 5, 2014): 116–28, https://doi.org/10.1016/j.rse.2013.10.027.

[14] Martha C. Anderson et al., "The Evaporative Stress Index as an Indicator of Agricultural Drought in Brazil: An Assessment Based on Crop Yield Impacts," *Remote Sensing of Environment* 174 (March 1, 2016): 82–99, https://doi.org/10.1016/j.rse.2015.11.034.

[15] David Gómez-Candón, Joaquim Bellvert, and Conxita Royo, "Performance of the Two-Source Energy Balance (TSEB) Model as a Tool for Monitoring the Response of Durum Wheat to Drought by High-Throughput Field Phenotyping," *Frontiers in Plant Science* 12 (April 16, 2021): 658357, https://doi.org/10.3389/fpls.2021.658357.

[16] J. L. Monteith, "Solar Radiation and Productivity in Tropical Ecosystems," *Journal of Applied Ecology* 9, no. 3 (1972): 747–66.

term that quantifies the rate of the conversion of absorbed radiation into biomass[17]. As suggested by Gitelson (2006)[18], chlorophyll content can be used a proxy for LUE and fAPAR and be directly utilized, in conjunction with PAR, to estimate GPP. Vegetation indices (VI) may be used as a proxy for chlorophyll and have been directly used to quantify LUE and fAPAR [19]. Various chlorophyll related VI have been tested with varying degrees of success to estimate chlorophyll content [20] and GPP [21]. Vegetation indices using broad band wavelength intervals such as Normalized Difference Vegetation Index (NDVI) or Green Chlorophyll Index have been applied relatively successfully for GPP simulations (Gitelson et al., 2012, 2006; Rossini et al., 2012). Indeed, Dong et al. (2020)[22] applied a NDVI-based LUE model using Landsat imagery and was able to effectively capture the spatial and inter-annual variability of winter wheat yields. LUE can also be retrieved through hypeerspectral remote sensing using the photochemical reflectance index (PRI) [23], which exploits narrow variations around the 0.531 um point, and more recently, through estimations of sun-induced chlorophyll fluorescence signal (SIF). The SIF-GPP relationship has recently gained traction due to the direct link between photosynthesis and SIF [24].

2.2.3 Crop growth models with remote sensing data assimilation

Crop growth models allow to simulate and represent crop development and yield, considering different meteorological, agricultural management and soil conditions, among other variables. Agroecosystem modeling, which simulate the soil-vegetation-atmosphere continuum, incorporate both physiological processes of plant and their interactions with abiotic factors, but also different management scenarios [25].These models are highly useful to assess the multidimensional relationships between different factors affecting crop development and yield, including evaluating the effect of a changing climate and seasonal weather patterns, irrigation management and fertilizer application [26]. Some of the most widely used crop growth models, among others, include the Decision Support System for Agrotechnology Transfer (DSSAT[27], the World Food Studies (WOFOST[28]), Agricultural Production System sIMulator (APSIM[29]), Simulateur mulTIdisciplinaire pour les

