Funded by
the European Union

D3.5: Upscaling sensing tools II

cracksense.eu

Author(s)/Organisation(s)	Tarin Paz-Kagan (BGU), Victor Alchanati, Avi Sadka (ARO)
Contributor(s)	Noy Saraf, Idit Ginzberg, Uri Hochberg, Amnon Lichter, Georgia Nikolakopoulou, George Ntakos (AUA), Laurent Julhia, José Quero Garcia (INRAE), Andreas Winkler (LVGA).
Work Package	WP3
Delivery Date (DoA)	36M
Actual Delivery Date	30.1.26

Document Revision			
Date	**Version**	**Author/Contributor/ Reviewer**	**Summary of main changes**

Dissemination Level		
PU	Public, fully open, e.g., web	X
SEN	Sensitive, limited under the conditions of the Grant Agreement	
Classified R-UE/EU-R	EU RESTRICTED under the Commission Decision No. 2015/444	
Classified C-UE/EU-C	EU CONFIDENTIAL under the Commission Decision No. 2015/444	
Classified S-UE/EU-S	EU SECRET under the Commission Decision No. 2015/444	

Funded by the European Union

CrackSense			
Participant Number	Participant organization name	Short name	Country
1	THE AGRICULTURAL RESEARCH ORGANISATION OF ISRAEL - THE VOLCANI CENTRE	ARO	IL
2	LEIBNIZ-INSTITUT FUR AGRARTECHNIK UND BIOOKONOMIE EV	ATB	DE
3	GEOPONIKO PANEPISTIMION ATHINON	AUA	EL
4	INSTITUT NATIONAL DE RECHERCHE POUR L'AGRICULTURE, L'ALIMENTATION ET L'ENVIRONNEMENT	INRAE	FR
7	UAB BETA VIA	ART21	LT
8	FONDACIJA VIZLORE LABS	VLF	RS
10	LACHISH WORKER'S MOSHAV FOR COOPERATIVE AGRICULTURAL SETTLEMENT LTD	Tali grapes	IL
11	KIBBUTZ TSOR A	KT	IL
12	AGROTIKOS SYNETAIRISMOS POLISEOS XIRON KAI NOPON STAFYLION KIATOY KORINTHIAS PIGASOS	PEGASUS	EL
13	CENTRE TECHNIQUE INTERPROFESSIONNEL DES FRUITS ET LEGUMES	CTIFL	FR
14	EVENFLOW	EVENFLOW	BE
15	LEHR-UND VERSUCHSANSTALT FURGARTENBAU UND ARBORISTIK EV (LVGA)	LVGA	DE
16	BEN-GURION UNIVERSITY OF THE NEGEV	BGU	IL
17	MEHADRIN TNUPORT EXPORT LIMITED PARTNERSHIP	MTEX	IL
18	UDRUZENJE ZA PREDUZETNISTVO I INOVACIJE FOODSCALE HUB	FSH-RS	RS

Title:

High throughput real-time monitoring and prediction of fruit cracking by utilizing and upscaling sensing and digital data technologies

Type of action: HORIZON Research and Innovation Actions

Topic: HORIZON-CL6-2022-GOVERNANCE-01-11

Start date of project: January 1 2023 ● Duration: 48 months

Table of Contents

List of abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
BBCH	Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie (phenological scale)
CWSI	Crop Water Stress Index
DEM	Digital Elevation Model
ET_0	Evapotranspiration
EU	European Union
FDS	Fruit Daily Shrinkage
GNDI	Greenness Normalized Difference Index
GNDVI	Green Normalized Difference Vegetation Index
IPVI	Infrared Percentage Vegetation Index
LAI	Leaf Area Index
LIDAR	Light Detection and Ranging
MDS	Maximum Daily Shrinkage
ML	Machine Learning
MSAVI	Modified Soil-Adjusted Vegetation Index
NDRE	Normalized Difference Red Edge Index
NDVI	Normalized Difference Vegetation Index
NPK	Nitrogen Phosphorus Potassium
OSAVI	Optimized Soil-Adjusted Vegetation Index
PA	Plant Area
PAI	Plant Area Index
SAVI	Soil-Adjusted Vegetation Index
SDSS	Spatial Decision Support System
SIPI	Structure Insensitive Pigment Index
SWP	Stem Water Potential
TOMMY	Terrestrial Orchard Monitoring Method Yaw
UAV	Unmanned Aerial Vehicle
VI	Vegetation Indix

Executive summary

This deliverable (D3.5) demonstrates the successful upscaling of integrated sensing and modelling approaches for assessing PWS, yield, and fruit cracking across multiple perennial fruit crops, climatic regions, and management regimes. By combining UAV-based multispectral, thermal, and LiDAR data with ground-based physiological measurements and environmental observations, the project established a robust, harmonized framework for monitoring tree health and quantifying cracking-associated yield loss at the tree and orchard scales.

Across sites in Israel, France, Greece, and Germany, Random Forest (RF) models consistently outperformed alternative machine-learning approaches in predicting key PWS indicators (trunk growth, stem water potential, and stomatal conductance), as well as final yield and fruit cracking rates. Model performance was high across crops and years, despite substantial interannual climatic variability, including contrasting rainfall regimes and pronounced heatwave events. Feature importance analyses highlighted the dominant role of mid- to late-season PWS, particularly SWP and SC, alongside canopy structural attributes (PAI) and short-term atmospheric drivers in controlling both yield formation and cracking risk.

Importantly, clustering analyses based on seasonal physiological trajectories enabled the identification of functional stress-response groups within orchards. These clusters showed clear, statistically significant differences in the proportion of healthy fruit and cracking severity, demonstrating the value of physiological clustering as a decision-support tool for defining management zones. Spatial mapping of predicted yield and cracking further revealed strong within-orchard heterogeneity linked to irrigation regimes, climate stress exposure, and canopy structure.

Overall, D3.5 provides compelling evidence that multi-sensor UAV data, when coupled with physiological insight and explainable machine-learning models, can deliver scalable, interpretable, and operational tools for predicting fruit cracking and associated yield losses. The harmonized methodology and shared processing pipelines developed within the consortium ensure reproducibility and cross-country comparability, laying the groundwork for broader deployment within the CrackSense framework. The novelty of this work lies in its **explicit integration of plant water status physiology with orchard-scale remote sensing, machine learning, and fruit cracking outcomes**, rather than treating cracking as an isolated postharvest or phenological phenomenon. Unlike previous studies that rely on single indicators, short-term experiments, or plot-level averages, this project introduces a **multi-sensor, multi-temporal, and tree-resolved framework** that links dynamic water-stress trajectories (SWP, trunk growth, stomatal conductance) to both yield formation and cracking severity across seasons and management regimes. A key innovation is the **use of physiological clustering** to translate continuous sensor data into functionally meaningful stress-response groups, enabling spatially explicit identification of cracking-risk zones within orchards. Furthermore, the harmonized cross-country implementation and shared modelling pipelines represent one of the **first coordinated demonstrations of scalable cracking prediction across crops, climates, and irrigation strategies**. This combination of mechanistic insight, explainable machine learning, and operational scalability goes beyond existing remote sensing applications. It establishes a new paradigm for precision management of cracking-prone fruit crops under increasing climate variability.

Task 3.2: Remote sensing applications to assess tree health and estimate cracking-associated yield loss

To improve understanding of how climate variability and orchard management interact to control plant water status (PWS) and cracking-related yield losses, several remote sensing and integrated sensor approaches were tested to assess the impact of environmental conditions and horticultural management practices on plant water status and cracking-associated yield loss at the regional scale, across four cultivars and two growing seasons. This task investigated the environmental, biological, and management factors controlling tree-level yield variability and fruit cracking in orchards. Experimental plots at multiple locations were monitored using sensors mounted on UAVs and autonomous platforms, with data acquisition conducted in parallel with field measurements and continuous observations of plant and soil sensors. UAV-based sensing included multispectral, thermal, and 3D LiDAR systems. Although hyperspectral sensors were initially proposed, they were not implemented due to operational constraints, data-processing challenges, and difficulties in ensuring methodological consistency across countries and study sites. Indicators of tree health and PWS were derived from remote sensing data and validated against field measurements. PWS, as reflected in stem water potential (SWP), trunk growth (TG), and stomatal conductance (SC), was monitored indirectly using thermal and multispectral imagery to capture spatial variability in canopy water status. Multispectral data were used to compute vegetation indices (VIs) related to nutritional status, chlorophyll content, and water stress, while thermal imagery provided canopy temperature metrics primarily associated with plant water status. LiDAR-derived structural parameters, including tree growth metrics and plant area index (PAI), were used to examine their relationships with yield and yield loss caused by fruit cracking.

To quantify how targeted orchard management interventions can mitigate fruit cracking and associated yield losses under contrasting environmental conditions, multiannual, field-scale monitoring of cracking intensity across plots in diverse agro-climatic regions was conducted, enabling the development of environmental- and management-driven models for estimating cracking and its impact on yield loss. Management-related variables included irrigation regimes, mineral fertilization, and the application of plant growth regulators (PGRs). The sensing approaches and methodologies developed in Task 3 were applied to parameterize and scale remote sensing–based assessments at the regional level. In addition, annual yield data and detailed horticultural practices (including PGR treatments, irrigation levels, pruning timing, and pruning type) were collected from selected trees and growers. Advanced machine-learning and computer-vision models were applied to estimate total yield and the proportion of cracked fruit. These estimates were evaluated against final plot-level yield measurements, cracking intensity recorded on reference trees, and grower reports, enabling quantitative assessment of yield losses attributable to fruit cracking. UAV campaigns were conducted approximately monthly throughout the growing season and aligned with key physiological stages of crop development, ensuring comprehensive temporal coverage.

The methodology applied in this study was uniform across all research groups and participating countries, enabling standardized analysis and comparative evaluation of crop responses across different environments and management regimes. All datasets were harmonized and analyzed using a common framework to facilitate cross-crop and cross-country comparisons. At the same time, analyses within each country were carried out according to the specific data available, reflecting differences in crop type, sensor deployment, experimental design, and previously implemented monitoring protocols or methodological modifications. The modelling framework was developed by the Ben-Gurion University (BGU) research group, and a dedicated GitHub repository was established to distribute standardized data-processing pipelines and modelling tools to all partners (https://github.com/DubininMD2024/CrackSense). These tools were implemented at each site with site-specific adjustments, which were jointly agreed upon by the full research consortium, ensuring methodological consistency while accommodating local conditions and data constraints. Tables **1-4** summarize the datasets collected across all project sites and crops, including UAV flight dates, sensor types, and corresponding crop phenological stages, providing a transparent overview of data harmonization and supporting robust inter-site and inter-crop comparison.

Table 1. Data collection Israel- Citrus, Pomegranates, and Grapes

Date	Country	Crop	Plot location	Type of photography	Fligh	Physiological stage

				required (RGB, MS, TIR, HS, LiDAR)	t No.	
7.6.23	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, MS	1	Stage I
28.6.23	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, MS	2	Stage I-II
19.7.23	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, MS	3	Stage I-II
28.8.23	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, MS	4	Stage II
7.9.23	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, MS	5	Stage II
20.9.23	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, MS	6	Stage II
1-5.10.23	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, MS	7	Stage III- green
5-9.11.23	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, MS	8	Stage III- color break
9.5.23	Israel	Pomegranate	Tsor'a	TiR, Lidar, RGB, MS	1	Full bloom
21.6.23 (S1)	Israel	Pomegranate	Tsor'a	TiR, Lidar, RGB, MS	2	Peel color changes to red, green
23.7.23 (S2)	Israel	Pomegranate	Tsor'a	TiR, Lidar, RGB, MS	3	Green fruit
15.8.23 (S3)	Israel	Pomegranate	Tsor'a	TiR, Lidar, RGB, MS	4	Color break
11.9.23 (S4)	Israel	Pomegranate	Tsor'a	TiR, Lidar, RGB, MS	5	Reddish color
09.10.23 (S5)	Israel	Pomegranate	Tsor'a	TiR, Lidar, RGB, MS	6	Red fruit
31.10.23 (S6)	Israel	Pomegranate	Tsor'a	TiR, Lidar, RGB, MS	7	Harvest-dark red
11.6.23	Israel	Grapes	Lachish	TiR, Lidar, RGB, MS	1	Post fruit set
5.7.23	Israel	Grapes	Lachish	TiR, Lidar, RGB, MS	2	Pre-veraison
6.8.23	Israel	Grapes	Lachish	TiR, Lidar, RGB, MS	3	Post-veraison
7.9.23	Israel	Grapes	Lachish	TiR, Lidar, RGB, MS	4	Mature (pre-harvest)
1.5-5.5.24	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, MS	1	Stage I
30.5-3.6.24	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, MS	2	Stage I-II
1.7-3.7.24	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, MS	3	Stage I-II
1.8-5.8.24	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, MS	4	Stage II
1.9-5.9.24	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, MS	5	Stage II
1.10-5.10.24	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, MS	6	Stage II
1.11-5.11.24	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, MS	7	Stage III- green
1.12-5.12.24	Israel	Citrus	Kfar Habad	TiR, Lidar, RGB, HS	8	Stage III- color break
-	Israel	Pomegranate	Tsor'a	TiR, Lidar, RGB, MS	1	Full bloom
18.6.24 (S1)	Israel	Pomegranate	Tsor'a	TiR, Lidar, RGB, MS	2	Peel color change red to green
22.7.24 (S2)	Israel	Pomegranate	Tsor'a	TiR, Lidar, RGB, MS	3	Green fruit
19.8.24 (S3)	Israel	Pomegranate	Tsor'a	TiR, Lidar, RGB, MS	4	Color break
16.9.24 (S4)	Israel	Pomegranate	Tsor'a	TiR, Lidar, RGB, MS	5	Reddish color
14.10.24 (S5)	Israel	Pomegranate	Tsor'a	TiR, Lidar, RGB, MS	6	Red fruit
4.11.24 (S6)	Israel	Pomegranate	Tsor'a	TiR, Lidar, RGB, HS	7	Harvest-dark red
25.6.24	Israel	Grapes	Lachish	TiR, Lidar, RGB, MS	1	Post fruit set
25.7.24	Israel	Grapes	Lachish	TiR, Lidar, RGB, MS	2	Pre-veraison
25.8.24	Israel	Grapes	Lachish	TiR, Lidar, RGB, MS	3	Post-veraison
25.9.24	Israel	Grapes	Lachish	TiR, Lidar, RGB, MS	4	Mature (pre-harvest)

Table 2. Data collection France – Citrus, Sweet cherry

Date	Country	Crop	Plot location	Type of photography required (RGB, MS, TIR, HS, LiDAR)	Flight No.	Physiological stage
06.7.23	France	Citrus	FRCitrus2	RGB, MS	1	Stage I
27.7.23	France	Citrus	FRCitrus2	RGB, MS	2	Stage I-II
11-13.09.23	France	Citrus	FRCitrus2	RGB, MS	3	Stage II
12.10.23	France	Citrus	FRCitrus2	Lidar, RGB, MS	4	Stage III- green
08.11.23	France	Citrus	FRCitrus2	Lidar, RGB, MS	5	Stage III- color break
06.7.23	France	Citrus	FRCitrus1	RGB, MS	1	Stage I
27.7.23	France	Citrus	FRCitrus1	RGB, MS	2	Stage I-II
11-12.09.23	France	Citrus	FRCitrus1	RGB, MS	3	Stage II
12.10.23	France	Citrus	FRCitrus1	Lidar, RGB, MS	4	Stage III- green
08.11.23	France	Citrus	FRCitrus1	Lidar, RGB, MS	5	Stage III- color break

Date	Country	Crop	Plot location	Type of photography required (RGB, MS, TIR, HS, LiDAR)	Flight No.	Physiological stage
17.07.24	France	Citrus	FRCitrus1	Lidar, RGB, MS, TiR	1	Stage I
08.08.24	France	Citrus	FRCitrus1	Lidar, RGB, MS, TiR	2	Stage I-II
10.09.24	France	Citrus	FRCitrus1	Lidar, RGB, TiR	3	Stage II
09-11.10.24	France	Citrus	FRCitrus1	Lidar, RGB, MS, TiR	4	Stage III- green
15.11.24	France	Citrus	FRCitrus1	Lidar, RGB, MS, TiR	5	Stage III- color break
18.07.24	France	Citrus	FRCitrus2	Lidar, RGB, MS, TiR	1	Stage I
14.08.24	France	Citrus	FRCitrus2	Lidar, RGB, MS, TiR	2	Stage I-II
10.09.24	France	Citrus	FRCitrus2	Lidar, RGB, TiR	3	Stage II
09-11.10.24	France	Citrus	FRCitrus2	Lidar, RGB, MS, TiR	4	Stage III- green
15.11.24	France	Citrus	FRCitrus2	Lidar, RGB, MS, TiR	5	Stage III- color break
23-24-29.04.24	France	Sweet cherry	FRsweetcherry	Lidar, RGB, MS, TiR	1	Stage I
16.05.24	France	Sweet cherry	FRsweetcherry	Lidar, RGB, MS, TiR	2	Stage I-II
05.06.24	France	Sweet cherry	FRsweetcherry	Lidar, RGB, MS, TiR	3	Stage II
24.06.24	France	Sweet cherry	FRsweetcherry	Lidar, RGB, MS, TiR	4	Stage III

Table 3. Data collection Greece - Grapes, Pomegranate

Date	Country	Crop	Plot location	Type of photography required (RGB, MS, TIR, HS, LiDAR)	Flight No.	Physiological stage
22.05.2023	Greece	Pomegranate	GRPOM	TiR, RGB, MS	2	Full bloom
15.06.2023	Greece	Pomegranate	GRPOM	TiR, RGB, MS	3	Peel color change red-green
03.07.2023	Greece	Pomegranate	GRPOM	TiR, RGB, MS	4	Green fruit
05.08.2023	Greece	Pomegranate	GRPOM	TiR, RGB, MS	5	Color break
01.09.2023	Greece	Pomegranate	GRPOM	TiR, RGB, MS	6	Reddish color
27.09.2023	Greece	Pomegranate	GRPOM	TiR, RGB, MS	7	Red fruit
20.10.2023	Greece	Pomegranate	GRPOM	TiR, RGB, MS	8	Harvest-dark red
15.06.2023	Greece	Grapes	GRGrapes	TiR, RGB, MS	1	Post fruit set
18.07.2023	Greece	Grapes	GRGrapes	TiR, RGB, MS	2	Pre-veraison
17.08.2023	Greece	Grapes	GRGrapes	TiR, RGB, MS	3	Post-veraison
10.09.2023	Greece	Grapes	GRGrapes	TiR, RGB, MS	4	Mature (pre-harvest)
25.05.2024	Greece	Pomegranate	GRPOM	TiR, RGB, MS	1	Full bloom
14.06.2024	Greece	Pomegranate	GRPOM	TiR, RGB, MS	2	Peel color change red-green
04.07.2024	Greece	Pomegranate	GRPOM	TiR, RGB, MS	3	Green fruit
01.08.2024	Greece	Pomegranate	GRPOM	TiR, RGB, MS	4	Color break
05.09.2024	Greece	Pomegranate	GRPOM	TiR, RGB, MS	5	Reddish color
17.10.2024	Greece	Pomegranate	GRPOM	TiR, RGB, MS	6	Red fruit
25.10.2025	Greece	Pomegranate	GRPOM	TiR, RGB, MS	7	Harvest-dark red
07.06.2024	Greece	Grapes	GRGrapes	TiR, RGB, MS	1	Post fruit set
01.07.2024	Greece	Grapes	GRGrapes	TiR, RGB, MS	2	Pre-veraison
18.07.2024	Greece	Grapes	GRGrapes	TiR, RGB, MS	3	Post-veraison
02.08.2024	Greece	Grapes	GRGrapes	TiR, RGB, MS	5	Post-veraison

Table 4. Data collection Germany - Sweet cherry

Date	Country	Crop	Plot location	Type of photography required (RGB, MS, TIR, HS, LiDAR)	Flight No.	Physiological stage
14.06.2023	Germany	Sweet cherry	LVGA_Cerry	RGB, MS	1	stage II - transition stage II/III
21.06.2023	Germany	Sweet cherry	LVGA_Cerry	RGB, MS	2	transition stage II/III - early stage III
28.06.2023	Germany	Sweet cherry	LVGA_Cerry	RGB, MS	3	early stage III - mid-stage III
05.07.2023	Germany	Sweet cherry	LVGA_Cerry	RGB, MS	4	early stage III - mid-stage III
12.07.2023	Germany	Sweet cherry	LVGA_Cerry	RGB, MS	5	mid-stage III - maturity
19.07.2023	Germany	Sweet cherry	LVGA_Cerry	RGB, MS	6	mid-stage III - maturity
26.07.2023	Germany	Sweet cherry	LVGA_Cerry	RGB, MS	7	maturity

Methodology:

Research approach

The research approach combined accurate local measurements of key plant water status (PWS) indicators, including trunk growth (TG), stem water potential (SWP), stomatal conductance (SC), and plant area index (PAI), from selected sample trees with high-resolution UAV imagery, environmental data, and irrigation records. Depending on crop type, site characteristics, and experimental design, the number of selected trees varied across studies, ensuring representative sampling while maintaining feasibility for intensive physiological monitoring. The overall workflow was divided into five main stages: **Stage 1: Field measurements**. Local PWS indicators (TG and maximum daily shrinkage, MDS) were continuously monitored on reference trees to provide ground-truth data for model calibration and validation. In addition, SWP, SC, and PAI were measured monthly on sampled trees, with the number of monitored trees varying by crop and site in accordance with experimental constraints and orchard structure. **Stage 2: Environmental monitoring**. Meteorological data, including air temperature, rainfall, vapor pressure deficit (VPD), soil moisture, growing degree days (GDD), and reference evapotranspiration (ET), were recorded using on-site weather stations, while irrigation inputs were documented for each treatment. These variables were used to quantify atmospheric demand, crop phenological development, and water supply conditions, providing essential environmental context for interpreting spatial and temporal variability in PWS. **Stage 3: UAV data acquisition and processing.** UAV flights equipped with multispectral, thermal, and LiDAR sensors were conducted over the orchards. LiDAR point cloud data were processed to classify ground returns, normalize tree height, and segment individual tree polygons. For each polygon, spectral reflectance values (Table 5) and vegetation indices (VIs; Table 5) were derived from multispectral imagery, while canopy temperature (Tc) and the crop water stress index (CWSI) were extracted from thermal data. PAI was derived directly from LiDAR data using leaf area density (LAD) profiles and projected canopy area and was evaluated against field measurements. **Stage 4: Machine-learning modelling.** Random Forest (RF) and complementary machine-learning models were trained using UAV-derived spectral, thermal, and LiDAR structural variables, along with climate and irrigation data, to estimate TG, SWP, SC, and yield and cracking at the orchard scale. Given the limited sample size available from the two monitoring seasons, machine-learning models were selected as the primary analytical approach to ensure robust model training and generalization; deep-learning models will be evaluated in a subsequent phase following integration of the full multi-site dataset. **Stage 5: Yield and fruit-cracking analysis.** PWS indicators derived for different phenological periods were analyzed in relation to yield and fruit-cracking rates using **machine-learning** models. By integrating climate dynamics, irrigation regimes, and PWS responses, this stage enabled the identification of critical stress windows and the evaluation of treatment effects at both temporal and spatial scales. A comprehensive summary of the variables used in the study, their data sources, and the corresponding calculation methods is provided in Table 5.

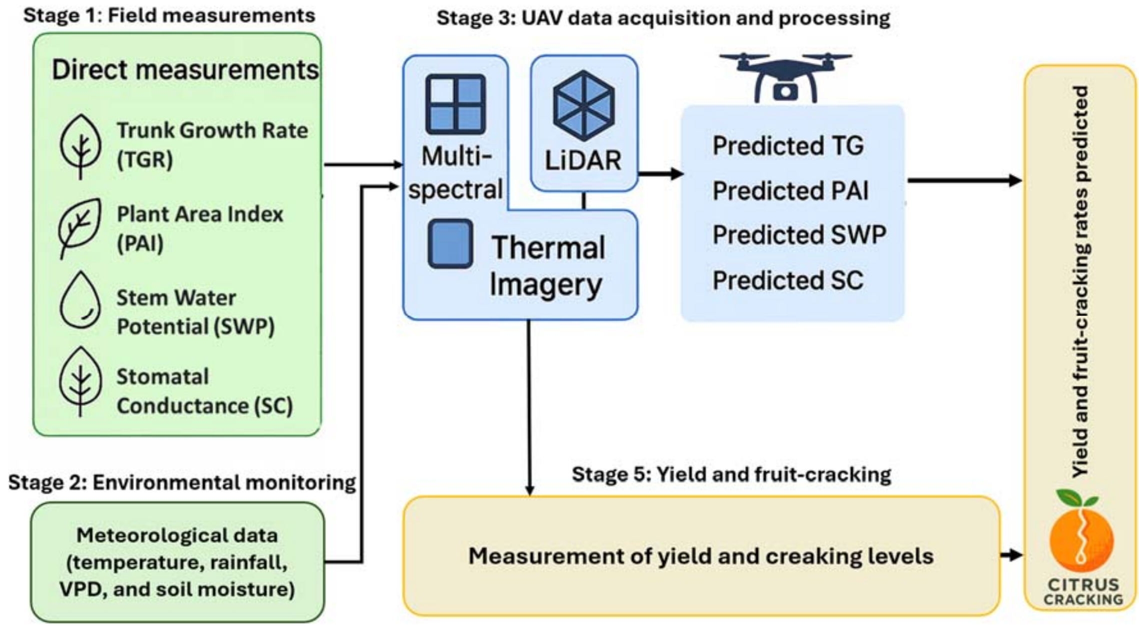


Figure 1. Flowchart of the research framework. Green frames represent field-collected physiological data; blue frames represent remote sensing data and pre-processing steps; yellow frames represent yield and fruit cracking data in the different crops and sites.

Table 5. Summary of variables used in the study, data sources, and calculation methods.

Category	Variable	Abbreviation	Unit	Data source	Calculation / Definition
Plant Water Status (PWS)	Trunk Growth Rate	TG	µm d^{-1}	Dendrometers	Daily change in trunk diameter; TG = D(*end) − D(*start)
	Trunk Max. Daily Shrinkage	MDS	µm	Dendrometers	MDS = D(*max) − D(*min) (daily cycle)
	Stem Water Potential	SWP	MPa	Pressure chamber	Midday stem water potential measured on bagged leaves
	Stomatal Conductance	SC	mol m^{-2} s^{-1}	Porometer (LI-6400XT)	Leaf-level stomatal conductance under standardized conditions
	Plant Area Index	PAI	m² m^{-2}	SunScan / LiDAR	Ratio of total plant surface area (leaf + woody) to ground area
Yield & Cracking	Total Yield	Yield	fruits tree^{-1}	Field harvest	Total number of fruits per tree at harvest
	Fruit Cracking Rate	Cracking	%	Field counts	(Number of cracked fruits / total fruits) × 100
Multispectral RS	Normalized Difference Vegetation Index	NDVI	–	UAV multispectral	(NIR − Red) / (NIR + Red)
	Green NDVI	GNDVI	–	UAV multispectral	(NIR − Green) / (NIR + Green)
	Red-Edge NDVI	NDRE	–	UAV multispectral	(NIR − RedEdge) / (NIR + RedEdge)
	Enhanced Vegetation Index	EVI	–	UAV multispectral	2.5 × (NIR − Red) / (NIR + 6Red − 7.5Blue + 1)
	Soil-Adjusted Vegetation Index	SAVI	-	UAV multispectral	((1+0.5)*(Nir -Red))/(Nir +Red+0.5)
	Modified Soil Adjusted Vegetation Index	MSAVI	-	UAV multispectral	(2* Nir +1 -√((2 * Nir +1)^2-8*(Nir - Red)))/2
	Chlorophyll Index	CI	-	UAV multispectral	(Nir /Green)-1
	Burn Area Index	BAI	-	UAV multispectral	1 / ((0.1 + Red)^2 + (0.06 + Nir))
	Blue Normalized Difference Index	BNDI	-	UAV multispectral	(Nir - Blue) / (Nir + Blue)
	Infrared Percentage Vegetation Index	IPVI	-	UAV multispectral	Nir / (Nir + Red)
	Structure Insensitive Pigment Index	SIPI	-	UAV multispectral	(Nir - Green) / (Nir + Red)
	Optimized Soil Adjusted Vegetation Index	OSAVI	-	UAV multispectral	(Nir - Red) / (Nir + Red + 0.16)

	Difference Vegetation Index	DVI	-	UAV multispectral	Nir - Red
	Green Index	GI	-	UAV multispectral	Green / Red
	Reflectance bands	Blue, Green, Red, RE, NIR	–	UAV multispectral	Surface reflectance after empirical line calibration
Thermal RS	Canopy Temperature	Tc	°C	UAV thermal	Radiometric conversion from DN to temperature
	Air Temperature	Ta	°C	Weather station	Measured at canopy height
	Crop Water Stress Index	CWSI	0–1	UAV thermal	CWSI = (Tc − Ta − T(*wet)) / (T*(dry) − T(wet))
LiDAR Structural	Crown Area	CA	m²	UAV LiDAR	Projected canopy area from the segmented tree polygon
	Tree Height	H	m	UAV LiDAR	Max height above ground from normalized point cloud
	Leaf Area Density	LAD	$m^2\ m^{-3}$	UAV LiDAR	Derived from gap fraction via Beer–Lambert law
	Digital Elevation Model	DEM	m	UAV LiDAR	Ground surface model after PMF filtering
	Plant Area Index		$m^2\ m^{-2}$	UAV LiDAR	Derived from the sum of LAD along the tree height.
Meteorological	Air Temperature	Ta	°C	Weather station	Daily mean/max
	Relative Humidity	RH	%	Weather station	Measured at standard height
	Rainfall	P	mm	Weather station	Daily precipitation sum
	Vapor Pressure Deficit	VPD	kPa	Derived	VPD = e(s)(T) − e(a)
	Growing Degree Days	GDD	°C d	Derived	$\sum_{i=Aprtil\ 1^{th}}^{Harvest\ date}\left(\frac{T_{i,min}+T_{i,maz}}{2}\right)+T_{base}$. T_{base} is a unique for each crop. In case if the exact value of T_{base} was unknown, we assign it to 0.
Soil & Irrigation	Soil Water Content	SWC	%	Soil probes	Volumetric water content at 15–60 cm
	Irrigation Amount	Irr	mm	Irrigation records	Applied water per treatment
	Irrigation Treatment	Trt	categorical	Experimental design	Control (100%/ regular treatment, over-irrigation (130%, 150%), deficit irrigation and the time (50% Early, 50% Late)
Temporal / Phenology	Month	Month	–	Calendar	Used as categorical/temporal predictor
	Phenological Stage	Pheno	categorical	Field observations	Fruit set, cell enlargement, harvest
Model Outputs (RF)	Predicted TG	TG (RF)	$\mu m\ d^{-1}$	RF model	ML-estimated trunk growth
	Predicted SWP	SWP (RF)	MPa	RF model	ML-estimated stem water potential
	Predicted SC	SC (RF)	$mol\ m^{-2}\ s^{-1}$	RF model	ML-estimated stomatal conductance
	Predicted Yield	Yield (RF)	$fruits\ tree^{-1}$	RF model	Yield prediction
	Predicted Cracking	Crack (RF)	%	RF model	Cracking prediction

Meteorological and soil moisture data

Meteorological data for the study areas were obtained from each site operated by the Ministry of Agriculture and Food Security of Israel, with data available at https://www.agrometeorology.co.il/. The dataset included air temperature (°C), relative humidity (RH, %), precipitation (mm day^{-1}), and growing degree days (GDD). Heatwave periods were identified according to the Israeli Meteorological Service (IMS) definition of the Eastern Mediterranean climate. According to this definition, a heatwave consists of at least three consecutive days with midday maximum temperatures of 32–35 °C, nighttime maximum temperatures of 28–29 °C, relative humidity above 50% at noon, and 80–90% at midnight. In addition to meteorological records, soil volumetric water content was continuously monitored using sensors installed at depths of 15, 30, 45, and 60 cm within the study plots (Phytech, LTD, Kfar Saba, Israel). Meteorological and soil moisture variables were subsequently integrated with UAV-derived indicators as inputs to the RF models, allowing the combined assessment of environmental drivers, PWS, yield, and fruit cracking.

PWS measurements

Trunk diameter was continuously monitored using dendrometers (Phytech© LTD, Kfar Saba, Israelhttps://www.phytech.com), installed on 24 trees across the four irrigation treatments. These measurements provided data on MDS and TG. Mid-day SWP was measured immediately after each UAV data acquisition session using a pressure chamber (Arimad 3000, MRC, Holon, Israel) in 72 sample trees, essentially as described (Nemera et al, 2023). SC was measured between 9:00 and 11:00 h using a portable photosynthesis system (LI-6400XT; LI-COR Inc., Lincoln, NE, USA). Measurements were performed on fully expanded leaves under controlled conditions: photosynthetic photon flux density of 1000 µmol m^{-2} s^{-1}, flow rate of 500 µmol s^{-1}, and leaf chamber temperature of 25 °C. PAI was measured using the SS1 SunScan Canopy Analysis System (https://delta-t.co.uk/product/sunscan/) on flight dates near midday (12:00). Measurements were conducted during clear periods, free from cloud cover, to eliminate the effect of clouds on the PAI readings (Ogunbadewa 2012). For each tree, four measurements were taken from the south, west, north, and east sides. The average of these four readings was used as the final PAI value, representing both the leaf (photosynthetic material) and the woody parts of the tree (non-photosynthetic material).

Yield and percentage fruit cracking data

Throughout the fruit-cracking period, which typically spans from September to December, cracked fruits on each measurement tree were manually counted every 3-4 weeks. To avoid double-counting, the counted fruits were either picked or removed from the tree's outer branches. During the commercial harvest, the total number of fruits per tree was recorded, and the final percentage of cracked fruit was calculated as the ratio of cracked fruits to the total fruit count. In addition, the total fruit weight per tree was measured at harvest.

Remote sensing UAV

We collected data from UAV flights conducted during two growing seasons (Table 1-4), acquiring multispectral and thermal imagery, as well as LiDAR point clouds. Flights were conducted around midday (12:00) at an altitude of 40 m above ground, resulting in images with a spatial resolution of 2–5 cm per pixel. The UAV platform (DJI-600) was equipped with three sensors: (1) Multispectral sensor (RedEdge-M, MicaSense, Seattle, WA, USA): records reflectance in five narrow bands, blue (475 nm), green (560 nm), red (668 nm), red-edge (717 nm), and near-infrared (NIR, 840 nm). These bands are sensitive to leaf pigment concentration, canopy vigor, and water content, making them particularly useful for vegetation indices. (2) Thermal sensor (A655sc thermal camera, FLIR® Systems, Billerica, MA, USA; spectral range 7.5-14 µm): provides canopy temperature (Tc) with high accuracy. Tc is a well-established proxy for plant water status, reflecting transpiration cooling and stomatal conductance. (3) LiDAR sensor (DJI H2O): captures 3D point clouds of the orchard, providing detailed canopy structure, including tree height, crown area, and plant area index (PAI).

Image processing involved empirical line calibration (Kizel, Bruzzone, and Benediktsson, 2017), which converts raw digital numbers from the multispectral camera into physically meaningful surface reflectance values. This empirical line approach corrects for illumination differences, atmospheric conditions, and sensor drift. This step minimized biases caused by seasonal changes in illumination or sensor response. Empirical line calibration was chosen over alternatives, such as radiometric panel calibration or atmospheric correction models (e.g., FLAASH, SEN2COR), because it directly utilizes in-scene reference targets, thereby improving accuracy under variable illumination and atmospheric conditions common in UAV campaigns (Fadi Kizel, Bruzzone, and Benediktsson 2017). Following calibration, we calculated 12 vegetation indices used as features in the RF model (VIs; Table 5), including indices sensitive to canopy chlorophyll and nitrogen content (e.g., NDVI and NDRE), as well as those related to water stress (e.g., GNDVI and red edge indexes). LiDAR-derived (PAI, crown area, height) features formed the multi-sensor dataset for machine learning models.

Thermal imagery was initially recorded as digital numbers (DN) values. These were converted to surface temperatures using an open-source thermographic analysis tool (Tattersall 2017), which applies standard radiometric conversion equations to transform DN values into physically meaningful canopy temperature (Tc). During this process, emissivity corrections and atmospheric normalization were used to ensure comparability across flight dates and environmental conditions. After conversion, the crop water stress index (CWSI) was calculated following Cohen et al. (2015) using Eq. 1:

$$\text{Eq. 1: } CWSI = \frac{(Tc - Ta) - Twet}{Tdry - Twet}$$

Where Tc = canopy temperature (°C), Ta = air temperature (°C), Twet = lower baseline (2.5th percentile of Tc − Ta), representing a well-watered condition, Tdry = upper baseline (97.5th percentile of Tc − Ta), representing a non-transpiring canopy. The CWSI provides a dimensionless indicator (ranging from 0 to 1) of PWS. Values close to 0 indicate well-watered conditions, with Tc near Twet, while values approaching 1 indicate severe stress, with Tc approaching Tdry. By normalizing Tc against theoretical upper and lower limits, CWSI accounts for varying atmospheric conditions and enables cross-date comparisons of canopy water status (Idso et al. 1981). Importantly, CWSI is closely correlated with stomatal conductance (SC), making it more physiologically meaningful than Tc alone and directly linking thermal imaging to plant water regulation (Idso et al. 1981; Blonquist, Norman, and Bugbee 2009).

The LiDAR point clouds underwent three main pre-processing steps: ground classification, height normalization, and tree segmentation. **Ground classification.** Ground points were identified using the Progressive Morphological Filter (PMF) method (Zhang et al. 2003), which iteratively removes non-ground points based on the window size and elevation thresholds. This step generated a reliable digital elevation model (DEM). **Height normalization.** A Triangulated Irregular Network (TIN) approach was applied to normalize heights (Roussel et al. 2020). This method uses linear interpolation within each triangle to correct elevation values, yielding a normalized point cloud in which heights are expressed relative to ground level. **Tree segmentation.** A Canopy Height Model (CHM) raster was created by subtracting the digital surface model (DSM) from the DEM. Individual trees were segmented from the CHM using a watershed algorithm (Roussel et al. 2020). Additional segmentation refinements were based on morphological features, including canopy area, mean tree spacing, and crown diameter. Manual corrections were applied to improve accuracy. The process involved defining minimal convex shapes that represented average citrus tree crowns and splitting connected canopies by identifying the longest axis and dividing it along the shortest vertical line, guided by the mean spacing, width, and crown area. This preprocessing pipeline provided precise tree-level delineation, enabling the extraction of canopy structural attributes, including tree height, crown area, and PAI, which were subsequently integrated with multispectral and thermal features to predict PWS.

In summary, UAV-based multispectral, thermal, and LiDAR data were preprocessed to produce a comprehensive feature set, including VIs, canopy temperature (Tc), CWSI, and structural metrics such as crown area and PAI. These features were aggregated at the tree level to ensure physiological and structural relevance for each monitored tree. By integrating spectral information on canopy spectral properties, thermal signals of transpiration dynamics, and LiDAR-derived structural attributes, the resulting dataset provides a multidimensional representation of PWS. This feature space was subsequently used as input to ML models to predict PWS indicators and to link them to yield and fruit-cracking outcomes.

ML-RF models for estimating SC, SWP, and TG

All data from ground-based measurements and UAV-derived features were combined into a single database, which was randomly split into training (70%) and validation (30%) subsets. Data from the control irrigation plots were reserved as an independent testing set to evaluate model robustness under field conditions. RF models were implemented to predict SC, SWP, and TG. RF is an ensemble learning algorithm that constructs multiple decision trees and aggregates their outputs, reducing overfitting while capturing nonlinear and high-dimensional relationships. It was selected over alternative ML methods because it is robust to noisy data, tolerant of multicollinearity, and capable of handling complex interactions among heterogeneous features (spectral, thermal, structural, and climatic) (Matsuki, Kuperman, and Van Dyke 2016). To optimize model performance, hyperparameters were tuned by varying the Number of trees (ntree) from 10 to 50, balancing computational efficiency with predictive stability. Number of variables at each split (mtry): 2–4, to regulate tree diversity and bias-variance trade-offs.

To minimize redundancy and improve parsimony, two complementary approaches were applied: (1) the Boruta algorithm (Kursa and Rudnicki 2010), a wrapper method based on RF that identifies all relevant variables by comparing original features against permuted "shadow" features, ensuring that informative predictors are retained. (2) Variance Inflation Factor (VIF) analysis (Groß 2003) applied with a threshold of 10 to detect and remove multicollinear predictors, reducing redundancy among spectral, thermal, and climatic variables. The best-performing models were selected based on the lowest root-mean-square error (RMSE) and the highest coefficient of determination (R^2) on the independent test dataset. For each model, variable importance (VIP) scores were calculated to quantify the relative contribution of spectral, thermal, LiDAR-derived, and meteorological predictors. The final predictor set included UAV features (VIs, Tc, CWSI, PAI, crown area), meteorological variables (air temperature [Ta, °C]), and topographical elevation derived from the LiDAR-based DEM. All RF models were developed in R (R Core Team, 2022) using the "caret" package (Kuhn, 2015), which provides a unified framework for model training, hyperparameter tuning, and performance evaluation in the R software platform (R Core Team, 2022).

ML-RF models for estimating yield and cracking-associated yield loss

All ground-based yield measurements, fruit cracking observations, and UAV-derived features were integrated into a unified database and randomly partitioned into training (70%) and validation (30%) datasets. Data from control irrigation plots and seasons not used during training were reserved as an independent testing dataset to evaluate model generalization and robustness under operational field conditions. RF models were implemented to estimate (i) total yield and (ii) cracking-associated yield loss at tree and plot scales. Model hyperparameters were optimized by systematically tuning the number of trees (ntree = 20–100) to ensure predictive stability while maintaining computational efficiency, and the number of variables randomly sampled at each split (mtry = 2-5) to balance bias-variance trade-offs. Model performance was evaluated on the independent test dataset using the coefficient of determination (R^2) and root-mean-square error (RMSE), and the best-performing configurations were retained. The final predictor set included UAV-derived variables (vegetation indices, canopy temperature [Tc], crop water stress index [CWSI], plant area index [PAI], crown area, and LiDAR-derived structural metrics), plant water status indicators on a monthly basis predicted by RF models (SC, SWP, TG), meteorological variables (air temperature [Ta, °C]), and management-related factors (irrigation regime and treatment indicators). Variable importance (VIP) scores were computed to quantify the relative contributions of spectral, thermal, structural, physiological, and environmental predictors to yield and cracking-loss estimation. All RF models were implemented in the R statistical environment (R Core Team, 2022) using the caret package (Kuhn 2015), which provides a unified framework for model training, hyperparameter optimization, feature selection, and performance evaluation.

Estimation of plant area index (PAI) using leaf area density (LAD)

LiDAR point clouds were masked using tree polygons obtained from the segmentation procedure. For each segment, we estimated the tree's leaf area density (LAD) profile, with the bottom height defined as 0.5 m above the ground (S. Li et al. 2017; Yao Wang and Fang 2020; Roussel et al. 2020). The LAD profile was derived from the gap-fraction profile, which quantifies the proportion of the sky visible through the canopy. Gap fraction values were converted into LAD using the Beer–Lambert law and an extinction coefficient (*k*) (Roussel et al., 2020). The value of *k* was estimated to be 0.4 based on a linear regression of ground-based PAI measurements (SunScan system) and LiDAR-derived PAI values (Roussel et al. 2020). In this study, we focused on estimating PAI rather than LAI because the available tools did not allow for excluding woody components (branches and stems) from the canopy. While PAI therefore includes both photosynthetic (leaves) and non-photosynthetic (woody) materials, it remains highly correlated with LAI (Fig. 2). It serves as a robust proxy for canopy structure, light interception, and transpiration potential (Fang et al. 2019).

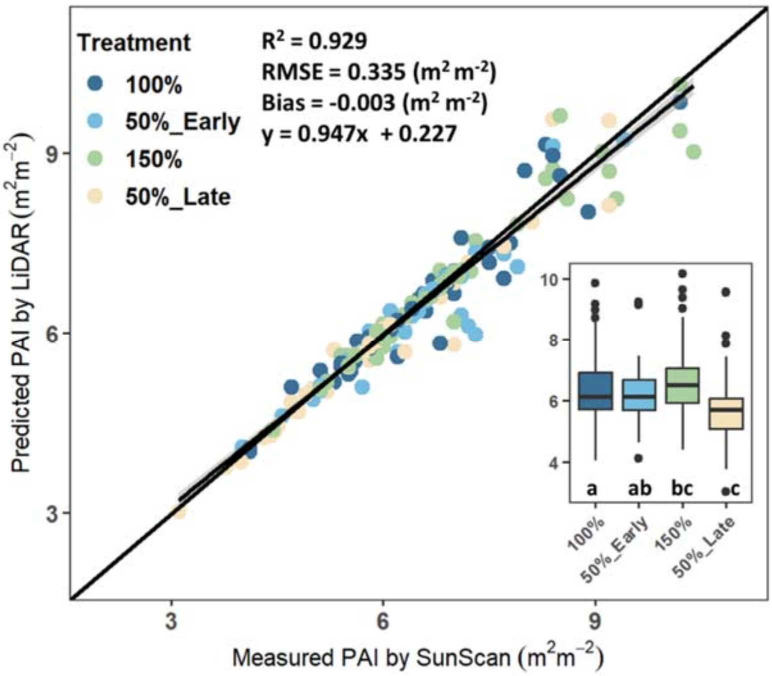


Figure 2. The correlation between estimated PAI derived from LiDAR and ground measurements from SunScan for the Israel citrus site in the 2023-2024 period is shown. The colors represent different irrigation treatments, and the small boxplot in the bottom-right corner of the graph shows differences in PAI values across treatments.

Clustering Analysis

Seasonal patterns of SWP, SC, and TG were derived from repeated measurements across the two growing seasons. These seasonal trajectories (session curves) capture the continuous physiological responses of citrus trees to environmental and management factors over time, including midday stress peaks, recovery phases, and growth fluctuations. Session curves provide a functional perspective on how PWS indicators evolve throughout critical phenological stages, rather than focusing only on isolated measurement points. To identify groups of trees with comparable seasonal trajectories, clustering analysis was applied. The optimal number of clusters was determined using the Calinski–Harabasz (CH) index (Caliński and Harabasz 1974), which maximizes separation between clusters relative to within-cluster variability. Candidate solutions ranging from 2 to 10 clusters were tested using the vegan package in R (Oksanen et al. 2022). Tree grouping was then performed using the Fuzzy C-Means (FCM) algorithm (Zhou, Fu, and Yang, 2014), which assigns each cluster a membership value between 0 and 1. This approach is particularly appropriate for physiological data, where water-stress responses are continuous rather than strictly discrete. FCM has been successfully applied in agricultural and environmental time-series studies (Ohana-Levi et al. 2021; 2022; Peeters et al. 2024). Cluster centroids, representing the average seasonal trajectory of each group, were extracted and compared against yield and fruit-cracking outcomes. Statistical differences among clusters were assessed using ANOVA with Tukey’s HSD test. All clustering analyses were performed in R using the ppclust package (Cebeci 2020). Importantly, clustering enabled the identification of management zones within the orchard, which are groups of trees with similar physiological stress responses. These zones can inform site-specific irrigation strategies to mitigate fruit cracking and optimize yield.

