International Journal of Applied Earth Observation and Geoinformation 146 (2026) 105140

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

# The yield strikes back: Enhancing the transferability of field scale wheat and barley yield models by leveraging Sentinel-1/2

Belen Franch [a,b,*], Italo Moletto-Lobos [a], Javier Tarín-Mestre [a], Lucio Mascolo [c], Eric Vermote [d], Natacha Kalecinski [b], Inbal Becker-Reshef [b], Alberto San-Bautista [e], Constanza Rubio [f], Sara San Francisco [g], Miguel Ángel Naranjo [g], Vanessa Paredes [h], David Nafria [h], Carlos Cantero-Martinez [i]

[a] Global Change Unit, Image Processing Laboratory (UCG-IPL). Universitat de Valencia, Spain
[b] Department of Geographical Sciences, University of Maryland, College Park MD, the United States of America
[c] Radarmetrics, Santander, Spain
[d] NASA Goddard Space Flight Center Code 619, Greenbelt MD, the United States of America
[e] Departamento de Producción Vegetal, Universitat Politécnica de València, Spain
[f] Centro de Tecnologías Físicas, Universitat Politécnica de València, Spain
[g] Fertinagro Biotech, Teruel, Spain
[h] Instituto Tecnológico Agrario de Castilla y León (ITACyL), Subdirección de Infraestructuras Agrarias, Área de Desarrollo Tecnológico, Valladolid, Spain
[i] Universitat de Lleida. Agrotecnio Center. Lleida, Spain

ARTICLE INFO

*Keywords:*
Yield
Wheat
Barley
Sentinel-1
Sentinel-2
GDD

ABSTRACT

Accurate and transferable crop monitoring from remote sensing remains challenging because vegetation signals are strongly affected by phenological asynchrony, climatic variability, and sensor-specific responses. Existing approaches rely on local calibrated relationships , limiting their effectiveness in data-sparse regions. This study investigates whether models calibrated on high-quality localized reference data can generalize to other regions by stabilizing sensor–biophysical relationships.

The proposed methodology integrates two components (i) thermal time normalization based on growing degree days (GDD) to reduce phenology-driven variability, and (ii) physically motivated optical and optical–SAR fusion indices designed within this normalized framework to enhance the consistency of learned relationships across contrasting environments.

The approach was evaluated through within-region cross-season, and cross-region experiments. Results show that GDD normalization improves performance relative to calendar-based approaches by up to 35%. In cross-season validation, fusion-based linear models achieved $R^2 > 0.60$ for wheat and 0.65 in barley. Cross-region validation shows reduced but meaningful transferability, with both crops reaching $R^2 \approx 0.45$, and fusion reducing RMSE by $\sim$ 200 kg $ha^{-1}$ compared to optical-only models. Machine-learning models did not improve generalization over simple parametric fits.

These findings confirm that stabilizing phenological and multi-sensor relationships is critical for transferring models from data-rich to data-limited areas, providing a foundation for scalable, global agricultural monitoring.

## 1. Introduction

Cereals, especially wheat and barley, dominate EU arable land, with wheat contributing nearly half of total output and barley a third (Kelly, 2019). In Spain—one of the top producers—these crops are concentrated in Castilla y León, Castilla-La Mancha, Aragón, and Andalucía. Being mostly rainfed, winter cereals are highly sensitive to temperature shifts, droughts, and extreme events (Cammarano et al., 2019); (Falloon and Betts, 2010) , and these risks are expected to intensify under future climate scenarios (COP21–COP29).

Earth Observation (EO) data are increasingly used to estimate wheat and barley yields at both field and within-field scales. While there is a

* Corresponding author.
*E-mail address:* belen.franch@uv.es (B. Franch).

https://doi.org/10.1016/j.jag.2026.105140
Received 10 September 2025; Received in revised form 8 January 2026; Accepted 27 January 2026
Available online 3 February 2026

variety of modeling approaches, most methods currently applied in the literature are empirical (including statistical and machine learning (ML) models). They establish a relationship between EO-based (or other climatic ancillary data) and ground-based yield data. Their dependency on the availability of training data, which has an associated uncertainty, can constrain their representativity and transferability across regions or seasons, and limit their performance under untrained conditions (Weiss et al., 2020). The primary challenge in precision farming, particularly for cereals like wheat and barley, is the need to effectively retrieve, process, and integrate high-quality ground truth data with advanced remote sensing time series. High-fidelity ground data is essential for calibration and validation, while satellite and drone remote sensing provide large, multi-temporal datasets on crop evolution (Becker-Reshef et al., 2026). The difficulty lies in managing the volume and heterogeneity of these data streams, performing complex data fusion and noise reduction, and developing robust algorithms to convert these diverse inputs into spatially and temporally granular yield predictions (Grassini et al., 2015). Next, we briefly describe the most recent works in this domain and their limitations focusing on wheat or barley.

Most current studies are based on Sentinel-2 (S2) and other data sources (Aslan et al., 2024); (Segarra et al., 2020) . However, their main limitation is that they are generally evaluated across a restricted number of seasons or regions, which may be correlated with the training data. Current literature on yield estimation mainly depends on site-specific or time-limited evaluations. This limits how widely the developed models can be used. Several studies have looked at multi-seasonal analysis in narrow geographic areas. For example, Zhao et al. (Zhao et al., 2020) focused on Australia, Sharifi (Sharifi, 2021) studied Iran, and Cavalaris et al. (Cavalaris et al., 2021) concentrated on Greece. While these studies, along with models from Switzerland (Perich et al., 2023); (Amin et al., 2024) , show strong performance locally, their focus on specific climate zones restricts their broader application. Importantly, despite using different methods—from simple statistical regressions to more complex machine learning models—performance tends to stay consistent. Relative errors typically fall between 10–25%. This indicates that current research, including spatial studies like those by Hunt et al. (Hunt et al., 2019) or district-level assessments by Li et al. (Li et al., 2024), has not yet solved the challenge of maintaining accuracy when models are used across different annual seasonal changes and varied regional landscapes.

Radar data are sensitive to plant structural, geometric and dielectric properties (linked to water content), which vary with crop phenology (Mandal et al., 2021). By operating independently of solar illumination and cloud cover, Synthetic Aperture Radar (SAR) provides high-resolution, all-weather, day/night imagery, offering a consistent data stream that is particularly suited to overcome the transferability limitations of optical sensors in regions with frequent cloud cover (Steele-Dunne et al., 2017). Polarimetric SAR (PolSAR) deals with microwave's vector properties, providing enhanced sensitivity to plant structure, and is widely used in agriculture, e.g., in phenology monitoring (Mascolo et al., 2025). Due to its sensitivity, PolSAR effectively estimates key biophysical variables like Leaf Area Index (LAI) and biomass (Mandal et al., 2021). However, while several studies have used SAR to identify yield variability (McNairn et al., 2004); (Smith et al., 2006) or estimate biomass as a proxy for yield (Wiseman et al., 2014); (Hosseini and McNairn, 2017) ; (Mandal et al., 2020) , the development of robust, cross-regional yield forecasting models remains under-explored. Most existing research focuses on site-specific correlations, leaving a gap in understanding how PolSAR-based models perform when transferred across different environmental conditions.

To address this gap, this study aims to evaluate the potential of PolSAR descriptors for wheat yield modeling, specifically focusing on their ability to provide stable and transferable features. By testing these models across six seasons for calibration and four regions for validation, we intend to demonstrate that PolSAR integration can significantly improve the reliability of yield forecasts in areas where optical data are inconsistent.

Recent studies have advanced crop yield estimation by integrating optical and radar (SAR) data, leveraging their complementary sensitivity. For wheat, Uribeetxebarria et al. (Uribeetxebarria et al., 2023) combined S2 vegetation indices with Sentinel-1 (S1) SAR backscatter using a CatBoost algorithm over 39 fields in Spain and a single season, outperforming S2 alone by 30%. Zhuo et al. (2019) assimilated S1-derived soil moisture and S2 NDVI into the WOFOST crop model in China, improving $R^2$ by 50%. Kalecinski et al. (Kalecinski et al., 2024) demonstrated that fusing C- and L-band SAR (from RADARSAT-2 (RS2), S1, and UAVSAR) with optical data (Planet, Landsat, and S2) through regression models or a Random Forest algorithm enhanced yield prediction for rice, soybean, and corn. Adding SAR parameters to the - Difference Vegetation Index (DVI) halved the prediction error compared to using DVI alone. For barley, Iranzo et al. (Iranzo et al., 2022) fused S1 data during the planting period with S2 vegetation indices, and climate data through linear regression models applied over a semi-arid Spanish farm. Their results highlighted the added value of the SAR data to detect the emergence of the crop combined with the temperature information. These recent works highlight that SAR–optical fusion improves the accuracy of either optical or SAR only models, however, the transferability of these models across seasons and regions is generally not studied given that most works analyze either different regions but over one season or vice versa.

While all these models present promising tools, there is a need to integrate the current technological capabilities in a more holistic way to improve the transferability of crop yield models for wheat and barley. Note that for a model to be considered generalizable, its yield estimates should closely match observed crop yields in growing seasons and locations it has not seen before or been trained on. However, the transferability of models across seasons and regions is generally not studied in the cited works given that most of them analyze different regions but over one season or vice versa. This need for transferability becomes even more critical considering that access to accurate and high-quality yield data is severely limited in many regions, particularly in developing countries or areas affected by armed conflict, where reliable yield forecasts are essential for food-security planning. Ideally, we should be able to train robust models in data-rich environments that can then be reliably applied to these challenging regions, and this paper contributes to understanding how such transferability could be achieved.

This work aims to advance the understanding of how currently available EO technologies can be leveraged to enhance the transferability of empirical models to infer wheat and barley yield at the field scale. Particularly, we examine the potential of the Growing Degree Day accumulation ($GDD_{accum}$) as a normalization factor to enhance the transferability of an empirical method across different regions and seasons (Franch et al., 2015); (Skakun et al., 2017) . The $GDD_{accum}$ parameter quantifies the cumulative sum of daily average air temperatures above a defined base temperature, below which crop development does not progress. Its normalization capacity arises from the physiological principle that cereal crops require the accumulation of a specific amount of heat to advance through their phenological stages. Previous studies have demonstrated the normalization potential of $GDD_{accum}$ across seasons at the subnational level (e.g. (Franch et al., 2021); 2015) or within a single region at field level (e.g. (Amin et al., 2024). However, its applicability for improving model transferability across both seasons and regions at the field level has not yet been systematically investigated.

Additionally, with the objective of integrating S1 and S2, we derive new vegetation indices based on GDDaccum phenological evolution that are sensitive to wheat and barley yields both based solely on optical data or combining optical (S2) and SAR (S1) data. These indices proposed through single regression models will allow us to understand which EO parameters are mainly related with the yield and how they change through each stage of the crop development. Finally, we investigate if wheat and barley could be integrated in a unique yield model or whether

they should be considered separately.

## 2. Materials

### 2.1. Study area

The study focuses on Spain's main cereal-producing regions—Castilla y León, Castilla-La Mancha, Aragón, Cataluña, Extremadura, and Andalucía. Fig. 1 shows the location of all parcels considered over the Köppen-Geiger climate classification map, highlighting that despite the relative short distance between regions, they are located across three different climate zones. Field-level reference data were collected near Valladolid (Castilla y León), serving as the calibration site. Reference plots (whose main statistics are summarized in Fig. 2) include detailed parcel-level wheat and barley yield data from direct post-harvest measurements across eight growing seasons (2017–2018 to 2024–2025). During the study period, the 2019–2020 season was wettest, and 2022–2023 the driest. The validation sites are sparsely located across four autonomous communities for both the 2021–2022 and the 2022–2023 seasons. Fig. 3 summarizes the main statistics of the validation dataset split by Autonomous Community.

In Spain, wheat and barley are winter cereals that are generally sowed in October, while Andalucía typically delays planting until November, and harvested by June. Both crops are generally rainfed and daily temperatures during the winter generally remain above 0°C. Thus, during the winter season the crops do not fully get into dormancy, but their phenological development moves on slowly.

### 2.2. Satellite and reanalysis data

The study uses 17 Sentinel-2 tiles, downloaded from 2017 to 2025 for the calibration site and from 2021 to 2023 for other regions.

For the purpose of this study, Single Look Complex (SLC) Interferometric Wide Swath (IW) dual-polarization VH-VV S1 SAR are used at a reference area of interest of approximately 43-45°. Note that the training dataset results in a time series with a revisit time of 6 days from 2018 to 2021, thanks to the combination of S1A/B; the revisit time then reduces 12 days from 2022 to 2024, due to the failure of S1B; and increases back to 6 days from April 2025 due to the availability of S1C.

Hourly air temperature is obtained from ERA5-Land (Muñoz-Sabater et al., 2021)at 0.1° resolution from 1979 to present.

## 3. Methods

The core methodology relies on the fusion of optical (S2) and radar imagery (S1) combined with meteorological data (ERA5-Land). The general workflow, summarized in Fig. 4, outlines the data curation process, the definition of spectral indices, and the subsequent implementation of linear and Gaussian models designed to evaluate crop dynamics over multiple seasons.