[17]Anatoly A. Gitelson et al., “Remote Estimation of Crop Gross Primary Production with Landsat Data,” *Remote Sensing of Environment* 121 (2012): 404–14.
[18]Anatoly A. Gitelson et al., “Relationship between Gross Primary Production and Chlorophyll Content in Crops: Implications for the Synoptic Monitoring of Vegetation Productivity,” *Journal of Geophysical Research: Atmospheres* 111, no. D8 (2006).
[19]Gitelson et al.; Gitelson et al., “Remote Estimation of Crop Gross Primary Production with Landsat Data”; M. Rossini et al., “Remote Sensing-Based Estimation of Gross Primary Production in a Subalpine Grassland,” *Biogeosciences* 9 (2012): 2565–84.
[20]P Zarco-Tejada, “Chlorophyll Fluorescence Effects on Vegetation Apparent Reflectance II. Laboratory and Airborne Canopy-Level Measurements with Hyperspectral Data,” *Remote Sensing of Environment* 74, no. 3 (December 2000): 596–608, https://doi.org/10.1016/S0034-4257(00)00149-8.
[21]Gitelson et al., “Remote Estimation of Crop Gross Primary Production with Landsat Data”; Gitelson et al., “Relationship between Gross Primary Production and Chlorophyll Content in Crops”; Rossini et al., “Remote Sensing-Based Estimation of Gross Primary Production in a Subalpine Grassland.”
[22]Jie Dong et al., “Estimating Winter Wheat Yield Based on a Light Use Efficiency Model and Wheat Variety Data,” *ISPRS Journal of Photogrammetry and Remote Sensing* 160 (February 1, 2020): 18–32, https://doi.org/10.1016/j.isprsjprs.2019.12.005.
[23]J. A. Gamon, J. Penuelas, and C. B. Field, “A Narrow-Waveband Spectral Index That Tracks Diurnal Changes in Photosynthetic Efficiency,” *Remote Sensing of Environment* 41, no. 1 (1992): 35–44.
[24]S. Wieneke et al., “Airborne Based Spectroscopy of Red and Far-Red Sun-Induced Chlorophyll Fluorescence: Implications for Improved Estimates of Gross Primary Productivity,” *Remote Sensing of Environment* 184 (2016): 654–67.
[25]W. A. Dorigo et al., “A Review on Reflective Remote Sensing and Data Assimilation Techniques for Enhanced Agroecosystem Modeling,” *International Journal of Applied Earth Observation and Geoinformation*, Advances in airborne electromagnetics and remote sensing of agro-ecosystems, 9, no. 2 (May 1, 2007): 165–93, https://doi.org/10.1016/j.jag.2006.05.003.
[26]Gustavo Ovando, Silvina Sayago, and Mónica Bocco, “Evaluating Accuracy of DSSAT Model for Soybean Yield Estimation Using Satellite Weather Data,” *ISPRS Journal of Photogrammetry and Remote Sensing* 138 (April 1, 2018): 208–17, https://doi.org/10.1016/j.isprsjprs.2018.02.015.
[27]J. W Jones et al., “The DSSAT Cropping System Model,” *European Journal of Agronomy*, Modelling Cropping Systems: Science, Software and Applications, 18, no. 3 (January 1, 2003): 235–65, https://doi.org/10.1016/S1161-0301(02)00107-7.
[28]C.a. van Diepen et al., “WOFOST: A Simulation Model of Crop Production,” *Soil Use and Management* 5, no. 1 (1989): 16–24, https://doi.org/10.1111/j.1475-2743.1989.tb00755.x.
[29]Dean P. Holzworth et al., “APSIM – Evolution towards a New Generation of Agricultural Systems Simulation,” *Environmental Modelling & Software* 62 (December 1, 2014): 327–50, https://doi.org/10.1016/j.envsoft.2014.07.009.

Cultures Standard (STICS[30]), the Model for Nitrogen and Carbon Dynamics in agro-ecosystems (MONICA[31]), Daisy model[32]and AquaCrop model[33]. These crop simulations models can dynamically ingest inputs and produce the outputs by updating the state variables, while model parameters can be calibrated depending on seasonal growth period and crop species [34]. As such, several efforts have been made to use remote sensing information to replace or adjust model state variables that affect vegetation development, such as LAI, biomass, chlorophyll content, water content, evapotranspiration, within crop growth models to improve the spatio-temporal dimensions of the outputs. Data assimilation techniques have been used to characterize the agro-ecosystem by combining data and information from various sources in different temporal and spatial scales. In general, remote sensing data is used to either **calibrate** model parameters or state variables, **forced** within the model by replacing a state variable or **updating** the model state variables whenever an observation is made [35]. However, since there are often important errors associated with remote sensing observations, data assimilation techniques need to also consider the error variance of these 'observations', through, for example, the use of Kalman filters [36]. For more details, refer to the comprehensive review on data assimilation techniques to combine remote sensing data with crop growth models by Dorigo et al. (2007).