Citrus germplasm and irrigation experiment plots

Israel study site-Citrus

The irrigation experiment was conducted in the central coastal region of Israel, an area known for its Mediterranean climate and citrus cultivation. The study site is a commercial orchard of 19-years old 'Ori' mandarin (*Temple mandarin (Citrus reticulata X C. Sinensis) X Kinnow mandarin (C. nobilis × C. deliciosa*) trees grafted on Sour Orange (*Citrus × aurantium*) rootstock, located near Kfar Chabad in the Lod Valley (31°59'00.4"N, 34°51'38.6"E; Fig. 3). The orchard, containing about 1300 trees covers 1.56 hectares, and is planted at a spacing of 6 m between rows and 3.5 m within rows. The region has a typical Mediterranean climate, with a long, dry summer and an average annual rainfall of 550 mm. Irrigation is usually applied between April and November. The experimental plot consists of 36 rows, with a variable number of trees per row, arranged into four irrigation treatments, each replicated across three randomized blocks (Fig. 3). Treatments included: (1) Control (100%) – irrigation at the recommended level for the area and soil type. (2) Over-irrigation (150%) – continuous application of 150% of the recommended level. (3) Early reduced irrigation (50% early) – 50% irrigation applied during the early season (21 May-30 June 2023 and 05.05.24 - 26.06.24). (4) Late reduced irrigation (50% late) – 50% irrigation applied during the late season (17 August–14 October 2023 and 29.08.24 - 27.10.24). Irrigation amounts in control treatment were determined using the Penman–Monteith reference evapotranspiration, following guidelines from Israel's Extension Service, Ministry of Agriculture and Food Security. Locations of sampled trees were precisely mapped using Real-Time Kinematic (RTK) GPS. Monthly UAV flights were conducted throughout the 2023 ("off-crop") and 2024 ("on-crop") growing seasons using an integrated multi-sensor system that combined LiDAR, thermal, and multispectral sensors.

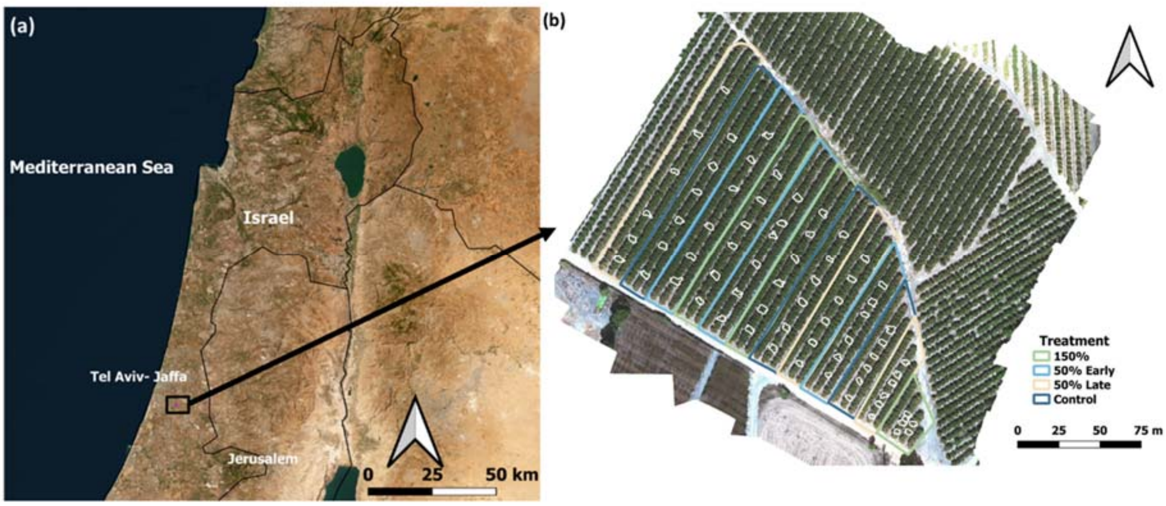


Figure 3. (a) Location of the study area in Israel; (b) layout of the irrigation experiment within the study plot. Coloured frames represent different irrigation treatments: yellow - overirrigation (150%), bright blue - early deficit irrigation (50% Early), yellow - late deficit irrigation (50% Late), and dark blue - control (100%). White frames indicate the studied trees.

Results

Climate and irrigation effects on the measured PWS

Seasonal climate conditions differed markedly between the two study years (Fig. 4). In 2023, total rainfall was lower (487 mm), and irrigation supply accumulated more slowly compared to 2024 (608 mm; Fig. 4), when water inputs were consistently higher across treatments. Despite the wetter conditions in 2024, the season was characterized by stronger and more persistent heat stress, with daily maximum temperatures frequently exceeding 35 °C and a total of 83 heatwave days compared to 65 in 2023 (Fig. 4). According to the Israeli Meteorological Service (IMS 2023) there were 18 and 28 distinct heatwave events in 2023 and 2024, respectively, with the majority concentrated in June and early July. Minimum temperatures followed the expected seasonal cycle, with cooler winter nights and pronounced summer-winter contrasts;

however, the frequency of extreme maxima in 2024 was noticeably greater. To mitigate extreme heat events, irrigation increased during summer 2024, resulting in a higher overall annual irrigation value (Fig. 4). These interannual climatic differences significantly influenced tree water relations. They underpinned the variability observed in PWS indicators, setting the stage for the yield and fruit cracking outcomes described in the following sections.

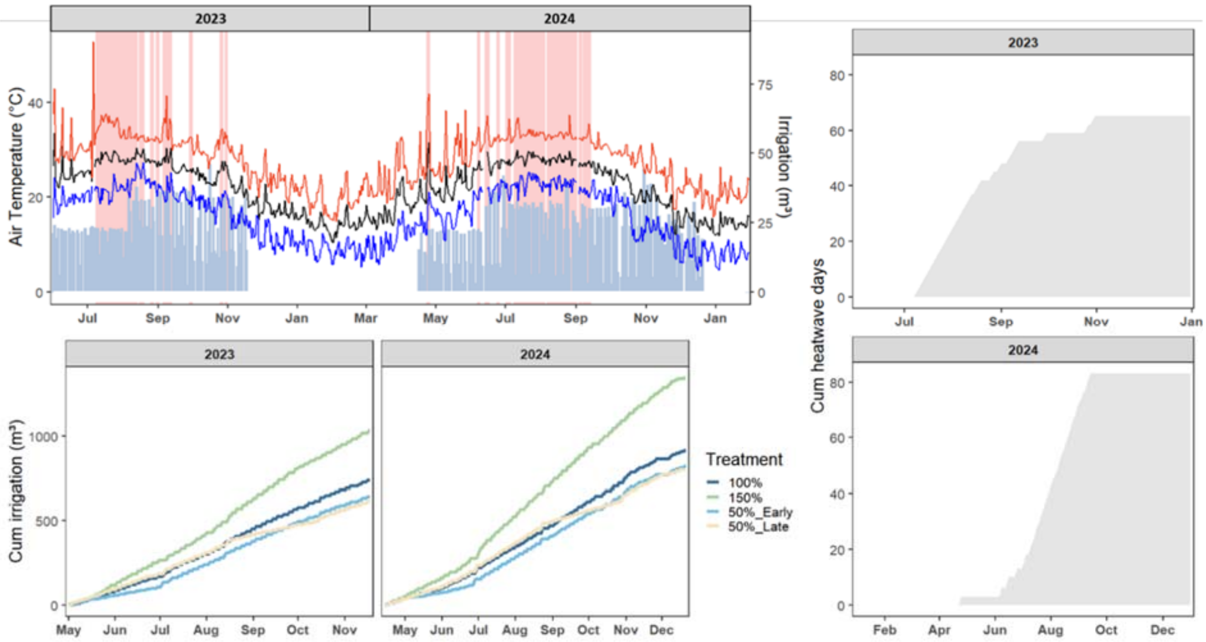


Figure 4. Seasonal meteorological conditions and irrigation dynamics during the 2023 and 2024 growing seasons, shown in relation to citrus phenological stages. (A) Soil moisture (light blue), maximum air temperature (red), mean air temperature (black), and minimum air temperature (blue). (B) Cumulative irrigation by treatment. (C) Cumulative number of heatwave days. Shaded vertical bars indicate key phenological phases: fruit set, cell enlargement, and harvest.

Plant water statues (PWS)

Tree physiological responses varied significantly across treatments and seasons (Fig. 5). TG steadily increased during 2023, reaching higher final values than in 2024, when growth was slower and more limited throughout the season. Regardless of irrigation intensity, this decline in cumulative growth coincided with the higher frequency of heatwaves and increased maximum temperatures observed in 2024, especially in the 50% irrigation early treatment, where negative values were detected (Fig. 5). Additionally, SWP reflected these differences, especially in the early deficit irrigation (50% Early) treatment, which exhibited the lowest SWP values. SC values remained low during the summer months and recovered toward autumn. In 2023, SC increased more uniformly across treatments (~0.04–0.06 mol m^{-2} s^{-1}), whereas in 2024, conductance was generally higher (~0.07–0.09 mol m^{-2} s^{-1}). In both years, SC showed the lowest values under deficient irrigation, with the early-deficit treatment exhibiting the lowest values in 2024, in line with the heat waves during this period (Fig. 5).

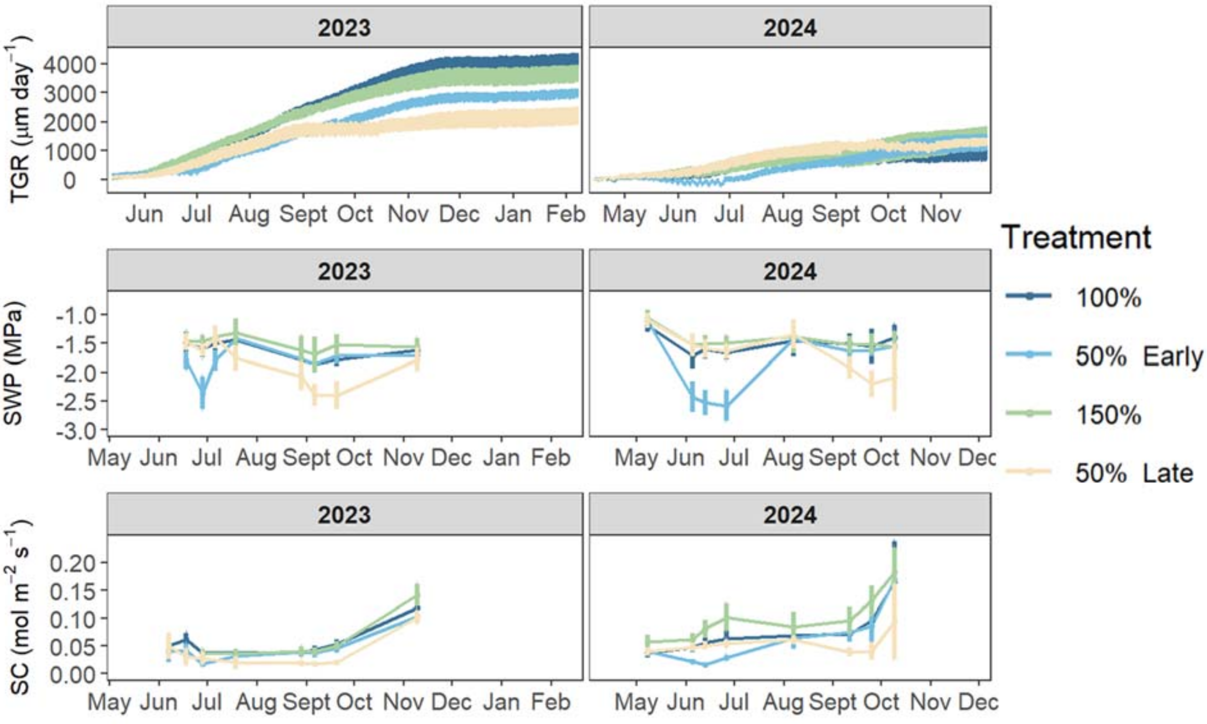


Figure 5. Seasonal dynamics of plant water status (PWS) indicators during the 2023 and 2024 growing seasons. (A) Cumulative trunk growth rate (TG). (B) Stem water potential (SWP). (C) Stomatal conductance (SC). Colors represent irrigation treatments: red - control (100%); purple – over-irrigation (150%); green - early deficit irrigation (50% Early); and pink - late deficit irrigation (50% Late).

UAV-based prediction of PWS indicators using ML models

Across all three physiological traits (TG, SWP, and SC), the Random Forest (RF) model outperformed XGBoost, ANN, and SVR, achieving the highest R^2 values (0.90–0.95) and the lowest RMSE and MAE values (Table 6). TG predictions showed strong agreement with observed measurements (R^2 = 0.90), while SWP and SC exhibited similarly high predictive accuracy (R^2 = 0.93 and 0.95, respectively). Although XGBoost and SVR achieved competitive results, both showed higher error rates and lower stability than RF. Overall, the results demonstrate that RF offers the most robust and reliable framework for estimating plant water-status indicators from UAV-derived features.

Table 6. Performance comparison of four machine-learning models: Random Forest, XGBoost, ANN, and Support Vector Regression (SVR), for predicting daily trunk growth (TG), stem water potential (SWP), and stomatal conductance (SC). Metrics include coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and cross-validation folds (CV). Random Forest consistently achieved the highest predictive accuracy across all traits, exhibiting low error rates and high model stability.

Model TG	**R^2 train mean**	**R^2 train std**	**R^2 test mean**	**R^2 test std**	**RMSE train mean**	**RMSE train std**	**RMSE test mean**	**RMSE test std**	**n_folds**
RF	**0.900**	**0.029**	**0.832**	**0.061**	**0.709**	**0.151**	**8.468**	**1.693**	**10**
ANN	0.963	0.039	0.431	0.351	1.631	1.503	13.379	3.526	10
XGBoost	0.914	0.023	0.803	0.073	0.699	0.170	9.190	1.787	10
SVR	0.908	0.026	0.808	0.069	0.704	0.158	8.832	1.706	10

Model (SWP)	**R^2 train mean**	**R^2 train std**	**R^2 test mean**	**R^2 test std**	**RMSE train mean**	**RMSE train std**	**RMSE test mean**	**RMSE test std**	**n_folds**
RF	**0.926**	**0.021**	**0.856**	**0.014**	**0.137**	**0.071**	**0.150**	**0.052**	**10**
ANN	0.874	0.064	0.524	0.252	0.352	0.516	0.663	0.269	10
XGBoost	0.952	0.014	0.793	0.063	0.174	0.087	0.220	0.120	10
SVR	0.942	0.055	0.680	0.204	0.246	0.335	0,215	0.190	10

Model SC	R^2 train mean	R^2 train std	R^2 test mean	R^2 test std	RMSE train mean	RMSE train std	RMSE test mean	RMSE test std	n_folds
RF	**0.945**	**0.009**	**0.851**	**0.024**	**0.013**	**0.052**	**0.030**	**0.020**	**10**
ANN	0.955	0.004	0.498	0.350	0.012	0.031	0.145	0.389	10
XGBoost	0.934	0.008	0.804	0.058	0.013	0.056	0.045	0.031	10
SVR	0.951	0.008	0.813	0.035	0.013	0.075	0.044	0.028	10

UAV-based modeling provided accurate predictions of PWS indicators across irrigation treatments, excluding the control plots, used to validate the model (Fig. 6). For TG (Fig. 6), the RF model achieved strong predictive performance (R^2Cal = 0.90, RMSECal = 0.709; R^2Val = 0.83, RMSEVal = 8.468; R^2Pred = 0.79, RMSEPred = 7.35). The most influential predictors included irrigation treatment and growing degree days (GDD), followed by relative humidity (RH), vegetation indices such as NDVI and EVI, red-band reflectance (668 nm), canopy temperature (Tc), and temporal features. For SWP (Fig. 6), accuracy was similarly high (R^2Cal = 0.92, RMSECal = 0.137; R^2Val = 0.85, RMSEVal = 0.15; R^2Pred = 0.82, RMSEPred = 0.12). The main predictors were mean NDVI, irrigation, 842 nm, and Tc, along with reflectance at 560 nm and in the red-edge (717 nm), highlighting the combined roles of climatic drivers and canopy reflectance features in shaping water potential. For SC (Fig. 6), predictive capacity was also robust (R^2Cal = 0.945, RMSECal = 0.013; R^2Val = 0.85, RMSEVal = 0.03; R^2Pred = 0.783, RMSEPred = 0.03). The strongest predictors included spectral features in the GNDI, GDD, red-edge (717 nm), elevation, and temporal variables. SWP predictions were slightly less accurate than TG, likely reflecting its higher short-term variability and dependence on environmental factors, including irrigation applied 2 to 3 days prior to the measurements. SC predictions highlighted the sensitivity of stomatal regulation to red-edge and NIR spectral information by linking UAV-derived features to TG, SWP, and SC.

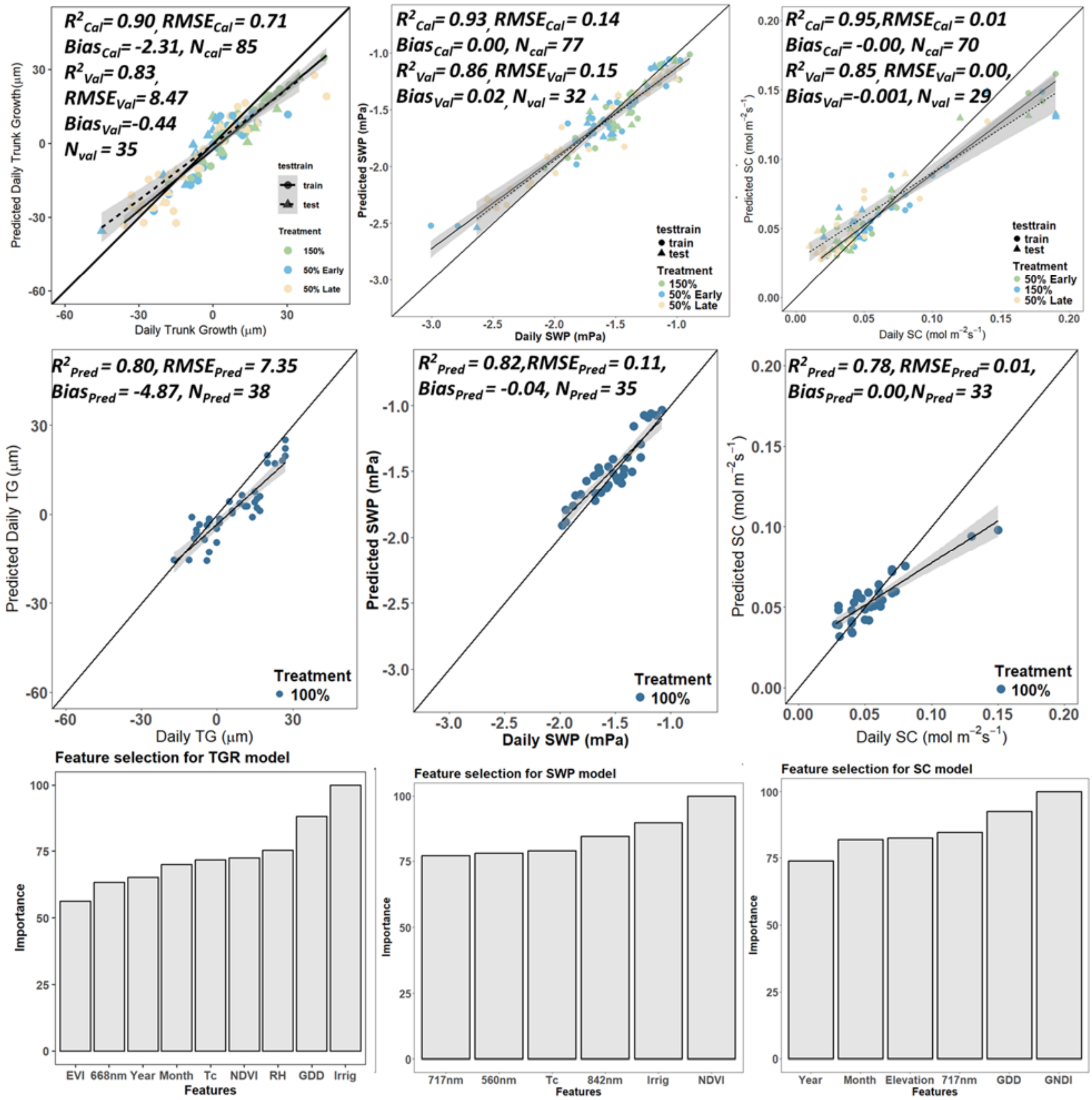


Figure 6. *Upper panels*: UAV-based prediction performance for plant water status (PWS) indicators: daily trunk growth rate (TG), stem water potential (SWP), and stomatal conductance (SC), across irrigation treatments (50% Early, 50% Late, 150%, and control). Observed versus predicted values are shown for calibration (colored points) and validation (gray triangles). The solid black line represents the 1:1 line, and the shaded band represents the model fit. Calibration and validation statistics (R^2, RMSE, Bias, and sample size N) are reported for each model. Middle panels: Prediction performance for TG, SWP, and SC using only the control plots, demonstrating UAV model robustness under standard irrigation conditions. *Lower panels*: Variable importance profiles (VIP) for the TG, SWP, and SC models, showing the relative contribution of spectral (e.g., 717 nm, NDVI), thermal, climatic (e.g., temperature, GDD), and management variables (e.g., irrigation) to model performance.

Yield and fruit-cracking

For yield estimation (Fig. 7), the RF model demonstrated very strong predictive performance, with high agreement between observed and predicted yields across calibration and validation datasets. Calibration results showed an R^2cal of 0.956 and an RMSEcal of 109.6 fruits, while validation performance remained high ($R^2 val$ = 0.913, RMSEval = 174.6 fruits), indicating good model generalization across irrigation treatments and seasons. Bias values were low and negative in both calibration (−16.8 fruits) and validation (−33.6 fruits), suggesting a slight underestimation at higher yield levels but no systematic distortion. Feature importance analysis revealed that SC and SWP in July were the dominant predictors of yield, followed by TGR. Structural canopy metrics, represented by PAI in July and September, also contributed to yield prediction but with lower relative importance. These results highlight the central role of mid-season PWS and growth dynamics in controlling final fruit load, consistent with the physiological coupling between canopy water relations, carbon allocation, and yield formation.

Table 7. Performance comparison of four machine-learning models: Random Forest, XGBoost, ANN, and Support Vector Regression (SVR), for predicting Yield and Cracking incidents. Metrics include coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and cross-validation folds (CV). Random Forest consistently achieved the highest predictive accuracy across all traits, exhibiting low error rates and high model stability

Yield Model	**R^2 train mean**	**R^2 train std**	**R^2 test mean**	**R^2 test std**	**RMSE train mean**	**RMSE train std**	**RMSE test mean**	**RMSE test std**	**n_folds**
RF	**0.956**	**0.075**	**0.913**	**0.099**	**109.595**	**63.183**	**174.613**	**73.522**	**10**
ANN	0.963	0.053	0.582	0.381	89.531	48.899	277.366	168.899	10
XGBoost	0.936	0.083	0.860	0.124	110.356	68.304	188.694	79.646	10
SVR	0.961	0.079	0.839	0.159	124.952	69.890	198.151	86.491	10

Cracking Model	**R^2 train mean**	**R^2 train std**	**R^2 test mean**	**R^2 test std**	**RMSE train mean**	**RMSE train std**	**RMSE test mean**	**RMSE test std**	**n_folds**
RF	**0.927**	**0.057**	**0.851**	**0.093**	**1.460**	**0.323**	**2.383**	**0.513**	**10**
ANN	0.959	0.043	0.382	0.329	1.248	0.287	6.287	3.415	10
XGBoost	0.919	0.059	0.843	0.096	1.464	0.328	2.391	0.528	10
SVR	0.924	0.057	0.80	0.102	1.469	0.332	2.408	0.531	10

For fruit cracking estimation (Fig. 7), the RF model exhibited strong predictive performance on both the calibration and validation datasets. Calibration results yielded an R^2cal of 0.927 and an $RMSC_{cal}$ of 1.46%, while validation performance remained robust (R^2val = 0.851, $RMSEval$ = 2.38%), indicating good generalization across irrigation treatments and seasons. Bias values were low in both calibration (−0.46%) and validation (−1.19%), suggesting stable, largely unbiased predictions across the observed range of cracking intensities. Feature importance analysis identified SWP as the dominant predictor of fruit cracking, particularly during September, followed by SWP in July. SC in July and September contributed moderately to model performance, while PAI in September showed a smaller but non-negligible influence. This hierarchy underscores the critical role of late- and mid-season plant water status in controlling fruit cracking incidence, highlighting the sensitivity of cracking to hydraulic conditions during advanced phenological stages. These results are consistent with the observed relationships between PWS indicators during July–September and cracking rates, supporting the physiological basis of the RF-based cracking prediction model.

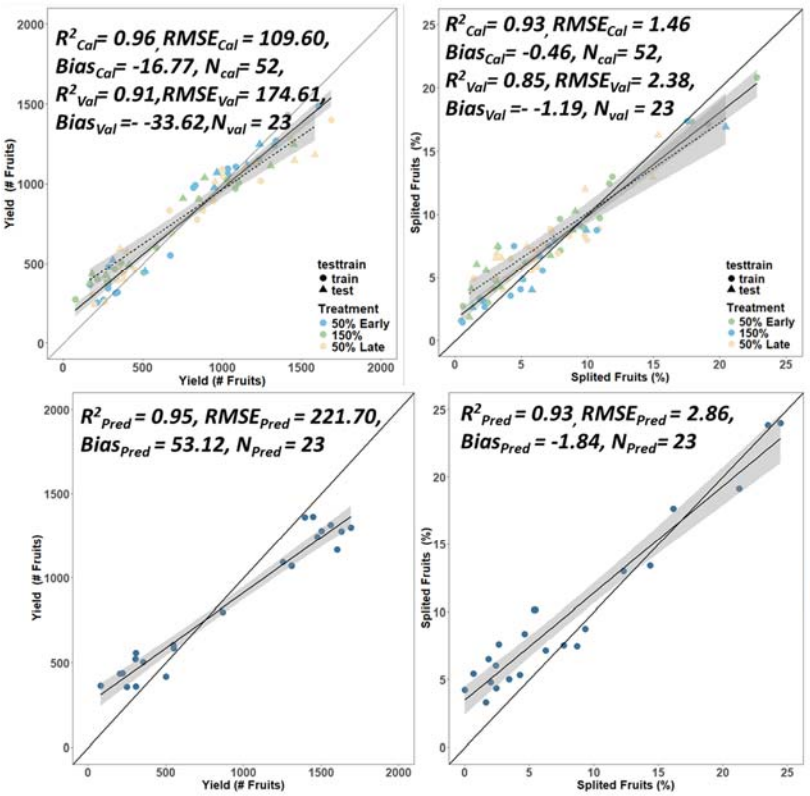
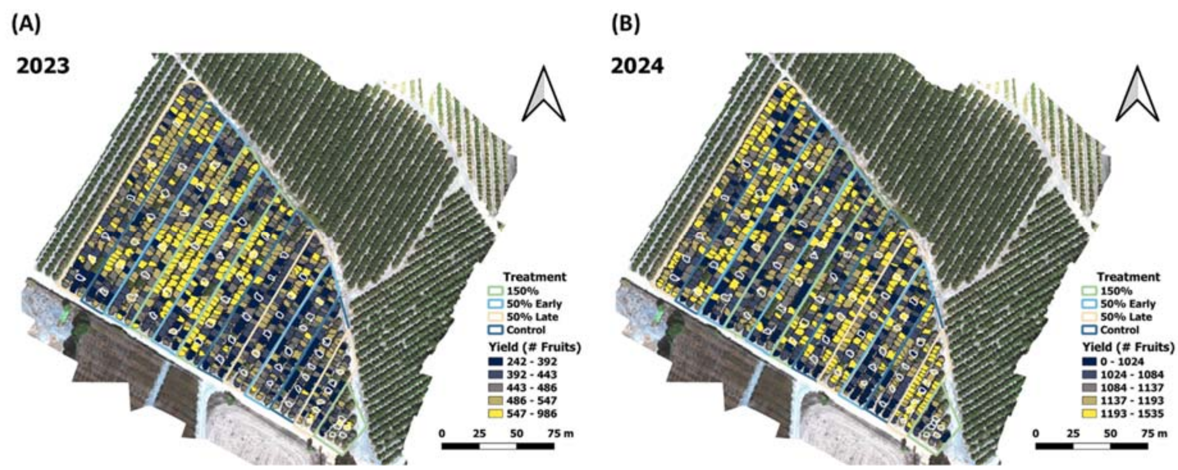


Figure 7. UAV-based prediction of citrus yield and fruit cracking under contrasting irrigation regimes (50% Early, 50% Late, 150%). (Left panels) Relationships between observed and predicted values for yield (top) and fruit cracking rate (bottom) for calibration (gray circles) and validation (gray triangles) datasets. The solid line denotes the 1:1 reference, and the dashed line represents the model fit. High predictive performance was achieved for both traits (Right panels). Feature importance analysis indicates that yield prediction was strongly influenced by July stomatal conductance (SC), July stem water potential (SWP), July trunk growth rate (TGR), and plant area index (PAI) in July and September. Fruit cracking prediction was driven primarily by July and September SWP and SC, and by September PAI. (Bottom-left and bottom-middle panels) Independent testing using data from the control plots demonstrates strong model generalization for both yield and cracking rate.



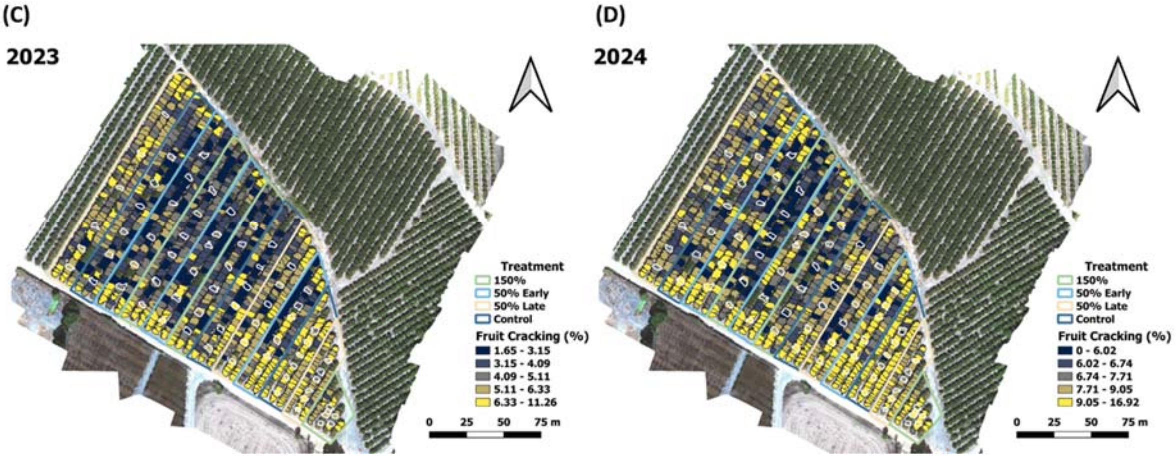


Figure 8. Spatial distribution of citrus yield and fruit cracking under different irrigation regimes (50% Early, 50% Late, 150%, and Control) in 2023 and 2024. (A–B) Yield (number of fruit trees¹) for 2023 and 2024. (C–D) Fruit cracking rate (%) for 2023 and 2024. Maps highlight strong treatment-driven heterogeneity: high irrigation (150%) consistently increased yield but was also associated with elevated cracking, whereas deficit irrigation (50% Early and 50% Late) reduced yield yet contributed to localized reductions in cracking severity.

France studies the citrus site.

We have produced data at a tree scale from flights over citrus and calculated vegetation indices, tree-morphology traits, and numerical products. We calculated 13 Vis:

Including NDVI, RVI, DVI, GNDVI, GRVI, Ctr2, CV, CCCI, CG, CL-green, and CL-red edge. In addition, we extracted structural data on the tree and orchard structures, including height, area, cover, volume, and growth, on a tree-by-tree basis using a segmentation approach. The orchard elevation' slop' aspect and some data calculations are still in progress. Figure 9 below presents examples of remote sensing and data acquisition. Since 2024, we have started a nitrogen nutrition experiment in citrus (Clementina) to develop a chlorophyll and nitrogen content model, as requested in the project. The same calculations for citrus in France were also performed for sweet cherries, with the data collected on the flight dates before. This is an example of the distribution of tree volumes estimated from flight data on different dates, as shown in Figure 9.

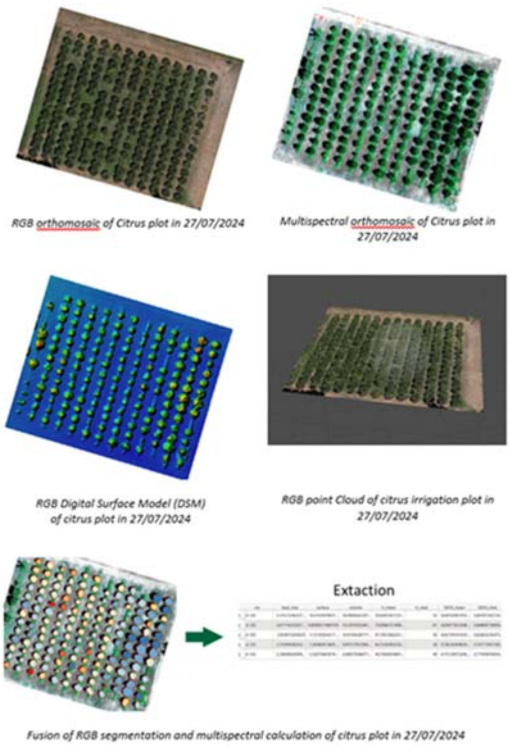


Figure 9. Example of image and data acquisition with remote sensing in citrus germplasm plot

Predictive Modelling based on UAV Remote Sensing and Weather data:

Study Area: Irrigation experiment – FRCitrus1

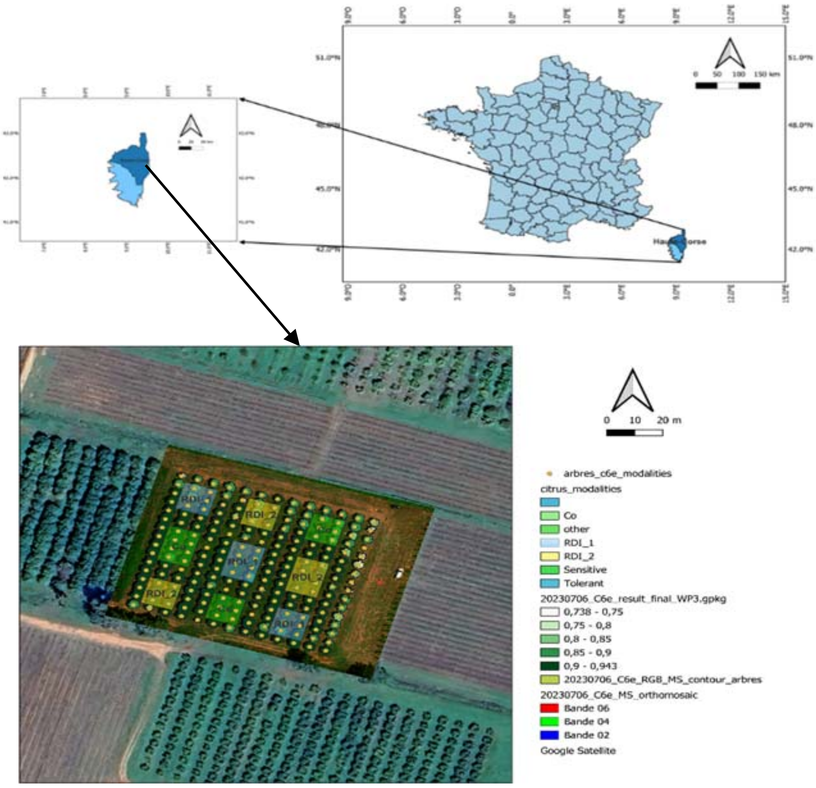


Figure 10. Study Area Map - FRCitrus1

The plot of the irrigation experiment - FRCitrus1 is located on the northeast coast of the island of Corsica, France (42.28181739595965, 9.521238561092073), a classical Mediterranean area for citrus growing. The study area map presents the design of our irrigation experiment with three modalities (Control, RDI_1, RD_2):

- **Control:** Irrigation at 100% ETP (Control in green)
- **RDI_1**: Deficit irrigation at 50% ETP from July (RDI > July in blue)
- **RDI_2:** Deficit irrigation at 50% ETP from August (RDI > August, yellow)

Stem Water Potential (SWP) Prediction:
Data context and SWP measurement conditions

The 2023 and 2024 data were collected under different conditions, which need to be taken into account when modelling SWP.

- **2023 campaign**

In 2023, the dataset included only multispectral (MS) imagery, as thermal infrared (IRT) acquisitions were not yet possible. Moreover, SWP measurements were not acquired on the same day as the UAV flights but within the same month, sometimes several days apart.

This temporal mismatch introduces additional uncertainty, since SWP can vary rapidly in response to short-term weather conditions.

- **2024 campaign**

In 2024, all flights included both MS and thermal (IRT) data, and SWP was measured either on the same day or within 1 day of the UAV acquisition, except in September, when a technical issue with the multispectral sensor prevented measurement. This provides much better temporal consistency between remote sensing inputs and physiological ground truth, making 2024 the highest-quality dataset for SWP prediction.

- **Implications for modelling**

Given these differences, the two years cannot be merged directly.

2023 increases sample size but suffers from lower measurement accuracy, while 2024 offers fewer samples but much higher reliability and includes thermal information.

For this reason, three datasets were used:

- **MS-only (2023 - 2024):** larger but noisier
- **MS-only (2024):** smaller, higher-quality labels with temporal alignment
- **MS + IRT (2024):** full feature set + temporal alignment

This dataset organization ensures a valid and unbiased evaluation of the impact of thermal features on SWP prediction performance.

Dataset construction and data preparation

The datasets used for modelling SWP were obtained by merging multispectral (MS), thermal (IRT), and weather data. Several preprocessing steps were required to ensure temporal consistency, data quality, and feature reliability.

Construction of SWP datasets (2023-2024)

After merging RS, weather, and SWP measurements, and applying same-day matching between UAV flights and SWP observations, the resulting dataset contains (before cleaning):

- 78 SWP samples* for 2023 **(MS only)**
- 105 SWP samples for 2024 **(MS + IRT)**
- 183 samples in total (2023–2024)

one sample = mean SWP of three leaves from one tree.

Weather feature engineering

Weather data were aggregated over short temporal windows preceding the flight, using a 3-day window for the main analysis. For each flight date, mean, sum, or maximum values were computed depending on the meteorological variable:

- Temperature, dew point, VPD, relative humidity (mean)
- Wind speed (mean and max)
- Rain (sum)
- Reference evapotranspiration ET0 (sum)
- Global radiation (mean)
- Daily thermal amplitude (mean)
- Growing Degree Days (GDD)

To capture seasonality, Day of Year (DOY) was also included. DOY indicates the position of the measurement within the annual cycle (1-365). Growing Degree Days (GDD) were calculated as the cumulative sum of daily thermal units, defined as the positive difference between the mean daily air temperature (computed as the average of the daily maximum and minimum temperatures) and a base temperature of 13 °C. This variable is primarily used in the following section to predict yield/cracking. A 3-day aggregation window was used because SWP responds to short-term atmospheric conditions. This window captures relevant variability in temperature, humidity, VPD, and ET_0 over the days preceding the flight, while avoiding the excessive smoothing effects of longer windows.

Remote sensing feature engineering

Multispectral data were enriched with additional vegetation indices derived from blue, green, red, red-edge, and NIR bands:

- NDWI, NIRv, WDRVI, and VARI (see the le below for formulas).
- **Classical indices:** NDVI, MSAVI, SAVI, EVI, GNDI, NDRE, SIPI, OSAVI, DVI.
- **Structural features:** canopy area, estimated canopy volume, DEM.

For the 2024 thermal dataset, canopy temperature (**Tc**) and Crop Water Stress Index (**CWSI**) were included.

Table 8. Vegetation indices derived from UAV multispectral imagery are used to characterize canopy water status, structure, and photosynthetic activity.

Index	Formula	Interpretation / Purpose
NDWI (Normalized Difference Water Index)	((G - NIR) / (G + NIR))	Sensitive to leaf and canopy water content; useful for detecting early water stress.
NIRv (NIR reflectance of vegetation)	(NIR x NDVI)	Combines greenness and NIR reflectance; relates to photosynthetic activity.
WDRVI (Wide Dynamic Range Vegetation Index)	(α × NIR - R) / (α × NIR + R), with α = 0.1	Reduces NDVI saturation in dense canopies, such as those of citrus.
VARI (Visible Atmospherically Resistant Index)	((G - R) / (G + R - B))	Describes vegetation greenness from visible bands and is robust to illumination changes.

Data cleaning and outlier removal

Several preprocessing steps were applied to ensure the reliability of the SWP modelling datasets:

- **Outlier detection (IQR method):**
  SWP outliers were detected per flight date using an interquartile range rule *(Q1 – 1.2×IQR*, Q3 + 1.2×IQR).*
  *(IQR = Q3 – Q1)**
- **Removal of missing or inconsistent records:**
  Rows containing missing values or invalid information were removed.
- **Preliminary feature filtering:**
  Highly correlated predictors were later removed using a threshold to avoid redundancy and multicollinearity (see next section).

Feature selection based on correlation analysis

To reduce predictor redundancy and prevent multicollinearity in the ML models, a correlation-based feature filtering step was applied. A Pearson correlation matrix was computed for all numerical features in each dataset (MS-only 2023-2024, MS-only 2024, and MS+IRT 2024). The corresponding matrices are presented in the following sections. Pairs of variables with very high absolute correlation (**$|r| > 0.90$**) were considered redundant, and only one variable from each pair was retained.

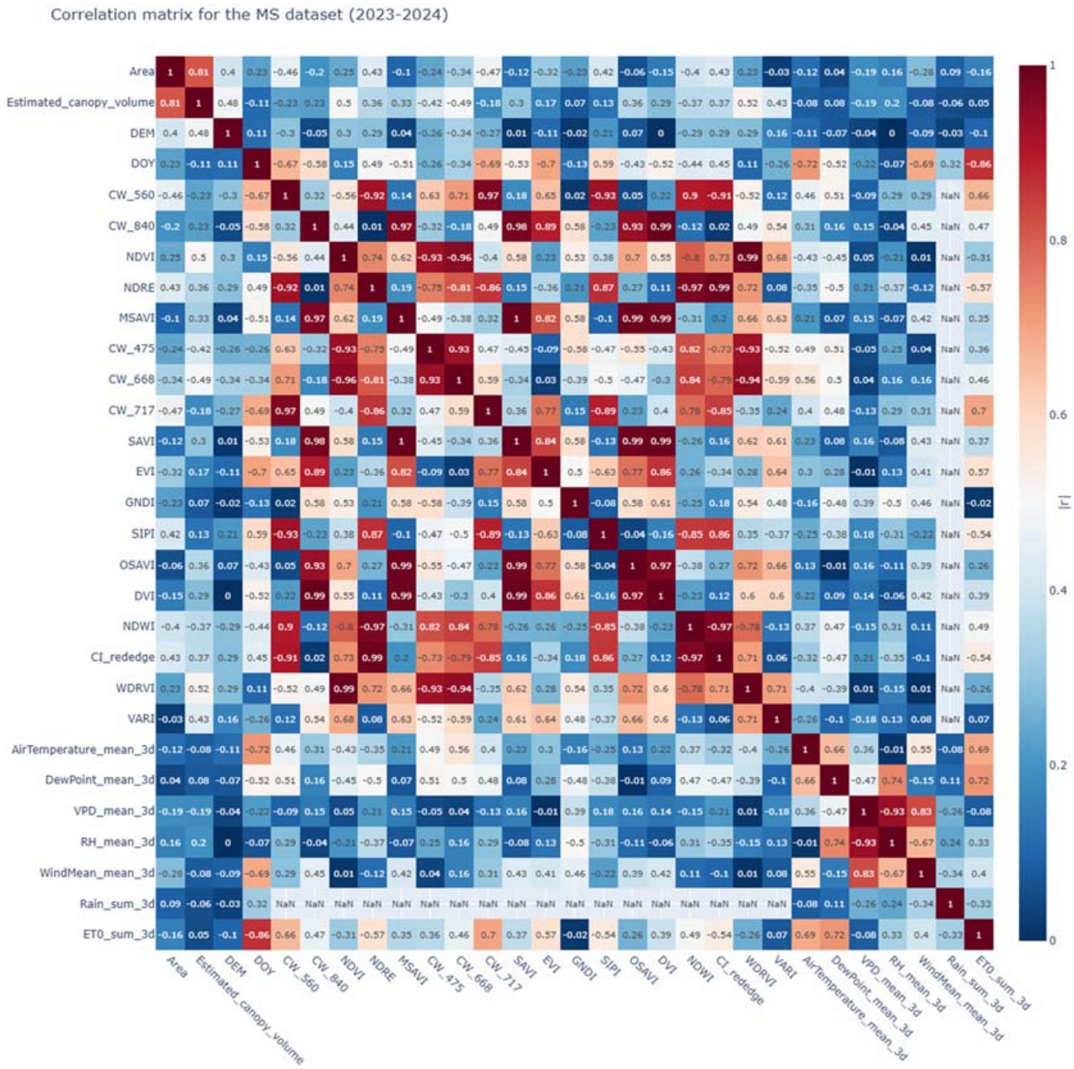


Figure 11. Correlation matrix for the MS-only dataset (2023–2024)

MS-only 2023-2024 dataset

The MS-only 2023-2024 correlation matrix revealed strong redundancy among vegetation indices and spectral bands (e.g., **NDVI & WDRVI, MSAVI & SAVI, NDRE & CI_rededge**) due to their shared reflectance components. Some weather aggregates were also strongly correlated, such as **VPD_mean_3d** and **RH_mean_3d**. **Rain_sum_3d** is equal to zero for most observations, resulting in near-zero variance. Therefore, linear correlations with vegetation indices cannot be computed and appear as **NaN values**. After applying the **|r| > 0.90** threshold, **16** features were retained out of **29**, ensuring that only non-redundant spectral and meteorological predictors were kept for modelling.