### 3.1. Sentinel-2 and Sentinel-1 processing

Atmospheric correction of the S2 data is performed using the Land Surface Reflectance Code (LaSRC) for S2 (Vermote et al., 2016). For cloud masking, the s2cloudless algorithm is considered with a 0.2 threshold. This algorithm (Zupanc, 2017)is selected due to its robust performance demonstrated in the Atmospheric Correction Intercomparison eXercise (ACIX) (Doxani et al., 2018). As s2cloudless does not generate cloud shadow masks, we use the cloud mask projection available within Google Earth Engine (Gorelick et al., 2017)to address shadowed areas.

The processing of the entire S1 dataset is implemented with the Sentinel Application Platform (SNAP) software provided by European

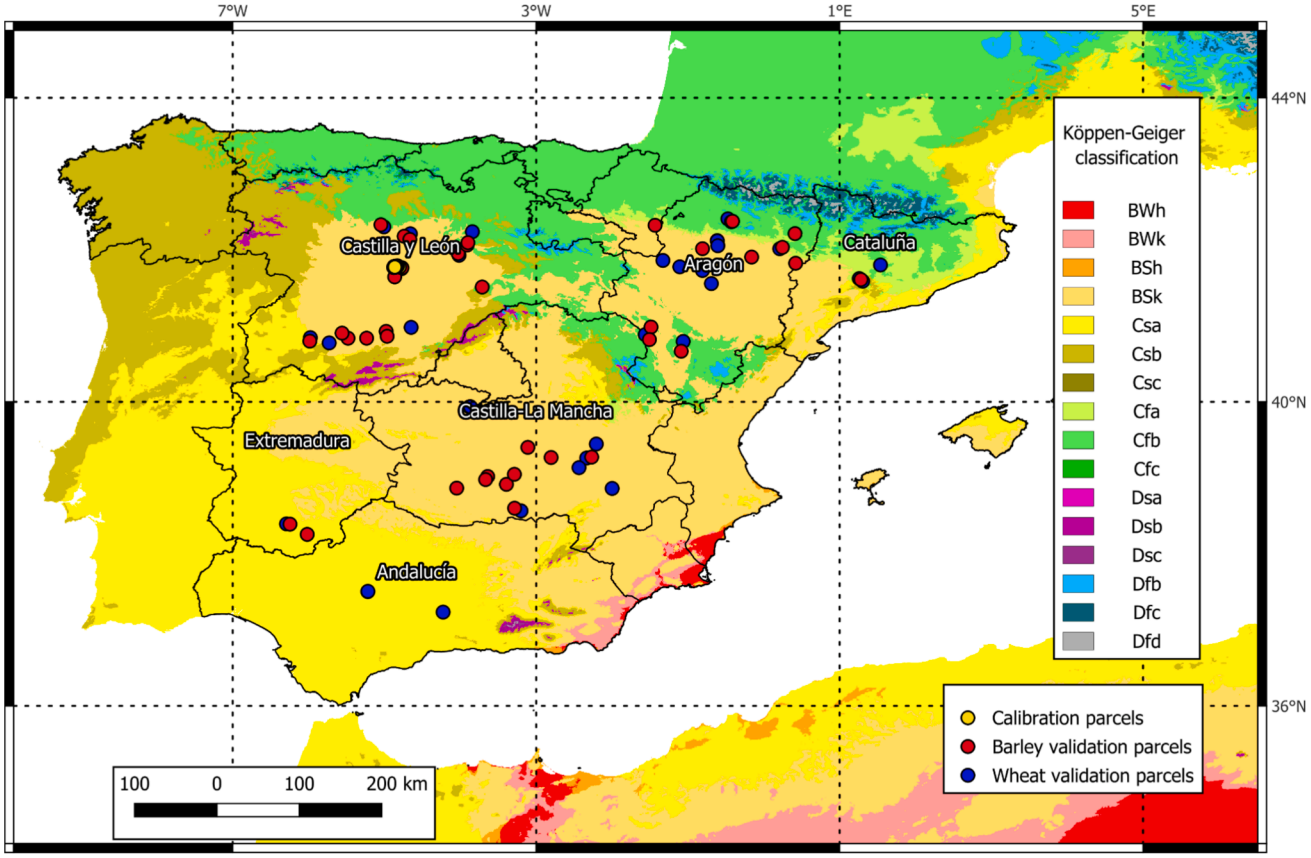



**Fig. 1.** Calibration and validation sites' location. The background shows the Köppen-Geiger classification map (Beck et al., 2018).

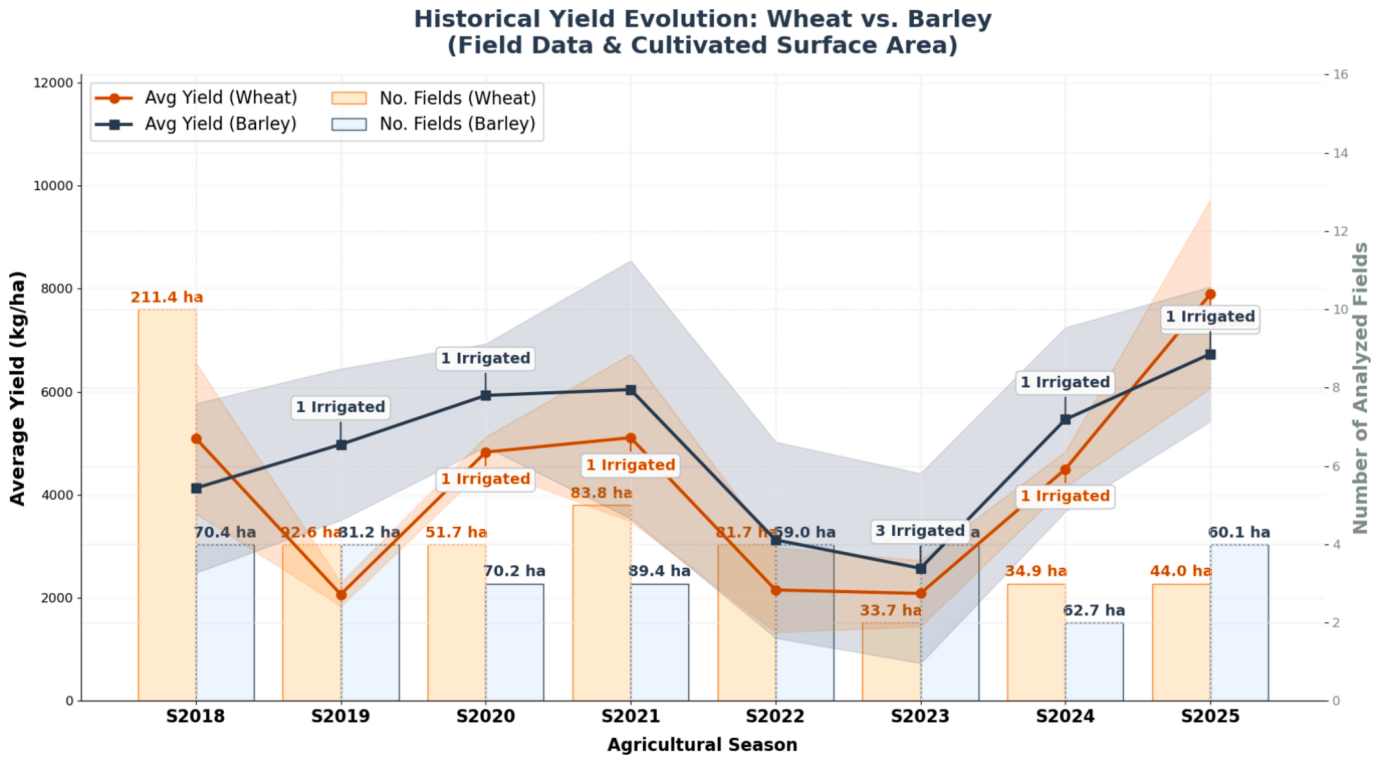



**Fig. 2.** Calibration site main statistics across seasons. The shaded area represents the standard deviation of yields in each season.Years shown in labels correspond to harvest year.

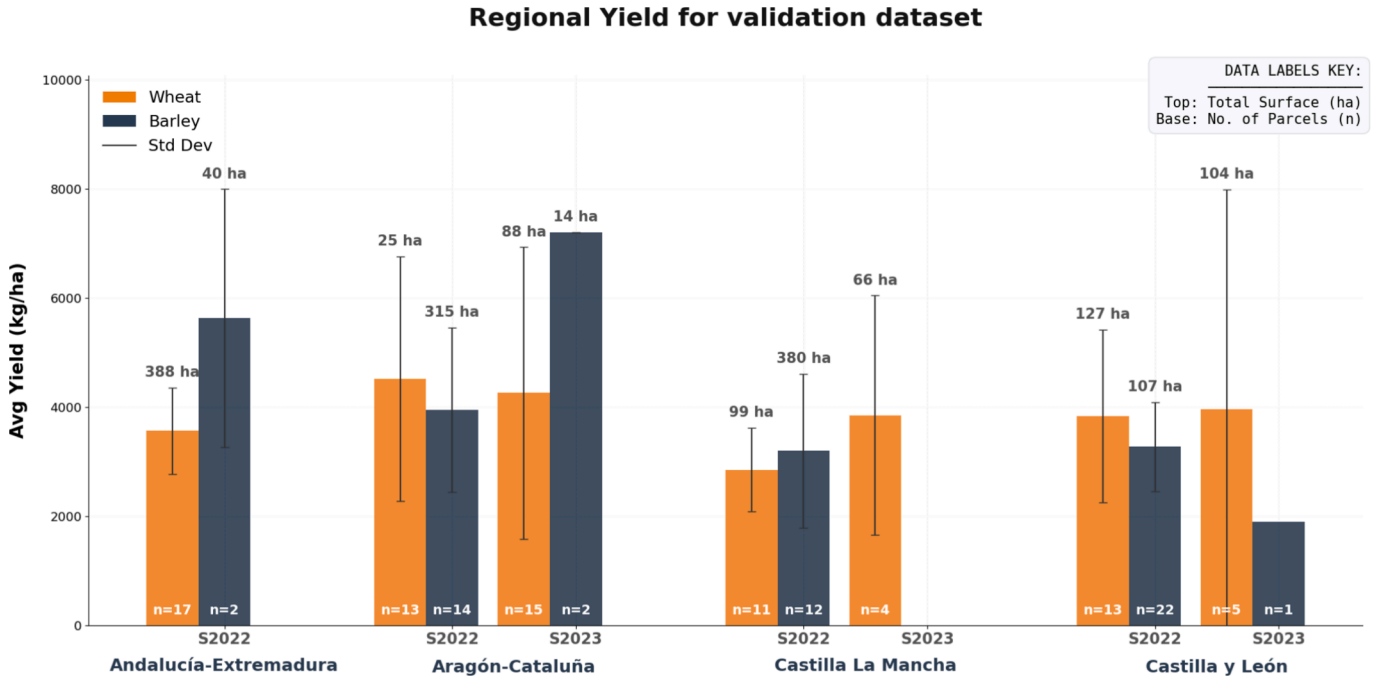



**Fig. 3.** Validation site average yield statistics and standard deviation (represented by the error bar). Years shown in labels correspond to harvest year.

Space Agency (Veci et al., 2017), and consists of the following steps:

- Apply Orbit File.
- Thermal Noise Removal on SLC IW dual-pol S1 data (Mascolo et al., 2022).
- Radiometric Calibration, followed by TOPSAR Deburst.
- Subset to area of interest.
- Generation of C2 matrix followed by Mulitlook (4 x 1) and 5 x 5 boxcar speckle filtering.
- Geocoding.