2.2.4 Hybrid Crop Growth Regression models

Due to the complexity and large data requirements needed to run crop growth models, they are not particularly well suited for large spatial scale applications, especially in data scarce regions, being computationally intensive and requiring site-specific information. However, alternative or **hybrid** approaches have been developed to use complex crop models to train simpler statistical models that relate yield with one or several remote sensing indicators used as predictors. In this way, these methods largely bypass the need for computationally intensive methods to calibrate crop growth models. For example, Sibley et al. (2014)[37] developed a simple linear model relating maize LAI and yield through simulations from a crop growth model (i.e. Hybrid-Maize model) and then applied the regression using both MODIS and Landsat imagery. The achieved high accuracy at predicting field scale maize yield using Landsat imagery, while this method also performed better than a more computationally expensive method using a crop growth model calibrated by satellite LAI (section 1.2.3) and a LUE modeling approach (section 1.2.2). Similarly, Lobell et al. (2015)[38] developed the Scalable satellite-based Crop Yield Mapper (SCYM) that associated crop yield with spectral VIs and weather co-variates. This approach consists in four main steps: 1) crop model simulations over a realistic range of soil, climate and management conditions, 2) establishing pseudo-observations by converting daily model outputs to indicators observable by remote sensing (i.e. LAI to VIs relationship), 3) training the statistical model and 4) applying the yield models using image acquisitions from satellites. Jin et al. (2017)[39] further improved the SCYM approach, using an ensemble of three

[30]N Brisson et al., "An Overview of the Crop Model Stics," *European Journal of Agronomy*, Modelling Cropping Systems: Science, Software and Applications, 18, no. 3 (January 1, 2003): 309–32, https://doi.org/10.1016/S1161-0301(02)00110-7.
[31]C. Nendel et al., "The MONICA Model: Testing Predictability for Crop Growth, Soil Moisture and Nitrogen Dynamics," *Ecological Modelling* 222, no. 9 (May 10, 2011): 1614–25, https://doi.org/10.1016/j.ecolmodel.2011.02.018.
[32]Per Abrahamsen and Søren Hansen, "Daisy: An Open Soil-Crop-Atmosphere System Model," *Environmental Modelling & Software* 15, no. 3 (March 1, 2000): 313–30, https://doi.org/10.1016/S1364-8152(00)00003-7.
[33]Dirk Raes et al., "AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: II. Main Algorithms and Software Description," *Agronomy Journal* 101, no. 3 (2009): 438–47, https://doi.org/10.2134/agronj2008.0140s.
[34]Karthikeyan, Chawla, and Mishra, "A Review of Remote Sensing Applications in Agriculture for Food Security."
[35]Dorigo et al., "A Review on Reflective Remote Sensing and Data Assimilation Techniques for Enhanced Agroecosystem Modeling."
[36]Karthikeyan, Chawla, and Mishra, "A Review of Remote Sensing Applications in Agriculture for Food Security"; Shangrong Wu et al., "Estimating Winter Wheat Yield by Assimilation of Remote Sensing Data with a Four-Dimensional Variation Algorithm Considering Anisotropic Background Error and Time Window," *Agricultural and Forest Meteorology* 301–302 (May 2021): 108345, https://doi.org/10.1016/j.agrformet.2021.108345.
[37]Adam M. Sibley et al., "Testing Remote Sensing Approaches for Assessing Yield Variability among Maize Fields," *Agronomy Journal* 106, no. 1 (2014): 24–32, https://doi.org/10.2134/agronj2013.0314.
[38]David B. Lobell et al., "A Scalable Satellite-Based Crop Yield Mapper," *Remote Sensing of Environment* 164 (July 2015): 324–33, https://doi.org/10.1016/j.rse.2015.04.021.
[39]Zhenong Jin, George Azzari, and David B. Lobell, "Improving the Accuracy of Satellite-Based High-Resolution Yield Estimation: A Test of Multiple Scalable Approaches," *Agricultural and Forest Meteorology* 247 (December 2017): 207–20, https://doi.org/10.1016/j.agrformet.2017.08.001.

crop models, including calibrating the phenology parameters in one the models, and using simulated biomass instead of yield to train the empirical model. They reported that simpler methods that relate crop biomass with yield, such as the constant harvest index, often outperform mechanistic simulations of grain formation. The integration of different remote sensing domains, such as thermal infrared or radar, to be used as predictive variables in these hybrid models remains largely unexplored, including the use of non-linear or machine learning statistical models.