MS-only 2024 dataset

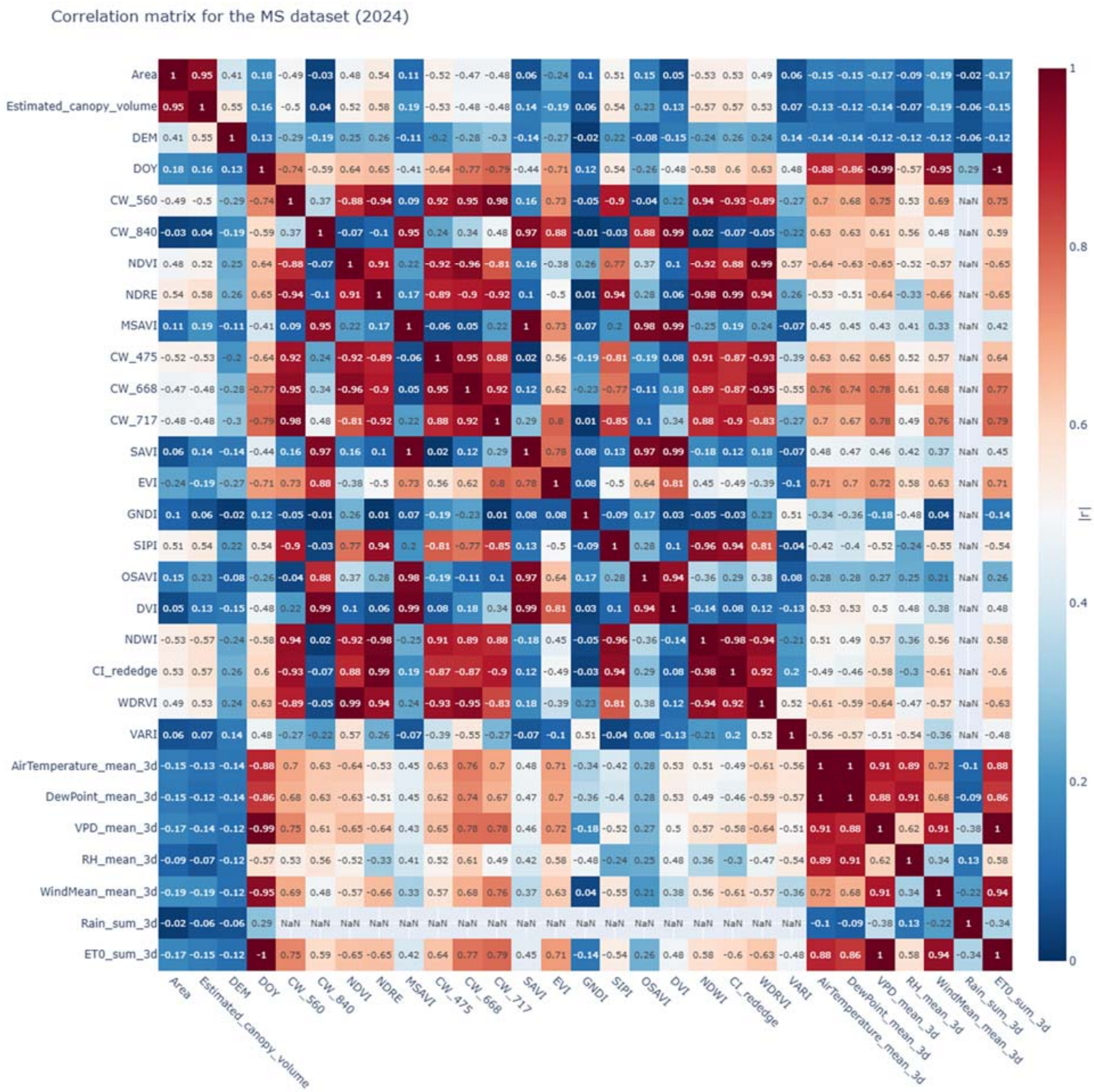


Figure 12. Correlation matrix for the MS-only dataset (2024)

The MS-only 2024 dataset displayed strong redundancy among vegetation indices and visible and near-infrared spectral bands, with several pairs exceeding |r| > 0.90. Some meteorological variables were also highly correlated, particularly temperature-related aggregates. As for the MS-only 2023-2024 dataset, the NaN values observed for Rain_sum_3d in the correlation matrix are due to its near-zero variance, since rainfall was zero for most observations. After filtering, 12 features were retained out of 29, preserving only the most independent spectral, structural, and short-term weather predictors for modelling.

MS+IRT 2024 dataset

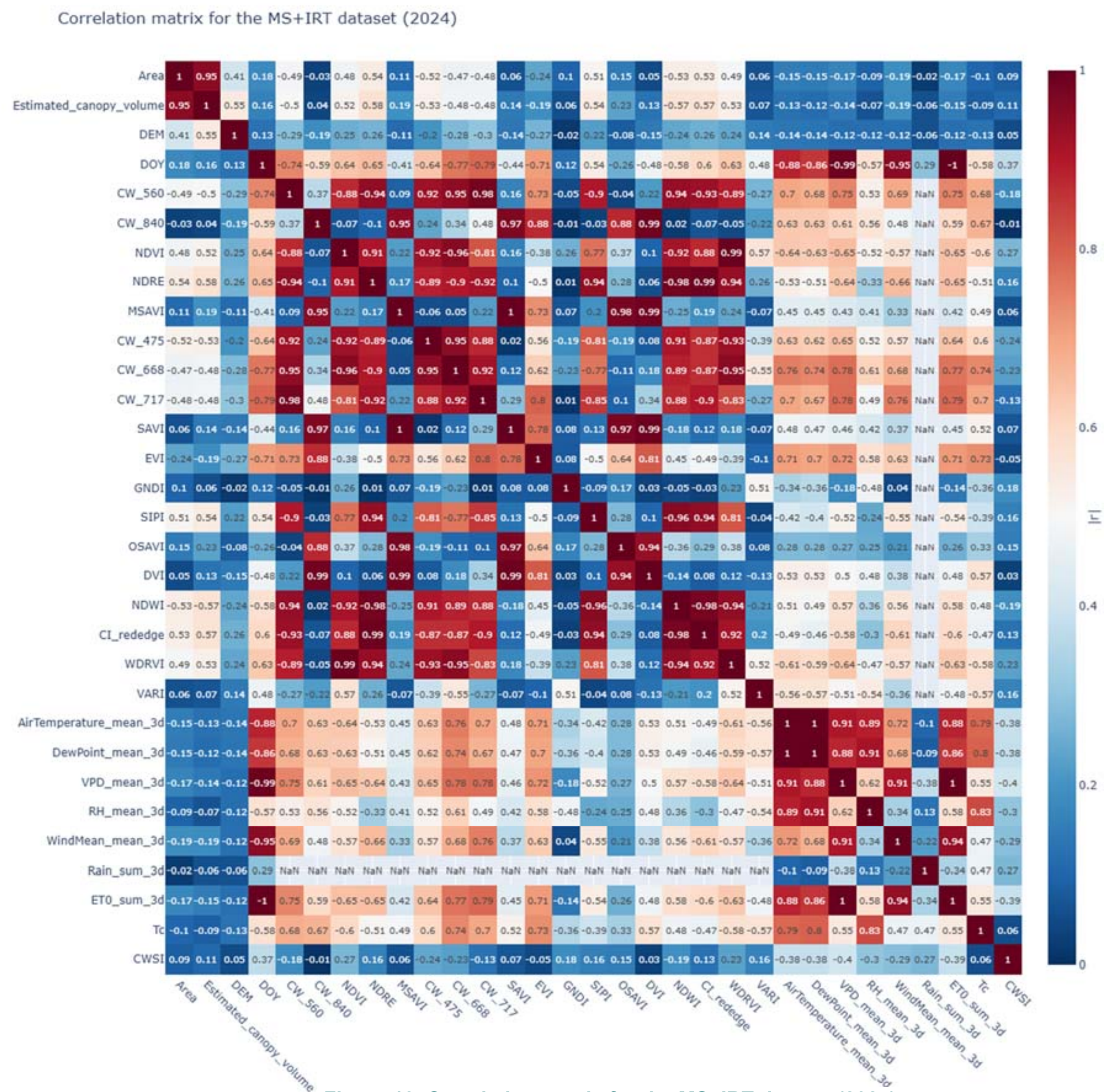


Figure 13. Correlation matrix for the MS+IRT dataset (2024)

Strong correlations were again observed among vegetation indices, visible and near-infrared spectral bands, and temperature-related weather variables. In contrast, the thermal predictors **Tc** and **CWSI** remained largely independent of both spectral and meteorological features, confirming their added value. After filtering redundant variables (**|r| > 0.90**), **14** features were retained, including both thermal indicators.

Machine learning modelling framework

To evaluate the ability of remote sensing and weather variables to predict SWP, a machine learning framework was implemented using the three complementary datasets: MS-only (2023–2024), MS-only (2024), and MS+IRT (2024). The modelling strategy was designed to:

- Ensure fair comparison between datasets
- Avoid information leakage due to repeated measurements on the same trees
- Have robust performance estimates despite the limited sample size.

Data splitting strategy: GroupKFold

Because multiple observations originate from the same trees, training and test samples cannot be treated as independent.
To correctly handle this structure, all evaluations were performed using **GroupKFold** with **TreeID** as the grouping variable. This guarantees that all measurements from a given tree are assigned exclusively to

either the training or the testing set, thereby preventing information leakage and ensuring that performance reflects the ability to generalize to new trees.

For each dataset, 3 outer folds were used, providing a good compromise between:

- sufficient data for training
- and enough independent trees in the test sets to obtain reliable performance estimates

Nested cross-validation for hyperparameter tuning

Within each training set of the outer GroupKFold, a nested cross-validation loop was applied to tune hyperparameters while preventing optimistic bias. The inner loop used GroupKFold (3 folds) restricted to the training subset. This nested approach ensures that:

- hyperparameters are selected independently from the test fold
- reported test metrics correspond to a true out-of-sample evaluation
- model complexity is controlled, and overfitting risks are reduced

Models evaluated

Four commonly used regression algorithms were tested:

- **Random Forest Regressor:** An ensemble method that averages predictions from multiple bootstrap-sampled decision trees, reducing variance and effectively capturing complex feature interactions while remaining robust to noisy and heterogeneous data.
- **Extra Trees Regressor:** A highly randomized tree ensemble where split thresholds are chosen at random, offering faster training and lower variance than Random Forest, often improving generalization on small or noisy datasets.
- **Gradient Boosting Regressor:** A sequential ensemble that builds trees to fit the gradient of the loss function at each step, enabling powerful modeling of complex nonlinear relationships but requiring careful regularization to avoid overfitting.
- **Support Vector Regressor (SVR):** A kernel-based margin method that constructs a regression function with maximal flatness while tolerating limited deviations, enabling flexible nonlinear modeling in high-dimensional feature spaces and performing particularly well on small to medium-sized datasets.
- **XGBoost Regressor:** An optimized gradient-boosting framework that uses regularized tree learning, efficient handling of missing values, and parallelized computation to achieve high predictive performance on structured tabular data.

All models were implemented within a unified scikit-learn pipeline, including:

- standardization of numerical variables
- one-hot encoding of categorical variables (Treatment)
- hyperparameter optimization via RandomizedSearchCV

Evaluation metrics

Model accuracy was assessed using:

- **Coefficient of determination (R^2):** Reflects the proportion of SWP variance explained by the model.
- **Root Mean Square Error (RMSE):** Quantifies the average prediction error in bar units.

For each model, metrics were reported for both:

- **training folds** (to inspect overfitting)
- **test folds** (true generalization performance)

Mean and standard deviation across folds were used as robust indicators of prediction stability.

Final model fitting and interpretation

For each dataset, the best-performing algorithm identified during cross-validation was refitted on the complete dataset using the optimal hyperparameters obtained from the nested CV procedure.

Model interpretation relied exclusively on SHAP values (SHapley Additive exPlanations).
SHAP provides a game-theoretic framework that quantifies the contribution of each predictor to individual model outputs, allowing both:

- **global interpretation** (which features the most influence on SWP overall),
- **local interpretation** (how each feature affects predictions for specific trees).

In this study, the focus is placed on **global SHAP importance**, which provides a clear and consistent understanding of the spectral, thermal, and meteorological variables that drive SWP prediction.

SWP prediction - Results

2023-2024 MS-only dataset

Model performance:

Table 9. Model performance comparison (MS-only 2023-2024)

Model	R^2 train mean	R^2 train std	R^2 test mean	R^2 test std	RMSE train mean	RMSE train std	RMSE test mean	RMSE test std	n_folds
ExtraTrees	**0.963**	**0.053**	**0.796**	**0.061**	**0.472**	**0.510**	**1.459**	**0.238**	**3**
RandomForest	0.934	0.021	0.767	0.085	0.833	0.134	1.554	0.303	3
XGBoost	0.986	0.010	0.763	0.080	0.0365	0.149	1.572	0.290	3
GradientBoosting	0.935	0.058	0.746	0.113	0.676	0.585	1.619	0.383	3
SVR	0.859	0.075	0.745	0.051	1.188	0.344	1.642	0.208	3

Across the MS-only 2023-2024 dataset, **ExtraTrees** achieved the strongest predictive performance, with the highest test **R^2 of 0.796** among all models. Overall, tree-based ensemble methods (ExtraTrees, Random Forest, XGBoost, Gradient Boosting) outperformed SVR on this multi-season (2023-2024) dataset.

Global feature importance (SHAP):

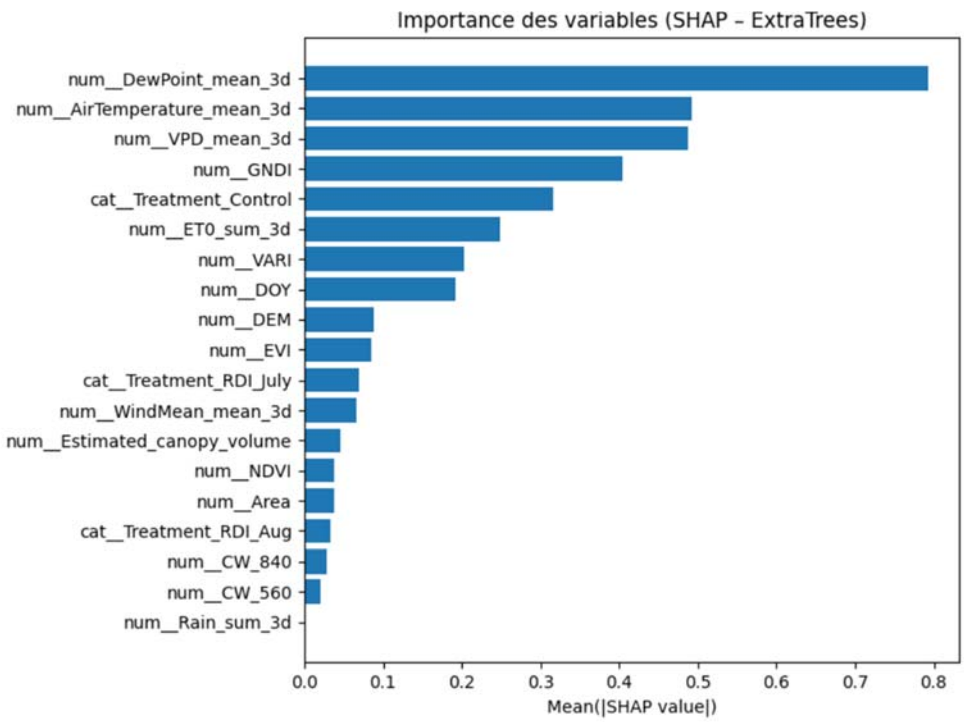


Figure 14. SHAP feature importance (ExtraTrees) on MS-only 2023-2024 dataset

SHAP analysis on the MS-only 2023-2024 dataset shows that **3-day averages** of **dew point**, **air temperature,** and **VPD** are the primary drivers of ExtraTrees predictions, while **treatment** effects (especially the **Control** group) and spectral indices (**GNDI**, **VARI**, **EVI**) play a secondary role, and the remaining structural and spectral bands contribute only marginally.

Prediction vs Observations:

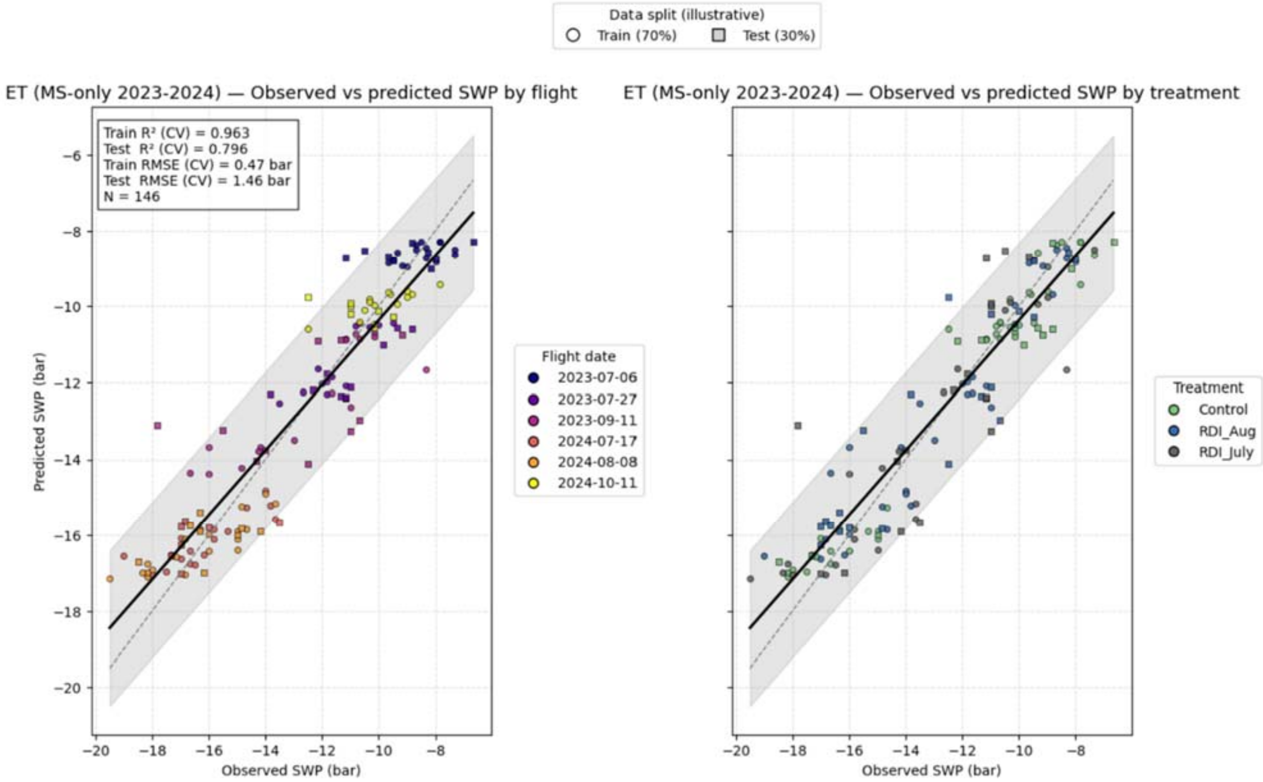


Figure 15. ExtraTrees (MS-only 2023-2024) - Observed vs predicted SWP using an illustrative 70/30 TreeID-based split

To visualize the model's behavior on the **MS-only 2023-2024** dataset, an illustrative 70/30 split stratified by TreeID was generated using the ExtraTrees hyperparameters obtained from the nested cross-validation. The resulting performance (**R^2 = 0.779**, **RMSE = 1.46 bar** on the test subset) remains closely aligned with the cross-validated test performance (**R^2 = 0.796**, **RMSE = 1.46 bar**), indicating that the model generalizes well to unseen trees. The separation between flight dates is clearly visible in the scatterplots, reflecting the seasonal trajectory of SWP across the two years.

2024 MS-only dataset

Model performance:

Table 10. Model performance comparison (MS-only 2024)

Model	R^2 train mean	R^2 train std	R^2 test mean	R^2 test std	RMSE train mean	RMSE train std	RMSE test mean	RMSE test std	n_folds
RandomForest	**0.913**	**0.044**	**0.798**	**0.080**	**0.893**	**0.246**	**1.377**	**0.290**	**3**
ExtraTrees	0.927	0.064	0.785	0.087	0.690	0.596	1.415	0.303	3
XGBoost	0.967	0.044	0.773	0.083	0.489	0.372	1.464	0.301	3
GradientBoosting	0.997	0.002	0.763	0.084	0.155	0.075	1.495	0.288	3
SVR	0.866	0.056	0.692	0.073	1.118	0.267	1.705	0.145	3

Across the **2024-MS-only** dataset, **Random Forest** provided the strongest predictive performance, yielding the highest **R^2=0.798** on the test folds and one of the lowest RMSE values. Tree-based ensemble models (RF, ET, XGBoost) outperformed Gradient Boosting and SVR, confirming their stability on this dataset.

Global feature importance (SHAP):

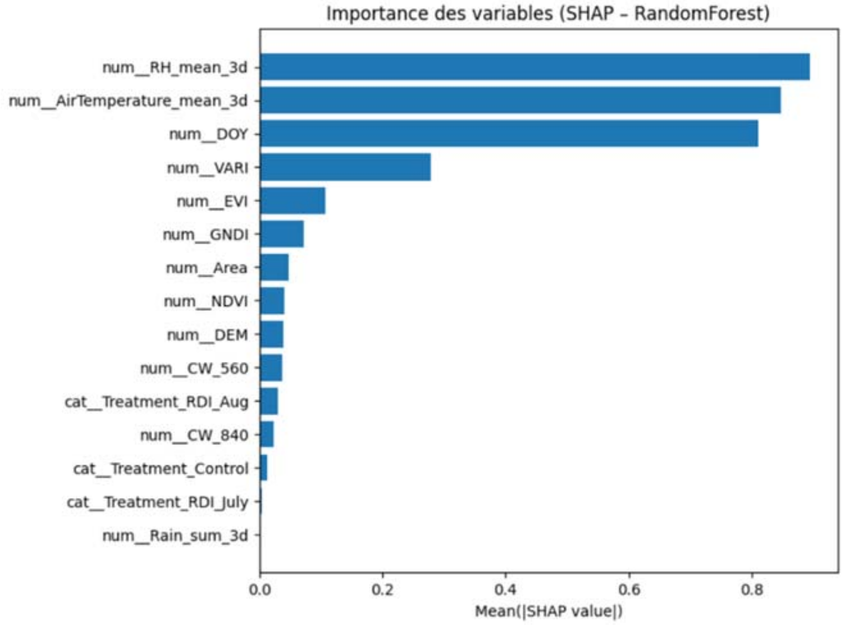


Figure 16. SHAP feature importance (Random Forest) on MS-only 2024 dataset

SHAP analysis shows that relative humidity, air temperature, and DOY (day of year) are the dominant drivers of the model’s predictions, while multispectral indices contribute only marginally, and treatment variables have very limited influence.

Prediction vs Observations:

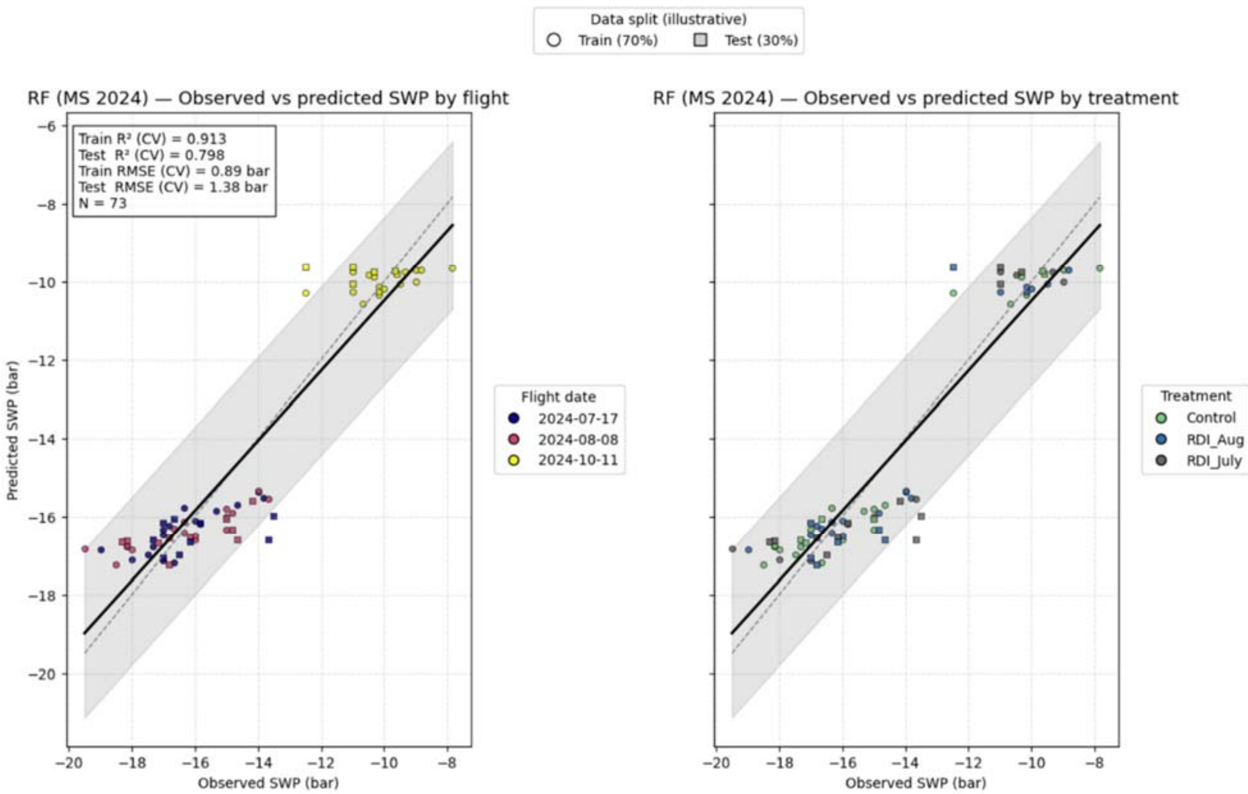


Figure 17. Random Forest (MS-only 2024) - Observed vs predicted SWP using an illustrative 70/30 TreeID-based split

Using the same illustrative 70/30 TreeID-stratified split procedure and the hyperparameters derived from the nested cross-validation. The RF model achieved R^2 = 0.752 and **RMSE = 1.34 bar** on the test subset, closely matching the cross-validated test performance (**R^2 = 0.798**, **RMSE = 1.38 bar**).
The scatterplots again reveal a clear separation between acquisition dates, consistent with the seasonal dynamics of SWP. The absence of the September multispectral acquisition is clearly visible. Although SWP measurements were available for that date, the missing MS features led to the removal of these

observations during preprocessing, reducing the usable dataset from 99 to 73 samples.

2024 MS+IRT dataset

Model performance:

Table 11. Model performance comparison (MS+IRT, 2024)

Model	R^2 train mean	R^2 train std	R^2 test mean	R^2 test std	RMSE train mean	RMSE train std	RMSE test mean	RMSE test std	n_folds
RandomForest	**0.906**	**0.021**	**0.810**	**0.069**	**0.948**	**0.105**	**1.335**	**0.261**	**3**
ExtraTrees	0.933	0.058	0.801	0.088	0.660	0.572	1.361	0.317	3
XGBoost	0.975	0.039	0.798	0.076	0.352	0.437	1.378	0.278	3
GradientBoosting	0.953	0.017	0.770	0.061	0.662	0.112	1.476	0.207	3
SVR	0.878	0.042	0.766	0.055	1.074	0.202	1.486	0.167	3

Across the MS+IRT 2024 dataset, Random Forest again achieved the best predictive performance, with the highest R^2 of 0.81 on the test folds among all models. Tree-based ensemble methods (RF, ET, XGBoost) consistently outperformed Gradient Boosting and SVR, confirming their suitability for small and nonlinear datasets.

Global feature importance (SHAP):

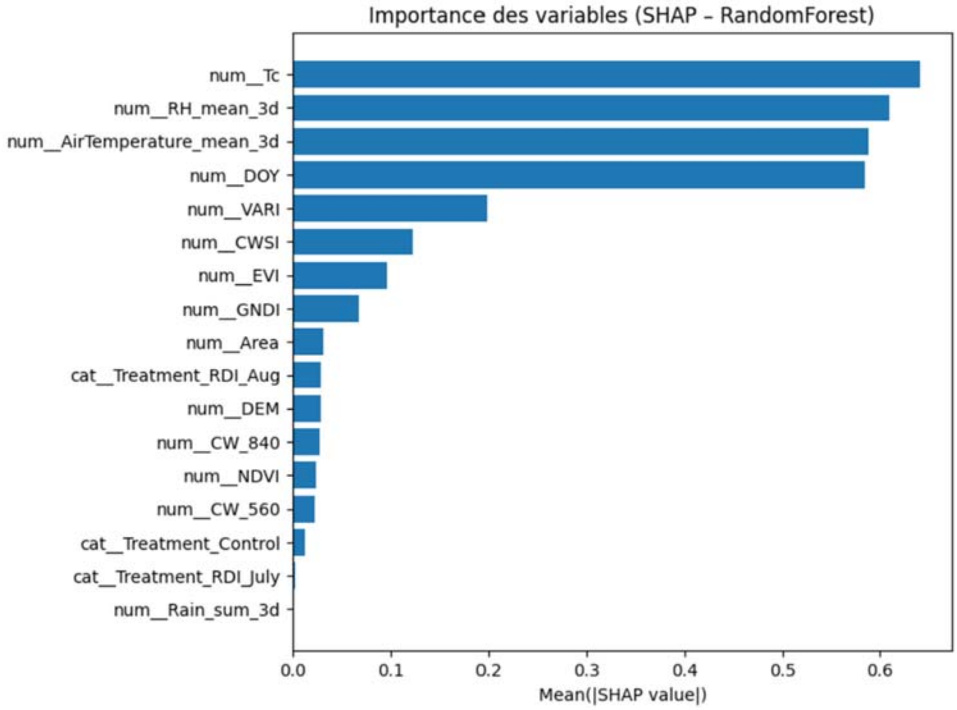


Figure 18. SHAP feature importance (Random Forest)

SHAP values indicate that SWP prediction is driven almost exclusively by short-term weather and thermal variables, especially **canopy temperature (Tc)**, **3-day humidity** and **air temperature**. **DOY** also contributes, whereas multispectral and structural features play only a minor role. This shows that weather-driven stress dominates SWP variation, with minimal contribution from spectral features.

Prediction vs Observations:

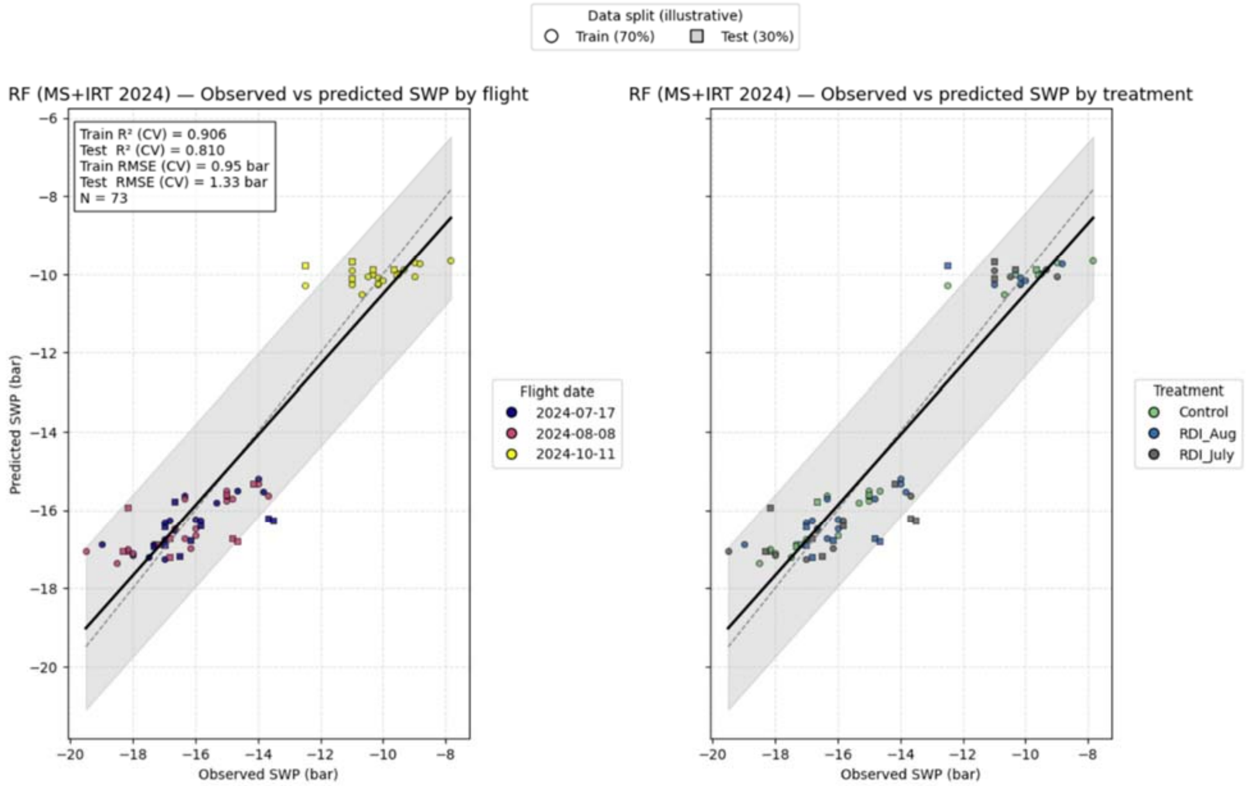


Figure 19. Random Forest (MS+IRT 2024) - Observed vs predicted SWP using an illustrative 70/30 TreeID-based split

Applying the same evaluation procedure as in the previous subsections, the model achieved test performance of R^2 = 0.749 and RMSE = 1.35 bar, which is consistent with the cross-validation results (R^2 = 0.81 ± 0.069). To ensure that the illustrative split was not atypical, 20 random 70/30 TreeID-based splits were evaluated. Test R^2 values ranged from 0.62 to 0.89, with a mean of 0.785 and a standard deviation of 0.065. Although the full range is slightly wider than the nested CV interval (0.81 ± 0.069), the mean and variability are highly consistent with the CV estimates. This confirms that the model's behavior under alternative train/test partitions remains stable and consistent with cross-validated performance.

Overall Summary of SWP Prediction Performance Across these Datasets (Cross-validation Results)

The comparison of the three datasets reflects the strong influence of acquisition protocols on SWP modelling. The MS-only 2023-2024 dataset achieved solid predictive performance (**R^2_test = 0.796**), despite the temporal mismatch between multispectral flights and SWP measurements in 2023. This indicates that multi-season multispectral data, combined with meteorological variables, can capture the main determinants of SWP even under imperfect measurement conditions. The MS-only 2024 dataset produced slightly higher performance (**R^2_test = 0.798**), consistent with its smaller sample size and the September missing multispectral acquisition, which reduced the usable dataset. However, the close temporal alignment between UAV imagery and SWP measurements enables reliable single-season predictions based solely on multispectral inputs.

The MS+IRT 2024 dataset achieved the highest accuracy (**R^2_test = 0.81**), with improved stability and reduced variability across validation folds. The addition of thermal infrared features - directly linked to canopy temperature and transpiration - provides complementary physiological information that strengthens model robustness, even though the overall gain in predictive accuracy compared to MS-only 2024 remains modest. SHAP analyses across all datasets consistently showed that meteorological variables dominate SWP prediction, while thermal information improves model interpretability and reduces prediction variance, and multispectral indices contribute more modestly.

Note: These results also indicate that SWP is primarily spatially predictable: models trained on a subset of trees within a given flight date generalize well to other trees from the same day. However, when training and testing are separated by flight date, and a GroupKFold is used based on Flight date, predictive performance collapses, in some cases yielding negative R^2 values. This suggests that the current datasets do not capture sufficient temporal variability to enable robust cross-date generalization, possibly due to the limited number of flight dates, the restricted range of environmental conditions sampled, and the strong day-to-day sensitivity of SWP to short-term weather dynamics.

Yield Prediction:

Data Context and Data Construction:

Yield in the experimental plot was measured over two consecutive seasons, 2023 and 2024. Each season included two harvest rounds (mid-November and mid-December). After data cleaning checks, the final dataset available for modelling contained:

- **2023**: 55 individual trees
- **2024**: 57 individual trees
- Combined dataset (**GLOBAL**): 112 trees

To evaluate the effect of remote sensing data availability on yield prediction performance, two seasonal information scenarios were considered for each year. These scenarios were defined by the timing of the last UAV acquisition, either in October or in November. For each scenario, remote sensing features were

Site	Study year	cutoff_oct	cutoff_nov	harvest_round1	harvest_round2
San Giuliano	2023	2023-10-12	2023-11-08	2023-11-15	2023-12-11
San Giuliano	2024	2024-10-11	2024-11-15	2024-11-26	2024-12-17

Figure 20. Dates of October and November UAV cut-offs

recalculated using only data available up to the corresponding date, enabling a direct comparison between an earlier (October) and a later (November) prediction context.

Remote sensing feature engineering

Remote sensing features were derived from multispectral UAV acquisitions collected during each growing season. For each tree and cut-off date (October or November), features were computed using only UAV flights available up to that date, ensuring temporal consistency. In addition to basic seasonal descriptors (mean, minimum, and maximum), seasonal slopes were calculated to capture temporal dynamics in vegetation indices and structural variables. Slopes were estimated by linear regression of variable values on acquisition dates, yielding a compact representation of canopy evolution over time. The extracted remote sensing variables included:

- **Vegetation indices** available in the multispectral dataset (NDVI, NDRE, GNDI, SIPI).
- **Structural metrics** such as **canopy area**, **estimated canopy volume**, and **DEM**.

Together, these features describe both the canopy condition and its seasonal evolution up to October or November, enabling a direct comparison between early- and late-season remote sensing data for yield prediction.

Exploratory correlations between canopy metrics and yield

October Cut-off:

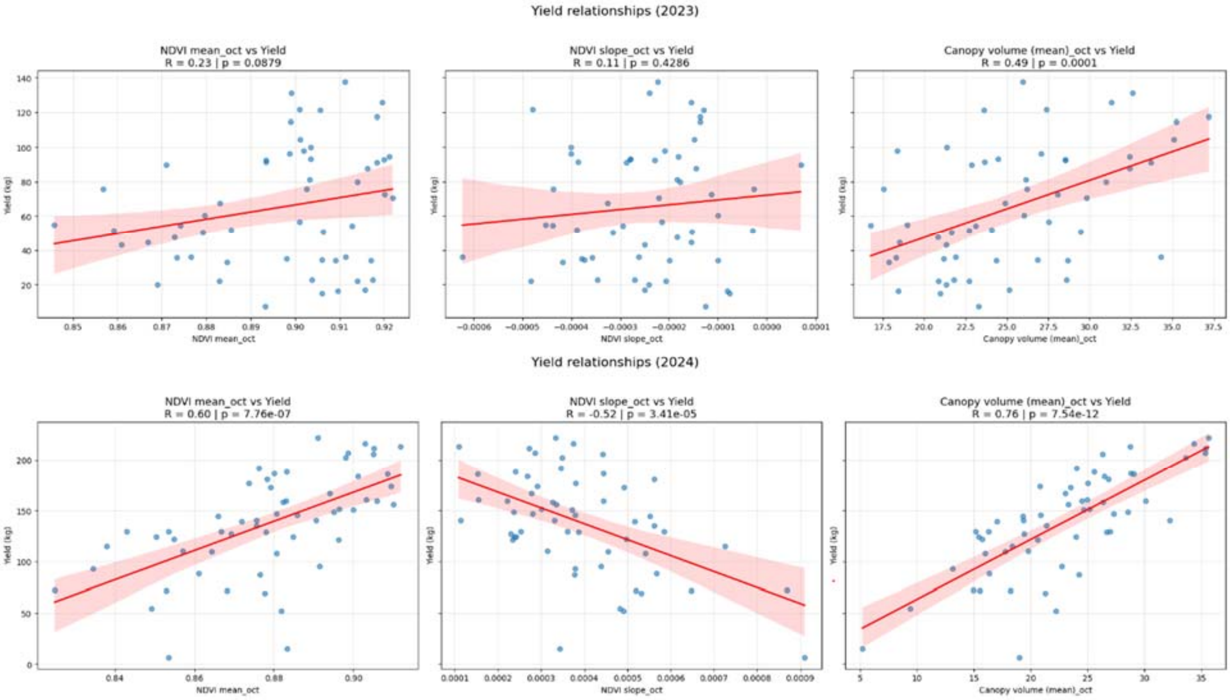


Figure 21. Correlations between canopy traits and yield (2023–2024) – October Cut-off

In 2023, correlations between *October* remote sensing variables and yield were generally weak. NDVI_mean and NDVI_slope showed low, non-significant relationships with yield, whereas canopy volume already exhibited a moderate, significant positive correlation (R = 0.49, p-value = 0.0001).

In 2024, clear relationships were observed at the October cut-off: yield was positively correlated with NDVI_mean (**R = 0.60**, **p-value = 7.76e-07**) and canopy volume (**R = 0.76**, **p-value = 7.54e-12**), and negatively correlated with NDVI_slope (**R = -0.52**, **p-value = 3.41e-05**). Among these variables, canopy volume showed the strongest correlation with yield.

November Cut-off:

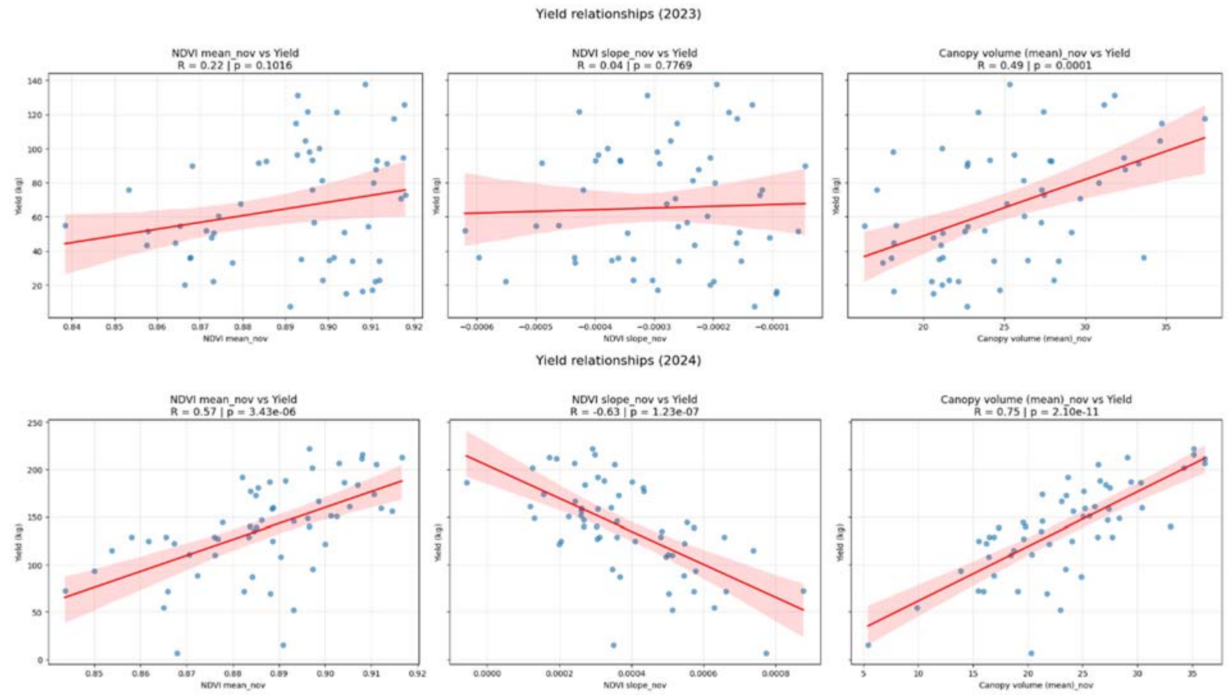


Figure 22. Correlations between canopy traits and yield (2023–2024) – November Cut-off

In 2023, including November data, vegetation-yield relationships did not substantially strengthen. NDVI_mean and NDVI_slope remained weakly and non-significantly correlated with yield (**R = 0.22** and R = 0.04, respectively), while canopy volume continued to show a stable and significant positive association

(**R = 0.49**, **p-value = 0.0001**).

In November 2024, correlations were further reinforced. NDVI_mean and canopy volume showed strong positive relationships with yield (**R = 0.57** and **R = 0.75,** respectively), while NDVI_slope remained strongly negative (**R = -0.63**), indicating that high-yield trees tended to maintain high canopy vigor earlier rather than increasing NDVI late in the season.

Note: Importantly, the negative NDVI_slope relationship suggests that trees whose NDVI increases toward the end of the season tend to produce lower yields, while trees whose NDVI stays stable or slightly decreases show much higher yields. This indicates that high-yield trees generally start the season with strong canopy vigor rather than gaining vigor late in the cycle.

These observations should be confirmed with the 2025 dataset to verify whether this pattern is consistent across years.

Meteorological and phenological features

Weather variables were aggregated over the productive season, defined as 1st April through the selected UAV cut-off date (October or November). This time window captures the climatic conditions influencing fruit development up to the point at which remote sensing data become available, ensuring temporal consistency between meteorological predictors and UAV-derived features.

The extracted descriptors included:

- **Seasonal weather indicators:** such as mean VPD, cumulative ET_0 and Rain_sum, calculated up to the corresponding cut-off date.
- **Thermal development:** GDD_sum_year and GDD_sum_last60d* **(same preceding each cut-off)**
- **Flowering window (15 April - 20 May):** Flowering_Rain_sum, Flowering_Tair_mean, Flowering_RH_mean, Flowering_VPD_mean

This flowering window was included because weather conditions during this period directly affect fruit set and early physiological drop, which can strongly influence final yield.

***GDD: Growing Degree Days (explained in the SWP section below).**

Final engineered dataset and feature selection

After merging structural, spectral, climatic, and yield information, the final dataset included 112 trees and 276 engineered features. Because this feature space was large and partially redundant, a targeted feature selection was applied to retain only variables with clear physiological relevance.
A reduced subset of 18 predictors was finally selected for each cut-off, capturing canopy structure, spectral changes, flowering conditions, and seasonal climate:

- **Canopy structure and vigor:** Area_mean, Estimated_canopy_volume, Estimatd_canopy_volume_slope, GNDI_slope, NDRE_slope, NDVI_slope, SIPI_slope
- **Seasonal climate and flowering conditions:** Daily Thermal Amplitude_mean_year, VPD_mean_year, ET0_sum_year, Rain_sum_year, Flowering_Rain_sum, Flowering_VPD_mean, Flowering_Tair_mean
- **Thermal development:** GDD_sum_year, GDD_sum_last60d
- **Other features:** Treatment, DEM_mean

This reduced set provides a clear, practical summary of how tree growth, stress, and weather conditions relate to final fruit yield.

Machine Learning Modelling Framework

The modelling workflow followed the same structure as the SWP prediction framework. Three dataset configurations were evaluated to assess robustness across years:

- a **GLOBAL** model using both seasons (**112 trees**)
- a **2023-only** model (**55 trees**)
- and a **2024-only** model (**57 trees**)

As in the SWP analysis, data splitting was performed using GroupKFold on TreeID, ensuring that each tree did not appear in both the training and test sets. A 3-fold cross-validation scheme was used in all scenarios. To obtain unbiased performance estimates, a nested cross-validation procedure was applied: an inner loop for hyperparameter optimization (Randomized Search) and an outer loop for model evaluation on unseen trees. Performance was assessed using R^2 and RMSE (kg/tree), providing complementary views of accuracy and prediction error. Four ensemble learning algorithms were tested under identical conditions: ExtraTrees, Random Forest, Gradient Boosting, and XGBoost. These models are well-suited for non-linear relationships and heterogeneous predictors. Unlike the SWP workflow, SVR was not included because preliminary tests indicated that tree-based models performed better.

Yield prediction – Results

Data cleaning and outlier removal:

The same IQR-based outlier-detection approach was applied to the data. Yield values (or "Total Weight" values in our dataset) were grouped by study year, and observations lying below Q1 - 1.2 × IQR or above Q3 + 1.2 × IQR were identified as outliers:

Exemples d'outliers détectés :

	_row_id	Country	Site	TreeID	Treatment	Study year	Total Weight	lower_bound	upper_bound
109	109	France	San Giuliano	FRCitrus1_L07	Control	2024	6.772	34.753	249.786
62	62	France	San Giuliano	FRCitrus1_C07	RDI_Aug	2024	15.207	34.753	249.786

Figure 23. Detected yield outliers

No yield outliers were detected for the 2023 season. In contrast, two individual trees were identified as yield outliers in 2024 using the IQR method. These two observations were **removed** from the dataset before model training.