In the case of dual-polarization SLC S1 SAR measurements, the main observable is the wave coherency matrix C2, which allows characterizing the polarization state of the received wave (Cloude, 2009). The C2

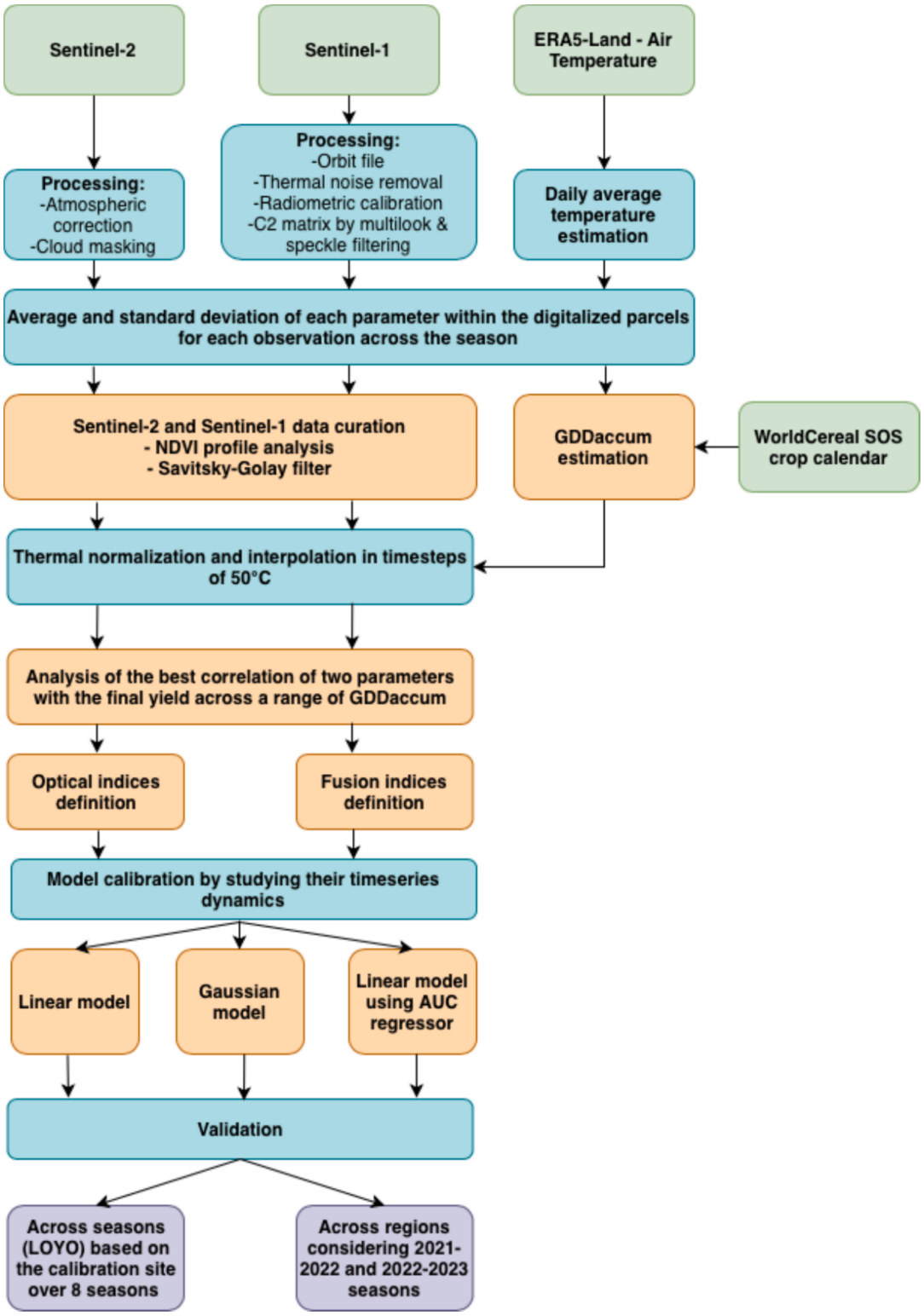



**Fig. 4.** General workflow of the methodology: inputs (green), subsections detailed in the methodology (blue), functions and analysis (orange), and outputs (purple).

matrix and the Stokes vector allow deriving all the polarimetric parameters associated with dual-pol S1 measurements. These parameters, used in this study to define the S1-based empirical models for yield forecasting, are grouped in four main sets: backscattering coefficients and their ratio; the outputs of the eigenvalues decomposition (Cloude, 2009) that is the wave entropy (H2), the average alpha angle ($\underline{\alpha}_2$, and the degree of polarization (DoP); the outputs of model-based decomposition for dual-pol SARs conceived in (Mascolo et al., 2022) that is $m_v$ and $m_p$ the powers of the volume and polarized terms, and the polarized-to-volume ratio $r_{pv} = m_p / m_v$.; radar vegetation indices RVI and DpRVI (Mandal et al., 2020).

### 3.2. Parcels digitalization and data curation

Each agricultural parcel is identified using the SIGPAC code (de Agricultura, 2025), which is the system developed by the Spanish Ministry of Agriculture for identifying agricultural plots within the framework of the Common Agricultural Policy (CAP). Remote sensing signals and the air temperature are extracted by considering the average and standard deviation of cloud free pixels within each polygon considering a 10 m inward buffer to avoid edges. The temporal evolution of average NDVI per parcel is used to assess EO data quality, excluding pixel or average per polygon observations with NDVI values $< 0$, $> 1$, or a standard deviation $> 0.2$.

Residual noise in the S2 time series, often from undetected clouds or shadows, is smoothed using a Savitzky–Golay filter (window = 5, order = 3) to ensure temporal consistency (Chen et al., 2004); (Savitzky and Golay, 1964) . Similarly, S1 data—affected by early-season soil moisture changes and radiometric noise—are filtered using the same method. Additionally, S1 parameters were normalized considering the Min-Max normalization according to (Singh and Singh, 2022) that is:

$$x_{i,norm} = \frac{x_i - min_i}{max_i - min_i}(max - min) + min \quad (1)$$

Where max and min are the desired maximum and minimum in the new range, in our case 0 and 1 to improve comparability with Sentinel-2.

### 3.3. Thermal normalization

This study integrates the Accumulated Growing Degree Days ($GDD_{accum}$) as a proxy to normalize the crops signature to the same temperature accumulation, which is a key driver of phenological development.

$$GDD_{accum}(day) = \sum_{d=biofixdate}^{day} GDD_d \quad (2)$$

Where $GDD_d$ is defined as the average daily maximum and minimum temperatures minus a base temperature of the crop, below which the crop won't progress. The biofix date is established based on the pixel-based WorldCereal Crop Calendars (Moletto-Lobos et al., 2025), that capture winter crops' dormancy periods. The base temperature is considered as 0 °C according to literature (Franch et al., 2015); (Bailey et al., 2020) . However, the results section analyzes the impact of a ± 1°C deviation in base temperature on $GDD_{accum}$ and the accuracy of phenological alignment through the Jeffries-Matusita distances (Khosravi et al., 2018).

All smoothed S2 and S1 parameters of each parcel are interpolated linearly considering timesteps of 50°C to have consistent datasets across the seasons.

### 3.4. Yield indices definition

To identify simple vegetation indices sensitive to wheat and barley yield, the coefficient of determination ($R^2$) of all band/parameter linear combinations is assessed against yield every 50°C $GDD_{accum}$ interval according to equation (4). Optical indices are defined using the full dataset (calibration + validation), while fusion indices are based only on the calibration site and considering 2017–2018 to 2020–2021 and 2024–2025 seasons, to preserve the optimal temporal resolution of S1 due to S1B failure in December 2021 and the availability of S1C since April 2025.

$$Yield_p = k1_G.P1_{p,G} + k2_G.P2_{p,G} + n_G \quad (4)$$

where $k1_G$ and $k2_G$ are the weighting coefficients for each $GDD_{accum}$ range, $G$, $P1_{p,G}$ and $P2_{p,G}$ are the S2 or S1 parameters averaged at parcel level, $p$, and $GDD_{accum}$ range, and $n_G$ is the intercept. Combinations using only S1 parameters are not included, as no added predictive value is observed when combining two S1 variables compared to using one.

The new indices are proposed by analyzing the variability of the $R^2$ of the modelled yield versus the actual yield for each GDD time step of 50°C across the season with the $GDD_{accum}$ representing the phenological evolution. Then, we extract the optimal $GDD_{accum}$ window when the $R^2$ is highest, and when the weighting coefficients ($k1_G$ and $k2_G$) are stable across a GDD range of a minimum of 150°C, to avoid singularities caused by any remaining noise in the timeseries. Finally, the indices are defined for a fixed $GDD_{accum}$ range where these two conditions are accomplished (highest correlation and temporal stability), by averaging the $P1_{p,G}$ and $P2_{p,G}$ along that range, applying equation (4) to infer the resulting $k1_G$ and $k2_G$, and normalizing the weighting coefficients based on their sum.

$$C_1 = \frac{k1_G}{k1_G + k2_G} C_2 = \frac{k2_G}{k1_G + k2_G} \quad (5)$$

### 3.5. Model calibration

Next, the temporal evolution of the newly defined vegetation indices and radar-derived parameters is analyzed across seasons in the calibration site in terms of $GDD_{accum}$ to extract yield-relevant temporal metrics. The following models to capture their temporal metrics that are then calibrated against the yield:

- Linear model considering a static period during the season when the signal is averaged and calibrated against the yield

$$ModelYield_f = m \cdot Index_{GDDinterval} + n \quad (6)$$

where m and n are the slope and intercept estimated within the optimal GDD interval.

- Gaussian fit:

$$Index(day) = A_f \cdot e^{-0.5 \cdot \left(\frac{GDD_{accum}(day) - B_f}{C_f}\right)^2} + D_f \quad (7)$$

where $A_f$, $B_f$, $C_f$ and $D_f$ are the fitting parameters that are calibrated against the yield of each field f.

$$ModelYield_f = s \cdot A_f + t \cdot C_f + u \cdot D_f \quad (8)$$

where s, t and u are the calibration coefficients. The Gaussian fitting yields four parameters ($A_f$, $B_f$, $C_f$, $D_f$), which describe the seasonal shape of the index: amplitude, temporal position, spread and baseline, respectively. Similar to the interpretation discussed in Franch et al. (Franch et al., 2021), not all of these parameters are equally informative for yield prediction. In our calibration, $A_f$ (peak magnitude), $C_f$ (curve width) and $D_f$ (soil baseline) whereas $B_f$ mainly captures the timing of the seasonal peak and exhibited no significant explanatory power for determine the yield evolution based on plant phenology. For this reason, the yield model is defined as a linear combination of $A_f$, $C_f$, and $D_f$ (Eq. (8), with s, t and u representing the regression coefficients.

- Area Under the Curve (AUC) predictor. Within this method we accumulate the index signal starting from the lowest $GDD_{accum}$ value that defines the range in Table 2 until the end of the season. Then, we calibrate the model analogously to equation (6).

$$ModelYield_f = a \cdot AUC(Index_{GDDIndex-GDDEOS}) + b \quad (9)$$

where a and b are the slope and intercept.

- Progressive Multi-Model Machine Learning Strategy

To complement the previous parametric approaches, a non-parametric Machine Learning (ML) benchmarking strategy is also implemented. While the parametric models rely on specific optimal GDD windows, the ML models utilize the full temporal resolution of the season, using as input dataset Sentinel-1, Sentinel-2, air temperature and accumulated precipitation. To manage dimensionality, the time series were aggregated into a single seasonal summary. Specifically, the mean value of each parameter was calculated over the entire active growing period (0–2500 GDD). This approach condenses the temporal variability into a robust static indicator representing the overall crop condition for the season. The final feature matrix is structured such that each row represents a unique parcel-season observation, and columns represent the sequential GDD windows. A diverse battery of regression algorithms are deployed to minimize model bias and evaluate the stability of predictions across different mathematical frameworks:

1. Tree-based Ensemble Models considered: Random Forest (Breiman, 2001), a bagging ensemble method robust to overfitting; and gradient boosting frameworks including XGBoost (Chen and Guestrin, 2016), LightGBM (Ke et al., 2017), and CatBoost (Prokhorenkova et al., 2018). These models are particularly effective at handling tabular data and capturing non-linear dependencies in remote sensing time series.
2. Deep Learning model considered: Multi-Layer Perceptron (MLP) neural network (Taud and Mas, 2017), with an architecture consisting of two hidden dense layers (64 and 32 neurons) with Rectified Linear Unit (ReLU) activation and a linear output layer.

### 3.6. Validation

Each forecast yield model is validated against final Yield by evaluating the RMSE, Relative RMSE (RRMSE), bias, sigma and $R^2$ at the field level (f) (equations 10–15). To assess seasonal transferability, a leave-one-year-out (LOYO) approach is used—each year is iteratively held out for validation while the rest are used for calibration. Due to reduced - S1 temporal resolution after December 2021, the 2021–2022, 2022–2023 and 2023–2024 seasons are excluded from calibration and used only for regional validation.

$$RMSE = \sqrt{\frac{\sum_{f=1}^{N}\left(ForecastYield_f - Yield_f\right)^2}{N}} \tag{10}$$

$$RRMSE = \frac{RMSE}{\frac{\sum_{f=1}^{N}\left(Yield_f\right)}{N}} \tag{11}$$

$$bias = \frac{1}{N}\sum_{f=1}^{N}\left(ForecastYield_f - Yield_f\right) \tag{12}$$

$$sigma = \sqrt{\frac{\sum_{f=1}^{N}\left(ForecastYield_f - Yield_{mean}\right)^2}{N}} \tag{13}$$

Second, the method transferability across regions is evaluated by applying the calibration coefficients ($m$ and $n$ in the case of the linear model; $s$, $t$ and $u$ in the case of the gaussian model, or a and b when considering the AUC predictor) to each area of the validation dataset and estimating the RMSE.

## 4. Results

Based on the methodology presented, we aim to answer the following science questions:

- SQ1: Does the GDD normalization enhance the transferability of models?
- SQ2: Which S2 spectral regions are mainly related to the yield and how does this change through each stage of the crop development?
- SQ3: Can wheat and barley be considered together to develop a crop yield model?
- SQ4: Which is the accuracy of a wheat and/or barley model based solely on S1 data or its added value when integrated with S2 compared to S2 alone?
- SQ5: Can an empirical model trained in a limited region with accurate data be transferable across seasons and regions?

To answer these questions, the section is structured in three subsections. The first section studies the capability of the accumulated GDD to normalize the signal across seasons and regions (SQ1). The second subsection analyzes the definition of the proposed indices (SQ2 and SQ3) based in the case of optical data on all the datasets considered, while the fusion indices definition is restricted to the data acquired during the seasons when S1 was at full capacity, that is 2017–2018 to 2020–2021 and 2014–2025 seasons in the calibration site. Finally, the third subsection is focused on analyzing the performance of the yield models across seasons and regions (SQ4 and SQ5).