2.3 Irrigation delimitation

Irrigated agricultural areas can be distinguished from adjacent agricultural parcels or natural areas either through:

1. Excess ET or crop vigour
2. Sudden increase in soil moisture and actual evapotranspiration which cannot be explained by other factors (e.g. change in weather or vegetation cover).

The first approach is more classical often and takes form of supervised or unsupervised land-cover classification[40]. It is particularly suited to regions with distinctive dry season during which there is sharp contrast between irrigated agriculture and rainfed agricultural or natural areas. The contrast can take form of increased vigour or greenness, often captured using NDVI, increased ET in the irrigated parcels compared to other areas, or increase in soil moisture with microwaves[41], or a combination of ET and microwave soil moisture[42]. It can also be observed in the structure of vegetation and detected with SAR data, such as the one acquired by Sentinel-1 satellites[43]. Regardless of the parameter used, or the combination of them, the first step of classification consists of creating temporal composites and indices, e.g. 25th percentile, median and 75th percentile of a parameter over the temporal compositing period. That period is usually monthly or seasonal. This data is then used in a classification model, such as random forest, to separate irrigated and non-irrigated agriculture. Depending on the regional conditions (e.g. temperate versus semi-arid climate) the classification can obtain an overall accuracy of between 50%-90%.

The second approach is based on identifying irrigation events through monitoring of changes in top-soil soil moisture[44] . This method can also be employed with evapotranspiration instead of soil moisture. The ratio of actual to potential ET ((ETa/p) should be used in order to avoid changes in ET due to changes in weather (e.g. increased wind speed) or crop cover (e.g. quick development of leaves) being attributed to irrigation. This ratio is closely related to root-zone water availability and therefore is mainly influenced by irrigation or rainfall events.

The method separates changes in the ratio of actual to potential ET due to either rain or irrigation. At the same time, it attempts to capture even small irrigation events (e.g. drip irrigation) while remaining robust to uncertainty in the estimation of ET. This is achieved by first calculating the change in ETa/p between the time

[40] Magidi, J.; Nhamo, L.; Mpandeli, S.; Mabhaudhi, T. Application of the Random Forest Classifier to Map Irrigated Areas Using Google Earth Engine. *Remote Sens.* **2021**, *13*, 876. https://doi.org/10.3390/rs13050876

[41] Dari, J.; Quintana-Seguí, P.; Escorihuela, M.J.; Stefan, V.; Brocca, L.: Morbidelli, R. Detecting and mapping irrigated areas in a Mediterranean environment by using remote sensing soil moisture and a land surface model. *Journal of Hydrology,* **2021,** 596, 126129. https://doi.org/10.1016/j.jhydrol.2021.126129

[42] Paolini, G.; Escorihuela, M.J.; Merlin, O.; Sans, M.P.; Bellvert, J. Classification of Different Irrigation Systems at Field Scale Using Time-Series of Remote Sensing Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **2022**, 15, pp.10055-10072. https://doi.org/10.1109/JSTARS.2022.3222884

[43] Pageot, Y.; Baup, F.; Inglada, J.; Baghdadi, N.; Demarez, V. Detection of Irrigated and Rainfed Crops in Temperate Areas Using Sentinel-1 and Sentinel-2 Time Series. *Remote Sens.* **2020**, *12*, 3044. https://doi.org/10.3390/rs12183044