October Cut-off

Model performance:

Table 12. Comparison of Machine Learning Models for Yield Prediction – October cut-off

Model	Data set	R^2 train (mean)	R^2 train (std)	R^2 test (mean)	R^2 test (std)	RMSE train (mean)	RMSE train (std)	RMSE test (mean)	RMSE test (std)	n_folds	n_samples
ExtraTrees	GLOBAL	0.976	0.041	**0.724**	0.011	4.969	8.607	28.811	0.822	3.000	110.000
ExtraTrees	2023	1.000	0.000	0.229	0.515	0.000	0.000	29.262	10.585	3.000	55.000
ExtraTrees	2024	1.000	0.000	0.589	0.090	0.000	0.000	27.435	0.819	3.000	55.000
RandomForest	GLOBAL	0.934	0.018	0.704	0.025	14.036	2.333	29.823	1.308	3.000	110.000
RandomForest	2023	0.839	0.112	0.269	0.408	13.250	5.052	28.742	8.853	3.000	55.000
RandomForest	2024	0.929	0.009	0.533	0.099	11.537	0.461	29.239	0.887	3.000	55.000
GradientBoosting	GLOBAL	0.994	0.005	0.647	0.078	3.949	2.174	32.412	2.697	3.000	110.000
GradientBoosting	2023	0.998	0.002	0.106	0.373	1.082	1.137	32.036	7.721	3.000	55.000
GradientBoosting	2024	0.996	0.004	0.574	0.123	2.328	2.074	27.788	2.111	3.000	55.000

XGBoost	GLOBAL	0.999	0.001	0.638	0.045	1.317	0.655	32.933	1.126	3.000	110.000
XGBoost	2023	0.970	0.052	0.271	0.456	3.714	5.791	28.581	9.695	3.000	55.000
XGBoost	2024	1.000	0.001	0.541	0.114	0.746	0.580	28.958	2.958	3.000	55.000

Overall, ExtraTrees provided the best and most stable performance for the October cut-off, especially on the GLOBAL dataset (**R^2_test = 0.724**), followed closely by Random Forest. Models trained on 2023-only data showed weak predictive skill, whereas 2024-only models achieved markedly higher accuracy, reflecting stronger vegetation-yield relationships that year. The consistently higher training than test R^2 across all models highlights the difficulty of yield prediction with limited sample sizes. Finally, the GLOBAL dataset produced the most robust predictions, confirming the benefit of combining multiple seasons for model generalization.

Global features of importance (SHAP):

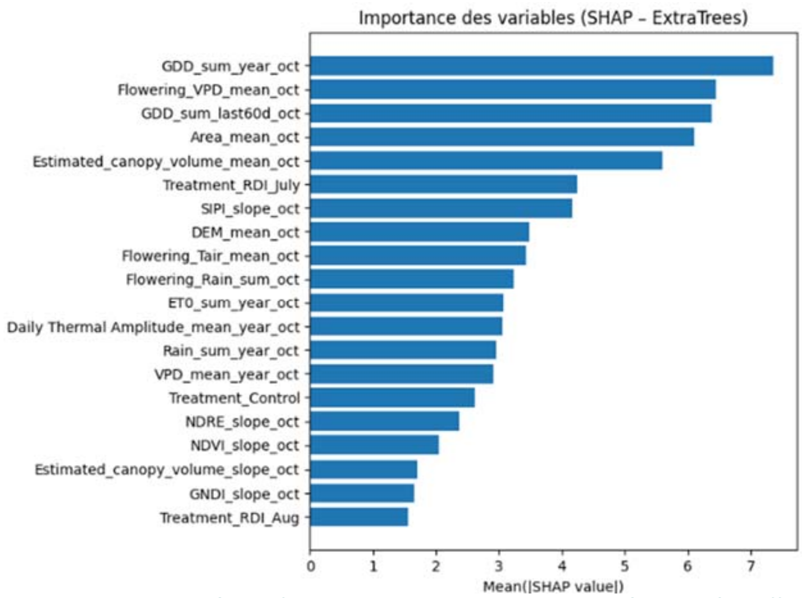


Figure 24. SHAP feature importance (ExtraTrees) – October Cut-off

SHAP analysis indicates that thermal accumulation (GDD) is the main driver of yield prediction at the October cut-off. Canopy structure variables, particularly canopy area and volume, remain major contributors, highlighting the importance of tree size. Atmospheric demand during flowering (VPD) also plays a key role, while flowering-period weather conditions and spectral indices provide complementary information on canopy vigor. Overall, yield prediction is primarily driven by early-season climate and canopy structure.

Prediction vs Observations:

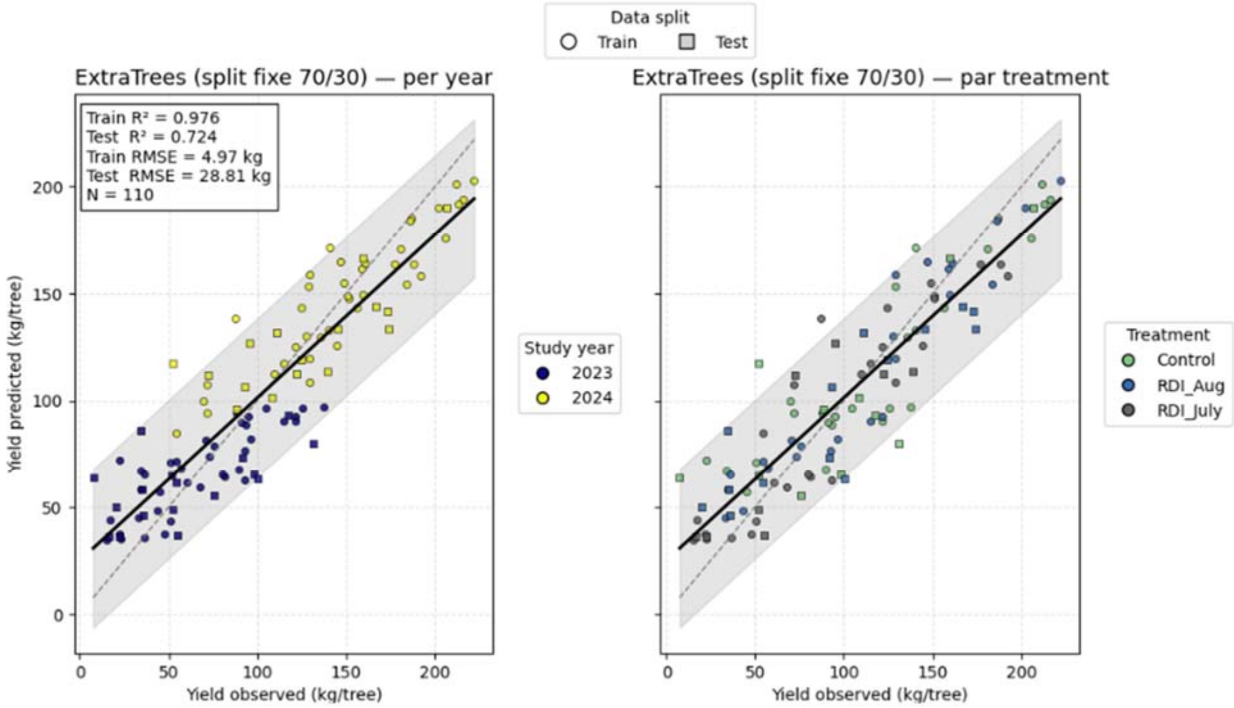


Figure 25. ExtraTrees - Observed vs predicted Yield using an illustrative 70/30 TreeID-based split – October cut-off

As we did in the SWP subsection, a fixed 70/30 train-test split stratified by TreeID was applied using the ExtraTrees model. On this split, the model achieved a test performance of **R^2 = 0.693** and **RMSE = 28.49 kg/tree**, slightly lower than the cross-validated performance for the October cut-off on the GLOBAL dataset (**R^2 = 0.724 ± 0.011, RMSE = 28.81 kg/tree**). This difference is expected given the limited sample size and the sensitivity of tree-based models to the specific composition of the test set in a single split. Cross-validation, therefore, provides a more robust estimate of model performance, while the fixed split remains useful for illustrative purposes. Overall, the results indicate reasonable generalization ability, albeit with noticeable prediction uncertainty, particularly for intermediate yield levels (100-150 kg/tree).

November Cut-off

Model performance:

Table 13. Comparison of Machine Learning Models for Yield Prediction – November cut-off

Model	Data set	R² train (mean)	R² train (std)	R² test (mean)	R² test (std)	RMSE train (mean)	RMSE train (std)	RMSE test (mean)	RMSE test (std)	n_folds	n_samples
ExtraTrees	GLOBAL	0.977	0.040	**0.742**	0.016	4.864	8.425	27.844	1.698	3.000	110.000
ExtraTrees	2023	0.869	0.122	0.191	0.502	10.185	8.987	30.116	10.152	3.000	55.000
ExtraTrees	2024	1.000	0.000	0.652	0.073	0.000	0.000	25.236	1.399	3.000	55.000
RandomForest	GLOBAL	0.958	0.003	0.732	0.048	11.336	0.581	28.332	3.342	3.000	110.000
RandomForest	2023	0.872	0.071	0.191	0.362	12.031	3.655	30.456	7.715	3.000	55.000
RandomForest	2024	0.943	0.008	0.594	0.106	10.312	0.496	27.218	3.247	3.000	55.000
GradientBoosting	GLOBAL	0.990	0.010	0.715	0.084	4.867	3.339	29.170	5.307	3.000	110.000
GradientBoosting	2023	0.766	0.085	0.127	0.299	16.506	3.301	31.829	6.363	3.000	55.000
GradientBoosting	2024	0.997	0.001	0.588	0.167	2.454	0.285	27.068	3.472	3.000	55.000
XGBoost	GLOBAL	0.985	0.024	0.723	0.076	5.145	5.559	28.691	4.741	3.000	110.000
XGBoost	2023	0.968	0.055	0.204	0.400	3.838	6.012	30.106	8.471	3.000	55.000
XGBoost	2024	1.000	0.000	0.574	0.121	0.350	0.226	27.806	3.263	3.000	55.000

Overall, ExtraTrees also provided the best and most stable performance for the November cut-off, particularly on the GLOBAL dataset (**R^2_test = 0.742**), followed closely by Random Forest. As observed for October, models trained on 2023-only data exhibited weak predictive skill, whereas models trained on 2024-only data achieved substantially higher accuracy, reflecting stronger and more consistent vegetation-yield relationships in that season. Training R^2 values remained systematically higher than test R^2 across all models, underscoring the challenges of yield prediction with limited sample sizes. Finally, the GLOBAL dataset again yielded the most robust predictions, confirming the benefit of combining multiple seasons for improved model generalization.

Global features of importance (SHAP):

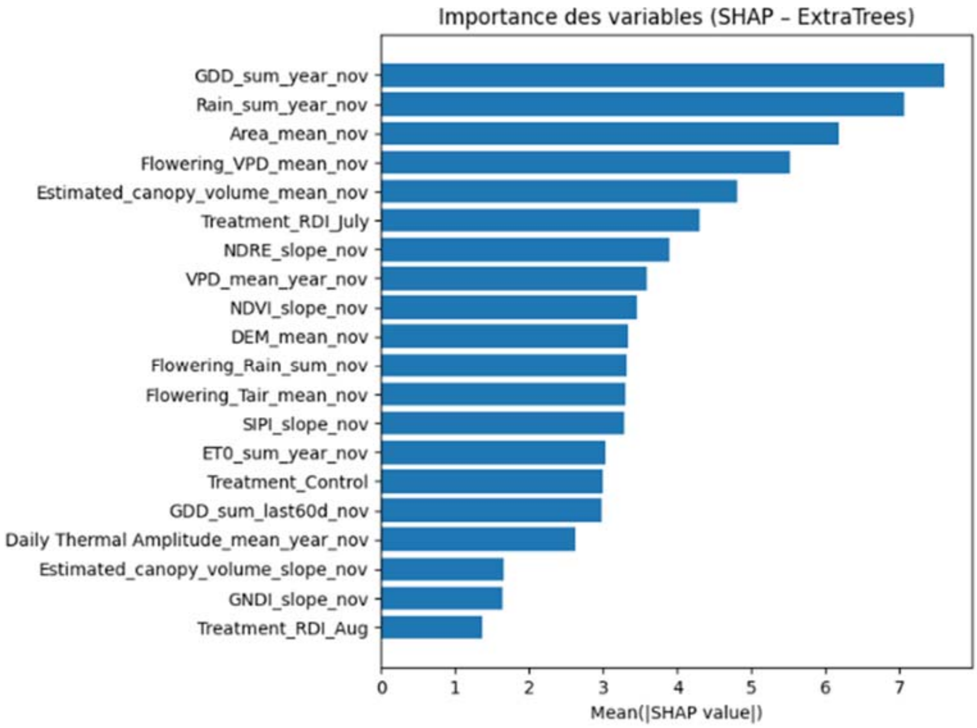


Figure 26. SHAP feature importance (ExtraTrees) – November Cut-off

SHAP analysis for the November cut-off shows a variable importance pattern largely consistent with that observed in October. Thermal accumulation remains a key driver of yield prediction, together with canopy structure variables such as canopy area and volume. Atmospheric demand during flowering also continues to contribute to model predictions, while additional weather and spectral variables provide complementary information. Overall, yield prediction in November is still primarily driven by seasonal climate conditions and canopy structure, with no major shift in the underlying drivers compared to the October scenario.

Prediction vs Observations:

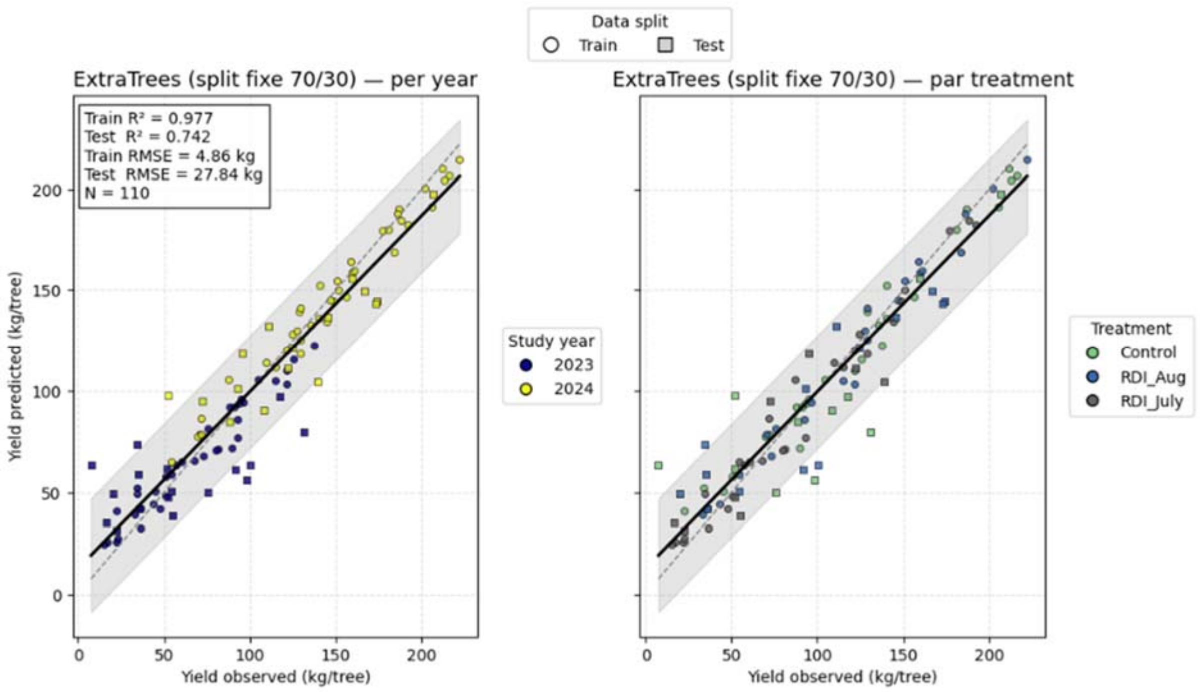


Figure 27. ExtraTrees - Observed vs predicted Yield using an illustrative 70/30 TreeID-based split – November cut-off

Following the same approach, a fixed 70/30 train-test split stratified by TreeID was applied using the ExtraTrees model for the November scenario. On this split, the model achieved a test performance of **R^2 = 0.746** and **RMSE = 25.94 kg/tree**, closely matching the cross-validated performance obtained on the GLOBAL dataset (**R^2 = 0.742 ± 0.016**, **RMSE = 27.84 kg/tree**). This stronger agreement between fixed-split and cross-validated metrics suggests a more stable model behavior in November, likely reflecting the availability of additional late-season information. Overall, observed and predicted yields showed good agreement across years and irrigation treatments, although prediction uncertainty remained noticeable, particularly for intermediate yield levels.

In summary:

- The ExtraTrees model achieves acceptable and robust predictive performance, with consistently higher accuracy for the 2024 season than for 2023, regardless of the cut-off date.
- Model performance is slightly higher for the November cut-off than for October (R^2 = 0.74 vs 0.72), indicating a modest but consistent gain from late-season remote sensing and weather information. However, the October scenario already provides reliable yield predictions, demonstrating that yield can be reasonably predicted approximately one month earlier.
- Some degree of overfitting is observed across all scenarios, which is expected given the model complexity and the limited sample size.
- Future improvements should focus on expanding the training dataset with additional seasons (notably 2025), refining temporal remote sensing descriptors, and exploring simpler or more regularized modelling approaches to improve generalization.

Cracking Prediction:

Dataset construction & feature engineering

Remote sensing & canopy structure

Remote sensing flights from the 2023-2024 seasons were aggregated at the tree level. Only the flights acquired before the cracking evaluation period (**July**, **August,** and **September**) were used to build predictors. This ensures that the models rely on pre-cracking canopy signals rather than on late-season responses that may already reflect the consequences of cracking. For each spectral narrow-band channel: (CW_475, CW_560, CW_668, CW_717, CW_842) and for each vegetation index: (NDVI, SIPI,

GNDI, CI, IPVI, DVI, SAVI, MSAVI, OSAVI, EVI, NDRE), two descriptors were extracted:

- the pre-September seasonal mean
- a temporal slope reflecting how the canopy evolves before cracking begins.

These slopes (NDVI_slope, SIPI_slope, CW_560_slope...) capture the dynamics of the canopy before the cracking period, such as loss of vigor or progressive stress.

In addition, structural variables were included:

- Area_mean
- Estimated_canopy_volume_mean
- DEM_mean

Weather features

Meteorological variables were aggregated at the **Study year** level. Four periods were defined to match the actual cracking assessment schedule in the orchard:

- **P1** - the first cracking evaluation: 7-day window preceding the first field scoring date (21 September 2023, 7 October 2024).
- **P2** - the second cracking evaluation: 7-day window preceding the second scoring date (approximated as 2 weeks after P1 when the exact date was not recorded).
- **P3** - early harvest: 7-day window preceding the early harvest date.
- **P4** - late harvest: 7-day window preceding the late harvest date.

For each period and for each variable (**air temperature**, **VPD**, **relative humidity**, **rain**, **thermal amplitude**, **ET0**, **GDD**), three descriptors were computed:

- the mean
- the sum (for rain, ET0, and GDD)
- a temporal slope, defined as the linear trend of the variable over the corresponding 7-day window.

These features jointly capture the climatic background and short-term dynamics experienced by the trees immediately before each cracking evaluation.

Heat-shock risk features (cracking-specific)

To inject agronomic knowledge directly into the model, a dedicated module computed **heat-shock risk indicators** based on the working hypothesis:

prolonged heat or dryness → sudden heavy rainfall → fruit cracking

Daily meteorological data were first aggregated at the site level and a rolling window of 4 days was applied to capture sustained stress conditions, defined through:

- **T_mean4**: 4-day mean air temperature
- **VPD_mean4**: 4-day mean vapor pressure deficit
- **deficit_mean4** = ET0 − Rain: 4-day mean climatic water deficit

Stress conditions were defined as:

- **heat_risk**: T_mean4 > 25 °C or VPD_mean4 > 0.8
- **dry_risk**: deficit_mean4 > 2 mm

A day was classified as at risk when either heat or dryness stress was detected.
To account for delayed physiological responses, heavy rainfall events (≥ 20 mm) were considered rainfall alerts when occurring within a short lag period (up to 3 days) following a stress day. From these daily time series, the following descriptors were extracted per site and year over the critical window (late

August to late October):

- **nb_days_risk:** total number of stress days
- **risk_streak_max:** longest consecutive sequence of stress days
- **nb_rain_alerts:** number of heavy rainfall events occurring shortly after a stress period
- **max_T_mean4, max_VPD_mean4, max_deficit_mean4:** maximum intensity of thermal and water stress during the season
- **mean_T_risk, mean_VPD_risk, mean_deficit_risk:** average climatic conditions during stress days
- **rain_sum_lag3d_max**: maximum cumulative rainfall occurring in the days following a stress episode, used as a proxy for rainfall shock intensity

Together, these variables encode stress–rainfall shock patterns known to favor fruit cracking: trees exposed to prolonged heat or water deficit followed by intense rainfall events are hypothesized to be at higher risk of cracking.

The plots below illustrate the meteorological conditions observed during the critical cracking window (from late August to late October), including:

- daily air temperature and VPD
- rolling 4-day averages used to detect sustained heat or water stress conditions
- periods classified as "at risk" (highlighted in red)
- and heavy rainfall events occurring shortly after a stress day (rainfall alerts).

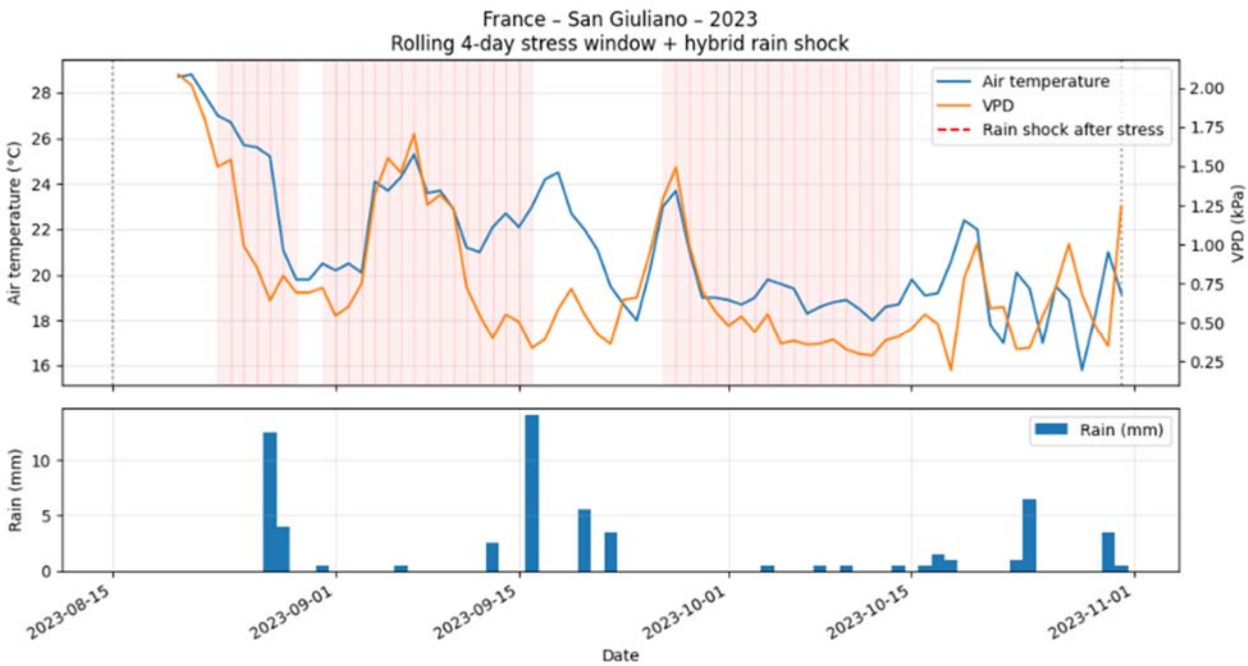


Figure 28. Stress-rainfall cracking pattern for 2023

In 2023, **no rainfall alerts** were detected, as precipitation during the stress period did not exceed the 20 mm threshold required to define a rain shock.

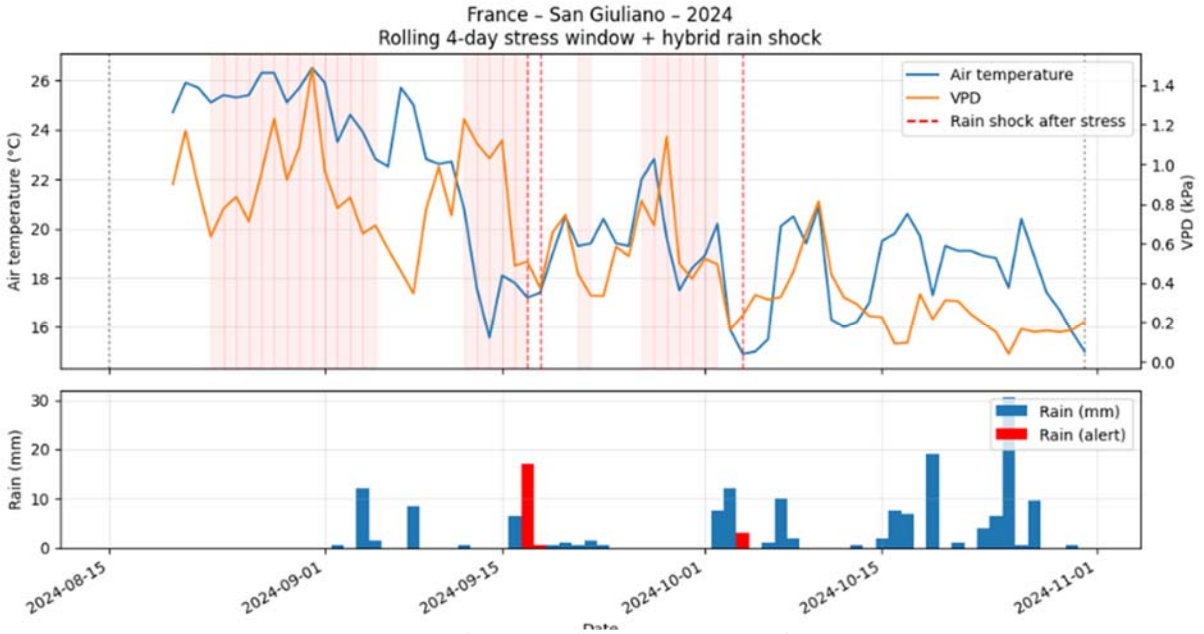


Figure 29. Stress-rainfall cracking pattern for 2024

In contrast to 2023, several rainfall alerts were detected in 2024 (17-18 September and 4 October). These events correspond to heavy rainfall occurring shortly after sustained heat or dryness stress, suggesting rainfall-stress sequences that may be conducive to fruit cracking.

From regression to classification

After merging all components, the remote sensing features (spectral bands and vegetation indices, expressed as mean and temporal slope), the canopy structure metrics, the site-level weather summaries for periods P1 to P4, and the heat-shock risk descriptors, a complete cracking dataset was obtained. It contains 112 trees from the 2023 and 2024 seasons and approximately 96 predictors before any filtering. Highly collinear variables were then removed using a correlation-based pruning step, while all heat-shock features were retained by construction.

Initial regression trial

The first modelling strategy approached cracking as a continuous outcome using regression. The objective was to predict the splitting percentage for each tree directly. However, even after testing a wide range of tree-based regressors and tuning strategies, the models continued to return negative test R^2 values and showed unstable behavior across cross-validation splits. These outcomes point to structural limitations of the dataset: the sample size is very small (112 trees), and cracking is highly variable, strongly influenced by rainfall event timing and fine-scale environmental differences.

Conclusion: With the available data, the continuous splitting percentage is not reliably predictable.

Reformulation into a 3-class problem

To stabilize the task, the continuous Splitting variable was discretized into three-level categories:

- **low:** the lowest third of the splitting percentage
- **mid:** intermediate third
- **high:** highest third

This three-class setup aligns better with practical field categories (low, moderate or high cracking) and proved more learnable than the direct regression: models started to capture reproducible patterns, although performance remained moderate.

Alternative formulation: binary “extremes-only” classification

To further reduce label noise and focus on the clearest extremes, an alternative binary formulation was evaluated in parallel.

The same percentile-based cuts were used, but only extreme groups were kept:

- trees in the 33%: class **low**
- trees in the 33%: class **high**
- trees in the 33%: **excluded from modelling**

By restricting the analysis to the two opposite ends of the splitting distribution, the classification task becomes more coherent and better aligned with the contrast of interest (very low and very high cracking levels). The middle group is difficult to interpret because it mixes trees with mild and moderate symptoms. It also tends to overlap with both ends of the distribution. Removing this intermediate class reduces label noise and increases the separation between the two remaining classes.

On this reduced dataset:

- 76 trees (38 low, 38 high)
- perfectly balanced classes
- same features and same nested GroupKFold pipeline

models' performances increased markedly (see Section 3.3).

Machine learning modelling

Models evaluated

Tree-based models were selected as the core modelling family because they handle non-linear relationships well and are robust on small datasets. The models tested were **ExtraTrees**, **Random Forest**, **Gradient Boosting**, and **XGBoost**. Two linear baselines were also included: **Multinomial Logistic Regression** (or **Logistic Regression** in the **binary** case) and **Linear SVC**, to provide simpler reference models. All algorithms were run through the same preprocessing pipeline to ensure comparable results.

ML pipeline and cross-validation strategy

As in the previous SWP/Yield modelling framework, but adapted for a classification task, all models relied on the same preprocessing pipeline: numerical variables were standardized with StandardScaler, categorical variables were encoded using OneHotEncoder, and both were combined. A grouped cross-validation strategy was applied to avoid data leakage: GroupKFold with TreeID. Model evaluation followed a nested validation design. The outer loop (3-fold GroupKFold) provided unbiased performance estimates, while the inner loop performed hyperparameter tuning with RandomizedSearchCV using another 3-fold GroupKFold. Hyperparameters were optimized using the macro-averaged F1-score, chosen for its balanced treatment of all classes.

F1-score.
The F1-score is the harmonic mean of precision and recall. It measures how well a model identifies a class while avoiding both false positives (precision) and false negatives (recall). It is particularly relevant when classes are imbalanced or when both types of errors matter.

Macro-averaged F1-score.
In multi-class classification, the macro F1-score computes the F1-score for each class independently and then averages them. This approach:

- gives equal weight to each class

- gives equal importance to minority and majority classes
- evaluates the balance of performance across all classes rather than overall accuracy

Cracking Prediction – Results

Three-class problem results (low, mid, high)

Table 14. Per-class mean F1-score (3-class problem)

Model	F1 macro_mean	F1 macro_std	F1-high	F1-low	F1-mid
ExtraTrees	**0.484**	**0.045**	**0.522**	**0.516**	**0.414**
RandomForest	0.450	0.019	0.471	0.513	0.366
GradientBoosting	0.458	0.030	0.525	0.537	0.311
LogisticRegression	0.350	0.035	0.426	0.411	0.212
LinearSVC	0.395	0.053	0.458	0.425	0.303
XGBoost	0.423	0.038	0.481	0.481	0.307

Across all models, macro F1-scores ranged between 0.35 (LogisticRegression) and 0.48 (ExtraTrees), indicating moderate predictability in the 3-class scenario.
Tree-based models performed best, with ExtraTrees reaching the top score (0.48), followed by Gradient Boosting, RF, and XGBoost.

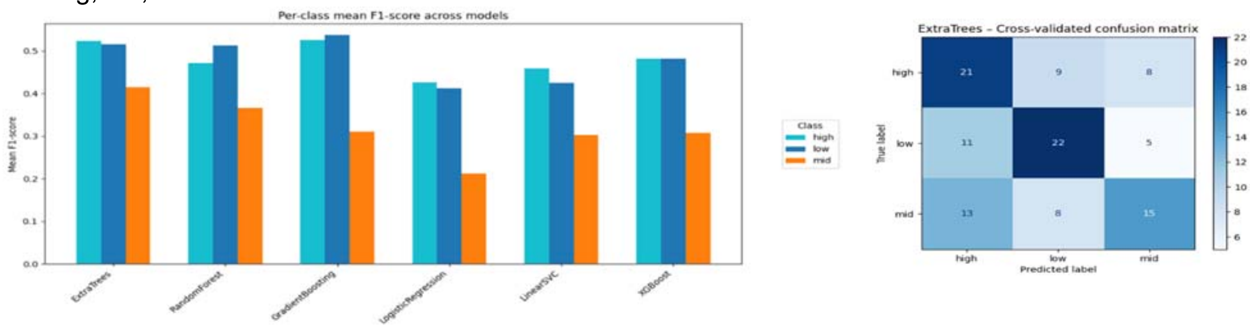


Figure 30. Per-class mean F1-score across all models, and ExtraTrees confusion matrix

A consistent pattern appears across all models, both in the table and in the visualizations above:

- The mid-class is by far the hardest to predict, with F1-scores often below 0.41.
- This class shows overlap with both low and high, explaining misclassification.
- The low and high classes are more separable, with better per-class recall in several folds.
- Linear models (Logistic Regression, Linear SVC) underperform, confirming that cracking depends on nonlinear patterns in the data.

These results justify the transition to the binary-extremes-only formulation discussed in the previous section.

Binary extremes-only problem (low vs high) results

Table 15. Binary classification (low vs high)

Model	F1-macro_mean	F1-macro_std	F1-high	F1-low
ExtraTrees	**0.684**	**0.105**	**0.685**	**0.683**
RandomForest	0.657	0.092	0.655	0.659
GradientBoosting	0.630	0.012	0.611	0.650
Logistic Regression	0.656	0.107	0.645	0.667
Linear SVC	0.654	0.129	0.637	0.671
XGBoost	0.615	0.060	0.619	0.612

Model performance

Across all models, macro F1-scores ranged from 0.61 (XGBoost) to 0.68 (ExtraTrees), indicating a clear improvement over the 3-class setup. The best-performing models were ExtraTrees, RandomForest, LogisticRegression and LinearSVC, all reaching F1-score **= (0.65 - 0.67)**, while Gradient Boosting-based methods performed slightly lower **(=0.61 - 0.63).** These results confirm that very low and very high cracking levels are substantially easier to distinguish from each other than the full set of three categories (low, mid, high). Removing the ambiguous mid-class reduces noise and yields higher, more consistent predictive performance across model families.

Confusion matrix analysis

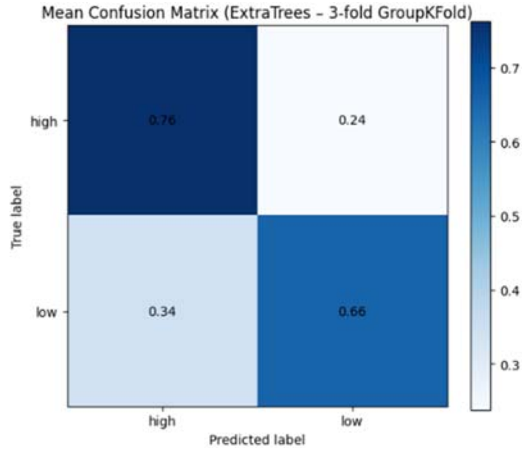


Figure 31. Mean Confusion Matrix (3-folds) - ExtraTrees

In the binary setting, the confusion structure is clearly improved. High-cracking trees are correctly classified in 76% of cases, while low-cracking trees are correctly identified in 66% of cases. Misclassifications remain moderate and occur in both directions, mainly reflecting ambiguity near the low-high cracking threshold. Overall, these results confirm that separating cracking into two extreme classes is substantially more reliable than the original three-level classification.

Feature importance

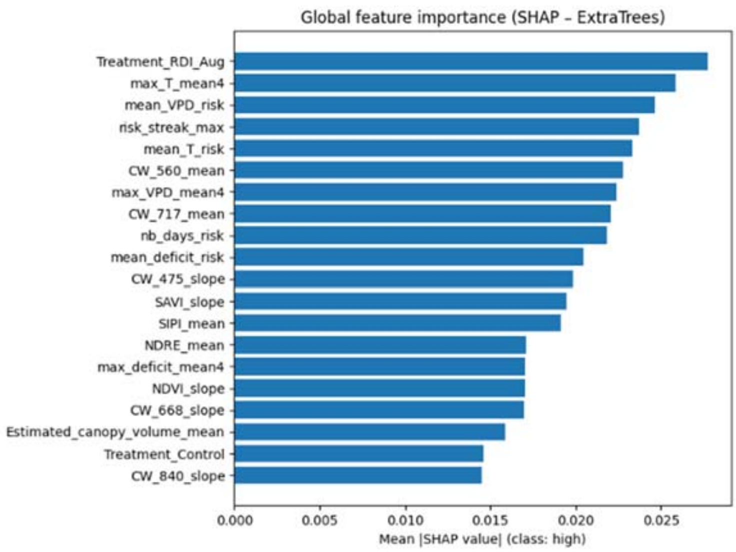


Figure 32. feature importance – ExtraTrees

SHAP analysis indicates that high fruit cracking is primarily driven by intense, persistent late-season climatic stress rather than by average seasonal conditions. Variables capturing stress magnitude (e.g., **peak 4-day mean temperature** or **VPD**) and stress duration (**number and length of stress periods**) are among the most influential predictors, highlighting the importance of cumulative exposure.
Remote sensing features and canopy structural metrics provide complementary information by reflecting

tree physiological status and response to stress, while irrigation treatments indicate that orchard water management modulates cracking risk by influencing tree sensitivity to stress-rainfall events. Overall, fruit cracking results from the combined effects of climatic stress, rainfall shocks, and plant physiological status, rather than from a single dominant driver.

Pomegranate irrigation experiment plots

Pomegranate irrigation treatment in Israel

Cracking incidence in pomegranate is driven by complex interactions among irrigation practices, atmospheric demand, temperature variability, and plant physiological status. Understanding and predicting these interactions is therefore critical for improving yield stability and resource-use efficiency. Within CrackSense, the pomegranate case study aimed to: (1) evaluate the effects of irrigation timing and amount on plant water status (PWS), yield, and fruit cracking; (2) develop and validate UAV-based multispectral, thermal, and LiDAR approaches for estimating PWS at tree and plot scales; and (3) integrate remote sensing data with machine-learning models to predict yield and cracking incidence under varying environmental and management conditions.

Tsor'a Study site

The experiment was conducted in a commercial pomegranate orchard (*Punica granatum* L.) cv. 'Wonderful', located at Kibbutz Tsor'a in the central lowland region of Israel (31°45′37″ N, 34°57′48″ E). The experimental design included four irrigation treatments arranged in a randomized block design, with six biological replicates per treatment collected in two seasons (Fig. 33). Each experimental block consisted of three adjacent rows, with ten trees per row. The central tree in each block (Fig. 33) was instrumented with the Phytech sensing system and continuously monitored for trunk and fruit growth parameters. The irrigation treatments were defined as follows: (i) Control (BK), representing the conventional management protocol, with irrigation initiated after full bloom; (ii) Early irrigation (R), with irrigation commencing prior to flowering;
(iii) High-water treatment (Y), receiving approximately 20% more water than the control; and
(iv) Low-water treatment (BE), receiving approximately 30% less water than the control.
Irrigation was applied every other day throughout the growing season. These treatments were designed to induce contrasting plant water status conditions while maintaining realistic commercial management.

Figure 33. ***(a)*** **Location of the pomegranate study site at Kibbutz Tsor'a, central lowland region of Israel.** ***(b)*** **Layout of the irrigation experiment within the study orchard. Colored frames indicate the four irrigation treatments: BK – control irrigation (irrigation initiated after full bloom), R – early irrigation (irrigation initiated prior to flowering), Y – high-water treatment (approximately +20% relative to the control), and BE – low-water treatment (approximately −30% relative to the control treatment).**

Fruit samples were collected for subsequent analyses of fruit skin tissue at six developmental stages, representing key phenological phases of skin color development: green young fruit at 50 and 85 days after full bloom (DAFB), color break at 110 DAFB, reddish coloration at 135 DAFB, and fully mature dark-red fruit at 185 DAFB (Fig. 34). These stages encompass the critical period from early fruit development through physiological maturity, during which fruit growth, water relations, and cracking susceptibility are known to vary.

Figure 34. Representative images illustrating progressive fruit development and coloration across the season, spanning early growth to full maturity. Numbers above each panel denote the developmental stage index (60-185), corresponding to increasing fruit size and peel pigmentation. This visual sequence highlights the transition from immature green fruit to fully colored mature fruit, providing phenological context for analyses of yield, cracking, and plant water status.

UAV campaigns were conducted approximately once per month throughout the pomegranate growing season, from flowering (late May) to harvest (October). Multispectral, thermal, and LiDAR data were acquired in parallel with ground measurements of physiological parameters and environmental variables. Meteorological data were obtained from a nearby weather station, ensuring consistent characterization of atmospheric forcing throughout the study period.

Meteorological conditions and irrigation

Climatic data for the region where the experimental plots were located were obtained from the Meteorological Services of the Ministry of Agriculture and Rural Development (Fig. 35). The dataset covers the entire pomegranate growing period, from May to 7 November 2023, spanning 190 days. Although a comprehensive set of meteorological variables was available, four key parameters considered most relevant to pomegranate growth and water relations are presented here: maximum air temperature, diurnal temperature range (the difference between daily maximum and minimum temperatures), total incoming solar radiation, and minimum relative humidity (RH). Hot and dry conditions characterized the 2023 growing season. During the monitored period, 53 out of 190 days recorded maximum temperatures exceeding 34 °C (Fig. 35). Large diurnal temperature fluctuations were also observed, with 16 days exhibiting a day–night temperature difference greater than 18 °C (Fig. 35a, blue line). In addition, atmospheric dryness was pronounced, with 68 days showing minimum RH values below 35% (Fig. 35b, brown line). Total incoming radiation declined seasonally from August onward (Fig. 35b, blue line), reflecting the transition toward late summer and autumn conditions.

In 2024, climatic conditions during the pomegranate growing season showed broadly similar seasonal patterns but with several notable differences compared to 2023 (Fig. 35). The season was again characterized by prolonged warm conditions, with extended periods of elevated maximum temperatures during summer, although the frequency and persistence of extreme heat events differed. Maximum air temperatures frequently exceeded 34 °C, particularly during mid to late summer, contributing to sustained atmospheric demand. Diurnal temperature ranges remained pronounced, especially during summer months, indicating continued exposure of trees to large day-night thermal gradients. Atmospheric dryness was also evident in 2024, with recurrent periods of low relative humidity, reinforcing high vapor pressure deficit conditions during critical phenological stages. In contrast to 2023, the accumulation of heatwave days progressed more gradually early in the season but reached comparable cumulative levels by late summer (Fig. 35, lower panels), indicating sustained thermal stress over time rather than short, isolated extremes. As in the previous year, incoming solar radiation declined toward late summer and autumn, reflecting seasonal changes in solar geometry and day length. Overall, the 2024 climatic conditions imposed persistent evaporative demand on the orchard, creating a contrasting yet complementary environmental context to 2023 for evaluating irrigation treatments, PWS dynamics, and their implications for fruit cracking and yield.

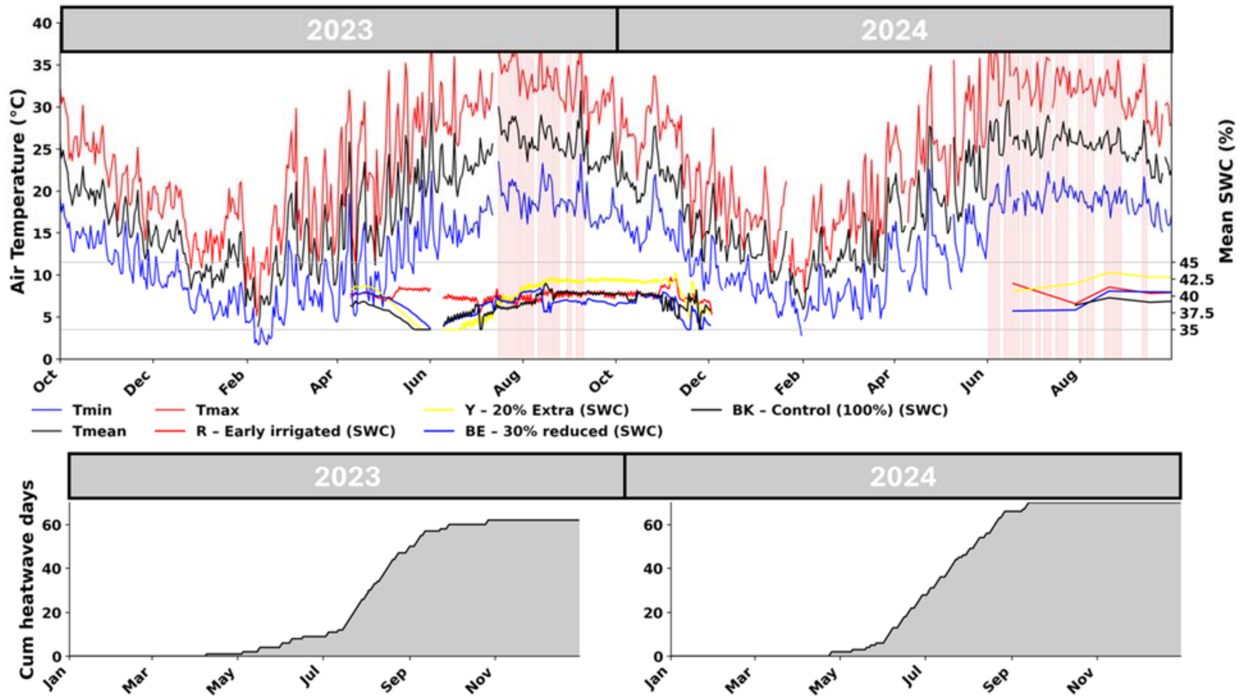


Figure 35. Climatic conditions during the 2023 and 2024 pomegranate growing seasons at the experimental site. The top panel shows daily minimum (Tmin), mean (Tmean), and maximum (Tmax) air temperature, together with mean soil water content (SWC) under different irrigation treatments. Bottom panels show the cumulative number of heatwave days for each year. Shaded areas indicate periods of the pomegranate growing season. The figure illustrates interannual variability in thermal conditions and evaporative demand, providing environmental context for interpreting plant water status dynamics, irrigation responses, and patterns of fruit cracking.