### 4.1. Thermal normalization (SQ1)

First, Table 1 shows the impact of ± 1°C deviation in base temperature on $GDD_{accum}$ based on the Jeffries-Matusita distances of the normalized NDVI curves. The results show that the minimum distance (lower separability) is achieved with base temperature equal to 0°C, excepting the seasonal wheat normalization, where the three base temperatures provide very similar JM separability. Therefore, hereafter the $GDD_{accum}$ is estimated using a base temperature of 0°C.

See Fig. 5 (top) shows in wheat the impact of thermal normalization across seasons in the calibration area, and (bottom) the impact across regions, compared to using the Day Of the Year (DOY). Barley plots can be found in Fig. A1. After normalization, the start of the season and the seasonal peak are better aligned, especially across seasons.

See Figs. A2-A4 (Appendix) illustrate the seasonal evolution of $R^2$ for the top-performing S1, S2, and S1 + S2 combinations as a function of GDDaccum, while Figs. A5-A7 shows the corresponding evolution with respect to DOY. When comparing GDDaccum and DOY as temporal references, DOY typically reaches its maximum $R^2$ values near the peak but only within a narrow time window of approximately 15 days. In contrast, GDDaccum provides a more stable and accurate temporal framework, exhibiting high $R^2$ values over a substantially longer period, exceeding 30 days (600°C). This improvement is especially notable in S2 setups calibrated with national-level data, where the use of GDDaccum enhances performance by up to 35% compared to DOY. For fusion indexes using more localized calibration data, the impact is less pronounced, around 10%. Consequently, due to its superior stability and performance metrics, the subsequent analysis focuses exclusively on GDDaccum.

Across the different data sources, S2 alone yields $R^2$ values of approximately 0.45 for wheat and 0.50 for barley (1500–1900°C GDDaccum), highlighting band combinations such as B11–B8 for both crops and B6–B7 for wheat. Integrating S1 and S2 data consistently improves model performance, which is expected given the limited

**Table 1**
Jeffries-Matusita distances of the normalized NDVI curves when considering different base temperatures (Tbase). *Seasonal* refers to the variability of the calibration site across all seasons (2017–2025) while *regional* refers to the different validation regions considered in the study.

| *Tbase (°C)* | *Wheat seasonal* | *Wheat regional* | *Barley seasonal* | *Barley regional* |
|---|---|---|---|---|
| −1 | 0.64 | 0.60 | 0.64 | 0.63 |
| 0 | 0.64 | 0.55 | 0.59 | 0.60 |
| 1 | 0.63 | 0.58 | 0.62 | 0.60 |

**Table 2**
$GDD_{accum}$ ranges (in °C) considered to define each index, equivalent range in DOYs (note that it represents the average range of DOYs when the GDD ranges happen) and relative weighting coefficients $C_1$ and $C_2$ that define each new index.

| *Crop* | *Data source* | *Definition* | *C1* | *C2* | *GDD range* | *DOY equivalence* |
|---|---|---|---|---|---|---|
| Wheat | S1 + S2 | $WyFI_{B11\text{-}ratio}$ = c1*B11 + c2*VH/VV | −0.87 | 0.13 | 550–1950 | 341–152 |
| | | $WyFI_{B5\text{-}VH}$ = c1*B5 + c2*VH | −0.91 | −0.09 | 900–1300 | 51–101 |
| | S2 | $WyOI_{B2\text{-}B4}$ = c1*B2 + c2*B4 | 0.57 | −0.43 | 1300–1700 | 101–136 |
| | | $WyOI_{B11\text{-}B8}$ = c1*B11 + c2*B8 | −0.90 | 0.10 | 1500–1800 | 120–143 |
| | | $WyOI_{B11\text{-}B3}$ = c1*B11 + c2*B3 | −0.76 | 0.24 | 1400–1950 | 111–152 |
| Barley | S1 + S2 | $ByFI_{B11\text{-}ratio}$ = c1*B11 + c2*VH/VV | −0.88 | 0.12 | 1000–1800 | 64–143 |
| | | $ByFI_{B3\text{-}ratio}$ = c1*B3 + c2*VH/VV | −0.89 | 0.11 | 1300–1500 | 101–120 |
| | S2 | $ByOI_{B11\text{-}B8}$ = c1*B11 + c2*B8 | −0.98 | −0.02 | 1500–1700 | 101–136 |
| | | $ByOI_{B11\text{-}B4}$ = c1*B11 + c2*B4 | −0.73 | 0.27 | 1400–1700 | 111–136 |

spatial extent and number of seasons used in this analysis. For wheat, fusion models exceed $R^2 = 0.70$ within the 800–1400°C GDDaccum range when using combinations such as B5 + VH/VV or B11 + VH. For barley, fusion models achieve $R^2$ values close to or above 0.80 across a broader window (1000–1900°C GDDaccum), particularly with B11–VH/VV combinations.

### 4.2. Indices definition (SQ2 and SQ3)

Table 2 summarizes the definitions of the newly derived indices (optical and fusion), which are selected according to their optimal performance within the yield prediction models presented in the following section. Among the optical indices, B11 emerges as an important contributor for both crops, and its relative weighting coefficient underscores its relevance compared with other spectral bands or S1-derived parameters. Within the fusion indices, the VV/VH parameter is also consistently influential for both crops, whereas no polarimetric (C2-derived) parameter demonstrates substantial added value.. The table also reports the corresponding $GDD_{accum}$ ranges used to define each index. Although most indices are defined around the timing of the NDVI seasonal peak, wheat indices—particularly those derived from S1 + S2 fusion, as highlighted in the previous section—tend to span a broader phenological window than those defined for barley. Moreover, despite some differences in the optimal temporal windows between crops, the indices combining the same parameters display very similar relative weighting coefficients. This consistency constitutes a preliminary indication of the feasibility of integrating both crops into a unified yield prediction model (SQ3).

A sensitivity analysis is performed to evaluate the stability across seasons and regions of each index by shifting 0,05 the relative weighting coefficients from 0 to 1 (so that $C_2 + C_1 = 1$) and studying the variability of the $R^2$ between the resulting vegetation index and the yield. This is done individually in each season of the calibration site and in each region, to check the relative weight variability across seasons or regions, and globally considering all datasets together.

As an example, Fig. 6 presents the $R^2$ variability under varying relative weighting coefficients across seasons (represented by the shaded area around the black line) or regions focusing on the wheat indices $WyOI_{B11\text{-}B8}$ and $WyFI_{B11\text{-}ratio}$, and the barley indices $ByOI_{B11\text{-}B8}$ and $ByFI_{B11\text{-}ratio}$. The other indices' plots can be found in Fig. A8. The x-axis represents the normalized coefficient $C_1$, and the secondary x-axis indicates the corresponding value of $C_2$. Note that the left plot displays the case where both coefficients have the same sign and the left plot displays the case where they have opposite signs (independently on which has the negativity) given their equivalence in $R^2$. The peak $R^2$ value denotes the optimal balance of both inputs.

All regions show a similar profile of $R^2$, showing a consistent increase with greater weight on B11 against the weight of the other parameters. In the optical indices, the maximum including all regions is reached close to $C_1 = 1$, while there are some regions whose maximum occurs at lower C1 values such as the calibration site, Aragon-Cataluña or Castilla La Mancha. However, in the fusion indices, the $R^2$ shows a peak around 0.85 across most regions and when combining them.

See Fig. 7 presents the seasonal dynamics of the selected optical and fusion indices at parcel level against the $GDD_{accum}$. The other indices' plots can be found in Fig. A9. All indices generally show negative values to preserve the typical seasonal evolution of vegetation indices with lower values at the beginning and the end of the season and maximum values (peak) in the middle of the season. While both optical and fusion indices have a major contribution of B11, each index shows a different temporal profile. Clear stratification is observed, particularly between the lowest (500–2333 kg/ha) and highest (9666–11500 kg/ha) yield classes. This separation is most pronounced in the 1400–1800°C $GDD_{accum}$ range,which typically corresponds to phenological stages such as heading and early grain filling in Mediterranean wheat and barley systems.

See Fig. 8 illustrates the temporal evolution of the inverse of the VH/VV ratio (i.e., the VV/VH ratio), and the average alpha angle (Av.α). The temporal profiles typically exhibit a bimodal pattern, though the timing and shape of the second peak or dip varies among parameters. Av.α second peak is around GDD ~ 1700, aligning with grain filling when radar sensitivity to yield is highest. In contrast, VV/VH shows a peak at GDD ~ 1700, while its earlier minimum (~1500 GDD) likely reflects canopy structural changes such as flag leaf emergence, is less directly linked to yield. To ensure phenological relevance, we focus on the GDD ~ 1700 segment—either the second peak (Av.α, VV) or the minimum (VV/VH) and fit a Gaussian function, whose fitting parameters are then linked to the yield, as described in (8).

### 4.3. Model transferability (SQ4 and SQ5)

In this final section all indices and parameters are evaluated to model the crop yield.

#### 4.3.1. Transferability across seasons

See Fig. 9 presents two Taylor diagrams (Taylor, 2001)summarizing the main performance metrics of the wheat (left) and barley (right) models at the calibration site across all seasons using the LOYO validation approach. Note that the three seasons excluded from calibration due to the insufficient temporal resolution of S1 (2021–2022, 2022–2023, and 2023–2024) are also included in this seasonal validation to enhance its robustness. Both crops reach maximum $R^2$ values slightly above 0.7, with RMSE and sigma values around 1000 kg $ha^{-1}$. The best-performing models for both crops correspond to fusion indices combined with a linear model. One noticeable difference between crops is that the wheat diagram includes more models with relatively good performance than barley, since the diagrams just present models with reasonable performance. Optical indices generally perform better for wheat when using a linear model, whereas their performance in barley is comparatively lower. Models based solely on S1 exhibit clear underperformance, particularly for barley. Machine learning models tend to

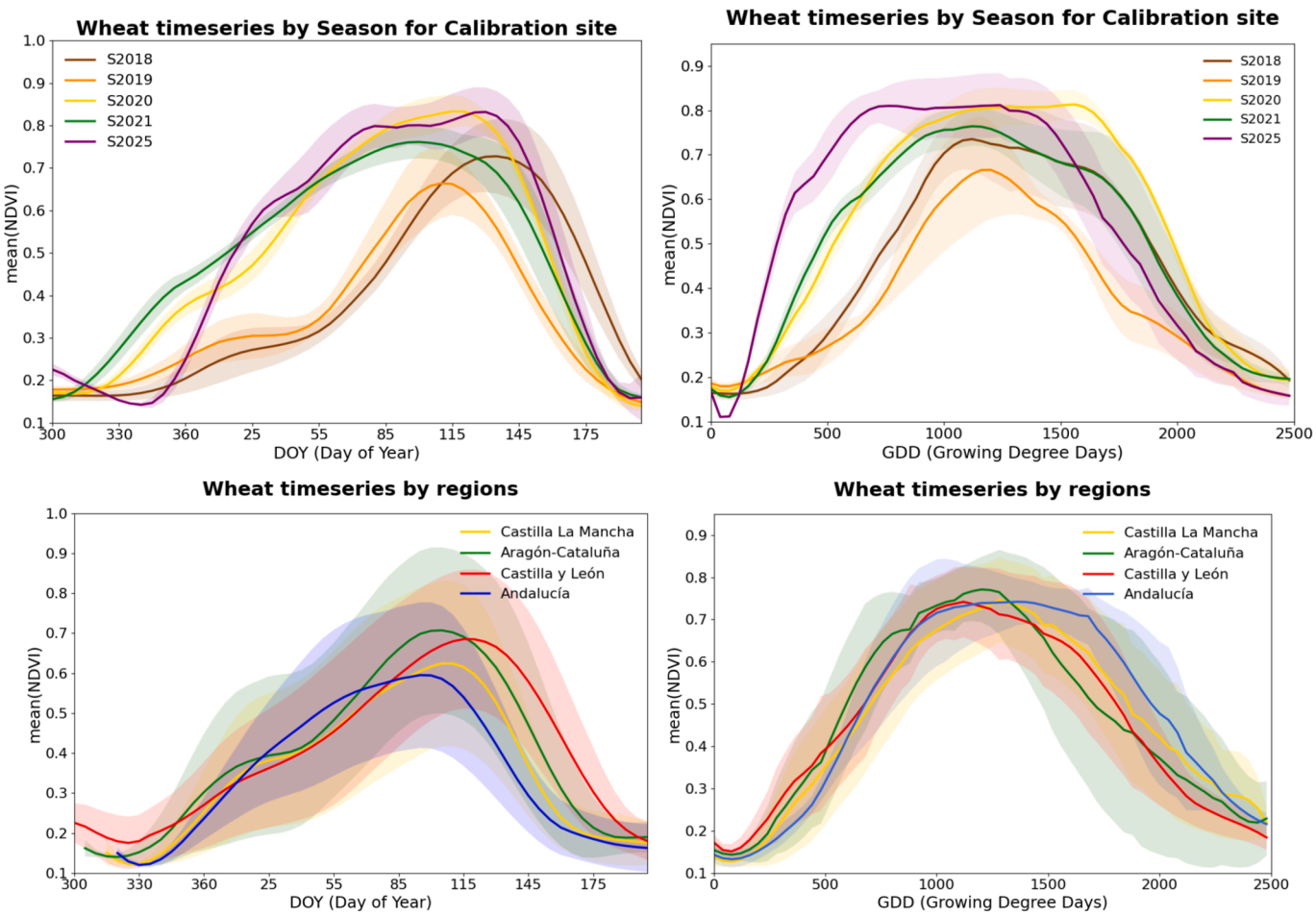



**Fig. 5.** Average NDVI wheat temporal evolution of each season in the calibration site (top) or region (bottom) against the DOY (left) and the $GDD_{accum}$ (right). The shaded area represents the standard deviation of all parcels considered in each case.

yield poorer results for wheat; however, for barley, the CatBoost algorithm achieves a notably low RMSE of 671 kg $ha^{-1}$, although its $R^2$ remains moderate at 0.5.