[44] Bazzi, H.; Baghdadi, N.; Fayad, I.; Zribi, M.; Belhouchette, H.; Demarez, V. Near Real-Time Irrigation Detection at Plot Scale Using Sentinel-1 Data. *Remote Sens.* **2020**, *12*, 1456. https://doi.org/10.3390/rs12091456

on which we irrigation is to be detect and most recent previous time on which ET estimates are available. This change is calculated both locally (i.e. at individual pixel level) and regionally (i.e. as an average change in all agricultural pixels within 10 km window). The local and regional changes are then compared to a number of thresholds to try to detect if:

a) There is no input of water into the soil (e.g. local ETa/p does not increase above a threshold)

b) There is input of water into the soil but due to rainfall (e.g. increase in regional ETa/p is over a threshold and larger or similar to increase in local Eta/p)

c) There is input of water to the soil due to irrigation (e.g. increase in local ETa/p is over a threshold and significantly larger than increase in regional ETa/p)

Detected irrigation events are further split into low, medium and high probability based on another set of thresholds. Since irrigation is normally applied on a larger area, the raster map with per-pixel irrigation events is cleaned up by removing isolated pixels in which irrigation was detected.

The second approach is reported to achieve overall accuracy of 80%-90%. Compared to the classification method, it has the benefit of being able to detect individual irrigation events. It is also able to perform even in periods and regions in which rainfall events are recorded at the same time as irrigation events. In addition, the second method can be considered a hybrid physical-empirical approach and while it might benefit from localised fine-tuning it does not require extensive local calibration or training. Finally, as its main inputs it uses data which is expected to be further improved in this project (actual and potential ET) and those improvements can then hopefully lead to improved delineation of irrigated areas.

2.4 Irrigation Accounting

Irrigation accounting is typically estimated by applying soil water balance (SWB) models and comparing them to EO observations. As a general form, the water balance can be described by Eq. 1:

$$\frac{d\,SM(t)}{d\,t} = [I(t)+P(t)] - D(t) - R(t) - ET(t) \tag{1}$$

where $\frac{d\,SM(t)}{d\,t}$ is the variation in time of soil moisture content, $I(t)+P(t)$ is the water entering in the soil in the form of irrigation (I) and/or precipitation (P), $D(t)$ is the drainage, $R(t)$ is the surface run-off, and $ET(t)$ is the actual evapotranspiration.

From Eq. 1, two main group algorithms can be classified, those that are mainly based on microwave soil moisture retrievals and those based on ET.

Microwave-based algorithms

Microwave algorithms basically compares the temporal differences between MW and SWB soil moisture. One of the most widely used approaches is based on the SM2RAIN[45] algorithm, which estimates the total water entering into the soil, and thus over irrigated regions it provides an estimate of rainfall plus irrigation. By

[45] Brocca, L.; Tarpanelli, A.; Filippucci, P.; Dorigo, W.; Zaussinger, F.; Gruber, A.; Fernández-Prieto, D.. How much water is used for irrigation? A new approach exploiting coarse resolution satellite soil moisture products. *International journal of applied earth observation and geoinformation*, **2018**, *73*, pp.752-766. https://doi.org/10.1016/j.jag.2018.08.023

knowing the rainfall from other sources (e.g. from meteorological observations or numerical weather models) the amount of irrigation water can be derived by subtraction[46]. To solve Eq. 1, SM2RAIN algorithm assumes that run-off is negligible, drainage is estimated as a semi-empirical function of soil moisture that requires calibration, and actual ET is usually linearly related to potential ET as function of S(t). SWB models need to be calibrated before, in order to obtain model parameter values related to soil water capacity or drainage. To that reason, these models are usually calibrated during rainfed conditions[46].

Several microwave products have been applied using this approach such as AMSR2-JAXA[45], SMAP, SMOS, ASCAT[47], or AMSR2-LPRM[46].

The microwave approach is able to provide an estimation of monthly irrigation with better performances in semi-humid climate and for larger irrigation rates[46]. However some major issues are found regarding his approach, which are related to the coarse spatial resolution of passive MW missions, microwave SM noise, and the confounding effect of vegetation that should be disentangled from the soil moisture contribution during the SM retrieval[45]. Furthermore, in order to convert volumetric SM into a corresponding water column depth of irrigation, various assumptions such as depth of soil, water capacity of the soil layer, and other empirical parameters are necessary which introduce additional uncertainties[48]

ET-based algorithms

The advantages of using ET over soil moisture is that ET is directly linked to plant transpiration reacting to irrigation whereas soil moisture produces an indirect estimate, especially since satellite MW systems only penetrate the topsoil (few cm)[49]. In addition, the spatial resolution of EO optical sensors is typically higher, with some orders of magnitude, than passive microwave.