Figure 36 summarizes the seasonal dynamics of TG, SWP, and SC under contrasting irrigation treatments during 2023 and 2024. In both seasons, accumulated TG increased steadily over time, with the control and higher irrigation treatments showing greater cumulative growth, while the reduced irrigation treatment (BE, −30%) consistently exhibited lower TG, indicating constrained vegetative growth under sustained water limitation. SWP showed clear treatment separation, particularly in 2023, with progressively more negative values toward late summer, reflecting increasing water stress during peak atmospheric demand. Reduced irrigation treatments exhibited the most negative SWP values, while the control and extra-irrigated treatments maintained comparatively higher water status. In 2024, SWP variability was lower, suggesting more moderate or stabilized water-stress conditions across treatments. SC responded rapidly to seasonal conditions and irrigation regime, with higher values early in the season followed by a decline toward mid–late summer. Differences among treatments were most pronounced during periods of elevated stress, highlighting SC's sensitivity to short-term water availability and atmospheric demand. Overall, the combined TG, SWP, and SC trajectories demonstrate coherent physiological responses to irrigation management and interannual climatic variability

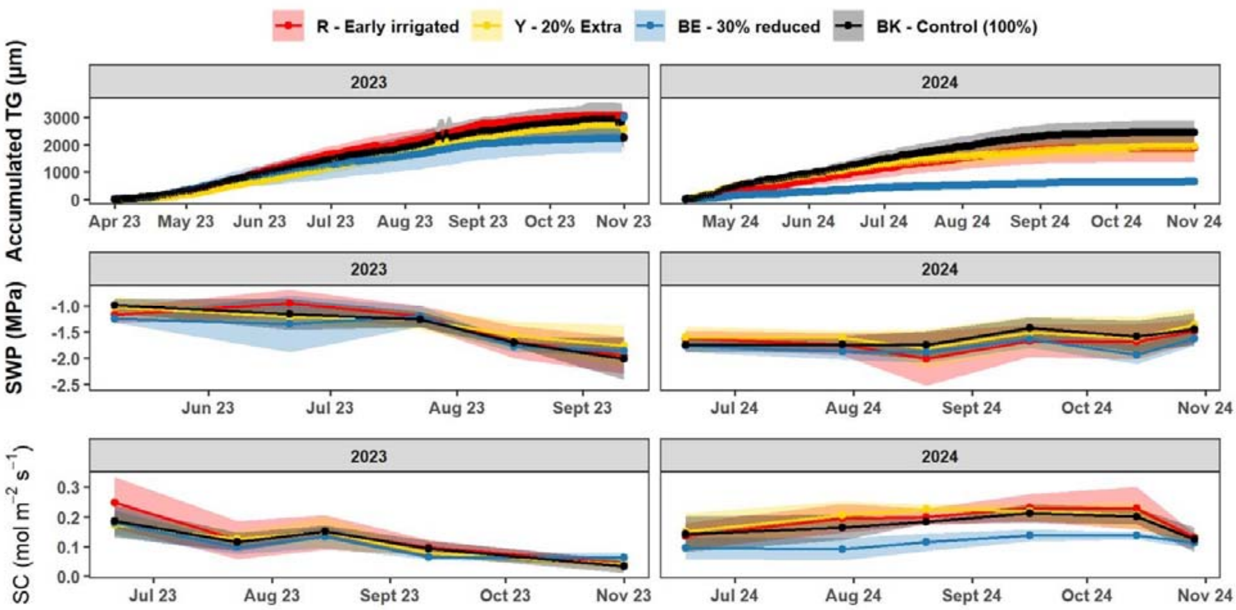


Figure 36. Seasonal dynamics of plant water status and growth indicators under contrasting irrigation treatments during the 2023 (left panels) and 2024 (right panels) growing seasons. The top panels show accumulated trunk growth (TG), the middle panels show stem water potential (SWP), and the bottom panels show stomatal conductance (SC). Lines represent treatment means, and shaded areas indicate variability among replicates. Results illustrate treatment-dependent differences in growth, hydraulic status, and stomatal regulation, as well as interannual variability in physiological responses to irrigation and climatic conditions.

Model performance for plant water status indicators

RF models demonstrated strong and consistent performance in estimating key PWS indicators, SC, SWP, and TGR, across two growing seasons and contrasting irrigation treatments (Fig. 37). Model evaluation against independent validation and test datasets confirmed the robustness and generalizability of the UAV-based approach under operational field conditions. For SC-RF predictions, high agreement with ground measurements was observed, with coefficients of determination exceeding ~0.85 and low RMSE across datasets. This indicates that the combination of multispectral and thermal UAV features, together with meteorological and temporal variables, effectively captured both spatial and temporal variability in canopy gas exchange. Model performance was consistent through irrigation treatments, demonstrating reliable differentiation between well-watered and water-stressed trees. Similarly, SWP was predicted with high accuracy, highlighting canopy temperature-based metrics and derived stress indices as strong proxies for tree hydraulic status. The RF model successfully captured both seasonal trends and irrigation-driven differences in SWP, with particularly strong performance during mid- to late-season periods, when atmospheric demand and water stress were most pronounced. This underscores the sensitivity of thermal indicators to changes in plant water balance under high evaporative conditions. Predictions of

TGR also performed well, despite the inherently higher noise associated with growth-based indicators. The RF model reproduced both daily and seasonal patterns in TGR, reflecting the integrated effects of water availability, temperature, and growth dynamics. Validation against independent data confirms that UAV-derived structural and spectral information, combined with temporal variables, can explain a substantial proportion of the variability in growth in mature pomegranate trees.

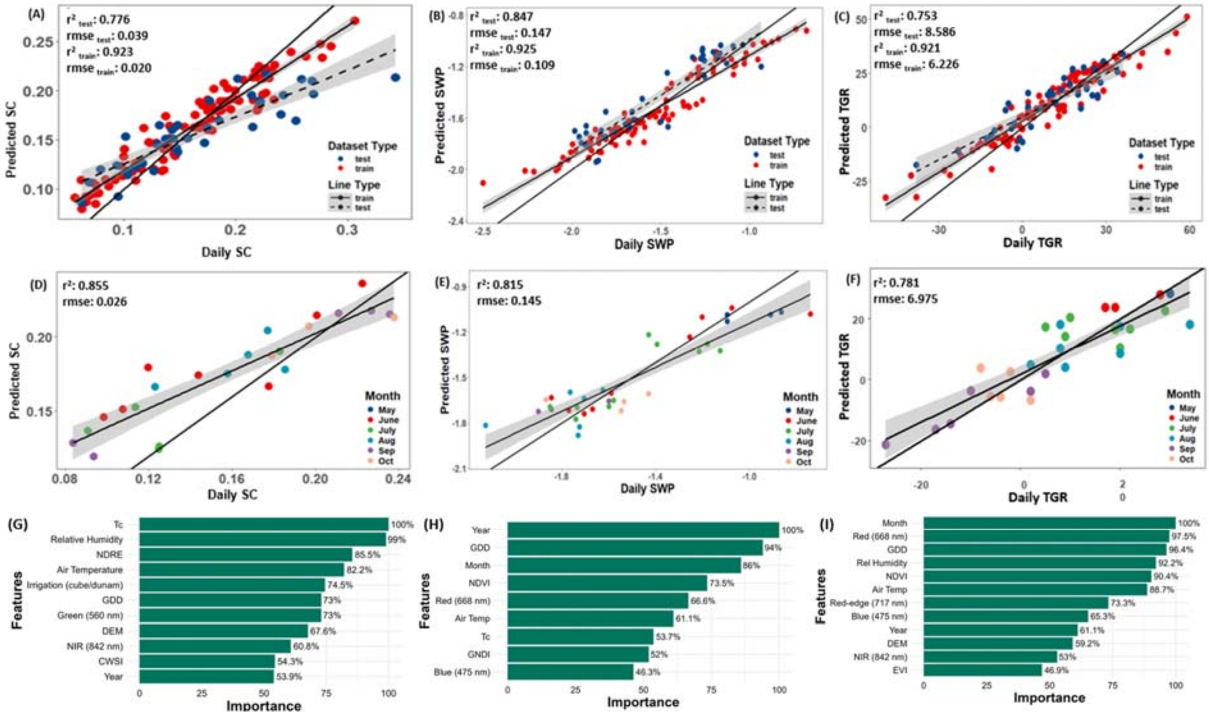


Figure 37. Random Forest (RF) model performance for the estimation of plant water status parameters: stomatal conductance (SC), stem water potential (SWP), and trunk growth rate (TGR). (A-C) Comparison between observed and RF-predicted values for the training and validation datasets. (D-F) Model performance was evaluated on the independent test dataset for the control irrigation treatment, comparing measured values to RF-predicted outputs; the future importance for the tree models (G-I).

Temporal dynamics and spatial patterns of predicted PWS

Spatial maps of predicted SC, SWP, and TGR generated at the beginning and end of each growing season revealed clear temporal and management-driven patterns. Early-season maps showed relatively homogeneous PWS across treatments, reflecting adequate soil moisture following winter rainfall and early irrigation. As the season progressed, spatial heterogeneity increased markedly. By late season, deficit-irrigation treatments exhibited lower predicted SC and SWP and reduced trunk growth, whereas early- and high-irrigation treatments maintained higher PWS values. These patterns were consistent across both study years, despite interannual differences in climatic conditions, indicating that the models captured fundamental physiological responses rather than year-specific artifacts. Importantly, the spatial predictions highlighted within-plot variability, demonstrating that individual trees responded differently to identical irrigation regimes. This intra-plot variability underscores the value of tree-level UAV monitoring for precision management, as plot-averaged measurements would obscure these fine-scale responses.

The results clearly showed that irrigation timing and amount had distinct and measurable effects on PWS. Early irrigation treatments consistently resulted in higher SC and less negative SWP values during critical phenological stages, particularly during flowering and early fruit development. In contrast, low-water treatments led to progressively increasing stress signals later in the season, coinciding with periods of high evaporative demand. High-water treatments did not always yield proportionally higher PWS compared to the control, suggesting diminishing returns beyond a certain irrigation threshold. This finding aligns with the hypothesis that irrigation timing, rather than total water volume alone, plays a dominant role in regulating pomegranate PWS and subsequent stress responses.

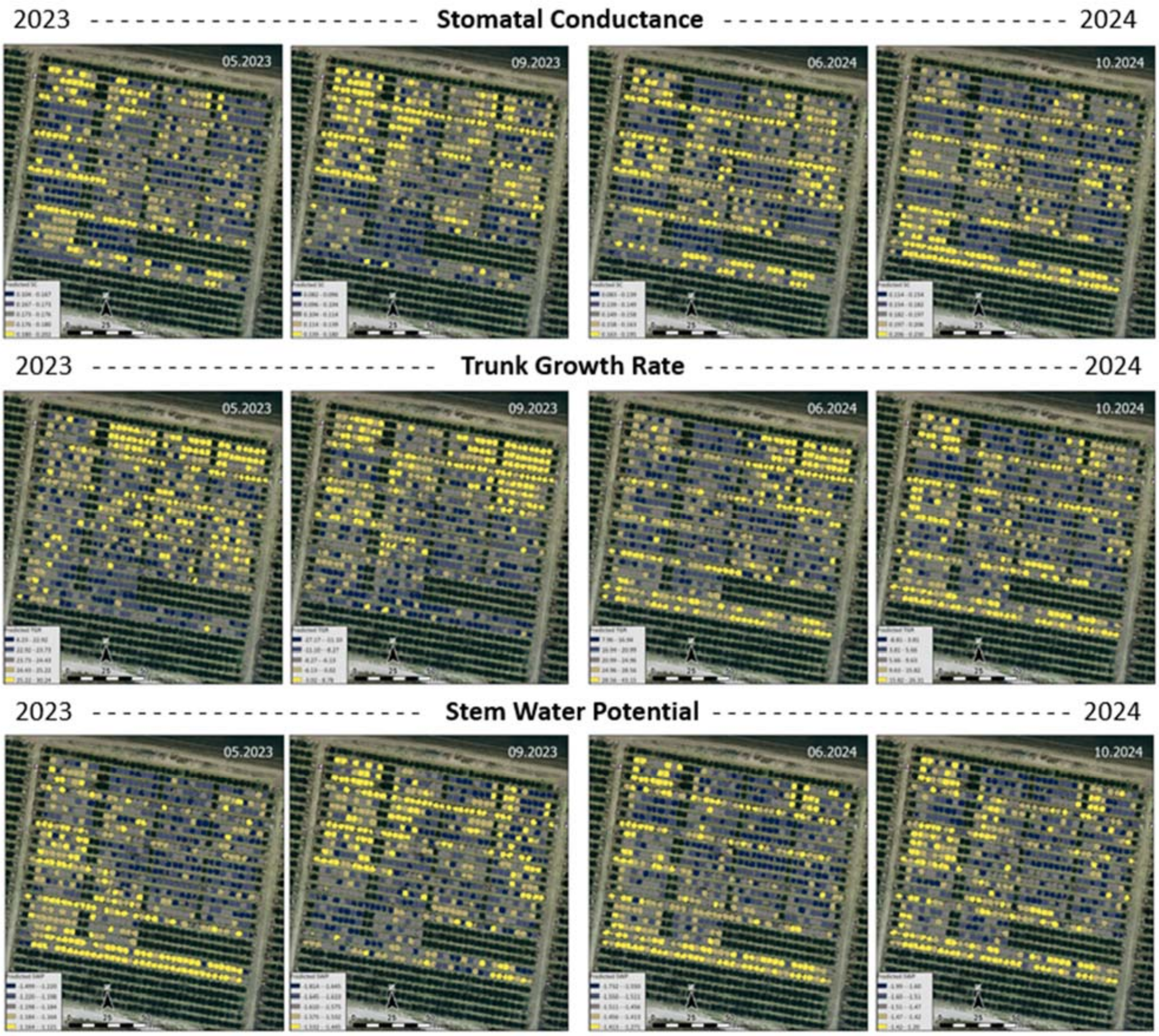


Figure 38. Spatial visualization of predicted plant water status (PWS) parameters, stomatal conductance (SC), trunk growth rate (TGR), and stem water potential (SWP), at the plot scale. Maps are shown for the beginning and end of each growing season for both study years. UAV flights were conducted approximately monthly from May to September 2023 and from June to October 2024, covering the pomegranate-growing season in Israel from flowering (late May) to harvest (October).

The effect of irrigation on fruit cracking intensity

Figure 39 summarizes fruit cracking incidence across irrigation treatments based on tree-level harvest measurements (top panels) and packinghouse records (bottom panels) for the 2023 and 2024 seasons. At the tree scale, cracking rates consistently increased under higher irrigation amounts, particularly in the extra-irrigated treatment, while reduced irrigation resulted in lower cracking incidence relative to the control. These treatment-dependent differences were observed in both seasons, indicating that the fruit-cracking response pattern is influenced by irrigation intensity and environmental factors. Packinghouse records showed the same overall pattern: higher cracking rates with increased irrigation and lower rates with reduced irrigation, confirming that the treatment effects observed at the tree level translated into commercially relevant outcomes. The close agreement between tree-based assessments and packinghouse data strengthens the reliability of the results. It demonstrates that irrigation-driven differences in cracking are detectable both at harvest and during postharvest sorting. To the best of our knowledge, these findings provide the first field-based evidence in Israel linking irrigation amount directly to fruit cracking incidence in commercial pomegranate orchards.

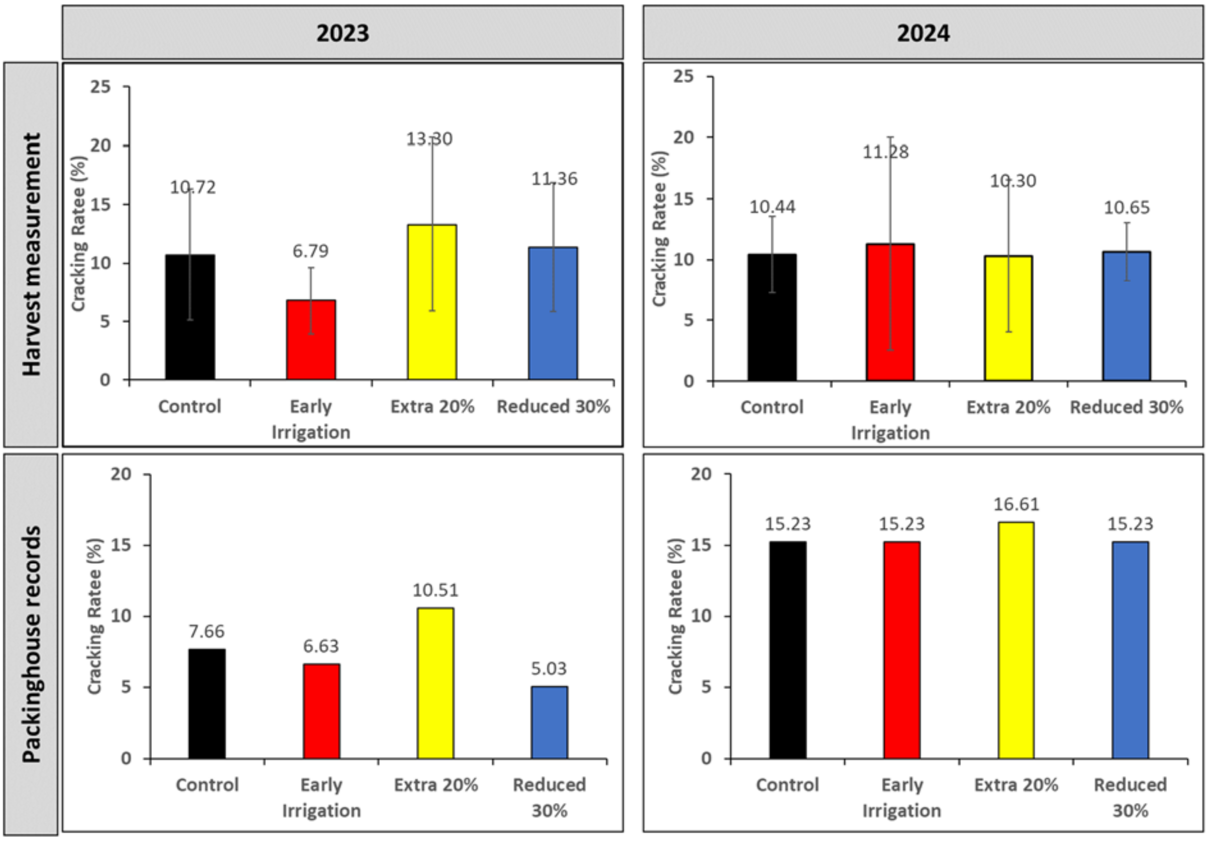


Figure 39. Fruit cracking rates (%) under four irrigation treatments (Control, Early irrigation, +20% irrigation, and −30% irrigation) for two contrasting seasons (2023 and 2024). Upper panels show cracking rates measured at harvest, and lower panels show cracking rates recorded at the packinghouse. Bars represent treatment means, with error bars indicating variability among trees.

Estimation of yield and fruit cracking with RS

Daily trunk growth (TG) estimates from tree dendrometers increased linearly with plant area estimated with LiDAR -UAV across irrigation treatments and years (Fig. 40). Based on June data, plant area explained a large proportion of the variability in TG ($R^2 = 0.90$), with a positive slope (1.34 μm m^{-2}), indicating that trees with larger canopy/plant area exhibited higher daily growth rates. This strong relationship was consistent across irrigation regimes and seasons, highlighting plant area as a robust structural proxy for tree growth potential and water-driven growth dynamics during early mid-summer.

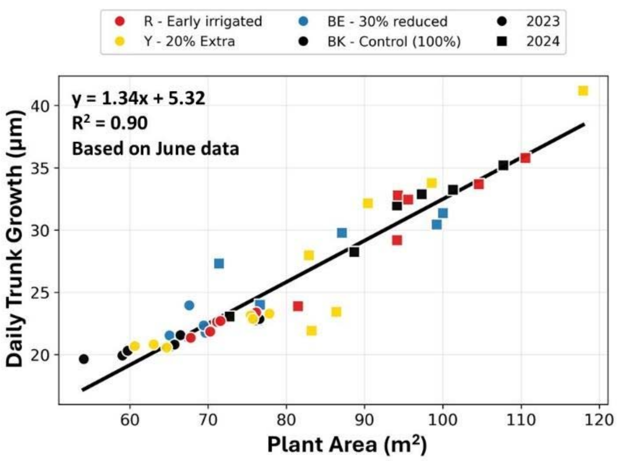


Figure 40. Relationship between plant area (m²) and daily trunk growth (TG, μm) during June across irrigation treatments and study years (2023–2024). Points represent individual trees under different irrigation regimes, and the solid line shows the fitted linear regression (TG = 1.34 × Plant Area + 5.32; R² = 0.90), indicating a strong positive association between canopy size and growth rate.

Plant area (PA) showed a strong and consistent relationship with both yield and fruit cracking across irrigation treatments and years (Fig. 41). Yield increased linearly with plant area when June data were used, explaining a large proportion of the variability in fruit number ($R^2 \approx 0.83$), indicating that canopy development during early mid-season is a robust predictor of final yield. A similarly strong relationship was observed between yield and PAI, although with slightly greater scatter, reflecting structural and density-related effects within the canopy and between growing seasons.

In contrast, cracked fruit counts exhibited a clear negative, non-linear relationship with plant area and PAI when analyzed using September data. Trees with smaller PA showed substantially higher cracking incidence, whereas larger, well-developed canopies were associated with markedly lower numbers of cracked fruits ($R^2 \approx 0.87$). The predictive model for cracking rate further confirmed this pattern, showing good agreement between measured and predicted cracking rates ($R^2 = 0.67$, RMSE = 3.06). Together, these results demonstrate that canopy size and structure are key integrators of water status and stress history, linking early-season vegetative development with final yield and late-season cracking risk.

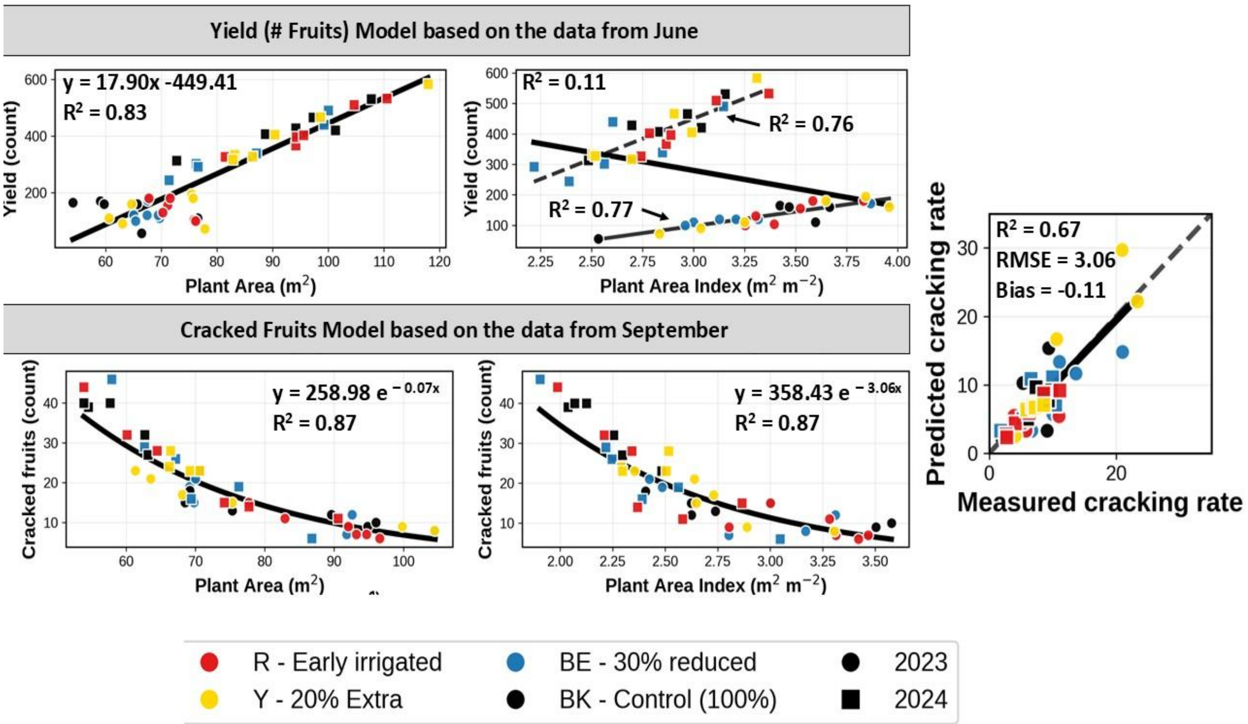


Figure 41. Relationships between canopy structure and fruit production and cracking in pomegranate trees. The upper panels show the relationships between plant area and yield (fruit number) based on June data, and between plant area index (PAI) and yield, with fitted regression models and coefficients of determination. The lower panels show negative exponential relationships between plant area and cracked fruit number, and between PAI and cracked fruit number, based on September data. The right panel presents the comparison between measured and model-predicted cracking rates. Points are colored by irrigation treatment (early irrigation, 20% extra, 30% reduced, and control), and symbols indicate study year (2023 and 2024).

Pomegranate deep learning yield and loss in Israel

Methodology Overview

A second approach was tested to assess yield and loss in pomegranate. This study aims to develop and evaluate a deep-learning (DL) framework for the automated estimation of pomegranate yield and fruit loss from UAV-based imagery, apply the framework to commercial orchard datasets, and validate its performance through statistical comparison with field-based ground-truth measurements. The methodology comprises two interconnected components that together form an automated system for estimating yield and loss from UAV-acquired RGB imagery. The overall workflow begins with a data preparation stage, in which still images are manually annotated, augmented, and pre-processed to train two task-specific object detection models: one for detecting healthy fruits (yield) and a second for detecting and counting defective or cracked fruits (loss; Fig. 42). These models are trained independently to address the strong class imbalance between healthy and faulty fruits and to improve detection robustness. Building on the detection stage, the second component integrates the trained detectors into a multi-object tracking framework for pre-processed UAV video sequences. Using a tracking-by-detection approach, detections in consecutive frames are temporally associated, allowing each fruit to be assigned a unique identity as it appears along the UAV flight path. This enables reliable counting of individual fruits while the UAV moves between orchard rows. The combined pipeline processes each tree from both filmed sides, generating estimates of healthy fruit yield and defective fruit loss that are directly comparable to ground-truth observations collected in the orchard. Together, these two components establish a scalable, non-destructive, and automated framework for assessing yield and fruit loss in commercial pomegranate orchards.

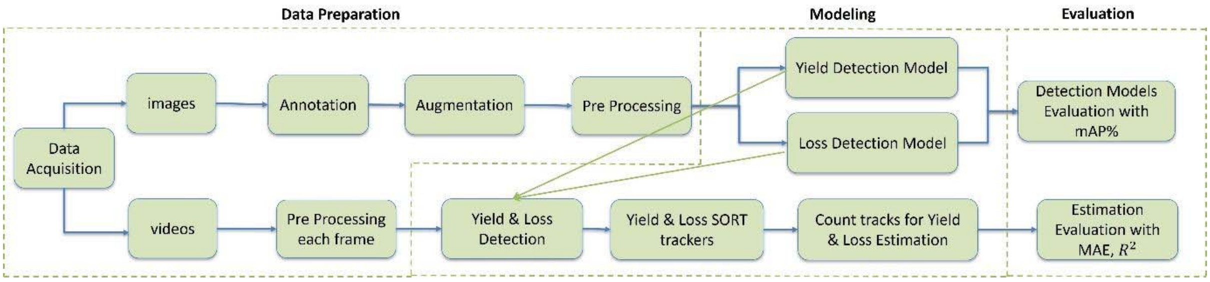


Figure 42. Workflow of the deep learning–based methodology for automated estimation of pomegranate yield and fruit loss from UAV imagery. The pipeline includes (i) data preparation and annotation of UAV-acquired RGB images, (ii) training of two task-specific object detection models for healthy fruits (yield) and defective fruits (loss), and (iii) integration of the trained detectors within a multi-object tracking framework applied to UAV video sequences, enabling tree-level estimation of yield and fruit loss.

Data acquisition

Image acquisition was conducted during three sampling campaigns between 2023 and 2025, each in October, coinciding with the pomegranate ripening and harvest period. The 2023 and 2024 campaigns were conducted in a commercial pomegranate orchard located at Kibbutz Tzora, Israel (31°45′32.7″ N, 34°57′17.9″ E). A consistent set of 30 sample trees was monitored throughout the study. High-resolution RGB images were captured from multiple viewing angles using two camera systems: a Xiaomi 2201116TG smartphone camera (4000 × 3000 pixels) and a Canon EOS 200D DSLR camera (6000 × 4000 pixels). These images were primarily used to build and train the object-detection datasets. In 2024, UAV-based video data were additionally acquired using a DJI Mavic 2 Pro platform. UAV flights were conducted along orchard rows using a forward, between-rows viewing geometry, in which two adjacent tree rows appeared at the edges of the frame while the ground surface remained visible in the center. The camera was positioned approximately 1 m above ground level, providing a consistent perspective on canopy structure and fruit distribution along each tree, as well as clear visibility of the ground for detecting pruned or fallen defective fruits. This acquisition geometry was deliberately selected to minimize double-counting, as trees located behind the filmed row are not visible in the field of view. Consequently, the two rows of trees were captured in a single pass, ensuring that each tree was observed and counted once per flight direction, as illustrated in Figure 43.

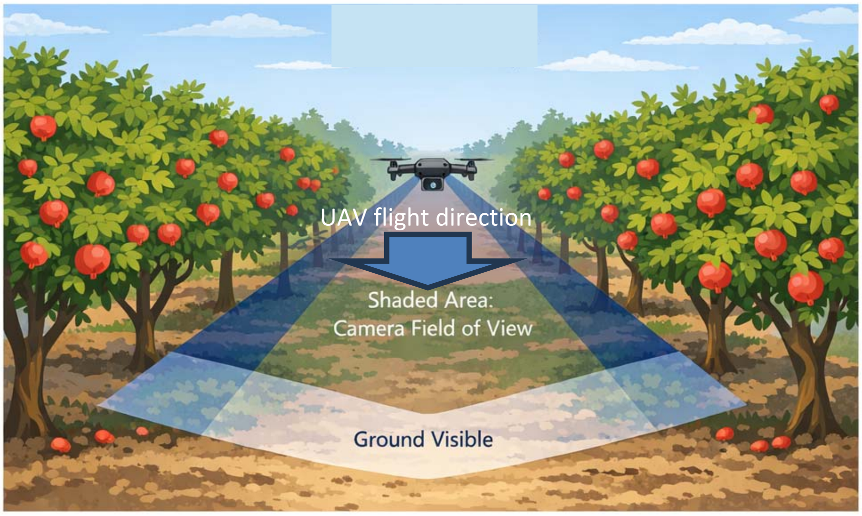


Figure 43. (a) Schematic illustration of forward-facing UAV data acquisition between two pomegranate orchard rows. The shaded triangular area represents the camera's field of view, capturing both adjacent rows while maintaining ground visibility in the center.

The 2025 data collection campaign was expanded to include three additional orchard plots that were not used for model training to evaluate model performance under previously unseen environmental and management conditions. These plots included a further section of the Tsor'a orchard (31°45'32.4" N, 34°58'36.2" E) and two adjacent commercial orchards in Mishmar HaNegev (31°22'52.5" N, 34°42'53.1" E). UAV data for model evaluation were acquired using the same forward-facing, between-row flight configuration applied in the pomegranate validation experiments. For each measuring tree, UAV video data were collected, and ground-truth yield was determined by manual fruit counting. Yield loss, including cracked and fallen fruit, was quantified by manually annotating the UAV video recordings, providing an independent reference dataset for model validation.

Measuring trees were defined as the individual trees for which both UAV imagery and ground-truth measurements were available, and they served as the validation set for assessing the performance of yield and yield-loss estimation. The number and distribution of validation trees across plots are reported in Table 16. To ensure accurate spatial alignment during video cropping and dataset construction, the start and end of each measuring tree were physically marked using red and white tape in the pomegranate experiments. Each orchard row was filmed from both sides, resulting in two independent viewing directions per measuring tree. This dual-view acquisition was implemented to reduce occlusion effects and improve robustness of the yield estimates, while avoiding systematic double-counting. In addition to the forward-facing recordings used for validation, a supplementary UAV video was acquired over a single row in the Tsor'a orchard from an altitude of approximately 5 m. This dataset was used exclusively for training purposes to improve recognition of the tree boundary markers. It was not included in the validation analysis to maintain consistency with the pomegranate validation protocol.

Figure 44. UAV flight-level view of a marked measuring tree segment within the pomegranate orchard. Red and white tape indicates the start and end points of each measuring tree, providing precise spatial boundaries for cropping UAV video segments and associating detections with the corresponding tree during yield-loss analysis.

Dataset for detection models

The detection dataset was constructed using 240 high-resolution still images collected during the 2023–2024 sampling campaigns in the Tsor'a pomegranate orchard, as described in Section XX. These images captured both healthy and defective fruits under a wide range of illumination conditions, viewing angles, canopy densities, and tree geometries, representative of operational orchard variability (Fig. 45). All images were manually annotated using the Roboflow platform and labeled into two classes: (i) healthy fruits and (ii) defective (cracked) fruits. Defective fruits typically appeared smaller, darker, and desiccated, reflecting exposure to solar radiation after pericarp cracking and partial drying on the tree or on the ground surface. The annotated dataset was split into training (70%), validation (15%), and testing (15%) subsets. To enhance model generalization and robustness to field variability, an extensive data augmentation pipeline was applied only to the training subset.

Augmentation operations were designed to simulate realistic variations encountered in UAV and ground-based imagery while preserving fruit geometry and class identity. These operations included random spatial transformations, such as random cropping with a maximum zoom factor of 25%, and image rotations uniformly sampled between −10° and +10°, accounting for camera tilt and UAV motion. Photometric augmentations were also applied, including exposure adjustments in the range of -12% to +12%, to mimic changes in illumination and shadowing. In addition, Gaussian noise with a per-channel standard deviation of 0.1 was added to improve robustness to sensor noise, and a Gaussian blur with randomly selected kernel sizes of 3 or 5 pixels was applied to simulate motion blur and defocus. Collectively, these augmentation steps increased the effective size of the training dataset by approximately 6-fold, yielding 852 augmented training images. No augmentation was applied to the validation and test sets, which were retained as representative, unaltered samples for unbiased model evaluation.

A preprocessing stage was further implemented to enhance visual consistency across images acquired under heterogeneous conditions. This stage consisted of contrast-limited adaptive histogram equalization (CLAHE) to improve local contrast, followed by saturation enhancement to increase the separability of fruit colors from foliage and background elements. The same preprocessing pipeline was applied uniformly across training, validation, and test datasets. In addition to the primary dataset, a supplementary set of 21 images was extracted from a UAV video at 5 m altitude. Frames from this recording provided clear visual examples of the red-and-white marking tape used to delimit measuring trees, which occasionally appeared within the UAV's field of view during operational flights. These images were included to ensure that the detection models learned to classify the tape as background, thereby preventing confusion with defective fruits. The supplementary images underwent the same augmentation and preprocessing pipeline as the main dataset, expanding from 21 raw images to 124 augmented samples, and were incorporated exclusively into the training set. Together, these dataset construction and augmentation steps ensured that the detection models were exposed to a broad range of visual conditions while preserving a strict separation between training, validation, and testing data, thereby supporting robust and unbiased performance evaluation.

Figure 45. Representative examples of images used to construct the detection dataset, illustrating variability in viewing angles, illumination conditions, canopy structure, and fruit appearance encountered across the pomegranate orchard

Video dataset for tracking validation

The tracking validation dataset was generated from UAV video recordings acquired in 2025, in accordance with the data acquisition procedures described in Section xxx. Using the red and white tape markers placed at the start and end points of each measuring tree, all UAV videos were cropped to include only the temporal segments in which the target tree was visible. This ensured precise spatial and temporal alignment between video data and ground-truth measurements. Because each orchard row was filmed from both sides, the cropping process produced two independent video segments per measuring tree, corresponding to the left-side and right-side viewpoints. This dual-view acquisition strategy enabled evaluation of the detection and tracking framework under varying viewing geometries and occlusion conditions. Across the three orchard plots, 28 measuring trees were included in the validation dataset. Each tree, therefore, contributed two cropped video segments, yielding 56 viewpoint-specific video samples for evaluating the combined detection and tracking pipeline. The distribution of the trees across the study plots is presented in Table 16. All cropped video segments were catalogued in a structured database, which included metadata for the tree identifier, orchard plot, and viewing side. This metadata framework ensured that, during automated tracking and yield estimation, model predictions were consistently assigned to the correct measuring tree and that only detections appearing on the designated side of each video were included in the final yield and loss counts.

Table 16. Distribution of Measuring Trees Across the Study Plots

Plot	Number of measuring trees	Tree IDs
Mishmar HaNegev plot 11	12	1–12
Mishmar HaNegev plot 7	4	8, 10, 11, 12
Tzora	12	1–12
Total	28	-

Yield and loss detection

Automated detection of healthy fruits (yield) and defective fruits (yield loss) was performed using the YOLO11 object detection framework (Glenn Jocher, 2024). YOLO11 was selected for its favorable balance between detection accuracy and computational efficiency across diverse field conditions, including variable illumination, complex canopy structure, and heterogeneous background textures (Jegham et al., 2025). To address the pronounced class imbalance between healthy and defective fruits, two independent single-class detection models were trained rather than a single multi-class detector. One model was trained exclusively to detect healthy fruits, while the second model was trained to detect defective or cracked fruits representing yield loss. This strategy reduced class competition during training and improved detection robustness for both targets, particularly for defective fruits, which are typically smaller, darker, and less frequent than healthy fruits.

Both detection models were based on the YOLO11 small (YOLO11s) architecture, initialized with pretrained weights to leverage general visual features learned from large-scale datasets. YOLO11 follows a single-stage, anchor-free detection architecture, in which feature extraction, object localization, and classification are performed jointly within a unified network. This design enables real-time inference while maintaining high detection accuracy, making it well-suited for UAV-based orchard applications. The models were fine-tuned on orchard-specific imagery using supervised learning. Training was performed with fixed input resolution, mini-batch stochastic optimization, and early stopping based on validation performance to prevent overfitting. Separate training runs were conducted for the yield and loss models, each optimized independently to maximize detection performance for its respective target class. This training strategy ensured robust convergence under class imbalance and preserved sensitivity to small, visually subtle defective fruits.

Model architecture and training strategy

Both detection models were based on the YOLO11 small (YOLO11s) architecture and were initialized with pretrained weights (*yolo11s.pt*) obtained from training on the Common Objects in Context (COCO) dataset (Lin et al., 2014). This initialization enabled transfer of generic visual features prior to fine-tuning on orchard-specific imagery. Two separate models were trained: a yield model, designed to detect healthy fruits, and a loss model, designed to detect defective or cracked fruits representing yield loss. Each model was implemented as a single-class detector, such that only one annotation type was included in the corresponding training dataset. For the yield model, only healthy fruits were annotated, while defective fruits and all other objects were treated as background. Conversely, for the loss model, only cracked or defective fruits were annotated, with healthy fruits treated as background. This design avoided severe class imbalance in a single detection task and empirically improved per-class detection accuracy and training stability.

Both models were fine-tuned for 100 epochs with a batch size of 16 and an input image size of 640 × 640 pixels, using the AdamW optimizer in the Ultralytics YOLO training framework. Model convergence and generalization were monitored using validation performance. In addition to the initial training, the loss model underwent a second fine-tuning stage using the extended dataset. In this stage, the augmented dataset derived from the 5-m overhead UAV flight, consisting of 124 images generated from 21 raw frames, was combined with the original 852 augmented training images. The combined dataset was used to further train the pretrained loss model for 15 additional epochs, with a batch size of 8 and an initial learning rate of 5×10^{-4}. To preserve the general feature representations learned during the initial training phase, the first ten backbone modules were frozen, while only the higher-level feature layers and detection heads were updated. This strategy reinforced the model's ability to correctly treat the red and white tree-marking tape as background while maintaining accurate detection of defective fruits.

Yield and loss tracking and counting

Fruit counting from UAV video data was performed by integrating object detection with multi-object tracking. Each measuring tree was filmed from both sides, and detection and monitoring were conducted independently for each viewing direction to account for differences in canopy structure, occlusion, and illumination. Prior to detection and tracking, each video frame underwent the same image preprocessing pipeline, consisting of contrast-limited adaptive histogram equalization (CLAHE) followed by saturation enhancement. After preprocessing, the two trained object detectors, the yield model, and the loss were applied to each frame to generate fruit detections at the frame level. These detections were subsequently passed to the tracking module to enable temporal association of individual fruits across consecutive frames, forming the basis for reliable per-tree yield and yield-loss estimation.

Tracking with SORT

Following object detection, temporal association of fruit instances was performed using the Simple Online and Realtime Tracking (SORT) algorithm (Bewley et al 2016). SORT was selected due to its computational efficiency, robustness, and suitability for tracking multiple small objects in sequential UAV video data acquired under field conditions. SORT is a tracking-by-detection framework that relies on a lightweight motion-based model architecture. For each detected object, SORT maintains a state vector describing the object's position, scale, and velocity, which is propagated across frames using a Kalman filter. This motion model predicts the location of each tracked object in the subsequent frame, assuming smooth, approximately linear motion between frames, which is appropriate for fruit observations under steady UAV flight.

Detections in the current frame are associated with predicted track positions using the intersection-over-union (IoU) metric. The resulting IoU cost matrix is passed to a linear assignment solver (Hungarian algorithm), which computes the optimal global matching between existing tracks and new detections. Associations with IoU values below a predefined threshold are rejected. Unmatched detections initiate new track identities, while tracks that fail to be associated with detections for a predefined number of consecutive frames are terminated. This design enables stable maintenance of object identities while minimizing track fragmentation and false persistence. SORT was preferred over more complex appearance-based tracking methods, such as DeepSORT, for several reasons. First, the tracked objects (fruits and defective fruit losses) exhibit limited, highly variable appearance features, with frequent changes in illumination, partial occlusion, and motion blur, thereby reducing the reliability of appearance embeddings. Second, the forward-facing UAV acquisition geometry and continuous camera motion ensure that objects typically reappear across consecutive frames, making motion cues sufficient for reliable association. Finally, SORT offers substantially lower computational complexity, facilitating scalable processing of long UAV video sequences without requiring additional deep feature extraction networks, which is advantageous for operational deployment. The parameters governing the tracking process, including IoU thresholds, track activation criteria, and track persistence settings, are summarized in Table 17.

Table 17. Parameters used for SORT-based tracking of yield and loss objects

Tracker	Parameter	Description	Value
Yield Tracker	minimum_consecutive_frames	Minimum number of consecutive frames an object must be tracked before being considered a valid track	4
Loss Tracker	minimum_consecutive_frames	Minimum number of consecutive frames an object must be tracked before being considered a valid track	3
Yield Tracker, Loss Tracker	minimum_iou_threshold	Intersection-over-union (IoU) threshold for associating detections with existing tracks	0.1
Yield Tracker	track_activation_threshold	Detection confidence threshold required to initialize a new yield track	0.4
Loss Tracker	track_activation_threshold	Detection confidence threshold required to initialize a new loss track	0.65
Yield Tracker, Loss Tracker	lost_track_buffer	The number of frames a track is retained without detection before being terminated	10

Explanation of tracking parameter selection

Tracking parameters were tuned separately for yield and loss objects to reflect their distinct visual and spatial characteristics. Yield fruits are generally larger and more consistently visible across frames, allowing a higher track confirmation requirement (minimum_consecutive_frames = 4). In contrast, loss objects are typically smaller and may appear intermittently on the ground, motivating a slightly lower confirmation threshold (minimum_consecutive_frames = 3) to avoid premature track rejection. A relatively low IoU threshold (0.1) was selected to accommodate moderate apparent motion caused by UAV forward movement and perspective changes between consecutive frames. Different track activation thresholds were applied to the yield and loss models, with a higher confidence threshold for loss objects (0.65) to reduce false positives arising from visually similar background elements, such as soil texture or marking tape. The lost track buffer was set to 10 frames to allow temporary occlusions or missed detections without immediate track termination, supporting stable identity assignment across the video sequence.

Yield and loss estimation

The tracking process maintained stable fruit identities across consecutive video frames, enabling reliable counting of individual fruits. Unique yield tracks corresponded to healthy fruits and were used to estimate yield for the filmed side of each tree, whereas unique loss tracks corresponded to defective fruits and were used to estimate yield loss. Because the forward-facing UAV acquisition geometry captures only one side of a measuring tree at a time, the per-tree loss estimate was obtained by summing the defective fruit tracks from both viewing directions. Most defective fruits were located on the ground, where occlusion effects are minimal; therefore, combining detections from both sides does not introduce systematic double counting under this viewing configuration. In contrast, yield estimation required a more conservative aggregation strategy. Fruit visibility differed between the two sides of each tree due to variations in canopy density, branch architecture, occlusion, and illumination conditions. In addition, depending on pruning intensity and canopy openness, the same fruits could be visible from both viewing directions, leading to double-counting when summing counts from both sides. To mitigate this effect and obtain a stable per-tree estimate, the maximum yield count observed across the two sides was used as the final yield estimate for each tree.

Detection evaluation

To assess the performance of the yield and loss detection models, standard evaluation metrics commonly used in object detection were applied. Precision and recall quantify detection accuracy and completeness and are defined as:

$$Precision = \frac{TP}{TP + FP} Recall = \frac{TP}{TP + FN}$$

where *TP* denotes true positives, *FP* false positives, and *FN* false negatives, detection performance was further summarized using the F1 score, which represents the harmonic mean of precision and recall:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

In addition, Average Precision (AP) was used as an aggregate performance metric. AP corresponds to the area under the precision–recall curve, defined as:

$$AP = \int_0^1 p(r)\, dr$$

where $p(r)$is precision as a function of recall. Precision, recall, F1 score, and AP were reported separately for the yield-detection and loss-detection models.

Counting evaluation metrics

Counting performance was evaluated using metrics that capture both absolute accuracy and systematic deviation between predicted and observed counts. The Mean Absolute Error (MAE) was used to quantify the average magnitude of prediction errors:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$$

where $y_i$represents the observed count and $\hat{y}_i$the corresponding predicted count. To assess systematic bias, the mean signed error (Bias) was computed as:

$$Bias = \frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)$$

Positive bias values indicate overestimation, whereas negative values indicate underestimation. Model explanatory power was further assessed using the coefficient of determination (R^2), which quantifies the proportion of variance in observed counts explained by the model.

Table 18. Summary of evaluation metrics used for detection and counting

Task	Metric	Description	Interpretation
Detection	Precision	Proportion of correct detections among all detections	High values indicate a low false-positive rate
Detection	Recall	Proportion of detected objects among all true objects	High values indicate a low false-negative rate
Detection	F1 score	Harmonic mean of precision and recall	Balanced detection performance
Detection	Average Precision (AP)	Area under the precision–recall curve	Overall detection quality
Counting	MAE	Mean absolute difference between predicted and observed counts	Lower values indicate higher accuracy
Counting	Bias	Mean signed prediction error	Indicates over- or underestimation tendency
Counting	R^2	Proportion of variance explained by the model	Higher values indicate stronger agreement

Results

Detection results

The performance of the detection models was evaluated on the independent test dataset. Quantitative results for both the yield and loss detection models are summarized in Table 19. The yield detection model, evaluated at an IoU threshold of 0.5 and a confidence threshold of 0.3, exhibited well-balanced detection performance. The model achieved a precision of 0.80, indicating that most predicted yield detections corresponded to true fruits, and a recall of 0.79, demonstrating a high ability to detect visible fruits across diverse canopy and illumination conditions. These values yielded an F1 score of 0.795, indicating a strong balance between detection completeness and reliability.

The model achieved a relatively high Average Precision (AP@0.5) of 0.848, indicating robust detection performance under moderate localization constraints. In contrast, the mean AP across IoU thresholds from 0.5 to 0.95 was lower, suggesting that while most yield fruits were correctly detected, precise bounding-box localization was more challenging in a subset of cases. These cases were primarily associated with partial fruit occlusion by foliage, fruit overlap within dense canopy regions, and heterogeneous illumination, all of which affect the exact delineation of fruit boundaries rather than detection itself. Overall, the yield detection results demonstrate that the model reliably identifies healthy fruits under realistic orchard conditions and provides a solid basis for downstream tracking and counting.

The loss detection model exhibited a different performance profile, reflecting the greater visual complexity of detecting defective fruit. Defective fruits are typically smaller, darker, less frequent, and often partially obscured by soil background, shadows, or plant residues. Despite these challenges, the loss model achieved high precision, indicating effective suppression of false positives, particularly those caused by confusion with background elements such as soil texture or marking tape. Recall values were lower than those of the yield model, reflecting the greater difficulty of consistently detecting all defective fruits. However, this conservative detection behavior was desirable, as false positives in loss estimation would directly inflate yield-loss predictions. The AP values for the loss model indicate that, under relaxed localization constraints, the detector successfully identified most visually distinguishable defective fruits, while stricter IoU thresholds exposed the inherent difficulty of accurately localizing small, irregularly shaped objects on heterogeneous ground surfaces. Taken together, the detection results demonstrate that both models achieved performance levels sufficient for reliable integration with the tracking framework, with the yield model prioritizing complete detection and the loss model prioritizing detection precision. This complementary behavior supports stable and robust estimation of both yield and yield loss in the subsequent tracking and counting stages.