See Fig. 10 displays the validation of the best performance models in each crop. The optical models show worse performance metrics, mainly as a consequence of a higher sigma, while the fusion indices show very good metrics, especially in barley.

### 4.3.2. *Transferability across regions*

See Fig. 11 presents the Taylor diagrams for the regional validation of wheat (left) and barley (right). As expected, model performance decreases compared to the seasonal validation: $R^2$ values are slightly above 0.5 for wheat and slightly below 0.5 for barley, with RMSE and sigma values of approximately 1400 kg $ha^{-1}$ for wheat and 1300 kg $ha^{-1}$ for barley. For both crops, the best-performing models (with lowest RMSE) rely on fusion indices combined with a Gaussian model, in wheat, or linear models, in barley. Notably, in wheat most optical models yield similar $R^2$ values but exhibit more variability in RMSE. As observed in the LOYO seasonal validation, barley displays a more restricted set of viable models compared to wheat. S1-only models consistently show poor performance. Finally, machine learning models fail to transfer across regions, with $R^2$ values close to zero.

See Fig. 12 displays the best performing models' regional validation. For wheat the added value of the fusion models compared with the optical is not statistically significant. In contrast, for barley the fusion index yields a notable improvement, reducing the RMSE by approximately 200 kg ha-1 and increasing R2 from 0.40 to 0.46 relative to the optical model.

See Fig. 13 summarizes regional validation metrics—RMSE, RRMSE sigma, bias, and $R^2$—for wheat and barley yield prediction of the models highlighted in Fig. 11 but extracting the statistics for each Autonomous Community in Spain separately. Note that Andalucia-Extremadura is not represented in barley since the sample just consisted of two data. Despite showing comparable overall performance metrics, at regional scale each model shows different metrics. Both crop types, using optical indexes $WyOI_{B2\text{-}B4}$ and $ByOI_{B11\text{-}B8}$ with linear fitting, show the best results in Castilla-La Mancha (highest $R^2$ above 0.9/0.7 and lowest RMSE ~ 1000/1500 kg/ha). Fusion indexes using the same band combination show similar performance, being the best correlation for $WyFI_{B11\text{-}ratio}$ in Aragón-Cataluña region ($R^2 \sim 0.6$ and RMSE $<$ 2000 kg/ha) and in Castilla-La Mancha for $ByFI_{B11\text{-}ratio}$ ($R^2 \sim 0.6$ and RMSE $<$ 1500/ha).

## 5. Discussion

Empirical yield models are fundamentally constrained by their dependence on training data, which may be spatially or temporally unrepresentative and thus difficult to transfer across regions or conditions. This study proposes a strategy to enhance model transferability by combining (1) a high-quality but localized calibration dataset, (2) thermal normalization through GDDaccum, (3) newly defined indices tailored to yield dynamics, and (4) the integration of optical and radar information from Sentinel-2 and Sentinel-1. Next, the science questions are addressed.

A crucial step in the methodology is the use of $GDD_{accum}$ to reduce seasonal or regional shifts driven by temperature variability and enhance transferability (SQ1). While DOY-based correlations' between

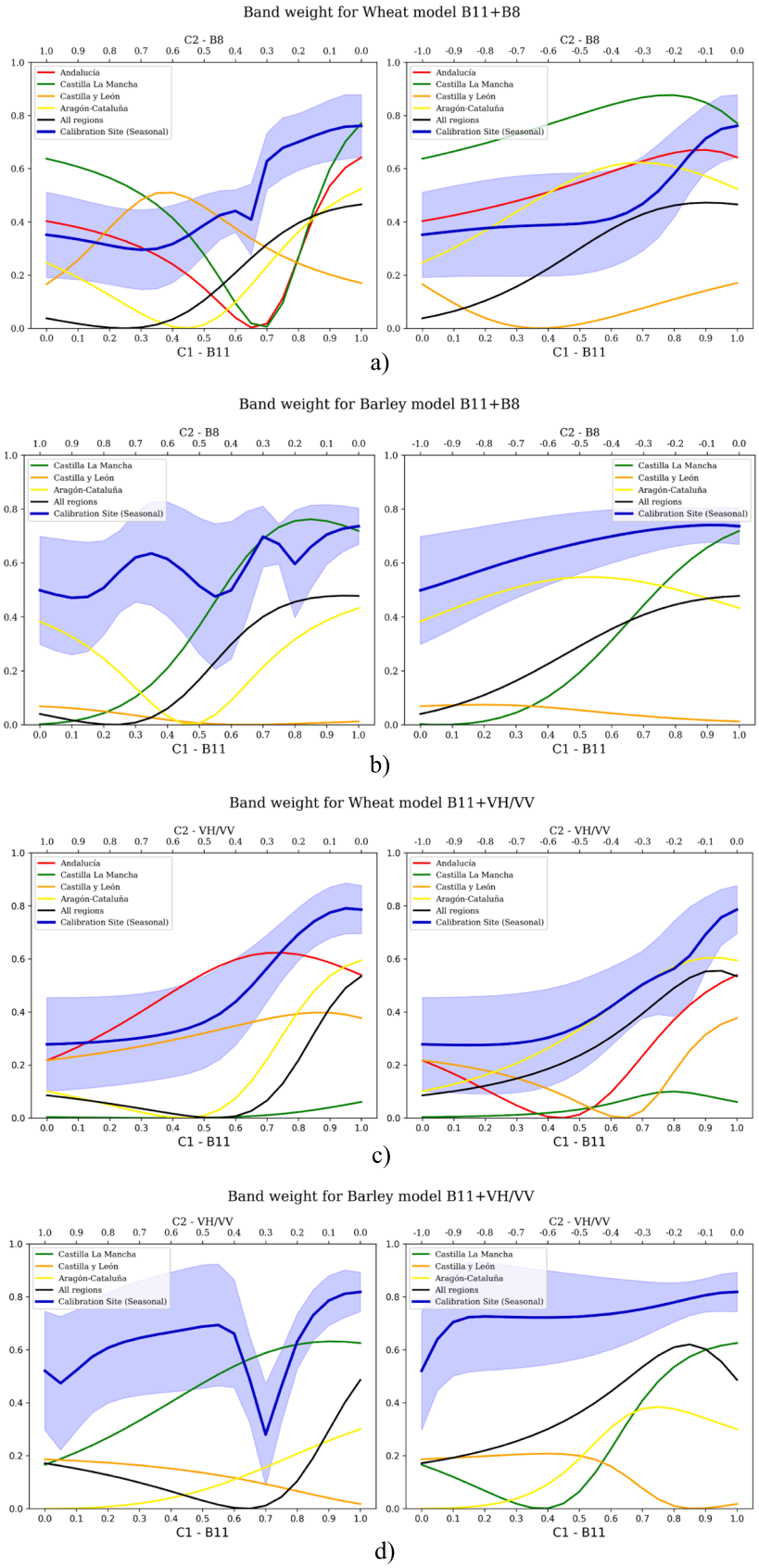



**Fig. 6.** $WyOI_{B11\text{-}B8}$ (a), $ByOI_{B11\text{-}B8}$ (b), $WyFI_{B11\text{-}ratio}$ (c) and $ByFI_{B11\text{-}ratio}$ (d) stability analysis by analyzing the $R^2$ variability when changing the relative weight of each S2 band considering both positive (left) or one negative and one positive (right) against the yield. The black line represents the average $R^2$ across all seasons in the calibration site and the shaded area represents its standard deviation.

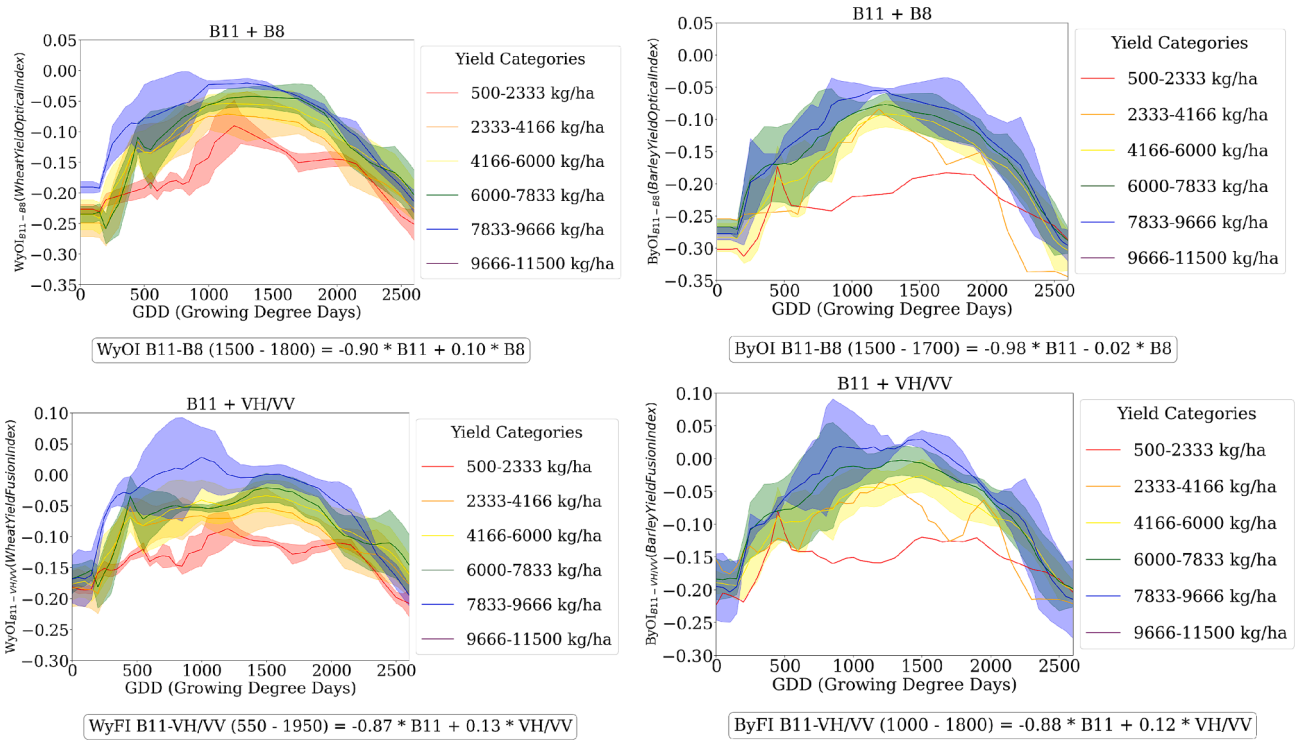



**Fig. 7.** Temporal evolution of the optical (top) and fusion indices (bottom), stratified by yield class (shaded area represents standard deviation of each class) in the calibration site.