Net irrigation is estimated based on the systematic evapotranspiration (ET) residuals between a remote sensing-based model and a calibrated hydrologic or SWB model that does not include an irrigation scheme[48,50]

$$I_{net}(t) = ET_{EO} - ET_{baseline} \qquad (2)$$

Some studies used simply a dense NDVI time series in order to monitor crop biophysical condition, and computing ET based on a crop coefficient[51] (Eq. 3).

$$I_{net}(t) = \frac{dS(t)}{dt} - P(t) + ET_{c,adj}(t) + D(t) \qquad (3)$$

[46] Jalilvand, E.; Tajrishy, M.; Hashemi, S.A.G.Z.; Brocca, L.. Quantification of irrigation water using remote sensing of soil moisture in a semi-arid region. *Remote Sensing of Environment*, **2019**, *231*, p.111226. https://doi.org/10.1016/j.rse.2019.111226

[47] Zaussinger, F.; Dorigo, W.; Gruber, A.; Tarpanelli, A.; Filippucci, P.; Brocca, L. Estimating irrigation water use over the contiguous United States by combining satellite and reanalysis soil moisture data. *Hydrology and earth system sciences*, **2019**, 23(2), pp.897-923. https://doi.org/10.5194/hess-23-897-2019

[48] Koch, J.; Zhang, W.; Martinsen, G.; He, X.; Stisen, S. Estimating net irrigation across the North China Plain through dual modeling of evapotranspiration. *Water Resources Research*, **2020**, 56(12), p.e2020WR027413. https://doi.org/10.1029/2020WR027413

[49] Kragh, S.J.; Fensholt, R.; Stisen, S.; Koch, J. The precision of satellite-based irrigation quantification in the Indus and Ganges basins. *Hydrology and Earth System Sciences Discussions*, **2022**, preprint, pp.1-29. https://doi.org/10.5194/hess-2022-307

[50] Garrido-Rubio, J.; Gonzalez-Piqueras, J.; Campos, I.; Osann, A.; Gonzalez-Gomez, L.; Calera, A. Remote sensing–based soil water balance for irrigation water accounting at plot and water user association management scale. *Agricultural Water Management*, **2020**, 238, p.106236. https://doi.org/10.1016/j.agwat.2020.106236

[51] Garrido-Rubio, J.; Calera, A.; Arellano, I.; Belmonte, M.; Fraile, L.; Ortega, T.; Bravo, R.; González-Piqueras, J. Evaluation of remote sensing-based irrigation water accounting at river basin district management scale. *Remote Sensing*, **2020**, 12(19), p.3187. https://doi.org/10.3390/rs12193187

Under this approach crop optimum conditions means that soil water content is not allowed to be lower than a predetermined threshold, avoiding water stress, or maintaining this water stress on a controlled way. However, larger overestimation errors in crops subject to controlled deficit irrigation, such as barley or grapevines, when assuming crop coefficient approach, since farmer-applied irrigation in those crops is lower than the calculated requirements during most of the campaign[50].

For that reason the usage of thermal infra-red to retrieve actual ET is a more robust approach. Several ET models or products have been used such as PT-JPL[48], FLUXCOM, NTSG[49] or WAPOR[46]. Indeed, implementing EO ET estimates into the SM2RAIN algorithm, Jalivand et al.[46] observed better results when using the thermal-based WaPOR ET product than using MW-based GLEAM product, likely due to the ability of thermal-based ET to better characterize root-zone SM conditions.

Finally it is worth noting that for validating, net irrigation is then converted into gross irrigation based on tabulated irrigation efficiency coefficient that depends on the transport and distribution system (e.g. open channel vs. pressurized pipe) as well as the application system (e.g. flood, sprinkler or drip irrigation)[51].