Table 19. Detection performance of yield and loss models on the independent test dataset

Model	Evaluation parameters	Precision	Recall	F1 score	AP@0.5	AP@0.5:0.95	AP@0.75
Yield	IoU = 0.5, confidence = 0.3	0.80	0.79	0.795	0.848	0.609	0.696
Loss	IoU = 0.5, confidence = 0.4	0.77	0.60	0.676	0.711	0.428	0.498

Tracking results

Counting performance reflects the combined effects of detection accuracy, temporal tracking stability, and UAV imaging geometry. Quantitative evaluation metrics for yield and loss estimation are summarized in Table 20. The yield estimation model achieved a mean absolute error (MAE) of 21.462 fruits and an R^2 value of 0.577, indicating moderate agreement between predicted and true yield counts. The average bias was negligible (-0.08 fruits), suggesting no systematic overestimation or underestimation across the dataset. However, the regression slope ($y = 0.787x + 26.020$) indicates compression of the dynamic range, with a tendency toward underestimation at higher yield levels and slight overestimation at lower yields. This behavior is consistent with dense canopy conditions, were fruit occlusion by branches and foliage limits visibility from individual viewpoints. In some cases, temporary occlusions may also contribute to localized overestimation when tracked fruits are momentarily lost and subsequently re-identified as new instances. Although each tree was recorded from two opposing sides, and the maximum count was used to reduce double-counting, portions of the canopy remained partially concealed, thereby constraining the accuracy of yield estimation.

In contrast, yield-loss estimation performed substantially better. The model achieved a low MAE of 4.259 fruits and a high R^2 value of 0.913, indicating strong agreement between predicted and observed loss counts. The small positive bias (3.741 fruits) suggests a slight overestimation, likely due to occasional misclassification of red and white marking tape as defective fruits. Despite this, the overall accuracy indicates that the proposed framework is particularly effective at quantifying cracking-related yield loss, where objects are primarily on the ground and are less affected by occlusion. Overall, the results indicate that while yield estimation is inherently constrained by canopy structure, the framework provides highly reliable yield-loss estimates, supporting its use for operational assessment of cracking-related losses in pomegranate orchards.

Table 20. Counting performance metrics for yield and loss estimation

Model	MAE	Bias	R^2
Yield	21.462	-0.077	0.577
Loss	4.259	3.741	0.913

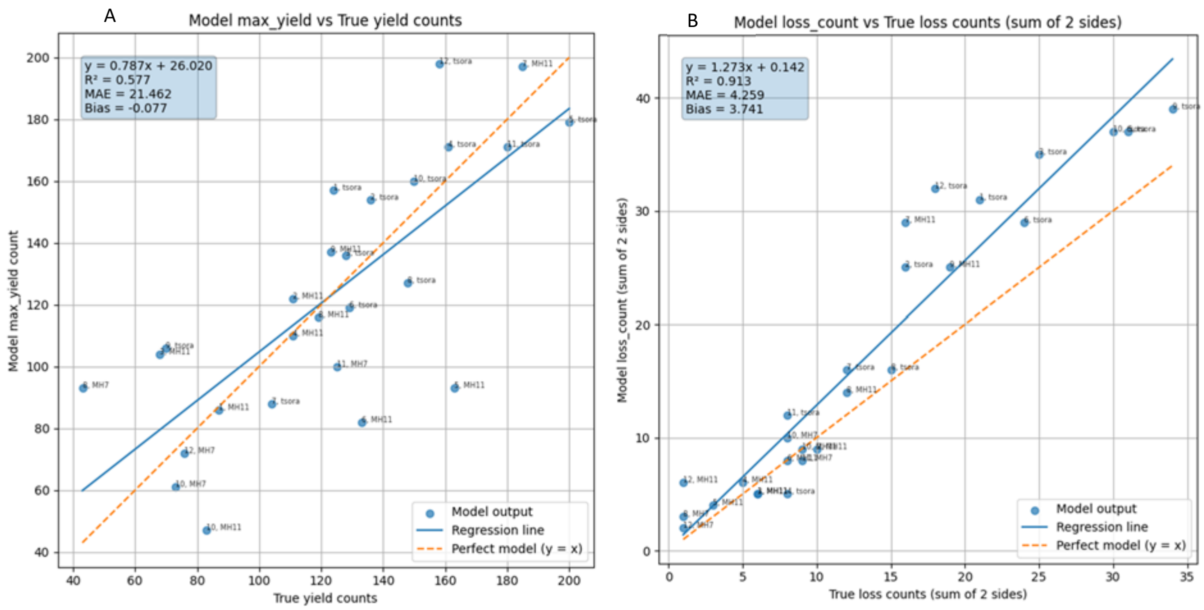


Figure 46. Comparison between model-predicted and ground-truth fruit counts at the tree level. (a) Linear regression of predicted yield counts versus observed yield counts, using the maximum value from the two viewing directions for each tree. Each point represents a single tree and is labeled by tree number and orchard. (b) Linear regression of predicted yield-loss counts versus observed loss counts, where loss estimates correspond to the sum of detections from both viewing directions for each tree. In both panels, the solid blue line represents the fitted regression, and the dashed line indicates perfect agreement (*y* = *x*).

Greece: Pomegranate Random Forest modeling results

Data acquisition

Image acquisition for pomegranate modelling in Greece was conducted during three consecutive cultivation seasons (2023-2025) in the Dalamanara area near Argos (37°38'59.8"N, 22°47'22.0"E) (more details in D3.2 & D3.4) (Fig. 47). In 2023 and 2024, UAV campaigns were performed over a 0.35 ha commercial orchard featuring the 'Wonderful' cultivar, with a consistent set of 9 trees monitored each year.

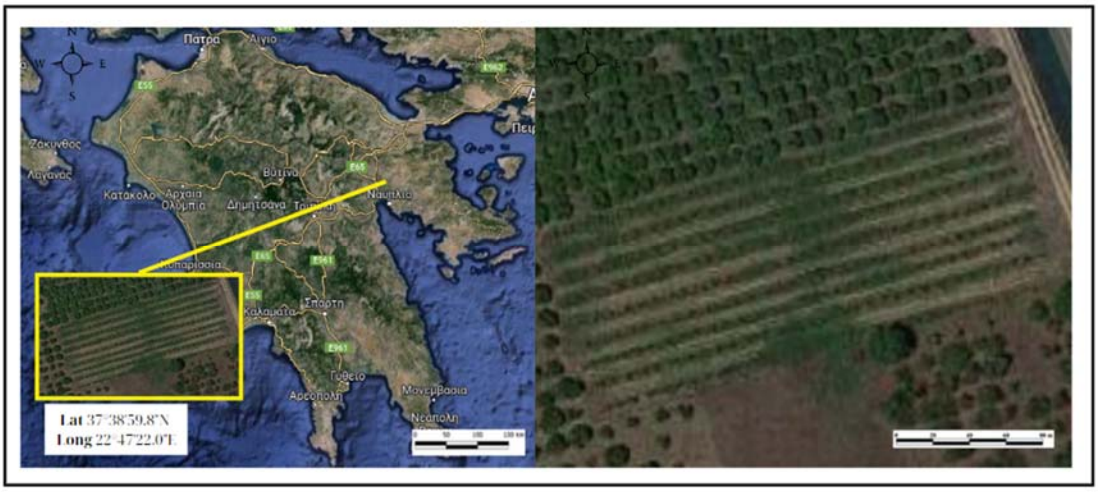


Figure 47. Experimental site location on a scale of 1:1,200,000 and the study area (Google Earth, 2024).

The experiment focused on three distinct irrigation treatments applied from March to October in both 2023 and 2024, including: (i) Standard irrigation (S), following the conventional local protocol (3 h at 40 m^3/h); (ii) Low Water Amount (LWA), representing a 30% reduction; and (iii) High Water Amount (HWA), a 20% increase over the standard protocol. These treatments were arranged in randomized blocks to explore the impact of water availability on yield and fruit cracking dynamics. In 2025, the study was expanded to three additional commercial fields (12 trees per field), providing increased spatial variability for model testing and validation (Fig. 48).

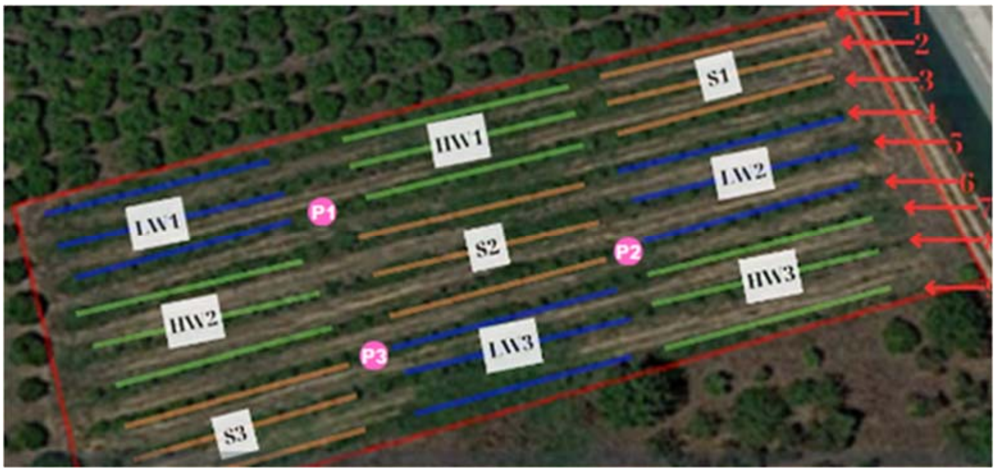


Figure 48. Experimental layout with randomized blocks for the three irrigation treatments, LWA (-30%, blue), HWA (+20%, green), and Standard (orange). Dots indicate the three additional trees monitored in 2025 as part of the first pilot site

UAV flights were carried out on 15 dates across the three seasons, targeting key phenological stages. Two UAV platforms were used: the DJI Mavic 3 Multispectral (for MS and RGB imagery) and the DJI Mavic Thermal. Flights were conducted between 11:00 and 13:00, ensuring solar elevation angles above 45° to minimize shadowing and illumination variability. For each mission, images were captured in nadir-view geometry using pre-calibrated multispectral bands and thermal sensors. Remote sensing data were complemented by in-field physiological and soil measurements, as well as meteorological data from the TOMMY platform, ensuring spatial and temporal alignment across datasets and enabling robust tree-level modelling of plant water status (PWS) indicators, including stem water potential (SWP), and stomatal conductance (SC) for predicting cracking rate and yield.

Dataset

The modelling dataset included 9 trees for training in each of 2023 and 2024. The modelling aimed to predict PWS indicators: SWP and SC, and by them to predict the yield, and fruit cracking rate incidence at the individual tree level using integrated physiological, meteorological, and remote sensing data.

Data inputs

As summarized in Table 21, the modelling incorporated four main input categories:

1. Vegetation indices: 16 spectral and derived indices (e.g., NDVI, SIPI, MSAVI, CW series).
2. Meteorological variables: daily weather parameters (e.g., air temperature, VPD, RH) from a local HC station.
3. Physiological measurements: SWP, SC, trunk and fruit growth, soil moisture and texture (30-60 cm).
4. Phenological data: recorded BBCH growth stage per sampling event.

All datasets were pre-processed for quality control, outlier removal, and temporal alignment across data modalities. Tree-level averages were calculated for each flight date and phenological stage for model training. UAV imagery was processed in Pix4D and QGIS to generate georeferenced orthomosaics and extract vegetation indices, aligned with per-tree shapefiles.

Table 21. Summary of the input features used for pomegranate yield and cracking prediction in Greek orchards (2023-2024).

Items	Features	Units	Type	Period
Vegetation Indices	NDVI, SAVI, GNDI, MSAVI, NDRI, IPVI, OSAVI, Clred, Clgreen, CW 475, CW 560, CW 668, CW 717, CW 842	-	Spatial Data	2023-2024
Meteorological data	HC Air Temperature (avg), Dew Point (avg), VPD (avg, min), HC Relative Humidity (avg), Precipitation (sum), Wind Speed (avg)	°C, °C, kPa, %, mm, m/s	Spatial Data	2023-2024
Physiological parameters	Stem Water potential, Soil Moisture, Sand (30cm),Silt (30 cm), Clay (30 cm), Sand (60 cm), Silt (60 cm), Clay (60cm), Fruit Average diameter, Trunk Diameter, Stomatal conductance	Bar Bar, % mm, mm C/S/m	Spatial Data	2023-2024
Phenological parameters	Growth Stage	-	Spatial Data	2023-2024
Output	Yield Fruit cracking SWP Stomatal conductance	Kg % bar C/S/m	Spatial Data	2023-2024

Results

The performance of the Random Forest model for pomegranate cracking rate (%), yield, SWP, and SC **estimation was evaluated using UAV data from 2023-2024.**

Stem Water Potential (SWP) prediction model - results

Model performance: SWP prediction was performed using a Random Forest (RF) model based on meteorological and multispectral inputs collected across the 2023 and 2024 seasons. The RF model yielded a test R^2 of 0.41 and an RMSE of 0.22 bar. Although overall predictive performance is moderate, this study demonstrates the feasibility of estimating tree-level SWP using remote sensing and weather data in the pomegranate context.The very high value of RMSE in the training set (10 Mpa) as opposed to the moderate value of RMSE in the test set (0.22 Mpa), implies that the data set needs further cleaning from extreme outliers (10 Mpa is not a physically acceptable value for SWP).

Table 22. Model performance summary for Random Forest-based SWP prediction (2023–2024).

Set	R^2	RMSE	n_trees	m_try
Train	0.31	10.07	10	5
Test	0.41	0.22	10	5

Feature importance (coefficient): Variable importance analysis (Figure 50) revealed that Relative Humidity, CW_560 reflectance, and Air Temperature were the dominant predictors of SWP variation. Other influential features included NDVI, Year, and bands CW_717 and CW_668, which relate to plant vigor and water status. Temporal and categorical variables, such as Month, type, and EVI, made limited contributions. This ranking reinforces the biological relevance of meteorological drivers in explaining short-term variation in stem water potential at the tree level.

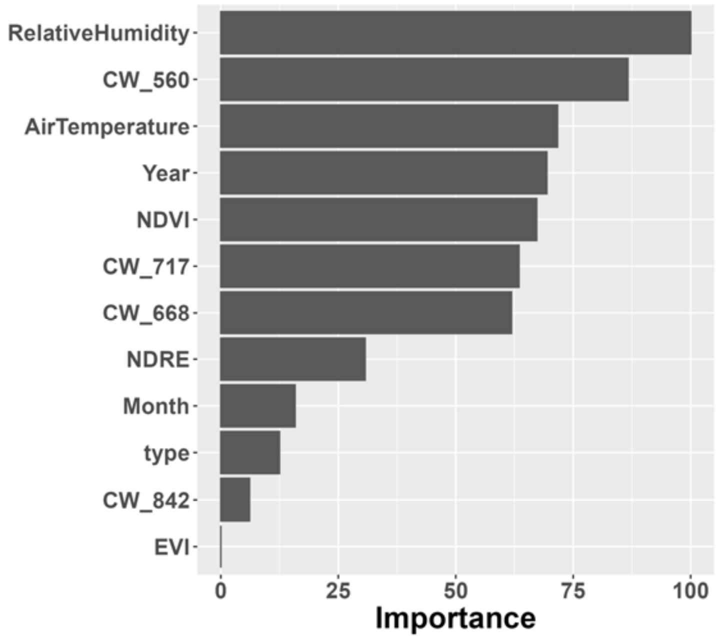


Figure 49. Variable importance plot from the Random Forest model trained to predict SWP using 2023-2024 data. Relative Humidity, CW_560, and Air Temperature are the top contributors

Summary of the SWP model:

Overall, this modeling effort highlights the feasibility of using meteorological and remote sensing data to estimate SWP at the tree level, providing a foundation for scalable monitoring of plant water status in commercial orchards. However, the model's moderate R^2 suggests that further improvements could be achieved by increasing the size and variability of the training dataset, integrating additional physiological variables, or exploring ensemble modeling approaches.

Stomatal conductance (SC) prediction model - results

Model performance: Stomatal conductance was predicted using a Random Forest model trained on meteorological and multispectral inputs collected during the 2023 and 2024 seasons. The model was trained on a dataset combining physiological observations and UAV-derived features and tested on an unseen subset of tree-level data. The Random Forest model yielded a test R^2 of 0.33, with an RMSE of 181.88 mmol/m²/s. The training performance was stronger (R^2 = 0.89, RMSE = 46.46 mmol/m²/s), indicating some overfitting. Despite this, the model shows potential for estimating SC at the tree level using remote sensing and weather data.

Table 23. Model performance summary for Random Forest-based SC prediction (2023-2024).

Set	R^2	RMSE	n_trees	m_try
Train	0.89	46.46	15	2
Test	0.33	181.88	15	2

Prediction vs Observations: The scatterplot (Figure 51) presents the relationship between observed and predicted SC values for both training and test datasets. The model generally follows the trend of SC variation but shows limited predictive capacity under extreme SC values. The clustering of points in the low-to-mid SC range indicates reasonable performance under common conditions, while a few high-value outliers reveal the model’s difficulty in capturing extreme physiological responses.

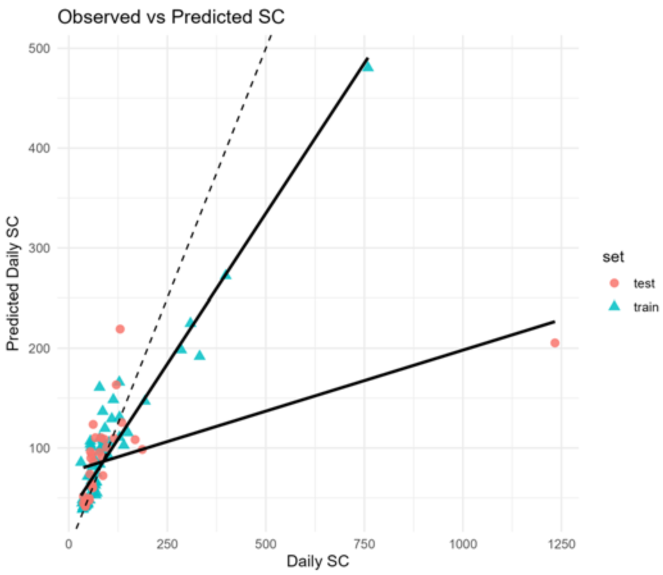


Figure 50. Observed vs. predicted SC values for training and test sets using the Random Forest model

Feature importance (coefficient): Variable importance analysis (Figure 52) highlights that CW_560 reflectance, Year, and EVI were the most influential predictors of SC. Additional relevant features include Air Temperature, Month, and NDRE, which are all indicators of environmental stress or vegetation status. In contrast to the SWP model, Relative Humidity played a smaller role in this model, possibly reflecting differences in physiological control mechanisms governing stomatal behavior.

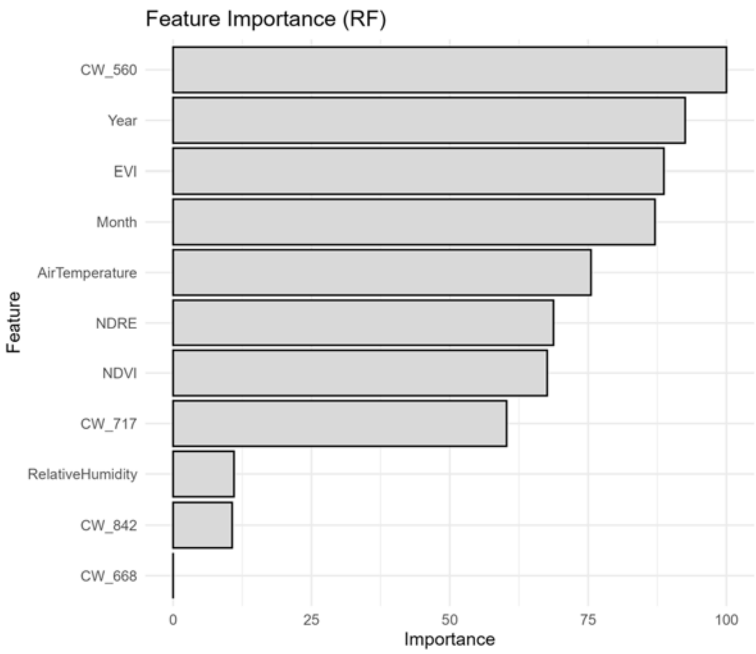


Figure 51. Variable importance plot indicating the most influential features for SC prediction. CW_560, Year, and EVI are top contributors.

Summary of SC model:

Overall, these findings support the potential of remote sensing and weather data to approximate SC dynamics but emphasize the need for larger and more balanced datasets, alternative modeling techniques, or the integration of physiological measurements (e.g., PAR, VPD) to improve model generalizability

Yield prediction model - results

Model performance: Yield prediction was conducted using a Random Forest model trained on UAV-based vegetation indices and meteorological variables collected during the 2023 and 2024 growing seasons in the Greek pomegranate orchards. The model was trained on multi-season data and evaluated on an unseen subset of trees to assess generalization capacity. The Random Forest model yielded a test R^2 of 0.089 and an RMSE of 15.30 kg, indicating low predictive power on unseen samples despite strong performance on the training data (R^2 = 0.85, RMSE = 6.79 kg). This discrepancy suggests potential overfitting, likely due to the complexity and variability of yield-related processes in the orchard environment.

Table 24. Model performance summary for Random Forest-based yield prediction (2023-2024).

Set	R^2	RMSE	n_trees	m_try
Train	0.85	6.79	15	2
Test	0.089	15.30	15	2

Prediction vs. Observations: The scatterplot (Figure 53) shows observed vs. predicted yield values for both the training and test subsets. While the model aligns well with the training data, there is considerable scatter in the test data, indicating limited ability to generalize yield predictions to unseen trees. The dispersion around the line of equality reflects variability in field conditions not fully captured by the input features.

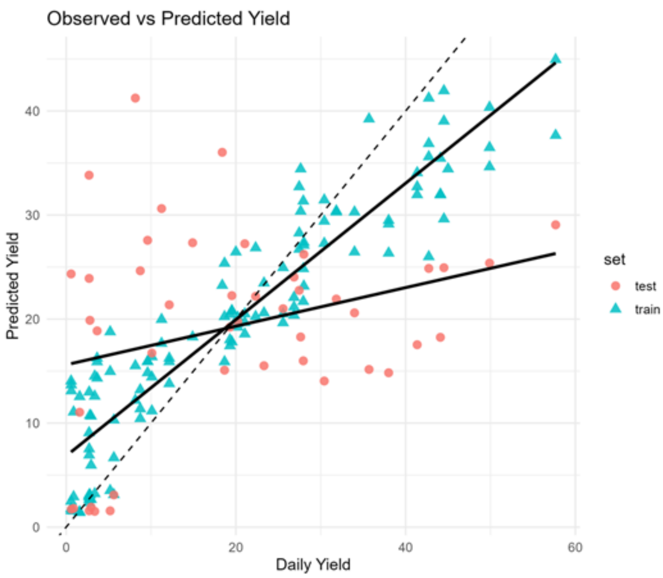


Figure 52. Observed vs. predicted yield values using the Random Forest model. The model fits the training data closely but shows substantial deviation on the test set, with increased error at higher yield levels

Feature importance (coefficient): Variable importance analysis (Figure 54) highlights that MSAVI, IPVI, and NDVI were the most influential predictors of pomegranate yield. These vegetation indices are known to reflect canopy vigor, greenness, and photosynthetic activity, aligning well with their relevance to fruit development. Other indices, such as OSAVI, GNDI, and SAVI, contributed to a lesser extent. NDRE had a negligible impact. These findings suggest that yield variability is more closely tied to structural and physiological canopy characteristics than to meteorological variation under the given conditions.

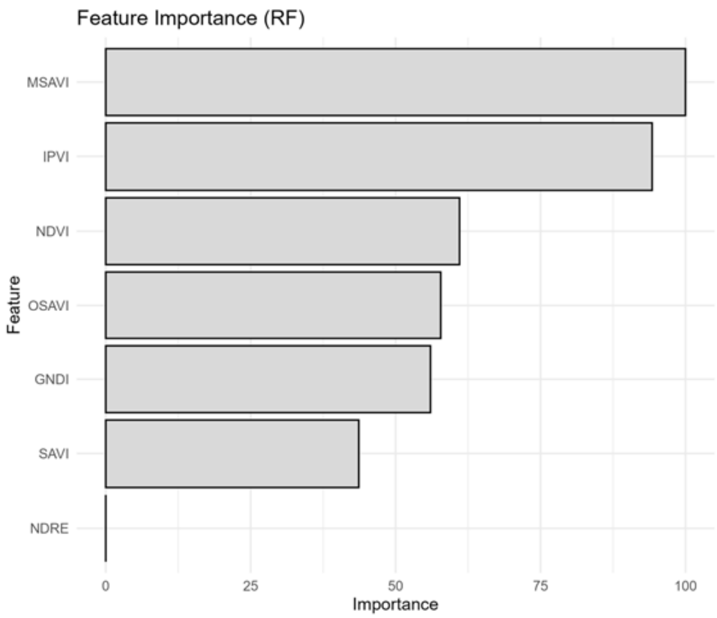


Figure 53. Variable importance plot from the Random Forest yield model (2023-2024). MSAVI and IPVI were the dominant predictors, followed by NDVI and other vegetation indices

Summary of yield model: The Random Forest model aimed to estimate fruit yield at the tree level using multispectral UAV indices and environmental data. While training performance was strong, test results showed limited generalization, as reflected by low R^2 and high RMSE values. This outcome indicates that, under the current input configuration, tree-level yield prediction remains challenging. The dominant predictors were MSAVI and IPVI, consistent with literature emphasizing the role of canopy density and vegetation vigor in yield formation. However, the relatively poor performance on the test set suggests that additional factors, such as management practices, soil fertility, or unmeasured physiological variables, may play a significant role in yield variation.

Overall, the model demonstrates the potential for remote sensing-based yield estimation but also highlights the need for further refinement, such as increasing sample size, enriching the feature set, and improving alignment between predictor and yield measurement dates.

Fruit cracking prediction model – results

Model performance: Fruit cracking prediction was performed using a Random Forest model trained on multi-source inputs, including physiological parameters (e.g., fruit and trunk diameter, stomatal conductance), meteorological data (e.g., rainfall, air pressure), and vegetation indices (e.g., NDVI, MSAVI, Clred, Clgreen). The model was developed using data from the 2023 and 2024 cultivation seasons in the Greek pomegranate use case and evaluated using an unseen subset of trees. The Random Forest model achieved a test performance of R^2 = 0.069 and RMSE = 13.31, indicating low predictive accuracy in generalizing to unseen samples, despite high performance on the training set (R^2 = 0.91, RMSE = 6.20). This suggests potential overfitting, which limits the model's applicability to broader, unseen datasets.

Table 25. Model performance summary for Random Forest-based fruit cracking prediction (2023-2024).

Set	R^2	RMSE	n_trees	m_try
Train	0.91	6.20	85	2
Test	0.069	13.31	85	2

Prediction vs Observations: Model predictions were compared with observed fruit cracking values for both training and test datasets. As illustrated in Figure 55, the scatterplot reveals that while the model captures the trend within the training set, it struggles to generalize to new data, reflected in the flattened regression line for test data.

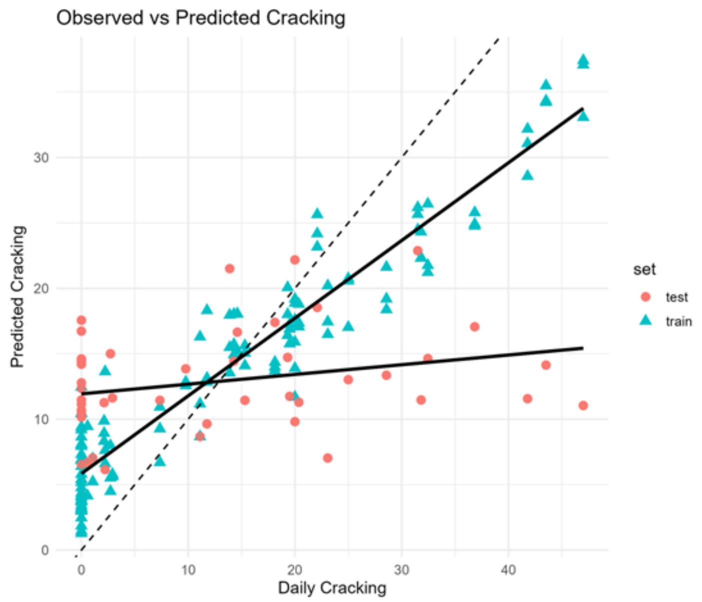


Figure 54. Observed vs. predicted fruit cracking using the Random Forest model (2023-2024). Scatterplot showing model predictions against observed values for both the training (blue triangles) and test (red circles) sets. The model captures general trends within the training data but underperforms on the test set, indicating limited generalization

Feature importance (coefficient): Variable importance analysis (Figure 56) identified fruit diameter and trunk diameter as the top predictors of fruit cracking, aligning with the physiological understanding of fruit development and susceptibility to cracking. Vegetation indices such as NDVI, Clred, IPVI, and MSAVI also contributed meaningfully, reflecting the plant's vigor and health status. Meteorological variables such as air pressure and rainfall, along with leaf temperature, further supported predictions, suggesting that environmental stress factors also influence cracking.

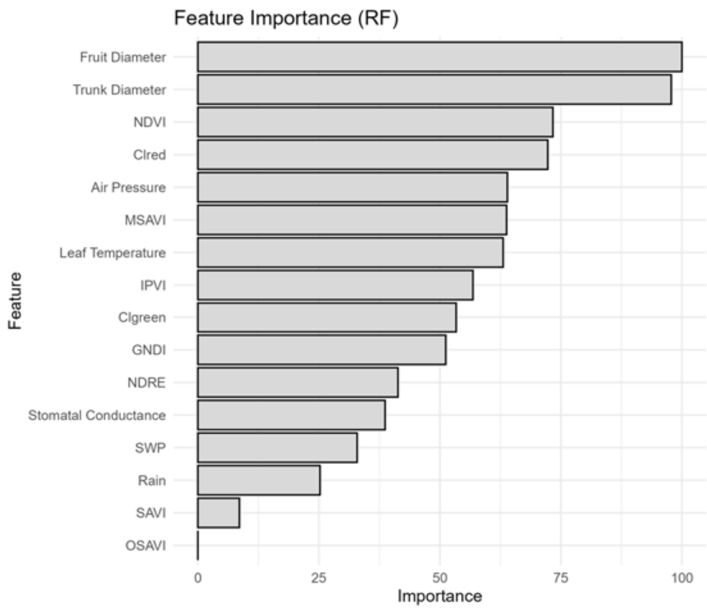


Figure 55. Feature importance plot from the Random Forest fruit cracking model. Fruit diameter and trunk diameter are the most influential features, followed by vegetation indices (NDVI, Clred, MSAVI), meteorological parameters (air pressure, rainfall), and physiological variables such as stomatal conductance

Summary of fruit cracking model: The Random Forest model developed for predicting pomegranate fruit cracking at the tree level demonstrated strong performance on the training data, yet only marginal generalization to unseen test data. The model relied heavily on structural and physiological parameters, particularly fruit and trunk diameter, indicating that plant-specific traits strongly affect cracking susceptibility. Additionally, reflectance-based vegetation indices and weather-related features provided auxiliary information on plant health and external stressors. However, the limited accuracy on the test set suggests challenges in generalizing the model, potentially due to dataset size, variability in field conditions, or a need for more temporally aligned observations.

Overall, while the current model outlines important biological and environmental predictors of fruit cracking, further efforts are required to improve generalization. This includes increasing the training dataset size, improving temporal alignment of variables, and exploring more robust or hybrid modeling techniques to capture the complex dynamics behind cracking incidence in pomegranates.

Sweet Cherry - Predictive Modelling based on UAV Remote Sensing and Weather data:

Study Area:

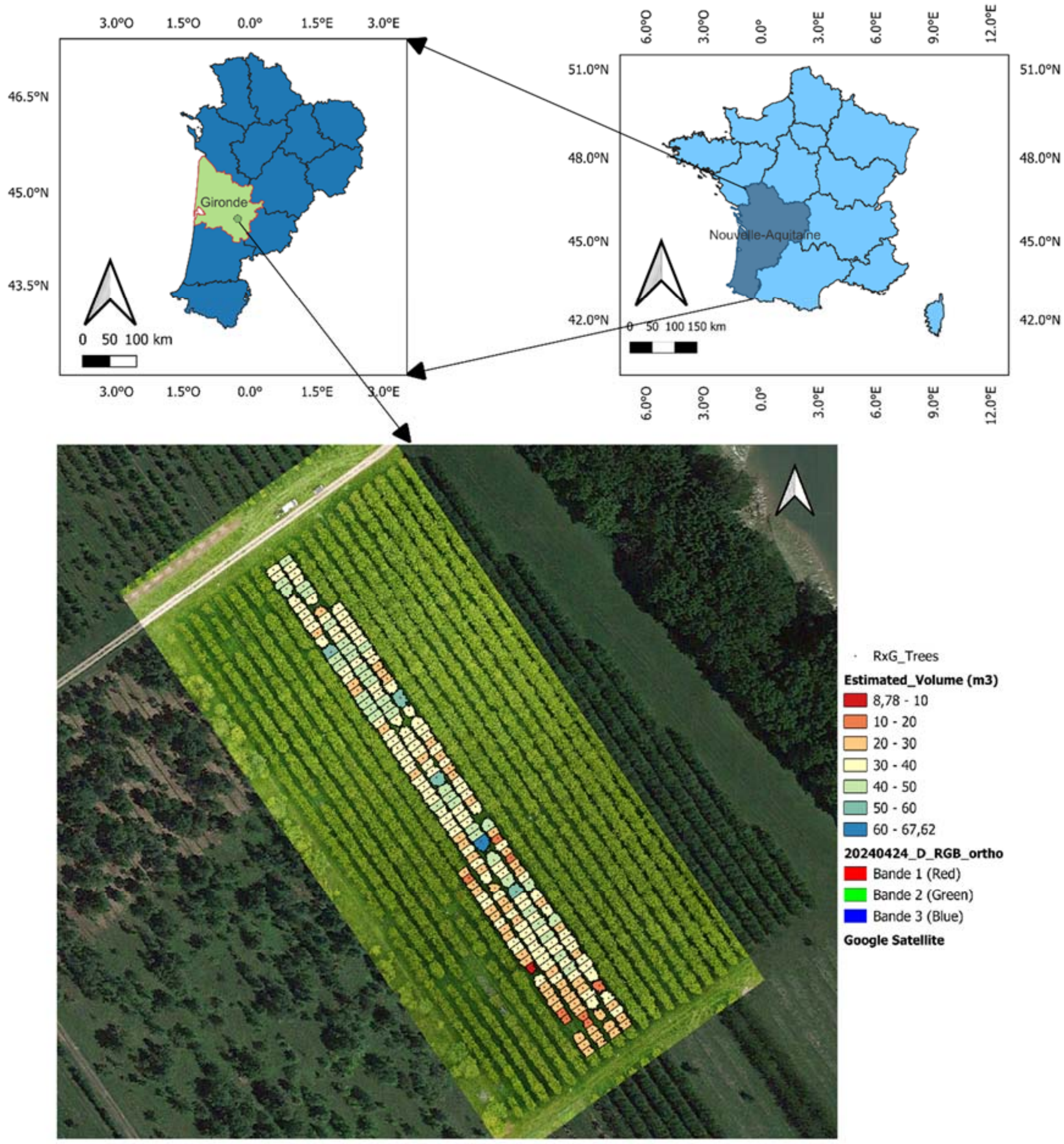


Figure 56. Study Area Map

The segregating populations plot experiment is located at Toulenne, 40 km south-east of Bordeaux (Gironde, Nouvelle Aquitaine). The site is very close to the Garonne River, in an area of grapevine production, specifically Sauternes liquor wine. The specificity of this area lies in the presence of the Ciron river, a small tributary of the Garonne, which contributes to frequent fog, favoring the development of noble rot on grapevines. This type of microclimate is also highly convenient for studying fruit cracking in sweet cherry. The plot contains several populations (F_1), but only one is studied within CrackSense: 'Regina' × 'Garnet', consisting of 117 full-sib hybrids planted in two blocks, along with the two parental accessions. The area under study is 3000 m^2.

Stem Water Potential (SWP) Prediction

For sweet cherry, we apply the same general approach as for citrus, using similar steps for data preprocessing, variable selection, and modelling.

Data context and SWP measurement conditions

- **2024 campaign**

In 2024, four flights were conducted between late April and late June, but only the last three included SWP measurements and are therefore used for the prediction analysis.

During these three flights, SWP was measured on 12 trees on the flight date and on 15 trees in each of the following two, resulting in a total of 42 samples.

Although this is a relatively small dataset for prediction and machine learning, it still provides a basis to explore the feasibility of SWP estimation in sweet cherry.

Dataset construction and data preparation

For the sweet cherry dataset, we used the same feature engineering approach as for citrus (see sections above), combining UAV remote sensing variables (MS + IRT) with weather features, including 3-day aggregated meteorological data before each flight. The only exception is the GDD (growing degree days) variable, which was removed because appropriate temperature thresholds for sweet cherry were not available in the literature.

The resulting sweet cherry dataset contains 42 samples and 72 variables. As with the citrus dataset, the first step is to remove outliers using the same IQR-based method.

Data cleaning and outlier removal

	Flight date	n	SWP_mean	SWP_std	Q1	Q3	IQR	lower_bound	upper_bound	n_outliers
0	2024-05-16	12.0	-11.913500	0.656401	-12.2505	-11.389	0.8615	-13.2843	-10.3552	0
1	2024-06-05	15.0	-17.816267	2.179062	-19.3810	-16.329	3.0520	-23.0434	-12.6666	0
2	2024-06-25	15.0	-19.107933	2.097843	-19.8850	-18.370	1.5150	-21.7030	-16.5520	2

Exemples d'outliers détectés :

	_row_id	Site	TreeID	Study year	Flight date	SWP	lower_bound	upper_bound
33	33	Toulenne	FRsweetcherry_RG68_X10Y65	2024	2024-06-25	-23.560	-21.703	-16.552
30	30	Toulenne	FRsweetcherry_RG67_X8Y38	2024	2024-06-25	-14.392	-21.703	-16.552

Figure 57. Results of outlier removal using the IQR method

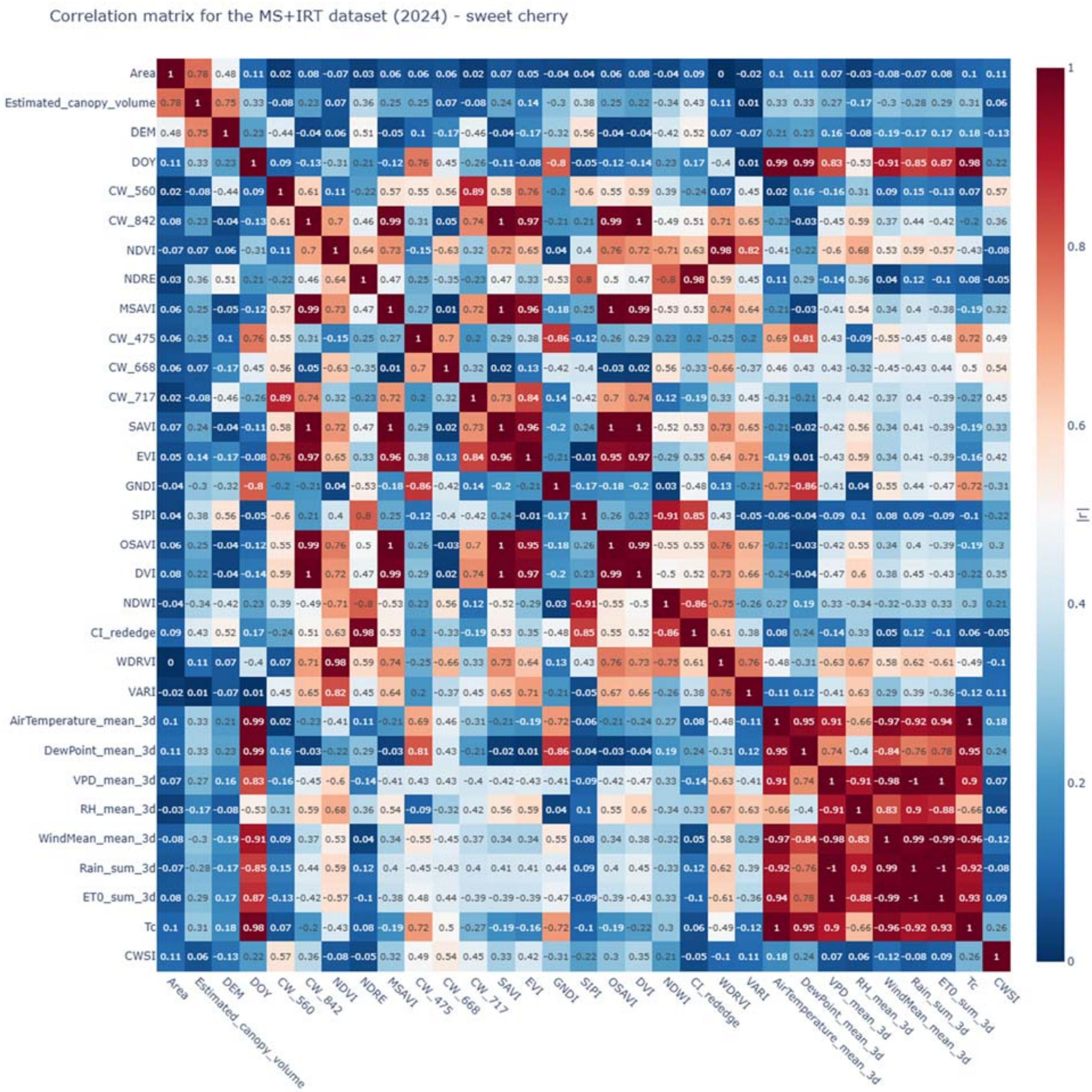
The IQR-based analysis shows that SWP values were generally consistent across flights. No outliers were detected for the first two dates, while two trees from the last flight fell outside the acceptable range and were removed.

Outliers were identified using the Interquartile Range (IQR) method with a threshold of **1.2 × IQR** (*the same as in the citrus analysis*), chosen to apply a slightly stricter rule given the small dataset. The acceptable interval was defined as **(Q1 - 1.2 × IQR, Q3 + 1.2 × IQR)**, with Q1 and Q3 being the first and third quartiles. Values outside this range were removed to avoid bias in model training and evaluation.

Feature selection based on correlation analysis

The same correlation-based feature filtering procedure as for citrus was applied. A Pearson correlation matrix was computed for all numerical predictors, and pairs of variables with very high absolute correlation ($|r| > 0.90$) were considered redundant, with only one variable retained in each pair.

The resulting correlation matrix is shown below:



After applying the correlation-based filtering procedure, 16 non-redundant features were retained for

Figure 58. Correlation matrix for the MS-only dataset (2023–2024)

modelling:

['Area', 'Estimated_canopy_volume', 'DEM', 'Tc', 'CW_560', 'CW_842', 'NDVI', 'NDRE', 'CW_475', 'CW_668', 'CW_717', 'GNDI', 'SIPI', 'VARI', 'VPD_mean_3d', 'CWSI']

This reduced set preserves the most informative predictors while eliminating highly correlated variables that could introduce multicollinearity.

Machine learning modelling framework

The same modelling workflow as in the citrus analysis was followed, but given the small size of the sweet cherry dataset, only a limited set of models was evaluated. In total, four algorithms were trained and assessed: one tree-based model (**Random Forest**) and three linear regularized models:

- **Lasso Regression:** a linear model using L1 regularization, which penalizes the absolute value of coefficients; this tends to push some coefficients exactly to zero, enabling feature selection.
- **Ridge Regression:** a linear model using L2 regularization, which penalizes the squared magnitude of coefficients; this shrinks them smoothly and helps stabilize the model when predictors are correlated.
- **Elastic Net:** a hybrid model combining both L1 and L2 penalties, offering a balance between sparsity (L1) and stability (L2).

These models were integrated into the same evaluation pipeline used for citrus to ensure full methodological consistency.

Table 26. Model performance comparison – sweet cherry

Model	R^2 train (mean)	R^2 train (std)	R^2 test (mean)	R^2 test (std)	RMSE train (mean)	RMSE train (std)	RMSE test (mean)	RMSE test (std)	n folds
Lasso	**0.860**	**0.032**	**0.803**	**0.047**	**1.261**	**0.172**	**1.494**	**0.262**	**3**
Ridge	0.834	0.035	0.793	0.095	1.372	0.172	1.513	0.444	3
ElasticNet	0.866	0.028	0.787	0.048	1.232	0.143	1.547	0.225	3
RandomForest	0.937	0.011	0.755	0.062	0.849	0.095	1.663	0.314	3

Model Performance

Across all models, performance remained consistent, with **Lasso** achieving the best overall generalization (highest mean **R^2=0.803** and lowest **RMSE=1.494** on the test folds). Although Random Forest obtained the highest training accuracy, its lower test performance suggests mild overfitting, indicating that simpler linear models may be better suited to the limited size of the sweet cherry dataset.

Feature importance (coefficients)

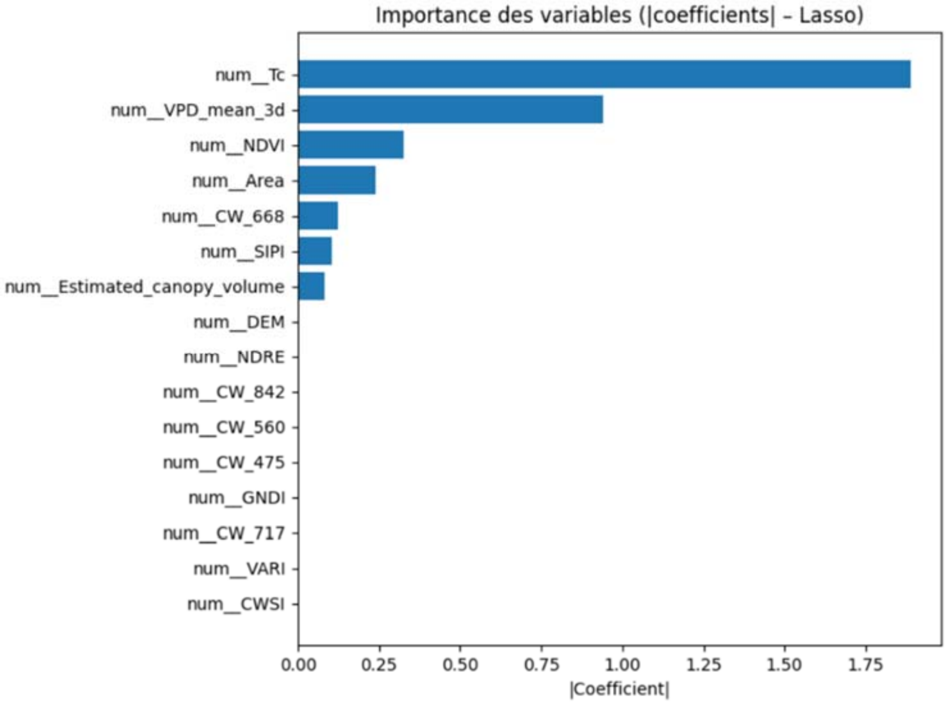


Figure 59. feature importance - LASSO Regression

Lasso identifies **Tc** as the strongest predictor of SWP, with **VPD_mean_3d** also contributing notably. A few spectral and canopy-related variables (e.g., **NDVI, Area, SIPI**) have moderate influence, while the remaining features show negligible coefficients, indicating limited relevance for SWP prediction in this small dataset.