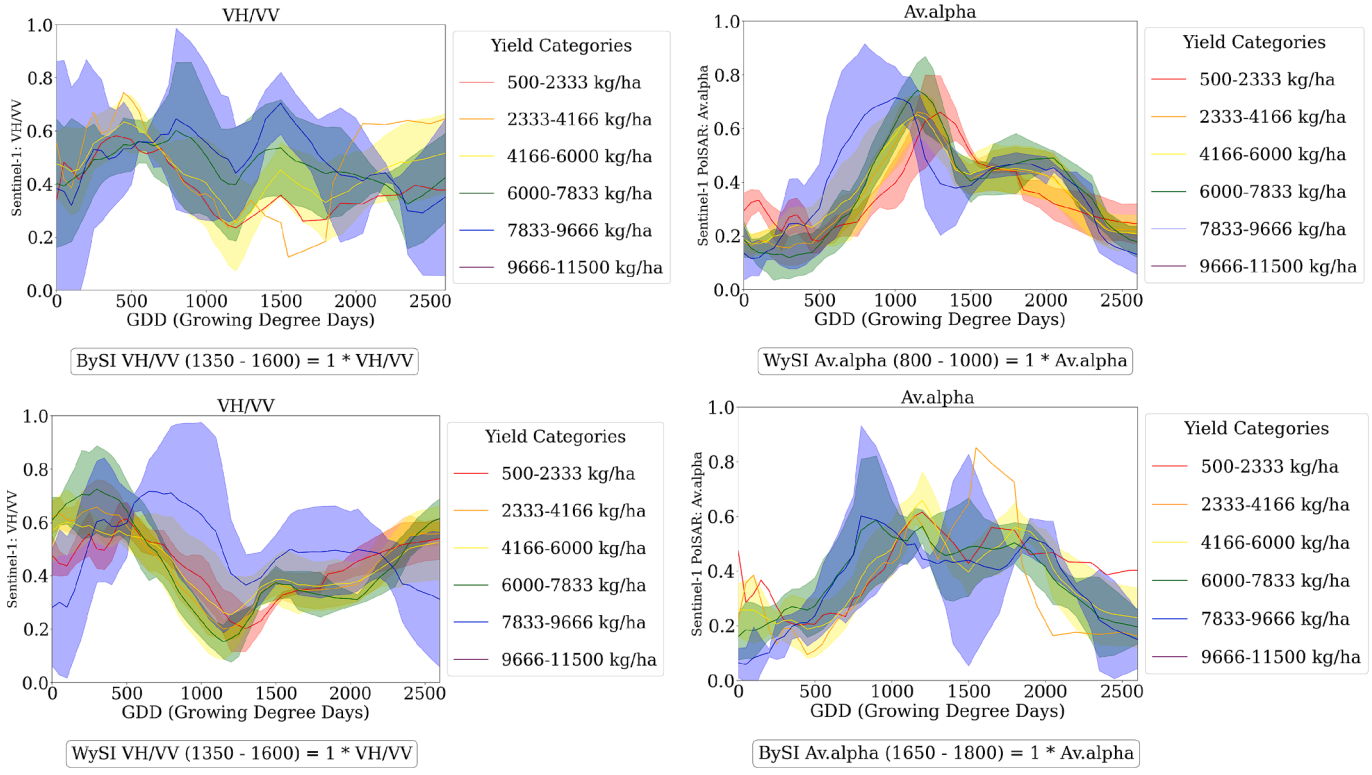



**Fig. 8.** Temporal evolution of the S1 parameters that are better correlated to the yield for wheat (right) and barley (left).

the EO parameters and yield peaked only within a narrow 10–20-day period around maximum NDVI, $GDD_{accum}$ maintained high $R^2$ values for more than 30 days. This broader stability enabled more reliable identification of phenologically relevant windows to define new indices (e.g., 1500–1900°C). However, while this parameter helps align especially the seasonal peak, the end of the season shows a wide standard deviation in the regional normalization. Thus, GDDaccum enhances transferability across seasons but only partially across regions, confirming the need for

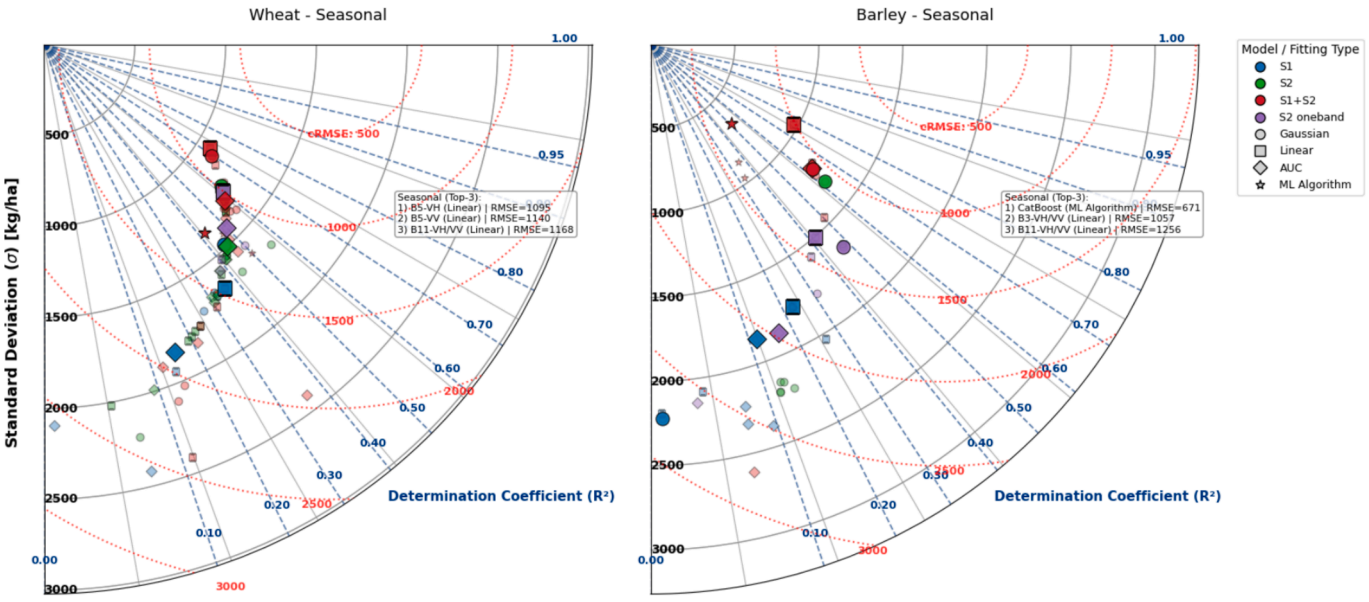



**Fig. 9.** Taylor diagram summarizing the seasonal validation's $R^2$, sigma and RMSE performance metrics of all models in wheat (left) and barley (right).

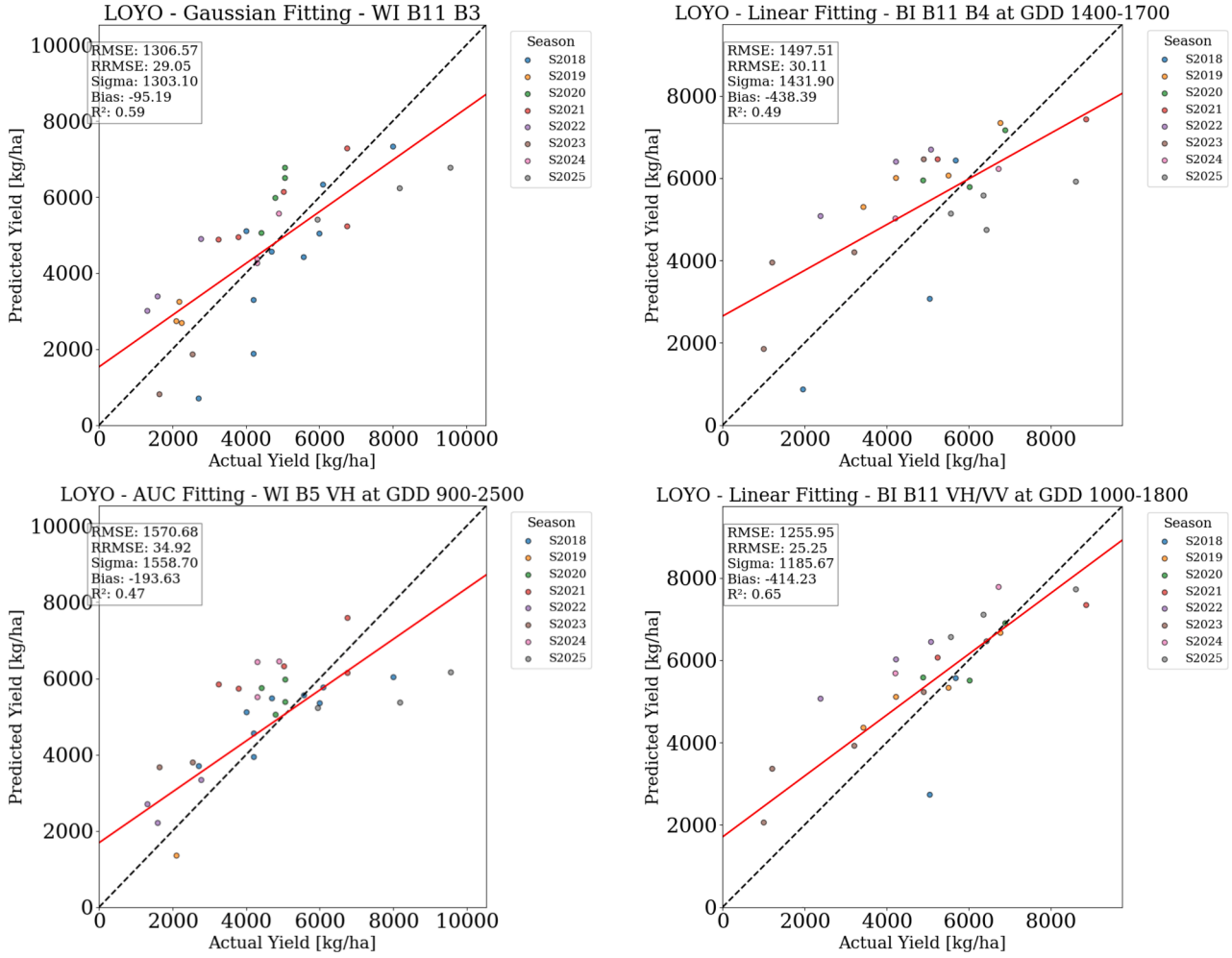



**Fig. 10.** Best performing wheat (left) and barley (right) best models based on optical (top) or fusion (bottom) indices of the calibration site's seasonal validation. The models with the lowest RMSE are highlighted in the boxplot.

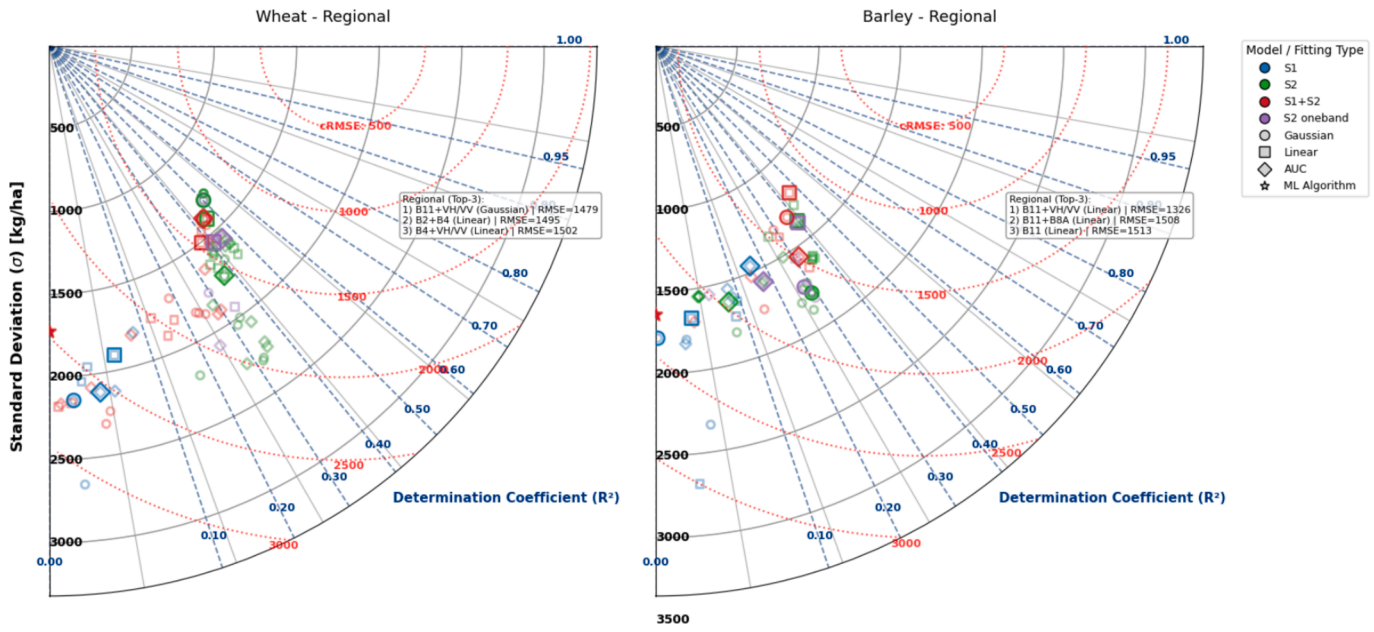



**Fig. 11.** Taylor diagram summarizing the seasonal validation's $R^2$, sigma and RMSE performance metrics of all models in wheat (left) and barley (right). The models with the lowest RMSE are highlighted in the boxplot.

additional normalization strategies in highly heterogeneous environments.

The proposed indices, using just two spectral bands and simple calculations, offer a scalable, interpretable approach to yield estimation. The indices definition reveals a consistent and biologically meaningful pattern that answers SQ2: B11 (shortwave infrared) emerges as the most influential band. In both crops, B11 (and its combination with other) shows the best performance during 1500-1900°C which captures the seasonal peak and the start of senescence. This strong contribution might be caused by the strong response of B11 to canopy water status, senescence, and internal structure, all of which are directly linked to yield formation, especially under Mediterranean rainfed conditions. This result aligns well with other works such as Xiao et al. 2024 who highlighted that the NIR and SWIR bands show highest correlations with yields, and other studies that underline that rainfed wheat and barley yields are mainly driven by water availability (e.g. (Del Pozo et al., 2016); (Gaona et al., 2022) .

Focusing on SQ3 (whether wheat and barley can be modeled jointly), the analysis of the best-performing indices shows that some provide highly consistent weighting coefficients across both crops, despite being calibrated separately. This is most evident for the fusion B11–VH/VV index, which not only exhibited nearly identical weights for wheat and barley, but also emerged as the top-performing predictor in both species, yielding comparable validation metrics. In contrast, barley generally displayed lower overall performance and a more limited set of reliable models. Taken together, these findings suggest that joint modeling of wheat and barley is feasible—but only when using spectral or radar–optical combinations that elicit a similar biophysical response to yield in both crops, such as the proposed B11–VH/VV fusion index.

Focusing on the value of SAR data (SQ4), the SAR-only time series highlights key limitations for empirical-based yield prediction which could be attributed to its high early-season soil moisture sensitivity, reduced late-season separability, and overall low correlation with yield. In fusion indices, SAR contributed only lower weights than optical bands (generally $< 0.15$), suggesting a limited direct role in yield prediction. Nonetheless, SAR provided complementary information under specific conditions—particularly in barley —leading to moderate improvements in RMSE during regional validation. However, reduced S1 temporal resolution during 2021–2024 constrained the potential benefits of radar integration.

Finally, to evaluate if empirical models can be trained in limited regions with good quality data across seasons and regions (SQ5), we tested several indices and empirical models, from parametric (gaussian, or linear) to non-parametric ML. Seasonal validation demonstrated that the models transferred well across years, with both wheat and barley reaching $R^2$ values above 0.7 and errors around 1000 kg/ha when using fusion indices in linear models. Optical indices alone performed reasonably for wheat but were less reliable for barley, while SAR-only models consistently showed poor predictive skill. Machine learning approaches did not outperform simple parametric models and, in most cases—especially for wheat—yielded weaker results. When applied across regions, model performance declined noticeably. Wheat reached $R^2$ values slightly above 0.5 and barley slightly below, with RMSEs around 1400 kg/ha and 1300 kg/ha respectively. Fusion indices provided the best results in both crops, although their advantage over optical indices was negligible for wheat. In barley, however, fusion offered a clear improvement, reducing RMSE by about 200 kg/ha and increasing $R^2$ from 0.40 to 0.46. Overall, regional transferability proved more challenging, and machine learning models failed entirely to generalize across regions.

Regional validation based on S2 indices shows strong generalizability, while fusion indices enhance performances in some areas. In wheat, the linear $WyOI_{B2\text{-}B4}$ performs best, especially in Castilla-La Mancha where $R^2$ reaches over 0.9 and RMSE below 1000 kg/ha. In Aragón-Cataluña the fusion index $WyFI_{B11\text{-}ratio}$ achieves strong results ($R^2 > 0.6$, RMSE $< 1500$ kg/ha) with a consistent bias of 100 kg/ha. For barley, validation results are generally weaker. The B11–B8 combination ($ByOI_{B11\text{-}B8}$, Linear model) yields the best overall metrics, while the fusion index $ByFI_{B11\text{-}ratio}$ performs best among fusion indices. Regionally, Aragon-Cataluña and Castilla La Mancha show better results with the $ByOI_{B11\text{-}B8}$. Finally, Castilla y Leon- the training region- shows poor metrics with all methods. Compared to wheat, barley yields are generally lower in rainfed conditions given that barley plants are more resistant and can be planted in regions with poorer conditions. Also, the wheat sample almost doubles the barley sample, which may drive the lower performance metrics of barley.

Our findings are consistent with past studies emphasizing the importance of SWIR and Red Edge bands for cereal yield estimation (e.g., (Xiao et al., 2024); (Sagan et al., 2021) . Similar to Zhuo et al. (Zhuo et al., 2019), integrating SAR and optical data produced modest

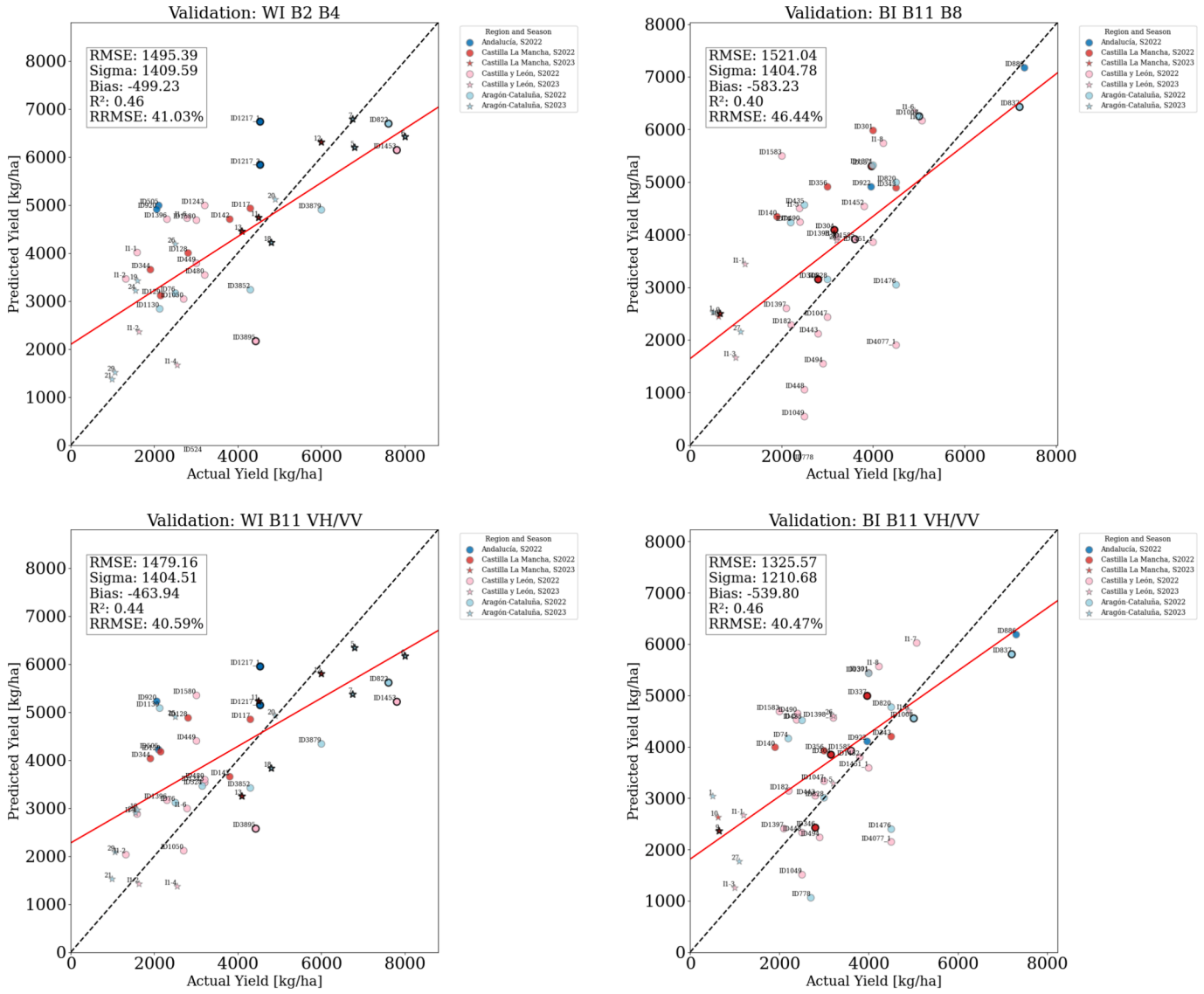



**Fig. 12.** Wheat (left) and barley (right) best performing models of the regional validation considering the linear model in the optical indices and the gaussian fitting model in the fusion indices.

improvements, although the magnitude in our study was smaller due to temporal gaps in S1 acquisitions after 2021. Seasonal validation metrics also resemble those reported in Switzerland (Amin et al., 2024)and Western Europe (Perich et al., 2023), while regional transferability aligned with earlier observations that empirical models degrade significantly when applied across agro-ecological gradients (Hunt et al., 2019).

Within this framework, the main contributions of our study are: 1) Thermal normalization applied at field scale across eight seasons is rarely explored in high-resolution yield studies, demonstrating its value for stabilizing spectral–yield relationships; 2) we propose newly defined indices using only two parameters to provide interpretable, computationally simple predictors; 3) we include a systematic assessment of S1 vs. S2 vs. fusion contributions, showing the limited but in some cases valuable role of radar for yield inference with empirical models; 4) a structured evaluation of transferability, combining LOYO and multi-region validation, which is still uncommon in empirical yield modeling literature; and 5) a crop-specific treatment of wheat and barley, showing why joint models are not advisable, despite similar EO signatures, though providing some insights on how they could be combined.

Several limitations must be acknowledged. First, regional transferability remained limited, confirming that a localized, high-quality calibration dataset cannot fully represent the diversity of environmental conditions found across Spain. Expanding calibration datasets to include multiple agro-climatic zones would likely improve robustness. Second, the reduced S1 revisit frequency during key years constrained the assessment of radar's full potential; future work should revisit this analysis once S1C ensures consistent 6-day coverage. Third, while simple indices offer interpretability and are well suited to current multispectral missions, the forthcoming Copernicus Hyperspectral Imaging Mission for the Environment (CHIME) will deliver far richer spectral information, requiring more advanced and efficient integration strategies beyond simple index formulations. Finally, future efforts should explore additional normalization strategies combining temperature, moisture, and crop-specific phenological models to further enhance regional transferability.

## 6. Conclusions

This study demonstrates that empirical yield models for wheat and barley can achieve strong seasonal transferability when supported by

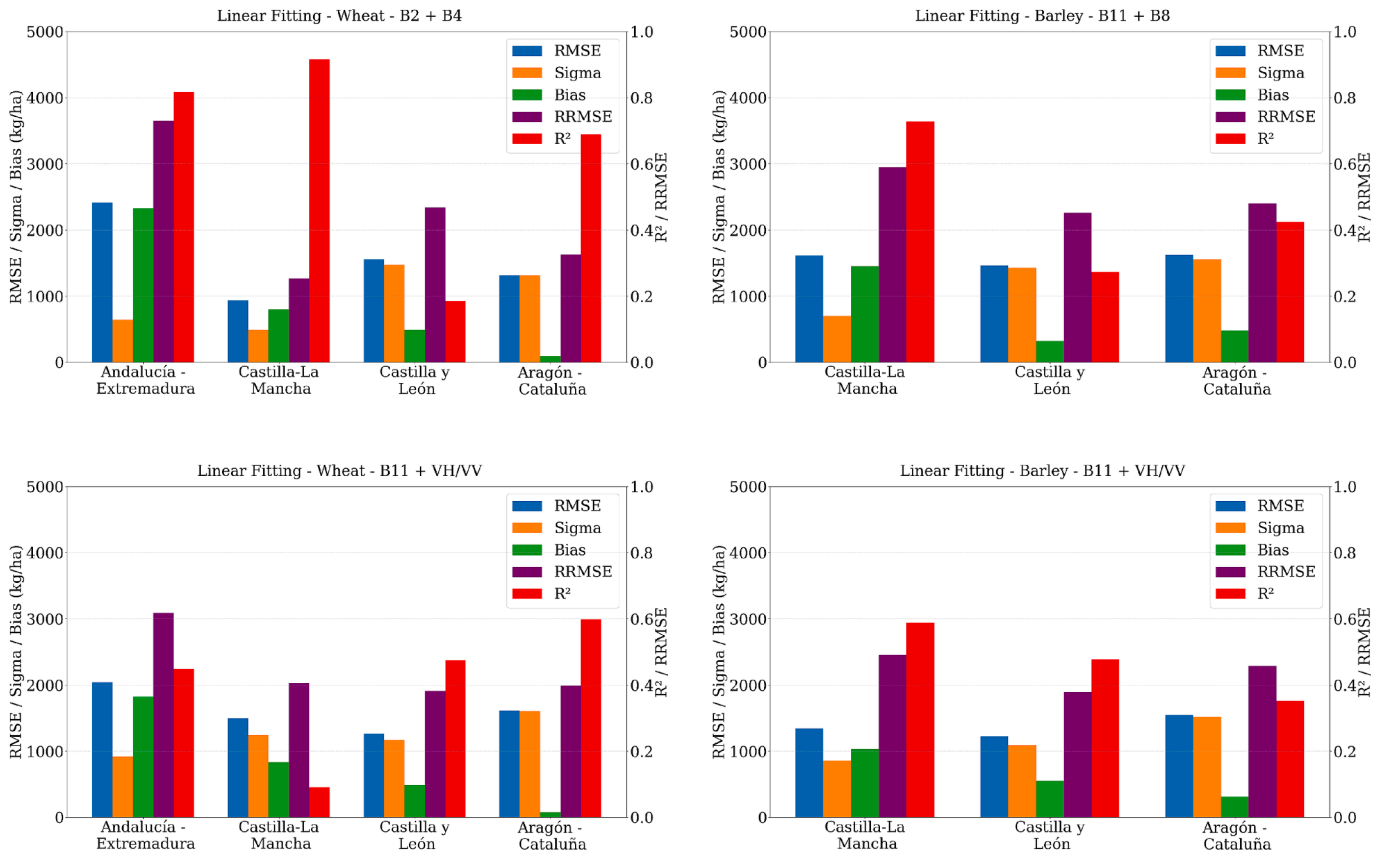



**Fig. 13.** Regional validation performance metrics (RMSE, standard deviation (sigma), bias and $R^2$) of the models highlighted in Fig. 12.

thermal normalization and simple, biologically meaningful indices derived from Sentinel-2. The SWIR band B11 is confirmed as the most yield-sensitive feature in both crops, while SAR adds only limited complementary information. Barley and wheat require species-specific calibration due to subtle but relevant differences in their spectral–yield relationships.