Prediction vs Observations:

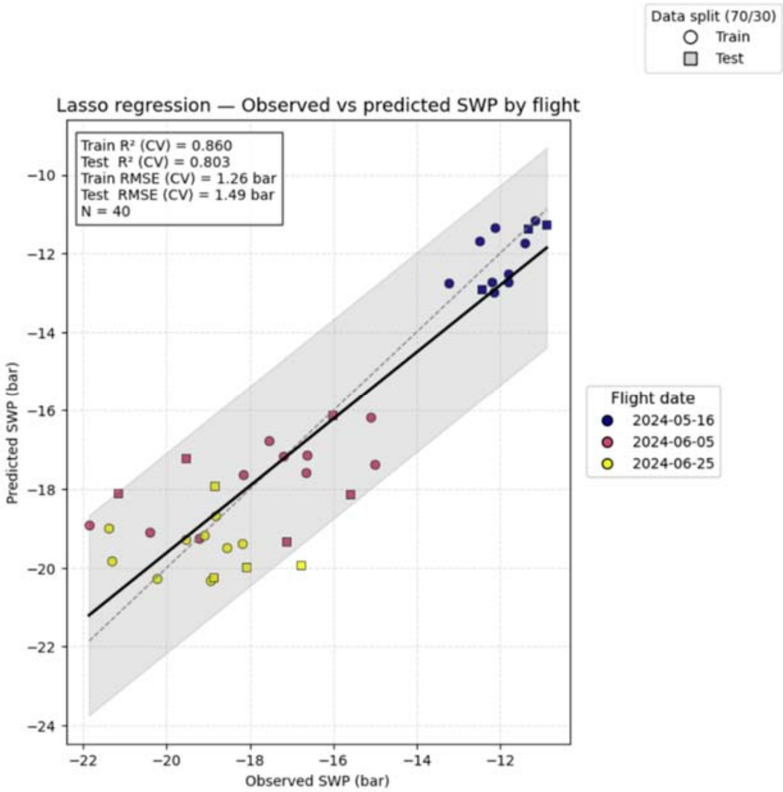


Figure 60. Lasso Regression - Observed vs predicted SWP using an illustrative 70/30 TreeID-based split

As in the citrus SWP subsection, we generated an illustrative 70/30 train-test split stratified by TreeID to visualise how the Lasso model behaves on an independent subset. This split is not intended for quantitative performance assessment but to show the relationship between observed and predicted SWP values. The scatterplot highlights that the model generally captures the overall trend, although some variability remains, especially in intermediate SWP ranges.

Cracking Prediction:

Reminder of the cracking evaluation protocol:

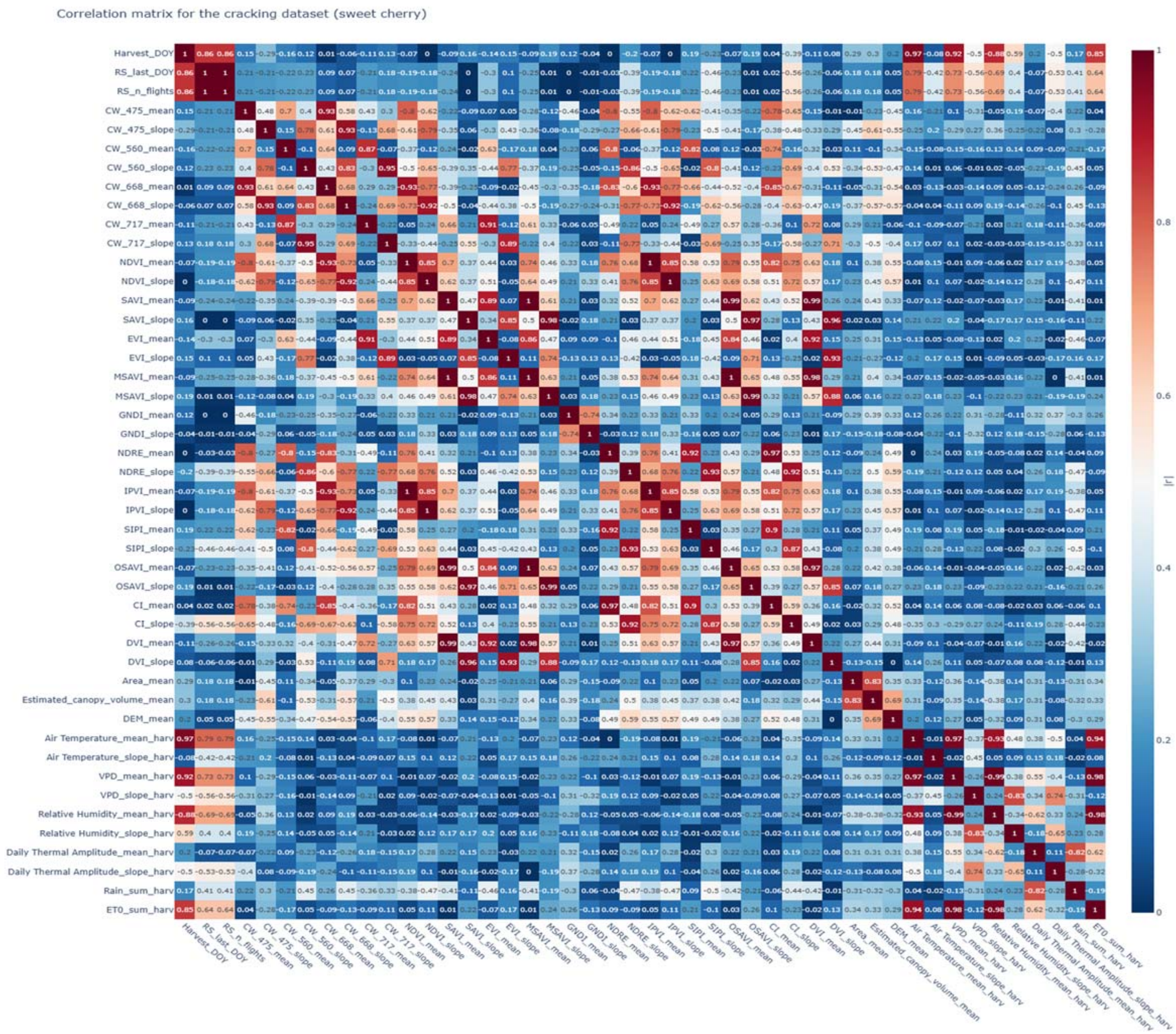
To evaluate sweet cherry cracking, we examined 44 trees and collected 50 fruits from each. The sample size was limited to 50 fruits per tree because only specific branches were protected with insect-proof nets against Drosophila suzukii. As a result, the cracking dataset comprises 44 tree-level observations, which is a relatively small sample size for statistical analysis and modelling. However, we nevertheless proceed to explore whether meaningful predictive patterns can still be identified.

Dataset construction & feature engineering

Remote sensing & weather data

Remote sensing features were computed from all UAV flights occurring **at least seven days** before each tree's harvest date, extracting both canopy condition (e.g., **NDVI, GNDI, MSAVI, structural features**) and its temporal evolution through slope indicators. Weather conditions were summarized over the 10 days preceding harvest, including variables such as air temperature, VPD, relative humidity, and rainfall, along with their short-term trends. All these features were merged with tree-level harvest data to produce a complete dataset for cracking prediction.

Feature selection based on correlation analysis



As in the previous sections, a correlation-based filtering step was applied to limit redundancy among predictors. All numerical features were included in a Pearson correlation matrix, and variables forming highly correlated pairs **(|r| > 0.85)** were considered redundant. By iteratively removing the second variable from each highly correlated pair, 20 features were retained from the initial 46, ensuring that only representative, non-collinear variables were retained for modelling.

From regression to classification

Initial regression attempt

```
===== RÉCAP GLOBAL (tous les modèles, cracking cerise) =====
          Model  R2_train_mean  R2_train_std  R2_test_mean  R2_test_std
0    ElasticNet       0.379903      0.160529     -0.373990     0.090744
1         Lasso       0.408668      0.148736     -0.510521     0.283931
2  RandomForest       0.744004      0.120179     -0.153142     0.086359
3         Ridge       0.445140      0.170895     -0.470337     0.223742
```

Figure 61. Model Performance - Regression

As we did previously for citrus, we first treated cherry cracking prediction as a regression problem, using the engineered remote sensing and weather features to model the number of cracked fruits per tree. Four algorithms were evaluated within a nested cross-validation framework (Lasso, Ridge, Elastic Net, and Random Forest). While Random Forest achieved the best fit on the training data, all models yielded negative mean R^2 values on the test folds. This indicates that, given the limited sample size (44 trees) and the noise inherent in cracking measurements, the regression models were unable to generalize reliably and failed to capture a stable quantitative relationship between predictors and the exact number of cracked fruits.

Binary "extremes-only" classification

Given the poor generalization of the regression models, the cracking prediction task was reformulated as a binary classification problem, following the same strategy as in the citrus analysis. Instead of predicting the exact number of cracked fruits, trees were grouped into two classes representing low and high cracking severity, based on the distribution of the cracking indicator (with intermediate values removed). This focuses the analysis on the most extreme cases and relaxes the requirement for precise numerical prediction, allowing us to assess whether the combined remote sensing and weather features still contain enough signal to distinguish between trees with low and high susceptibility to cracking.

Model performance

Table 27. Binary classification (low vs high)

Model	F1-macro_mean	F1-macro_std	F1-high	F1-low
RandomForest	**0.797**	**0.005**	**0.783**	**0.811**
LogisticRegression	0.696	0.174	0.693	0.698
LinearSVC	0.693	0.108	0.676	0.711
GradientBoosting	0.685	0.109	0.679	0.690
SVC_RBF	0.528	0.215	0.633	0.422

Across all tested models, classification performance remains moderate due to the small dataset (n = 30 trees). Random Forest clearly achieves the best generalization (F1-macro = 0.80), outperforming linear models and boosting approaches, and showing balanced performance across both classes. Linear models (Logistic Regression and Linear SVC) achieve similar intermediate accuracy (R^2=0.69), suggesting that part of the signal is linear, but the overall signal remains limited.

Feature importance (SHAP) – Random Forest

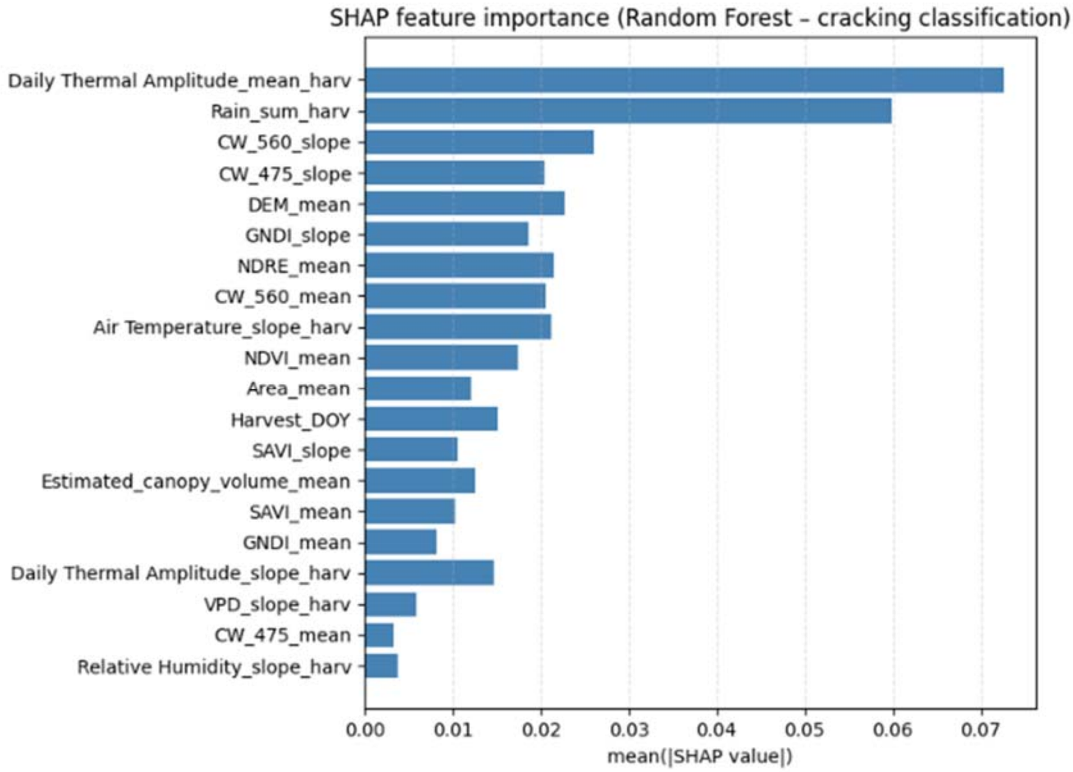


Figure 62. feature importance – RF

The SHAP analysis of the Random Forest model indicates that short-term weather conditions before harvest are the strongest predictors of cherry cracking. The most influential variable is the daily thermal amplitude during the days preceding harvest, followed by cumulative rainfall. While temperature fluctuations may contribute to fruit stress, rainfall remains the factor most clearly linked to cracking in the literature, as water uptake through the skin and microcracks increases internal pressure inside the fruit. Drone-based spectral features also play a role but remain secondary compared with these weather variables. Because the dataset is small (n = 30), these results should be treated as preliminary and need to be confirmed with larger datasets

Grapevine irrigation experiment plots

Grapevine irrigation experiment plots in Israel

In grapevines, cracking typically occurs near harvest, although it may also manifest postharvest, primarily due to storage conditions (Ramteke et al., 2017a). Abrupt changes in berry water status can lead to rapid increases in internal turgor pressure that exceed the mechanical tolerance of the berry skin and cuticle, resulting in fissuring and cracking (Griesser et al., 2024; Gariglio et al., 2024). These skin ruptures accelerate dehydration and microbial decay, necessitating costly sorting or culling and translating into tangible economic losses at both grower and supply-chain levels. Beyond direct yield penalties, cracking degrades sensory quality and shelf life, thereby affecting winery throughput and product consistency. Cracking arises from the interaction of genetic susceptibility (cultivar or clone), berry morphology and cuticle integrity, and dynamic environmental drivers, particularly water availability and microclimatic conditions. Rapid fluctuations in soil moisture, such as rainfall or irrigation following dry periods, high relative humidity, and large vapor pressure deficits (VPD), can all alter berry water influx and increase skin strain. The timing of these stresses relative to sensitive phenological stages, especially from version to harvest, is critical, as identical water inputs may result in markedly different outcomes depending on developmental stage and prior stress history (Matthews et al., 1987; Ramteke et al., 2017b).

This study investigates the impact of contrasting irrigation strategies on grapevine water status, growth dynamics, and fruit quality in a vineyard located in Israel. Because vine water status lies at the core of the cracking mechanism, irrigation management represents a key operational lever for regulating cracking risk. Approaches such as regulated deficit irrigation (RDI) and carefully managed post-deficit rehydration aim to minimize large oscillations in berry turgor pressure while maintaining vine physiological function. The effective implementation of such strategies depends on the availability of reliable, field-scale indicators of vine water stress. Among these, SWP, SC, and TG are widely regarded as the reference physiological metrics, as they integrate soil water availability, atmospheric demand, and plant hydraulic status. Complementary high-frequency proxies, such as TG dynamics (including maximum daily shrinkage), soil moisture profiles at multiple depths, and microclimatic variables (air temperature, relative humidity, and VPD), provide additional, actionable information for irrigation decision-making. Linking these physiological indicators to fruit outcomes is central to precision irrigation management. If seasonal patterns of vine water stress, rather than isolated measurements, can be quantified and related to cracking severity and yield, growers can identify at-risk vines earlier, stabilize berry development, and improve harvest quality. Accordingly, the objectives of this study were to: (1) Characterize seasonal PWS using SWP, SC and TG patterns under contrasting irrigation regimes; (2) Determine whether distinct irrigation-induced water status profiles are associated with differences in cracking severity (weight of berry (final), splitting % (healthy clusters, and severity index) and yield (based total weight and cluster per vine); and (3) Develop predictive models capable of classifying vines into water-stress clusters using sensor-based measurements alone, thereby providing a framework for early detection of cracking risk and improved irrigation management.

Israel study site-Grapevine

Plot location and experimental design – Lachish vineyard

The irrigation experiments were conducted over two consecutive growing seasons (2023 and 2024) in commercial table grape vineyards located in the Lachish region, southern Israel. Lachish is Israel's largest table grape production area. It is characterized by semi-arid Mediterranean climatic conditions, making it well-suited for evaluating irrigation-driven variability in grapevine water status, yield, and fruit cracking. For each season, a separate vineyard plot within the same production area was selected to implement the irrigation treatments. Although the exact plot location differed between years, both sites shared similar soil type, cultivar, trellising system, and management practices. Figure 65 presents an aerial image of the vineyard plots and a schematic layout of the irrigation treatments.

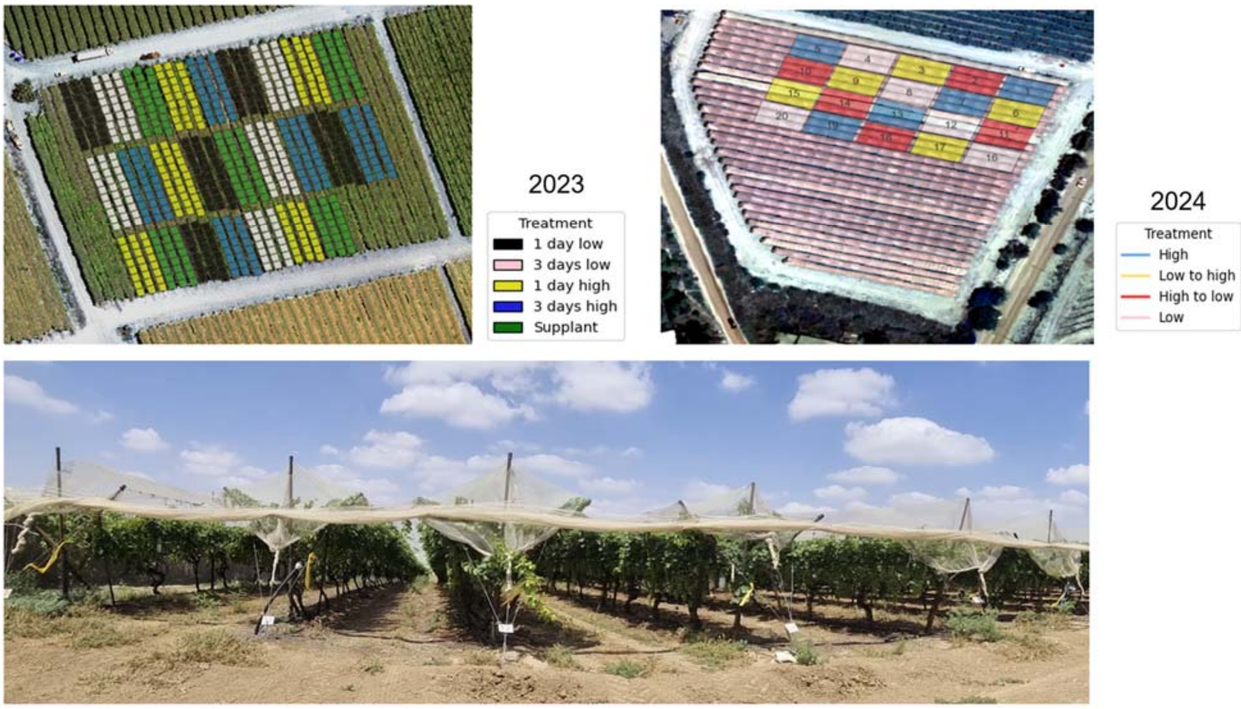


Figure 63 Experimental layout and irrigation treatments across two growing seasons. Top panels show UAV-derived orthomosaics of the experimental orchards in 2023 (left) and 2024 (right), with plots color-coded by irrigation regime. In 2023, treatments included short- and multi-day low- and high-irrigation pulses and a supplemental treatment, while in 2024 treatments represented contrasting seasonal water strategies (High, Low, Low-to-High, and High-to-Low). The bottom panel shows a ground-level view of the orchard infrastructure and tree rows, illustrating the plot arrangement and field conditions during the experiment.

2023 irrigation experiment

In 2023, irrigation was applied from 1 May to 20 October 2023. Five irrigation treatments were implemented, each replicated five times, to create controlled contrasts in irrigation frequency and water amount. The treatments were designed to disentangle the effects of high versus low seasonal water input and frequent versus infrequent irrigation events (1) 1d-high – High irrigation amount applied daily, corresponding to approximately 8,420–8,500 $m^3\ ha^{-1}\ year^{-1}$. (2) 3d-high – High irrigation amount applied every three days, totalling approximately 8,500 $m^3\ ha^{-1}\ year^{-1}$. (3) 1d-low – Low irrigation amount applied daily, totalling approximately 5,180 $m^3\ ha^{-1}\ year^{-1}$. (4) 3d-low – Low irrigation amount applied every three days, totalling approximately 5,210 $m^3\ ha^{-1}\ year^{-1}$. (5) SupPlant treatment – A variable irrigation regime determined by the SupPlant decision-support algorithm, which adjusted irrigation timing and amount dynamically based on sensor inputs and plant and environmental feedback. This experimental design allowed separating the effects of irrigation frequency (daily vs. every 3 days) from seasonal water volume and comparing them with a data-driven adaptive irrigation strategy.

2024 irrigation experiment

In 2024, the experiment was repeated in a different plot within the same Lachish region, maintaining comparable vineyard characteristics. The irrigation treatments were restructured to emphasize dynamic irrigation transitions over the season, thereby inducing contrasting temporal water-stress trajectories: (1) High – Consistently high irrigation applied daily throughout the season (approximately 796 mm $year^{-1}$). (2) High → Low – High irrigation applied daily during the early season, followed by a transition to irrigation every three days later in the season (approximately 532 mm $year^{-1}$). (3) Low → High – Reduced irrigation applied every three days during the early season, followed by a transition to daily irrigation later in the season (approximately 645 mm $year^{-1}$). (4) Low – Consistently low irrigation applied every three days throughout the season (approximately 376 mm $year^{-1}$). These treatments were designed to assess how early-season versus late-season water availability, as well as changes in irrigation regime, influence vine water status, growth dynamics, yield formation, and fruit cracking susceptibility.

Table 28. Irrigation treatments applied in the Lachish vineyard during the 2023 growing season

Treatment code	Irrigation strategy	Irrigation frequency	Seasonal water input (m^3 ha^{-1} yr^{-1})	Description/purpose
1d-high	High water amount	Daily	~8,420–8,500	High-frequency, high-volume irrigation
3d-high	High water amount	Every 3 days	~8,500	Lower frequency, same seasonal volume as 1d-high
1d-low	Low water amount	Daily	~5,180	High frequency, reduced seasonal volume
3d-low	Low water amount	Every 3 days	~5,210	Low frequency, reduced seasonal volume
SupPlant	Algorithm-based	Variable	Variable	Adaptive irrigation controlled by the SupPlant decision-support system

Experimental structure and data collection

Across both seasons, each irrigation treatment was replicated five times. Data collection was conducted at multiple temporal scales, including Continuous measurements of trunk growth dynamics and soil moisture, Monthly physiological measurements (e.g., SWP, SC), and UAV-based remote sensing campaigns, complemented by meteorological monitoring. This design created controlled and contrasting irrigation scenarios, enabling assessment of how sustained versus variable water-stress patterns affect grapevine physiology, yield, and cracking under commercial field conditions. A key limitation of the current experiment is that the vineyard plots were covered with protective nets, which partially obstruct the canopy and limit the direct visibility of vine structure and fruit from above. This netting affects the quality and interpretability of UAV-based remote sensing data, particularly for optical and LiDAR measurements, and constrains the ability to exploit UAV observations for detailed full-canopy analysis. As a result, UAV data in this experiment were primarily used to complement ground-based physiological measurements, while acknowledging reduced sensitivity relative to open-canopy conditions.

Table 29. Irrigation treatments applied in the Lachish vineyard during the 2024 growing season

Treatment code	Irrigation strategy	Irrigation frequency pattern	Seasonal water input (m^3 ha^{-1} yr^{-1}(	Description/purpose
High	Constant high irrigation	Daily	~7,960	Consistently high-water availability
High → Low	Decreasing irrigation	Daily early season → every 3 days later	~5,320	Early-season rehydration followed by a deficit.
Low → High	Increasing irrigation	Every 3 days, early season → daily later	~6,450	Early deficit followed by late rehydration
Low	Constant low irrigation	Every 3 days	~3,760	Sustained water deficit

Results

Climate and irrigation effects on the measured PWS

Climatic data were collected from a nearby (1 km) meteorological station. Here we present daily data on the minimum, maximum, and average temperature (Fig. 66: minimal, maximal, and average vapor pressure deficit (Fig. 66); and average solar radiation (Fig. 66) during the entire irrigation period. Climatic data for the grapevine experimental sites during the 2023 and 2024 growing seasons are presented in Figure 66, including daily minimum, mean, and maximum air temperature, soil moisture at 20 cm depth, and the cumulative number of heatwave days. Both seasons were characterized by pronounced summer heat and extended periods of high atmospheric demand, typical of the Mediterranean climate of the Lachish region. In 2023, maximum air temperatures frequently exceeded 35–40 °C during mid to late summer, with multiple heatwave events occurring between July and September, as reflected in the steady increase in cumulative heatwave days. Soil moisture showed strong treatment-dependent dynamics, with higher and more stable values under high-irrigation treatments and marked declines under reduced-irrigation treatments, particularly during peak heat periods. This combination of high temperature and declining soil moisture created conditions conducive to elevated vine water stress during critical phenological stages.

In 2024, heatwave accumulation began later in the season but increased rapidly from June onward, ultimately reaching cumulative levels comparable to or exceeding those observed in 2023. Maximum temperatures again reached extreme values during summer, while soil moisture exhibited sharper temporal fluctuations, reflecting both transitions in irrigation regimes and intensified evaporative demand. Compared with 2023, the 2024 season featured a more abrupt onset of thermal stress, providing a contrasting climatic context for evaluating vine water status responses and irrigation strategies. Overall, the two seasons encompassed distinct yet complementary climatic stress patterns, enabling a robust assessment of grapevine physiological responses, yield formation, and cracking risk under variable thermal and soil-moisture conditions.

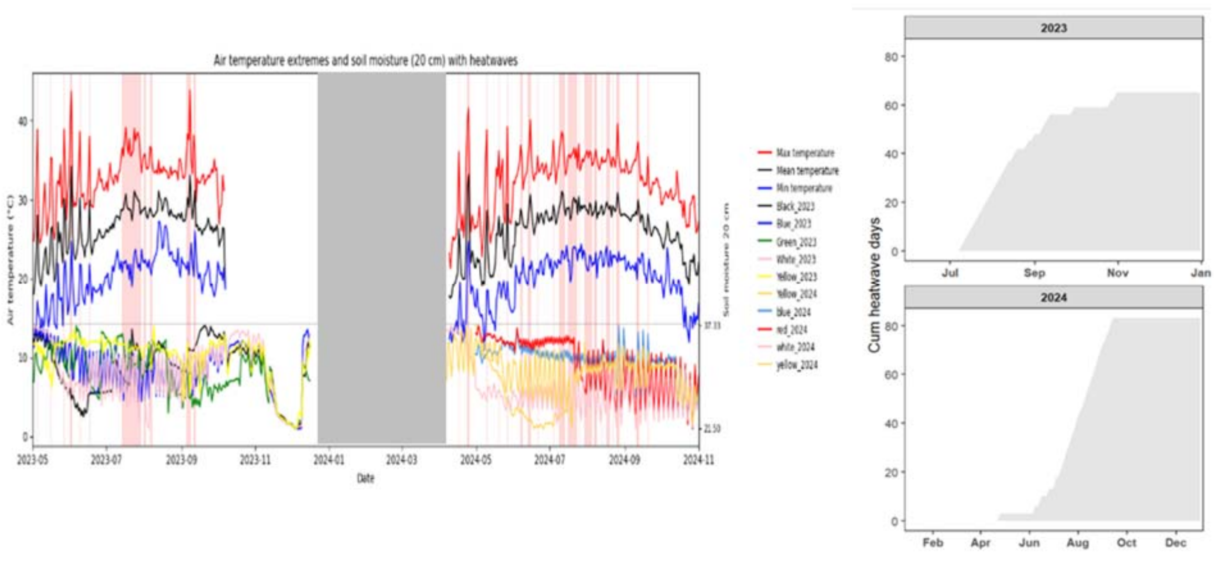


Figure 64 Seasonal patterns of air temperature, soil moisture, and heatwave occurrence during the 2023 and 2024 growing seasons. The left panel shows daily minimum, mean, and maximum air temperatures (°C) together with soil moisture at 20 cm depth for both years; shaded vertical bars indicate heatwave events. The central gray band denotes the non-growing (winter) period between seasons. Right panels show the cumulative number of heatwave days for 2023 (top) and 2024 (bottom), highlighting the substantially greater frequency and intensity of heatwaves in 2024. Together, the figure illustrates the contrasting climatic conditions between seasons and provides context for observed differences in plant water status, growth dynamics, yield, and fruit cracking.

The effect of irrigation on the vines' plant water status

The figure illustrates the seasonal evolution of SWP, TG, and SC under the different irrigation treatments applied in 2023 (top panels) and 2024 (bottom panels). Together, these indicators capture both the vine's integrated PWS and the short-term physiological response at the leaf level (SC). In 2023, clear and consistent differences in SWP were observed among irrigation treatments, reflecting contrasts in both water amount and irrigation frequency. The high-irrigation treatments (1-day high and 3-day high) maintained the least negative SWP values throughout most of the season, indicating comparatively low water stress. In contrast, the low-irrigation treatments (1-day low and 3-day low) exhibited progressively more negative SWP, particularly during mid-season, with the 3-day low treatment showing the strongest water stress signal. This pattern demonstrates that reducing seasonal water input had a stronger effect on vine water status than irrigation frequency alone. The SupPlant treatment generally tracked intermediate SWP values, adjusting dynamically to environmental conditions and avoiding extreme stress levels observed in the fixed low-irrigation regimes. The temporal variability in SWP, including short-term recoveries following irrigation events, highlights the sensitivity of SWP to both cumulative and short-term water supply. Stomatal conductance in 2023 followed trends consistent with SWP, though with greater temporal variability. High-irrigation treatments sustained higher SC values, particularly early in the season, indicating active gas exchange and limited stomatal restriction. Low-irrigation treatments exhibited marked reductions in SC, especially during periods of elevated atmospheric demand, reflecting stomatal regulation in response to declining vine water status. Notably, differences among treatments diminished toward late season, likely due to phenological constraints and partial stomatal closure as harvest approached.

In 2024, the irrigation treatments were designed to impose dynamic transitions in water availability, and this was clearly reflected in both SWP and SC trajectories. The constant high-irrigation treatment maintained relatively stable and less negative SWP values throughout the monitoring period. In contrast, the high-to-low treatment showed a pronounced decline in SWP following the reduction in irrigation frequency, indicating rapid vine sensitivity to late-season water limitation. Conversely, the low-to-high treatment exhibited partial recovery in SWP after irrigation intensity increased, demonstrating the vine's capacity to rehydrate when water supply is restored. However, recovery remained incomplete compared with the continuous treatment. The constant low-irrigation treatment consistently showed the most negative SWP values, indicating sustained water stress. Stomatal conductance patterns in 2024 mirrored these dynamics but with amplified short-term fluctuations. Treatments transitioning from high to low irrigation showed a sharp reduction in SC, while low-to-high treatments displayed transient increases in SC following rehydration. These responses underscore the rapid stomatal adjustment to changes in water availability and the lag between hydraulic recovery (PWS) and stomatal reopening.

Synthesis across seasons: Across both years, the results demonstrate a strong coupling between irrigation regime, vine water status, and stomatal behavior. Sustained or abrupt reductions in water supply consistently led to more negative SWP and reduced SC, while higher or restored irrigation supported improved physiological function. Importantly, the comparison between 2023 and 2024 highlights that temporal irrigation patterns (constant vs. transitional) can be as influential as total seasonal water input in shaping vine water status. These findings confirm that SWP and SC are sensitive indicators of irrigation-driven stress dynamics, providing a physiological basis for linking seasonal water-status trajectories to downstream outcomes, such as yield variability and fruit-cracking risk.

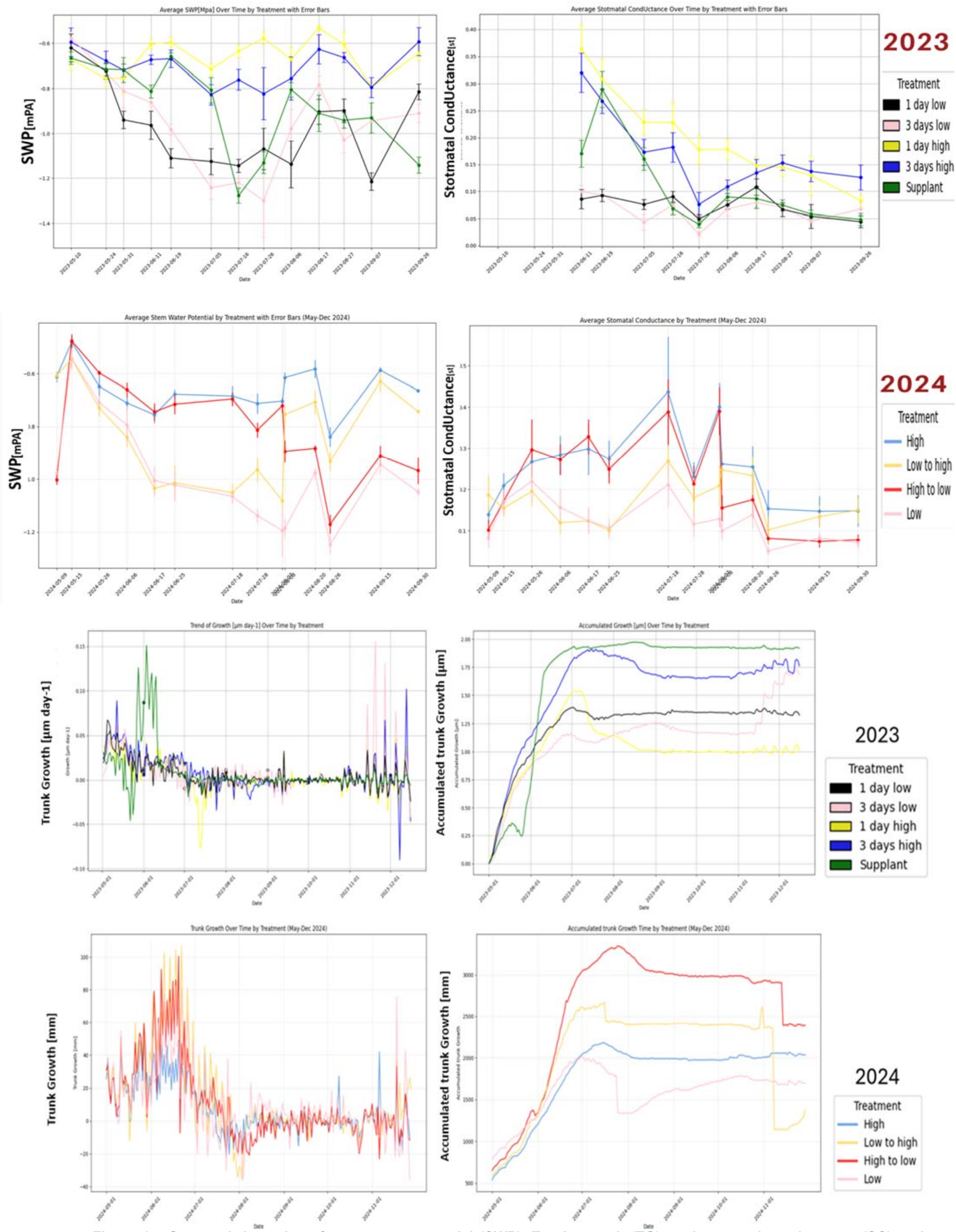


Figure 65. Seasonal dynamics of stem water potential (SWP), Trank growth (TG), and stomatal conductance (SC) under contrasting irrigation regimes in the Lachish vineyard during the 2023 (top panels) and 2024 (bottom panels) growing seasons. Left panels show average SWP (MPa) and right panels show average SC (mol m⁻² s⁻¹) for each irrigation treatment.

In 2023, treatments differed in both irrigation amount and frequency, while in 2024, treatments imposed dynamic transitions between high and low irrigation levels. Symbols represent treatment means, and error bars indicate ± standard error across replicates. Temporal patterns illustrate treatment-dependent differences in vine water status and stomatal regulation in response to sustained and transitional irrigation strategies.

Effects of irrigation regime on yield components and cluster health

Figure 68 summarizes the response of grape yield and quality parameters to contrasting irrigation regimes in 2023 (top panels) and 2024 (bottom panels), highlighting differences in total yield, cluster number, berry weight, and the proportion of healthy clusters. In 2023, irrigation treatments produced pronounced differences in the number of clusters per vine, which was the primary driver of yield variability. The high-irrigation treatments (3 days high and SupPlant) exhibited the highest cluster numbers, reaching approximately 75–80 clusters per vine, whereas the low-irrigation treatments showed substantially lower cluster counts. Total yield (kg vine^{-1}) followed a similar pattern, with the highest yields observed under 3 days of high and SupPlant irrigation, indicating that sustained water availability promoted both cluster initiation and retention. Berry weight showed a weaker response to irrigation compared with cluster number. While slightly higher berry weights were observed under high-irrigation regimes, differences among treatments were modest, suggesting that irrigation primarily influenced the number of sinks (clusters) rather than individual berry size. The comparison between the total number of clusters per vine and the number of real healthy clusters reveals an important quality dimension. Across all treatments, the number of healthy clusters was consistently lower than the total cluster count, indicating losses due to cracking, shriveling, or other defects. Notably, treatments with higher total cluster numbers (e.g., 3 days high and SupPlant) also exhibited larger absolute losses, reflecting increased susceptibility to cracking or quality degradation under high water availability. In contrast, low-irrigation treatments produced fewer clusters overall but maintained a relatively higher proportion of healthy clusters.

In 2024, irrigation strategies emphasized temporal transitions in water availability, and their effects were reflected in both yield and cluster health. The constant high and low-to-high treatments yielded higher total yield and cluster numbers than the constant low treatment, confirming the importance of adequate water supply during key developmental stages. However, the proportion of healthy clusters varied markedly among treatments. The high-to-low and low treatments exhibited the highest percentage of healthy clusters (exceeding ~80%), despite having fewer total clusters. In contrast, treatments that maintained or increased water supply late in the season (high and low-to-high) showed reduced healthy cluster proportions, indicating increased cracking or quality losses. This pattern suggests that late-season water availability, rather than total seasonal water input alone, plays a critical role in determining cluster integrity. Treatments that imposed controlled late-season water limitation limited excessive berry swelling and reduced cracking incidence, thereby improving cluster health despite lower absolute yields.

Synthesis across years: Across both seasons, the results demonstrate a clear trade-off between yield maximization and fruit quality. High or sustained irrigation increased cluster number and total yield but also elevated the risk of cracking and quality losses. Conversely, reduced or strategically timed irrigation lowered total yield but improved the proportion of healthy clusters. These findings underscore that irrigation timing and seasonal water-stress trajectories are as important as total water input in shaping grapevine productivity and fruit quality. Integrating physiological indicators of vine water status with yield and quality metrics provides a mechanistic basis for optimizing irrigation strategies to balance economic yield and marketable fruit quality. Among the evaluated irrigation strategies, the high-to-low irrigation treatment provided the most favorable balance between yield and fruit quality, combining moderate yield levels with the highest proportion of healthy clusters, and therefore represents the most promising management option for reducing cracking risk without substantial yield penalties.

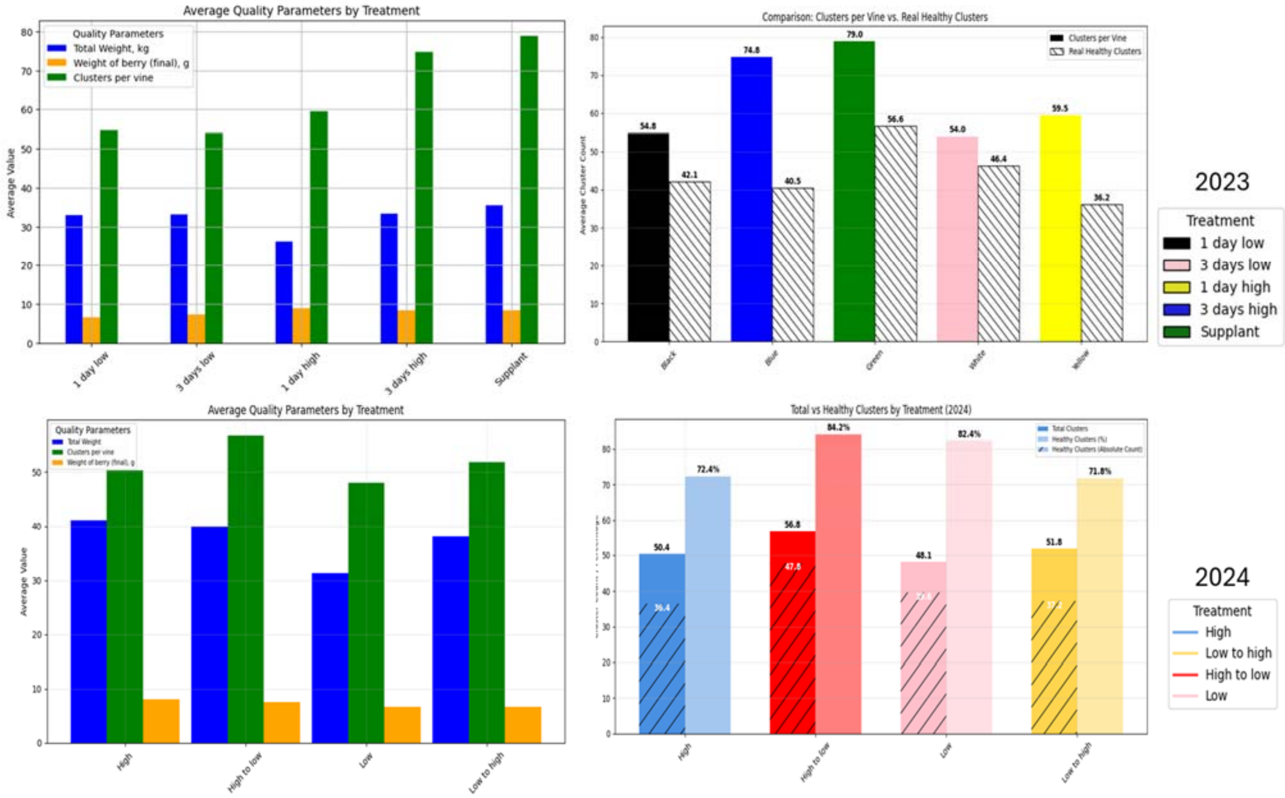


Figure 66. Effects of contrasting irrigation strategies on grape yield components and cluster health during the 2023 (top panels) and 2024 (bottom panels) growing seasons. Left panels show average total yield (kg vine⁻¹), cluster number per vine, and final berry weight across irrigation treatments. Right panels compare total cluster counts with the number (2023) or proportion (2024) of healthy clusters. Results highlight trade-offs between yield and fruit quality, with dynamic irrigation strategies, particularly high-to-low irrigation, achieving an improved balance between yield maintenance and reduced cracking incidence

Model performance for predicting SWP, TG, and SC

The performance of four machine-learning models for predicting stem water potential (SWP) was evaluated using the coefficient of determination (R^2), root-mean-square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). Overall, the results indicate that SWP can be reasonably estimated from the combined set of daily physiological, environmental, and thermal predictors, although clear differences in predictive performance were observed among the tested approaches. Among the models, XGBoost achieved the best overall performance, yielding the highest R^2 (0.76) together with the lowest RMSE (0.11), MAE (0.07), and MAPE (7.9%). This performance reflects the ability of gradient-boosted decision trees to capture nonlinear relationships between SWP and heterogeneous input features while maintaining robustness under field conditions. To further interpret the predictive behavior of the XGBoost model, SHAP (SHapley Additive exPlanations) analysis was applied to quantify feature-level contributions to model predictions (Figure 69). The SHAP results indicate that short-term atmospheric conditions, particularly relative humidity metrics at a 7-day temporal scale, were the most influential predictors of SWP, with higher humidity values generally associated with higher predicted SWP. Measures of air temperature, including short-term variability and mean values, also exhibited substantial contributions, highlighting the importance of thermal dynamics in driving vine water status. In addition, soil moisture at 40 cm depth, especially short-term mean values, contributed meaningfully to the model, emphasizing the role of subsurface water availability. Physiological indicators derived from fruit growth, trunk growth, and dendrometer-based measurements (MDS) further appeared among the top-ranked features, suggesting that short-term growth dynamics encode relevant information about vine water stress.The Random Forest and Gradient Boosting models demonstrated comparable but slightly lower performance ($R^2 \approx 0.70$-0.71), with similar error magnitudes, confirming the general suitability of ensemble tree-based methods for SWP prediction. In contrast, Linear Regression exhibited substantially weaker performance (R^2 = 0.21), underscoring its limited ability to represent the complex and nonlinear interactions governing plant water relations. Collectively, these results demonstrate that ensemble tree-based models, particularly XGBoost, not only provide superior predictive accuracy for SWP but also rely on physiologically meaningful drivers, lending confidence to their use for data-driven assessment of vine water status.

Table 30. Performance comparison of machine-learning models for predicting stem water potential (SWP).

Model	*R²*	*RMSE*	*MAE*	*MAPE (%)*	*N-train*	*N-test*
Random Forest	0.700	0.123	0.088	10.258	612	154
XGBoost	**0.756**	0.111	0.067	7.872		
Linear Regression	0.211	0.201	0.155	18.886		
Gradient Boosting	0.709	0.122	0.076	8.852		

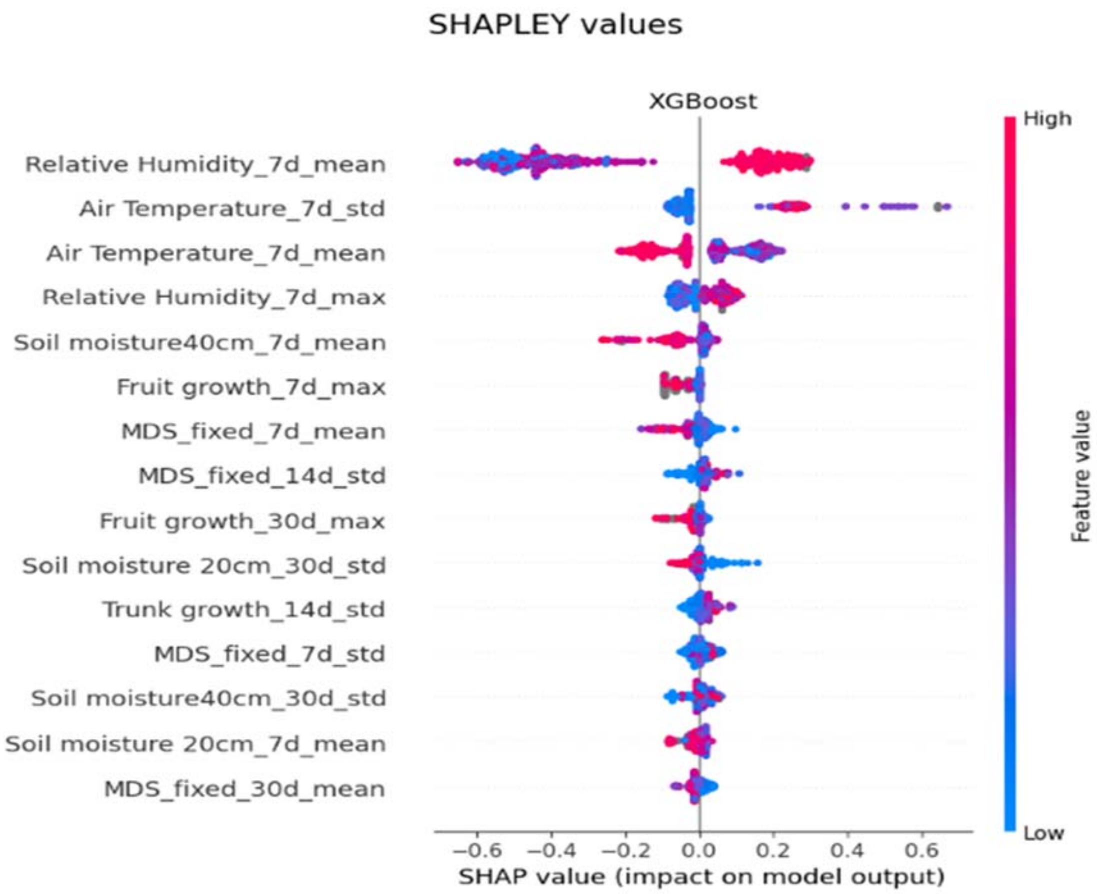


Figure 67. SHAP analysis for the Random Forest model for predicting stem water potential (SWP).

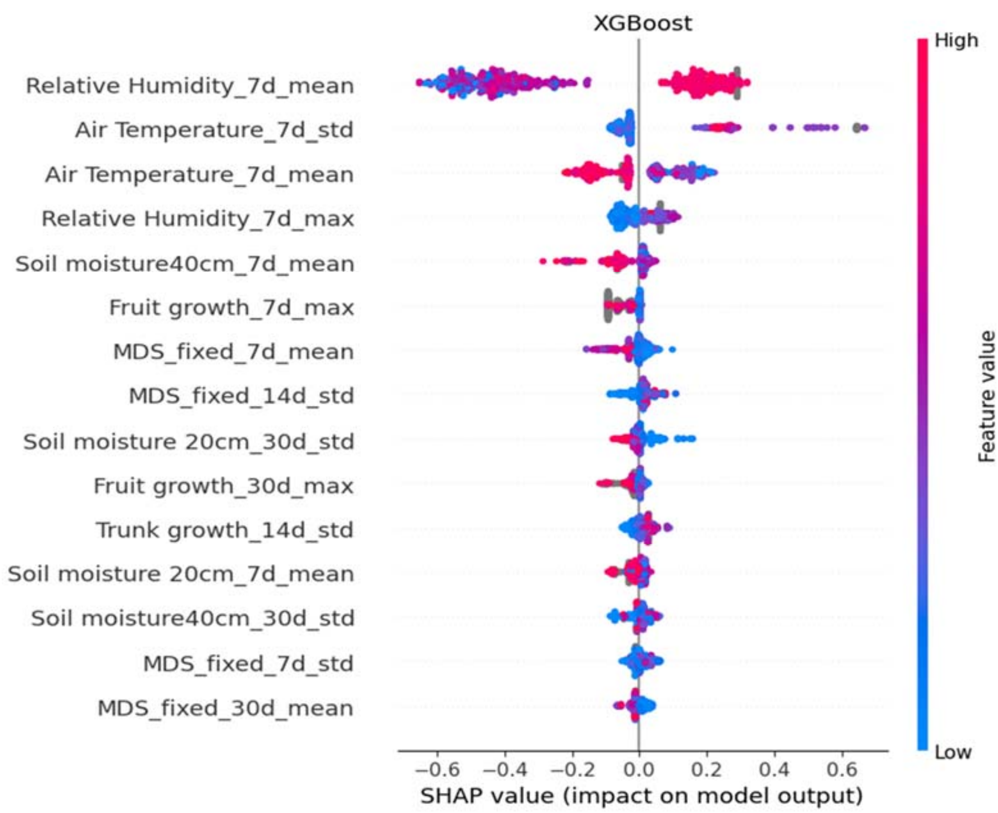
The performance of four machine-learning models in predicting stomatal conductance (SC) was evaluated using the coefficient of determination (R^2), root-mean-square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) (Table 31). Overall, the evaluated models demonstrated moderate predictive skill, with R^2 values ranging from 0.20 to 0.71, indicating that a substantial proportion of the observed variability in SC can be explained by the combined set of physiological, climatic, and soil-related predictors, albeit with higher uncertainty than observed for SWP. Among the tested approaches, XGBoost achieved the highest overall performance, yielding the highest coefficient of determination (R^2 = 0.71) together with the lowest RMSE (0.048) and MAE (0.030). This performance reflects strong sensitivity to nonlinear interactions and threshold-like relationships among predictors, which are characteristic of stomatal responses to short-term atmospheric and environmental drivers. The relatively lower MAE further suggests improved average prediction stability across a wide range of SC values.The Random Forest (RF) model performed comparably, achieving an R^2 of 0.70 and a similar RMSE (0.049), while exhibiting slightly higher MAE and substantially higher MAPE. This suggests that RF provided reliable absolute predictions but was less effective in minimizing relative errors, particularly for lower SC values. Given the inherently dynamic and noisy nature of stomatal conductance, RF's robustness to outliers and short-term fluctuations remains advantageous for field-scale applications. The Gradient Boosting (GB) model demonstrated slightly lower predictive performance (R^2 = 0.65), with modestly higher RMSE and MAE values, indicating reduced accuracy in capturing the full variability of stomatal responses under heterogeneous field conditions. In contrast, Linear Regression exhibited weak predictive performance (R^2 = 0.20) and substantially higher error metrics across all evaluation measures, highlighting its limited capacity to represent the complex, nonlinear controls governing stomatal regulation. Notably, all models exhibited relatively high MAPE values compared to those observed for SWP prediction, underscoring the greater intrinsic variability of stomatal conductance and its sensitivity to rapid environmental fluctuations. Overall, the results indicate that ensemble tree-based methods-particularly XGBoost and Random Forest-are best suited for predicting stomatal conductance in this study, offering a favorable balance between predictive accuracy and robustness when applied to multi-source physiological and environmental data. These findings further support prioritizing machine-learning approaches over linear models for modeling stomatal behavior under field conditions.

Table 31. Performance comparison of machine-learning models for predicting stomatal conductance (SC)

Model	*R²*	*RMSE*	*MAE*	*MAPE (%)*	*N-train*	*N-test*
Random Forest	0.697	0.049	0.037	34.491	612	254
XGBoost	**0.711**	0.048	0.030	24.906		
Linear Regression	0.204	0.080	0.064	65.403		
Gradient Boosting	0.651	0.053	0.034	28.106		

Figure 68. SHAP analysis for the Random Forest model for predicting stomatal conductance (SC)

Prediction of Fruit Firmness



Fruit firmness prediction was formulated as a regression task using the same feature-engineering pipeline applied for stem water potential modeling. Input features included atmospheric variables (air temperature, relative humidity, and dew point), soil moisture metrics at multiple depths, trunk growth indicators, fruit growth dynamics, and dendrometer-derived measurements aggregated over multiple temporal windows. Both linear and nonlinear regression models were evaluated, including Linear Regression, Random Forest (RF), Gradient Boosting, and XGBoost. Overall, tree-based ensemble models substantially outperformed linear regression, indicating strong nonlinear relationships between environmental conditions, plant physiological responses, and fruit firmness dynamics. Among the evaluated models, the Random Forest achieved the highest predictive performance, with an R^2 of 0.80 and the lowest RMSE (0.24), indicating a strong ability to capture variability in firmness across the dataset. XGBoost and Gradient Boosting followed closely, exhibiting comparable explanatory power ($R^2 \approx 0.78$) and similar error magnitudes, suggesting that boosted tree ensembles are also effective in modeling firmness responses. In contrast, Linear Regression showed markedly lower predictive performance ($R^2 = 0.62$, RMSE = 0.33) and higher MAE and MAPE values, confirming that firmness dynamics cannot be adequately described using linear assumptions alone. Model interpretability was assessed using feature-importance rankings and SHAP analysis for the Random Forest model (Figure 71). Both approaches consistently identified short-term atmospheric conditions, particularly relative humidity and air temperature metrics aggregated over 7–14 day windows, as the dominant drivers of firmness variability. Higher relative humidity values were generally associated with increased predicted firmness, while temperature-related features reflected the influence of thermal stress and evaporative demand. Soil moisture indicators, especially at deeper soil layers (40 cm), contributed secondary explanatory power, along with trunk growth and fruit growth variability metrics, highlighting the coupling between water availability, vegetative growth, and fruit mechanical properties. Collectively, these results emphasize the multifactorial nature of fruit firmness regulation, governed by the combined effects of atmospheric demand, soil water status, and plant physiological processes.

Table 32. Performance comparison of machine learning models for fruit firmness prediction

Model	*R^2*	*RMSE*	*MAE*	*MAPE (%)*	*N-train*	*N-test*
Random Forest	**0.796**	0.241	0.176	9.405	184	47
XGBoost	0.784	0.248	0.159	8.503		
Linear Regression	0.619	0.330	0.253	13.193		
Gradient Boosting	0.777	0.253	0.151	7.824		

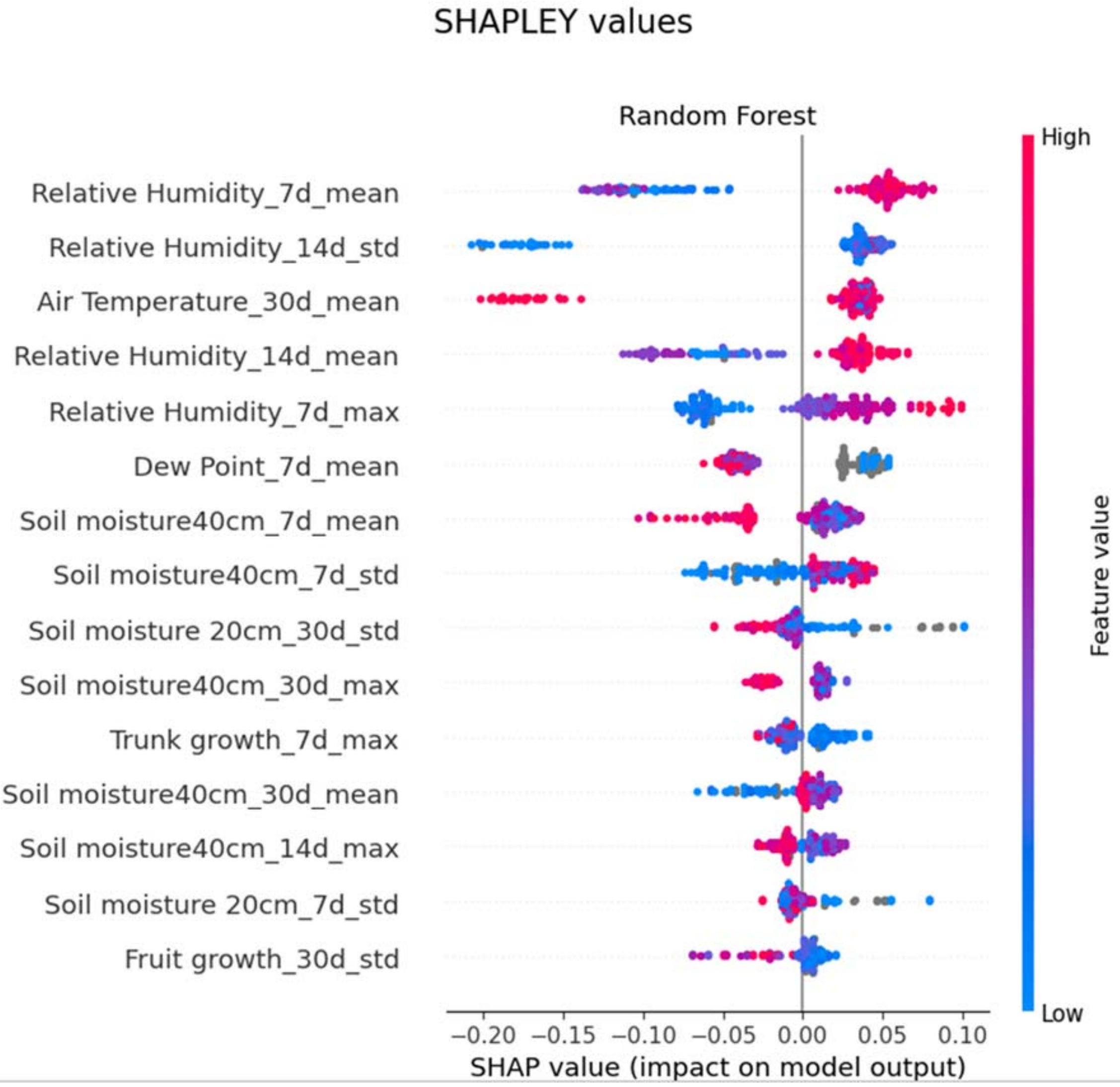


Figure 69. SHAP analysis for the Random Forest model for fruit firmness prediction

Predicting models for yield and cracking indicator on clustering

A clustering-based framework was applied to characterize seasonal physiological behavior patterns in grapevines and evaluate their relevance to cracking outcomes. Physiological measurements were clustered at the tree–season level to group vines with similar stress-response dynamics, without incorporating any cracking information during cluster formation. The resulting clusters were then evaluated for their association with cracking-related metrics, including the healthy cluster percentage and the cracking severity index, thereby assessing whether physiological patterns meaningfully discriminate against cracking susceptibility.

For stem water potential (SWP) and trunk growth, measurements were mapped to standardized weekly indices within each season and reshaped into seasonal curves using a pivot-table representation. Missing values were linearly interpolated, and outliers were filtered using the interquartile range (IQR) method. Clustering was performed directly on the resulting curves using Fuzzy C-Means, with the optimal number of clusters selected by maximizing the Calinski–Harabasz index. Accumulated trunk growth was processed similarly but without temporal interpolation, as it reflects cumulative rather than dynamic seasonal behavior. Firmness measurements were available at a limited and uneven number of time points per season, preventing direct curve alignment. Therefore, clustering was based on season-level summary features capturing the magnitude, variability, and temporal trend of firmness change. These features were standardized and clustered using Fuzzy C-Means, with the number of clusters selected based on the Calinski–Harabasz index, and clustering quality assessed using the Fuzzy Partition Coefficient.

Figure 70. Associations between physiological clusters and fruit cracking outcomes. Boxplots show healthy cluster percentage (top row) and cracking severity index (middle row) and fruit firmness across clusters derived from SWP, trunk growth, accumulated trunk growth, and fruit firmness. Significant separation is observed between SWP-clusters, whereas accumulated trunk growth shows limited discriminative power. Boxes represent interquartile ranges, center lines denote medians, and points indicate individual observations.

Physiological clusters derived from SWP, accumulated TG, and fruit firmness were evaluated against cracking outcomes using healthy cluster percentage and cracking severity index. Overall, cluster-based analyses revealed that indicators reflecting instantaneous PWS and fruit mechanical properties were more strongly associated with cracking outcomes than cumulative growth metrics. SWP-based clusters showed clear and statistically robust separation. Both healthy cluster percentage and cracking severity index differed significantly among clusters, indicating that trees maintaining more favourable SWP conditions were associated with higher proportions of healthy fruit clusters and reduced cracking severity. This highlights SWP as a sensitive integrator of soil–plant–atmosphere water relations during cracking-prone periods. Trunk growth-based clusters exhibited moderate differentiation. While ANOVA detected significant differences in both cracking metrics, these effects were not consistently supported by nonparametric tests, suggesting greater within-cluster variability and greater sensitivity to distributional assumptions. Nonetheless, lower trunk growth clusters tended to exhibit lower healthy cluster percentages and higher cracking severity. In contrast, accumulated trunk growth did not discriminate cracking outcomes. Neither healthy cluster percentage nor severity index differed significantly among clusters, indicating that cumulative growth integrates long-term vigor but is less responsive to short-term stress dynamics that trigger cracking.

Firmness-based clusters demonstrated the strongest and most consistent associations with cracking. Significant differences were observed for both response variables, with higher firmness clusters corresponding to substantially higher healthy cluster percentages and lower cracking severity. These results emphasize fruit mechanical resistance as a critical downstream outcome of PWS regulation. Collectively, the clustering analysis indicates that dynamic physiological indicators (SWP and firmness) outperform cumulative growth metrics in explaining variability in fruit cracking, supporting their use for early risk stratification and precision irrigation management

Table 33. Summary of physiological clustering performance in explaining fruit cracking outcomes for healthy clusters and the severity index

Measurement type	Clusters	Normality (SW)	ANOVA (p-val)	Kruskal p	Main pattern in Healthy Clusters
SWP	3	All clusters normal	**0.0021**	**0.0097**	Clear separation between clusters; Cluster 2 shows highest % healthy clusters
Trunk Growth	3	One cluster non-normal	**0.0374**	0.1485	Moderate separation; one low-performing cluster (Cluster 0)
Accumulated Trunk Growth	4	One cluster non-normal	0.7707	0.6233	No meaningful differences between clusters
Firmness	3	One cluster non-normal	**0.0005**	**0.0134**	Strong separation: Highest statistical significance

Measurement type	Clusters	Normality (SW)	ANOVA (p-val)	Kruskal p	Main pattern in the Severity Index
SWP	3	One cluster non-normal	**0.0020**	**0.0056**	Clear separation; one cluster shows substantially higher cracking severity
Trunk Growth (daily)	3	One cluster non-normal	**0.0209**	0.1405	Moderate separation; parametric significance only
Accumulated Trunk Growth	4	One cluster non-normal	0.7457	0.6045	No meaningful differences
Firmness dynamics	3	One cluster non-normal	**0.0008**	**0.0171**	Strong separation; one cluster shows highest severity

Spatial clustering patterns

The cluster maps reveal a clear and repeatable spatial organization of tree physiological states within the orchard across both seasons. In 2023, clusters formed contiguous blocks that largely followed irrigation treatments and local field conditions, indicating strong spatial coherence in plant water status and growth responses. In 2024, despite changes in plot location and experimental layout, similar cluster structures emerged, demonstrating the temporal robustness of the clustering approach. The persistence of specific clusters across seasons suggests that spatial variability in water status and growth is not random, but rather controlled by stable interactions between irrigation regime, soil properties, and microclimatic gradients. These spatial patterns provide a practical basis for zone-specific irrigation management and targeted mitigation of cracking risk.

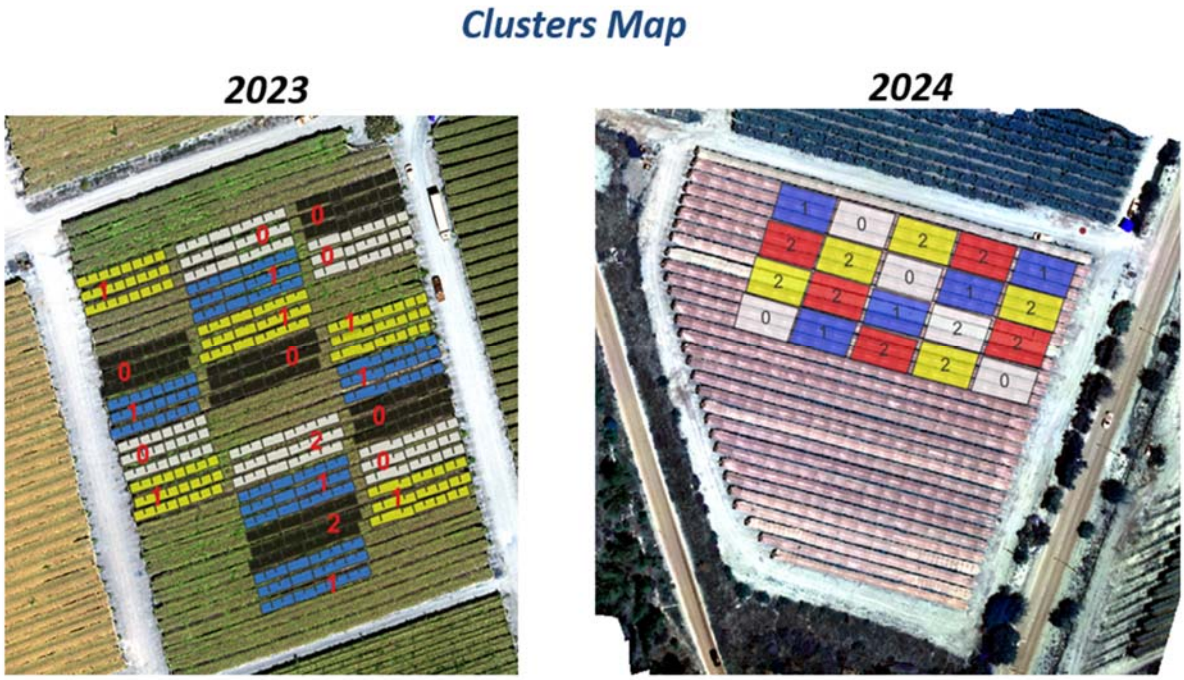


Figure 71. Spatial distribution of physiological clusters derived from UAV- and ground-based measurements for the 2023 and 2024 seasons. Clusters represent distinct tree-level physiological states, as defined by integrated indicators of PWS and growth. Colored blocks denote cluster membership at the plot scale, illustrating spatial coherence and treatment-driven patterns within each season. Despite differences in plot layout and experimental design across years, similar spatial structures emerged, highlighting the robustness of the clustering framework and its potential for site-specific irrigation and cracking-risk management

Greece: Table grape Random Forest modeling results

Data acquisition

Image acquisition and field measurements for the table grape pilot were conducted over three consecutive growing seasons (2023-2024) in vineyards located near Kiato, northeastern Peloponnese, Greece (37°56′50.7″N, 22°46′04.0″E), an area traditionally known for viticulture (Figure 74). The 2023-2024 experimental field included 32 vines monitored under controlled conditions.

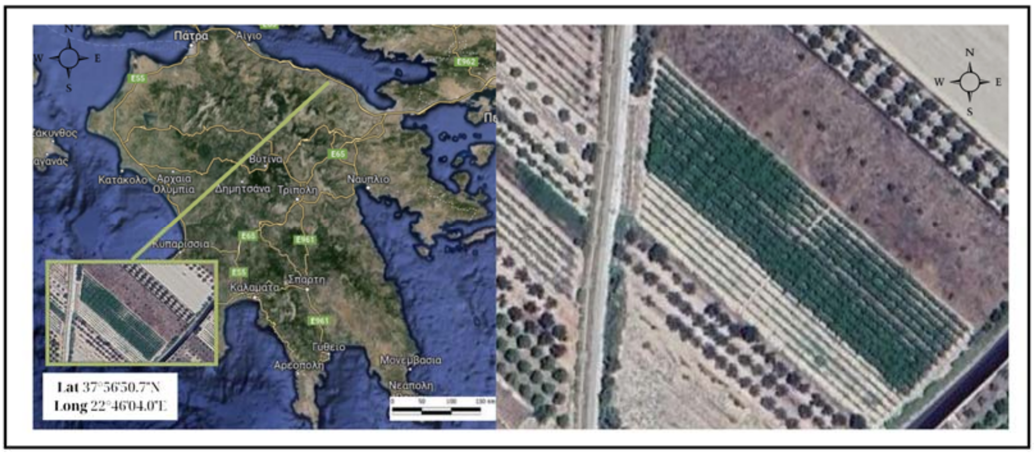


Figure 72. Experimental site location and aerial view of the vineyard plots used during the 2023-2024 table grape campaigns in Kiato, Greece

Throughout the monitoring period, UAV-based image acquisition was carried out using two platforms: the DJI Mavic 3 Multispectral and the DJI Mavic Thermal. Flights were performed between 11:00 and 13:00 under clear-sky conditions with solar elevation angles above 45°, ensuring stable lighting and minimal shadowing. All images were captured using nadir-view geometry, with pre-calibrated multispectral and thermal sensors.

Experimental design

As reported in D3.2, the experimental setup included three separate sub-trials to investigate the effects of irrigation, plant growth regulators (PGRs), and fertilization on vine physiology and yield. All three experiments followed a randomized block design to minimize environmental noise and enable reliable treatment comparisons:

Irrigation trial (2023-2024): A Randomized Complete Block Design (RCBD) was applied, comparing two irrigation strategies: GAW (High frequency, green markers): Weekly irrigation for 5 hours at 20 m^3/h, CAW (Low frequency, orange markers): Biweekly irrigation for the same duration and rate. Irrigation treatments were conducted from March 15 to August 1.

PGR trial: Conducted in an adjacent block, vines were sprayed with three Plant Growth Regulator treatments: GA_3 (3%, April 3): Enhancing cell elongation and berry size, CPPU 6 ppm (April 3): Stimulating cell division, and CPPU 4 ppm (April 3 & 13): Sustained berry firmness. Control vines were also included for comparison.

Fertilization trial: Involved three nutrient regimes: Balanced NPK (12-5-15, standard rate), Intensive NPK (25-5-35, high rate), and Reduced input (0-0-0, control).

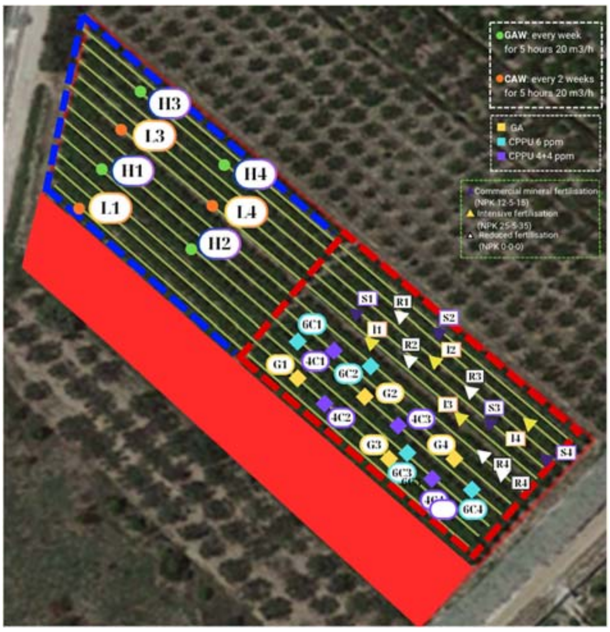


Figure 73. Experimental layout for irrigation, mineral fertilization and PGR treatments in Greek table grapes

Yield in Greek table grape vineyards was estimated using field, weather, and remote sensing data from 2023-2024. Due to consecutive heatwaves in recent years, no fruit cracking was observed in the Kiato (Peloponnese), allowing the focus to shift entirely to yield prediction. Data were collected at the final critical phenological stages before harvest (Pre Veraison, Harvest date), including measurements of stem water potential, trunk diameter, berry weight, soil moisture, and vegetation indices. Among the predictors, remote sensing features such as GNDI, MSAVI, SAVI, NDVI, Clgreen, NDRE, and Clred emerged as the most influential variables. The model demonstrated strong predictive performance, with a high correlation between observed and predicted yields, supporting its potential for site-specific vineyard management and early yield forecasting.

Data collection

Across all campaigns, tree-level measurements were conducted at four key phenological stages: BBCH 77 (Post fruit set), BBCH 81 (Pre-veraison), BBCH 83 (Post-veraison), and BBCH 85 (Final ripening), with sampling dates:

- 2023: June 15, July 18, August 17, September 10
- 2024: June 7, July 1, July 18, August 2

Measurements included stem water potential (SWP), trunk diameter, berry weight, soil moisture, and yield, complemented by detailed weather records (e.g., temperature, VPD, RH) from the TOMMY system. Vegetation indices (NDVI, GNDVI, MSAVI, etc.) This multimodal dataset was harmonized and spatially aligned with individual vine shapefiles for machine learning modelling.

Dataset characteristics

Given the absence of fruit cracking during the monitored seasons, attributed to sustained high temperatures and dry conditions, this pilot focused on yield prediction as the primary modelling target.

Results

The performance of the Random Forest (RF) model for SWP and yield estimation was evaluated using training data from 2023-2024.

SWP prediction model - results

Model performance: The model was trained on tree-level data that included UAV-derived spectral indices and environmental parameters, and evaluated using a stratified 70/30 train-test split. The model achieved a training R^2 of 0.897 and RMSE of 0.051, while the test R^2 reached 0.603 with RMSE of 0.093, demonstrating moderate to good performance. These results indicate the model's ability to capture short-term SWP dynamics across individual grapevines using remote sensing and weather-based inputs.

Table 34. Model performance summary for Random Forest-based SWP prediction (2023-2024).

Set	R^2	RMSE	n_trees	m_try
Train	0.897	0.051	30	2
Test	0.603	0.093	30	2

Prediction vs Observations: The scatterplot (Figure 76a) compares predicted and observed SWP values for both training and test subsets. The model successfully captures the main trend, with a consistent alignment between predicted and observed values, especially for the mid-range SWP levels. Some dispersion is observed at the extremes, likely due to variability in vine-level conditions or limited sample size.

The histogram (Figure 76b) shows that the distribution of SWP values across training and test datasets is similar, confirming that the dataset is well-suited for model training and validation.

(a) (b)

Figure 74. (a) Observed vs. predicted SWP values for the 2023 season using the Random Forest model. Points represent test (red) and train (blue) data; (b) Histogram showing the distribution of SWP values in training (blue) and test (red) sets

Feature importance (coefficient): Variable importance analysis (Figure 77) indicates that CW_717, Air Temperature, and Relative Humidity were the most influential predictors of SWP. These reflect the impact of both vegetation condition and short-term weather conditions on vine water status. Secondary contributors included Month, NDRE, and bands CW_668 and CW_560, which relate to canopy vigor and reflectance. Temporal and categorical variables like type and year also contributed moderately.

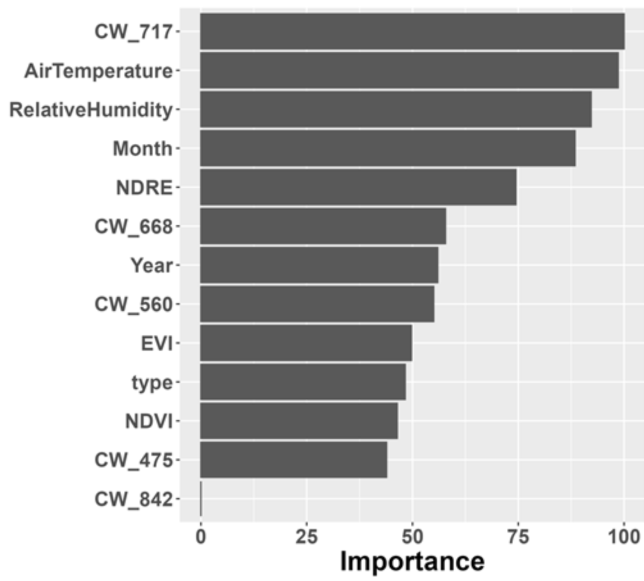


Figure 75. Variable importance ranking from the Random Forest model based on 2023 data. CW_717, Air Temperature, and Relative Humidity are the top contributors

Summary of SWP model: The test R^2 of 0.60 highlights the model's moderate accuracy in capturing SWP variability. The leading features, CW_717, Air Temperature, and Relative Humidity, underscore the relevance of both spectral and meteorological inputs in modeling short-term water status.

Overall, these results demonstrate the viability of scalable SWP estimation in vineyards and suggest opportunities for further improvement through data augmentation and multi-season training.

Yield prediction - results

Model performance: Yield prediction was carried out using a Random Forest regression model trained on meteorological, multispectral, and phenological variables. The training was performed on a tree-level dataset collected during the 2023 and 2024 season in the Greek table grape pilot, and tested on a hold-out test set. The model yielded an R^2 of 0.865 on the training set with an RMSE of 2.68, while the performance on the test set was lower, with an R^2 of 0.017 and RMSE of 5.67. These results indicate strong model fitting during training, but limited generalization ability to unseen data, suggesting potential overfitting or data inconsistencies between train and test subsets.

Table 35. Model performance summary for Random Forest-based yield prediction (2023-2024).

Set	R^2	RMSE	n_trees	m_try
Train	0.865	2.68	100	2
Test	0.017	5.67	100	2

Prediction vs. Observations: The scatterplot (Figure 78) shows the relationship between observed and predicted yield values for both the training and test subsets. The model captures the overall increasing trend in yield, but its performance on the test set is less consistent, particularly for higher-yield observations, where underprediction is evident.

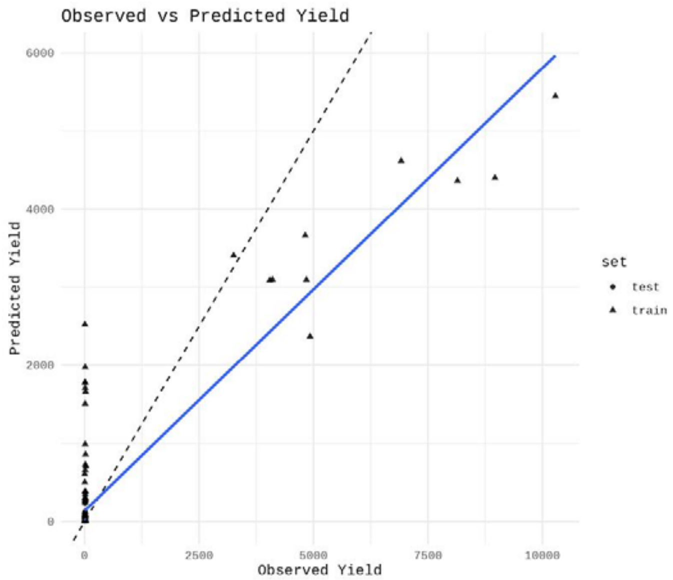


Figure 76. Observed vs. predicted yield values using Random Forest regression. The solid line shows the fitted model, and the dashed line represents the 1:1 reference. The separation between sets illustrates prediction accuracy differences.

Feature importance (coefficient): The variable importance analysis (Figure 79) indicates that Dew Point, NDRE, and OSAVI were the most influential predictors of yield variation, followed by GNDI, CIred, and maximum air temperature. Phenological features such as growth stage and date also contributed, along with several spectral indices (MSAVI, NDVI, CIgreen). This pattern suggests that both weather-driven physiological conditions and canopy structure are important for yield estimation in table grapes.

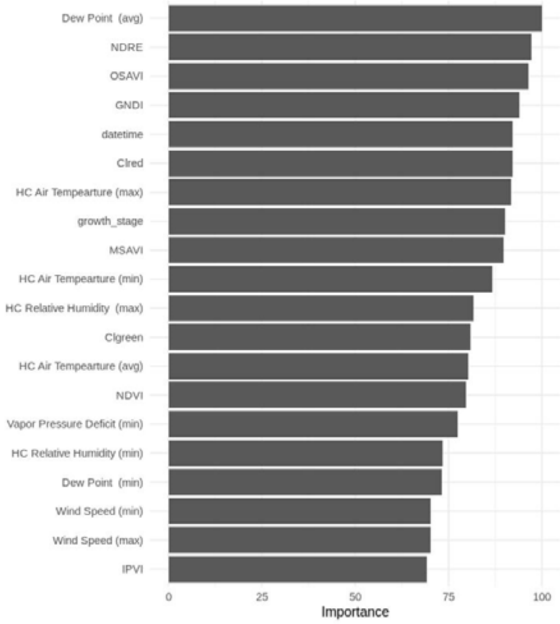


Figure 77. Variable importance plot for the yield model (table grapes), showing the relative contribution of input features. Dew Point, NDRE, and OSAVI rank as the top predictors.

Summary of yield model: The Random Forest model for yield prediction in the Greek table grape pilot displayed high explanatory power on the training set, but low generalization to the test data. This may be attributed to limited training sample size, data imbalance, or the complexity of yield determinants.

Overall, the most impactful variables combined weather metrics (dew point, air temperature), vegetation indices (NDRE, OSAVI, GNDI), and phenological data (growth stage, date), underscoring the multifactorial nature of grapevine productivity.

Summary

This section describes the process of ingesting data, running edge AI models, and processing results on the VLF edge controllers. The objective is to demonstrate the feasibility of implementing this pipeline and deploying edge intelligence in the field to achieve results as early as possible. The focus is on obtaining near-real-time tree segmentation on edge devices. The scope of this section is training object detection and segmentation models (Yolo11) using open-source Treetop data, describing the full pipeline for data preparation and model output postprocessing, along with technical information on how this setup is to be utilized in the context of the project.

Dataset & Data Preparation

Data collection was carried out using drone aerial photography, and the raw data included RGB, thermal, and multispectral images, as well as LiDAR data. The primary focus of this section is RGB images for segmenting treetops in the collected data. This dataset is unlabeled, which is why it was used only for inference, while training was conducted on open-source data [1] of a similar structure and with a similar inference goal.

Training Dataset

In order to edge models capable of treetop segmentation, a labeled dataset was needed. For this purpose, the OAM-TCD dataset [1], comprising 5072 2048x2048 images, was used. This is a diverse dataset comprising aerial images of treetops from around the world. In the context of the CrackSense project, aerial images would mostly show orchards, which are much more organized/structured than what you would typically find in nature. This favours training segmentation models, as they will generalize better and avoid focusing solely on rows of trees, as is often seen in orchards. The dataset was labelled using a combination of automatic labelling using a high-yield model. These initial labels were then provided to human annotators for review and modification in accordance with detailed instructions. This resulted in a dataset with labels for over 280k individual trees and 56k groups of trees, classified as “tree” and “canopy,” respectively. Examples of training images are shown in Figure 80.

Figure 78. Example images from the OAM-TCD dataset. They contain a broad set of examples, not just data collected from orchards

Inference Dataset

The inference dataset was collected by the Agricultural Research Organisation of Israel at a citrus orchard in Israel (dataset “0001_IL_KFH”). Top-down drone shots of the Citrus tree orchard yielded 734 images for processing by the edge controller. Example images are shown in Figure 81. These images are quite large, with a resolution of 8192x5464, which makes edge processing challenging. Consequently, certain preprocessing steps are required to segment the images at the edges. Image metadata includes additional information such as latitude/longitude and acquisition time. Based on this information, individual images could be stitched together to create a mosaic image, which depicts the entire field as a single image. The mosaic of the citrus orchard is shown in Figure 82.

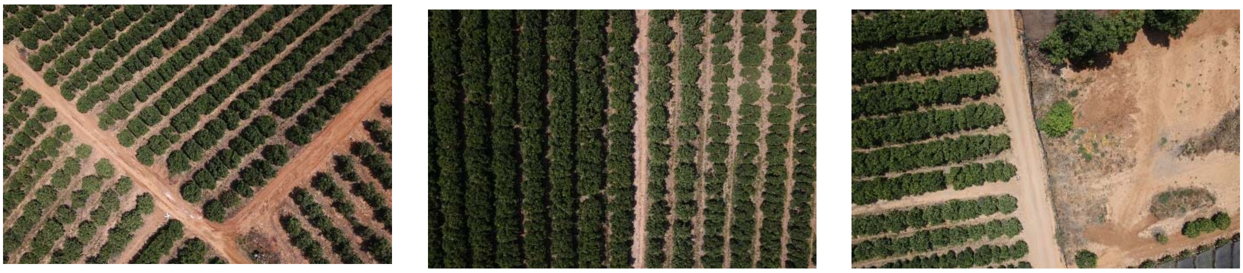


Figure 79. Example images from the “0001_IL_KFH” dataset

Figure 80. A composite orthomosaic processed by the edge device, effectively merging a sequence of aerial photographs acquired during drone flight operations

Data Preprocessing

Two distinct input pipelines are explored as part of the work on the edge units – individual images and the mosaic, each with its own preprocessing steps.

Single-image approach:

1. **Letterbox Resizing –** The image is first resized to a chosen resolution (set to *2048x2048*, which follows the native resolution of the training dataset), where any remaining area is padded with a neutral color to match the target aspect ratio without distortion.
2. **Tiling (Cropping) –** The resized image is then split into smaller and overlapping crops (*1024x1024*) with a sliding window approach. The model can then perform inference on an image with greater detail that might be lost if the image were downscaled. Overlap ensures that the objects at the

edges are fully captured and prevents them from being missed (duplicates are resolved in postprocessing).

Mosaic approach:

1. **Large-Scale Tiling –** The mosaic is segmented into large crops, which are calculated as the target model input size multiplied by a zoom factor using a non-overlapping sliding window. With this method, we achieved a "zoom out" effect, aligning the input resolution with the approximately 10 cm/px scale of the OAM-TCD training dataset. Tiles at the edges that are smaller than the calculated crop size are padded to maintain the aspect ratio before resizing.
2. **Content Filtering –** Background tiles with empty or solid color are filtered out based on pixel intensity variance. This ensures the model processes only regions that contain actual visual data, reducing inference time.
3. **Downsampling –** Each crop is resized to the model's native *1024x1024* resolution.

Model Architecture & Training

For instance, for segmentation, we chose the Yolo11 family of models [2] and adapted them to the CrackSense use case.

Yolo11 Models

The segmentation models are trained on the COCO dataset and fine-tuned on the OAM-TCD dataset for application to data from the Citrus orchard in Israel. Since the models need to be deployed on VLF edge units, they must be small, which is why the **nano** and **small** variants of Yolo11-seg were chosen. Key characteristics of the models in question are listed in Table 36.

Table 36. Some of the key characteristics of the Yolo11 nano and small models

Model	Size (pixels)	Speed CPU ONNX	Speed T4 TensorRT10	Params (M)	FLOPs (B)
YOLO11n-seg	640	65.9 ± 1.1	1.8 ± 0.0	2.9	9.7
YOLO11s-seg	640	117.6 ± 4.9	2.9 ± 0.0	10.1	33.0

Two training variations were conducted: one with the sole objective of segmenting the “tree” class, and the other with both the “tree” and “canopy” classes as targets.

Output Postprocessing

The YOLO models output segmentation masks with individual objects as polygons. The number of detected objects depends on the confidence level threshold set for the model. For lower confidence levels, duplicate detections may be introduced, which is why postprocessing of the output is needed. Postprocessing consists of up to four stages to refine the results, depending on the detection scenario:

1. **Non-Maximum Suppression (NMS) –** NMS is applied to the final stitched image, which tries to eliminate highly overlapping bounding boxes based on Intersection over Union.
2. **Overlap & Containment Filtering –** A custom filter that removes detections that are significantly overlapped by or fully contained within other detections. The set condition prioritizes higher confidence scores.
3. **Size-based Outlier Removal –** Detections are first analyzed statistically, and objects with areas much smaller than the mean are discarded as noise.

4. **Tree-in-Canopy Logic (Two-Class Specific) –** This filter is applied in scenarios where we have both "Tree" and "Canopy" classes. It removes "Tree" detections that are significantly overlapped by and smaller than a "Canopy" detection.

Training Environment

The YOLO models were fine-tuned on a single-node server equipped with an NVIDIA GeForce RTX 3080 Ti, an AMD Ryzen Threadripper 3970X 32-Core Processor, and 64GB of RAM**.**

Inference Workflow

The general inference workflow follows the steps outlined in Figure 83. Two main input pipelines were considered: one uses original drone-captured images, while the other starts with mosaics stitched from the entire set of aerial RGB images. Both input pipelines converge during preprocessing, producing images with a resolution of *1024x1024.* These images are fed into the Yolo11-seg variants of models. Tree detections depend on the confidence threshold configured in the model, which is why some postprocessing of the output is required. First, duplicate detections are removed, and then the size of each detection is considered. Candidate tree detections that differ greatly from the meaning are eliminated as false positives. Finally, when considering both the “tree” and “canopy” classes, individual tree detections that are contained within a greater canopy detection are also filtered out as duplicates.

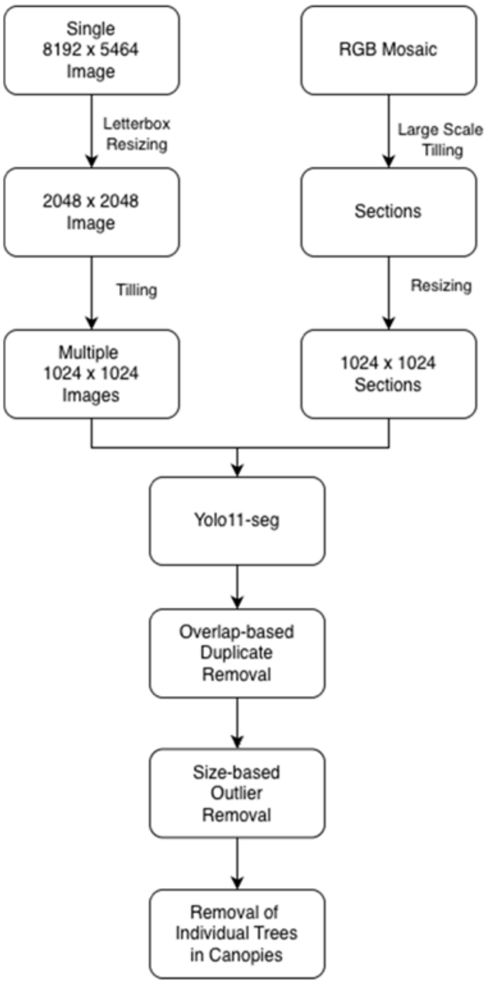


Figure 81. Steps taken during inference time for both single-image and mosaic-level inference. These pipelines converge by creating outputs of size *1024x1024*, which are

References

[1] Veitch-Michaelis, J., Cottam, A., Schweizer, D., Broadbent, E., Dao, D., Zhang, C., ... & Max, S. (2024). OAM-TCD: A globally diverse dataset of high-resolution tree cover maps. Advances in neural information processing systems, 37, 49749-49767.

[2] Jocher, G., & Qiu, J. (2024). Ultralytics YOLO11 (Version 11.0.0). Retrieved from https://github.com/ultralytics/ultralytics

Future Work

Building on these results, several key directions will be pursued in the next project phase:

Multi-site and multi-year model generalization: As additional seasons and sites are integrated, models will be retrained and validated across a larger climatic and management envelope, improving robustness and transferability between regions and crops (WP5 pilot part).

Integration of deep learning and hybrid models: As the dataset expands, deep-learning architectures (e.g., temporal CNN–RNN hybrids) will be evaluated alongside RF models better to capture nonlinear temporal dependencies and stress memory effects.

Earlier-season risk forecasting: Future analyses will focus on identifying early-season physiological and environmental indicators that can predict cracking risk weeks to months before harvest, enabling proactive management interventions.

Operational decision-support tools: Cluster-based management zones and cracking-risk maps will be translated into actionable irrigation and crop-load management recommendations, supporting precision agriculture workflows for growers and advisors.

Extension to additional quality traits and stresses: The established framework will be extended to other quality attributes (e.g., firmness, size uniformity) and to compound stress scenarios involving heat, water, and nutritional constraints.

Stakeholder-oriented validation and deployment: Close collaboration with growers, extension services, and industry partners will be used to validate model outputs under operational conditions and to co-develop user-friendly interfaces for CrackSense tools.

In summary, D3.5 marks a critical step toward operational, scalable sensing-based prediction of fruit cracking and yield loss. They demonstrated physiological relevance, statistical robustness, and spatial explicitness of the approach position CrackSense to deliver impactful, climate-resilient solutions for high-value perennial fruit production.