However, regional transferability remains limited, validating concerns that localized calibration alone cannot fully generalize across heterogeneous agro-ecological environments. Fusion models offer modest improvements in certain regions but do not overcome the inherent variability introduced by climate, management, and soil differences. Future efforts should therefore prioritize multi-regional calibration, improved SAR temporal consistency, and enhanced phenological normalization strategies.

Overall, the study provides a structured framework to improve transferability in empirical yield modeling and highlights both the opportunities and persistent challenges for operational EO-based yield estimation.

## CRediT authorship contribution statement

**Belen Franch:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Italo Moletto-Lobos:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation. **Javier Tarín-Mestre:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation. **Lucio Mascolo:** Writing – original draft, Methodology, Formal analysis. **Eric Vermote:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization. **Natacha Kalecinski:** Writing – review & editing, Validation, Formal analysis. **Inbal Becker-Reshef:** Writing – review & editing, Validation, Conceptualization. **Alberto San-Bautista:** Writing – review & editing, Funding acquisition. **Constanza Rubio:** Writing – review & editing, Funding acquisition. **Sara San Francisco:** Writing – review & editing, Resources, Funding acquisition. **Miguel Ángel Naranjo:** Writing – review & editing, Resources, Funding acquisition. **Vanessa Paredes:** Writing – review & editing, Resources. **David Nafria:** Writing – review & editing, Resources. **Carlos Cantero-Martinez:** Writing – review & editing, Resources.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This research has been funded by the PREDIC-PRO project SCPP2100C008733XV0 (CPP2021-008733) of the State Research Agency of the Ministry of Science, Innovation and Universities.

## Appendix

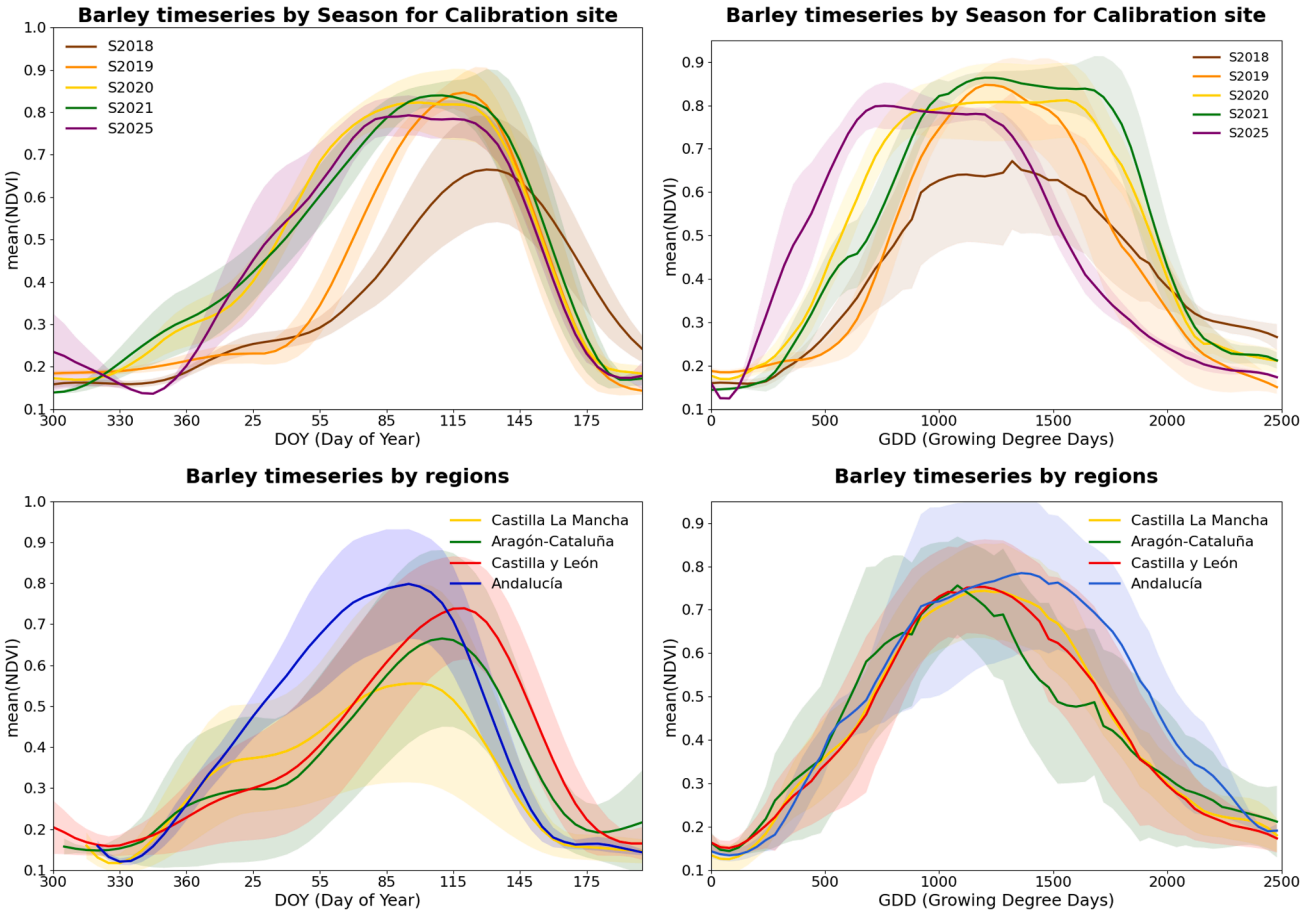



**Fig. A1.** Average NDVI wheat temporal evolution for each season (top) or for each region (bottom) against the DOY (right) and the $GDD_{accum}$(left). [1] The shaded area represents the standard deviation of all parcels considered in each case.

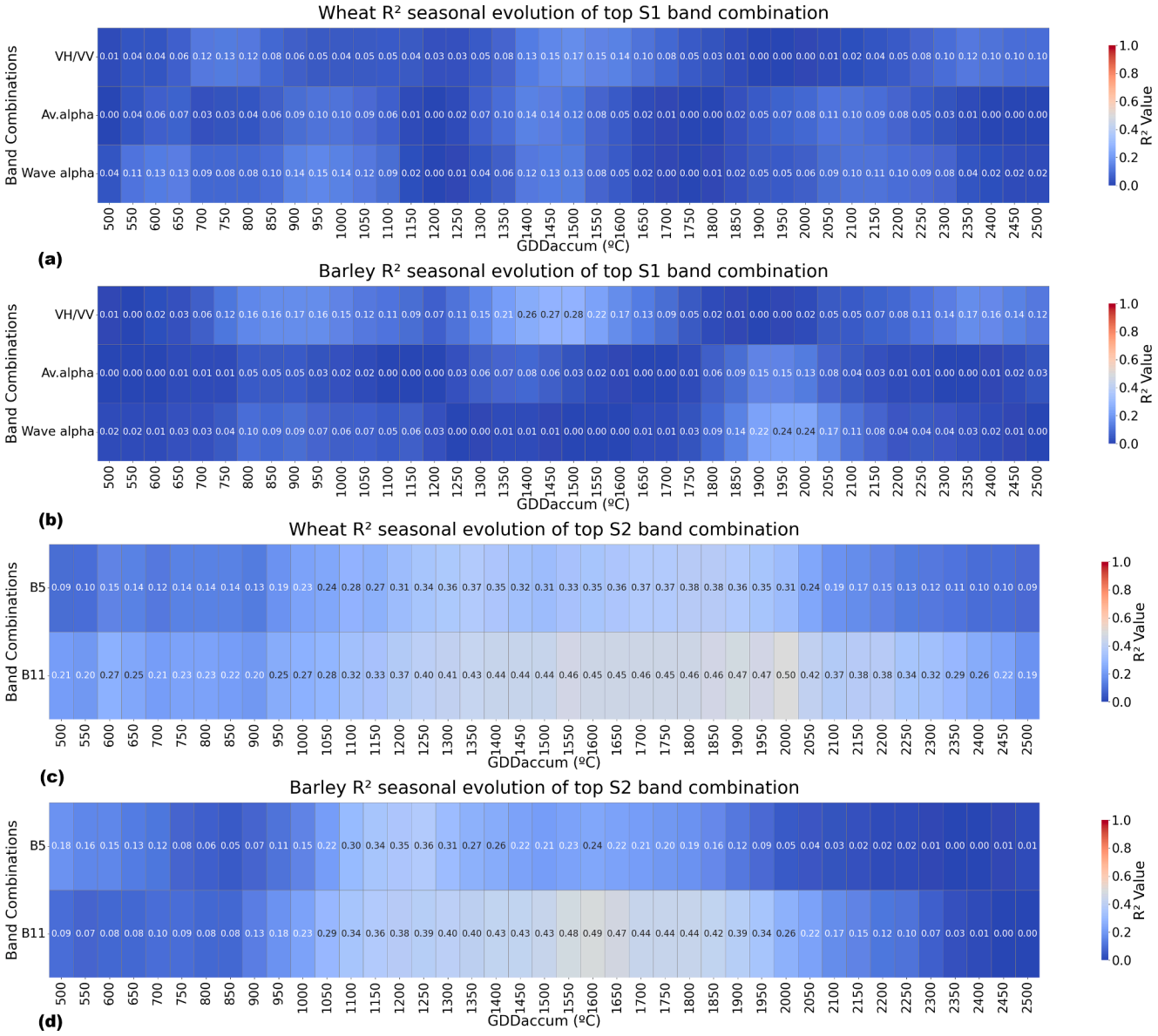



**Fig. A2.** Temporal evolution of the $R^2$ across the $GDD_{accum}$ considering one S1 (a, b) or S2 (c, d) spectral band against the yield of wheat (a, c) and barley (b, d) for all parcels.

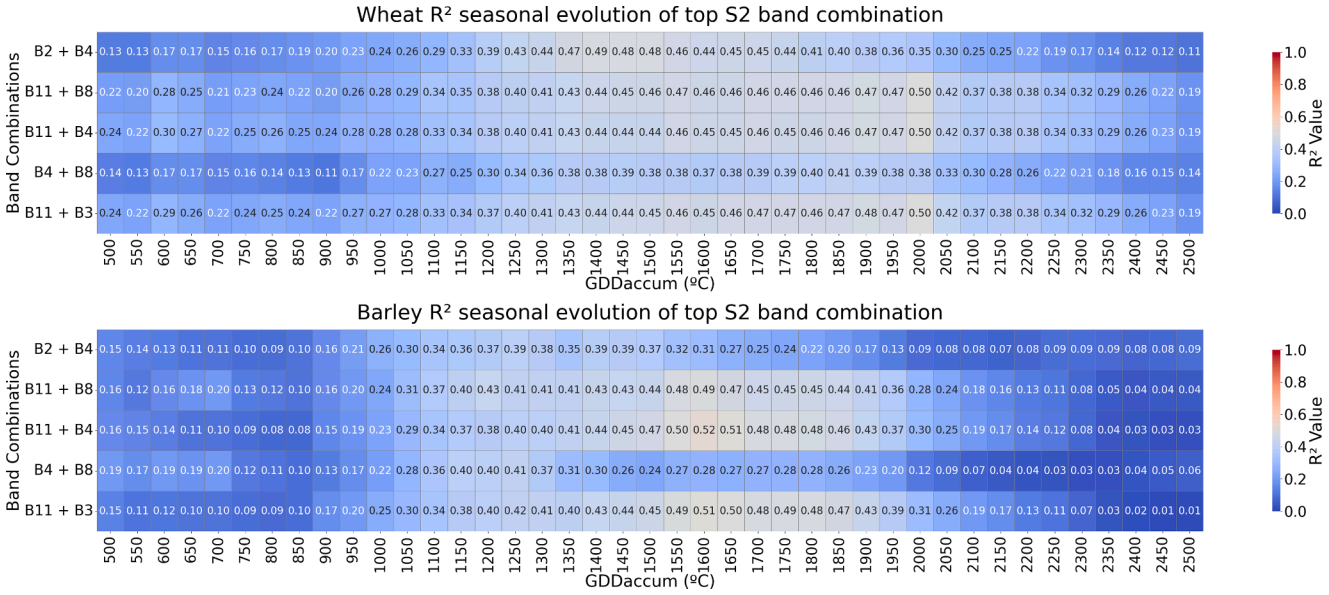



**Fig. A3.** Temporal evolution of the $R^2$ across the $GDD_{accum}$ considering the linear combination of two S2 spectral bands against the yield of wheat (top) and barley (bottom) for all parcels. Top combinations are selected based on the maximum seasonal $R^2$ achieved.

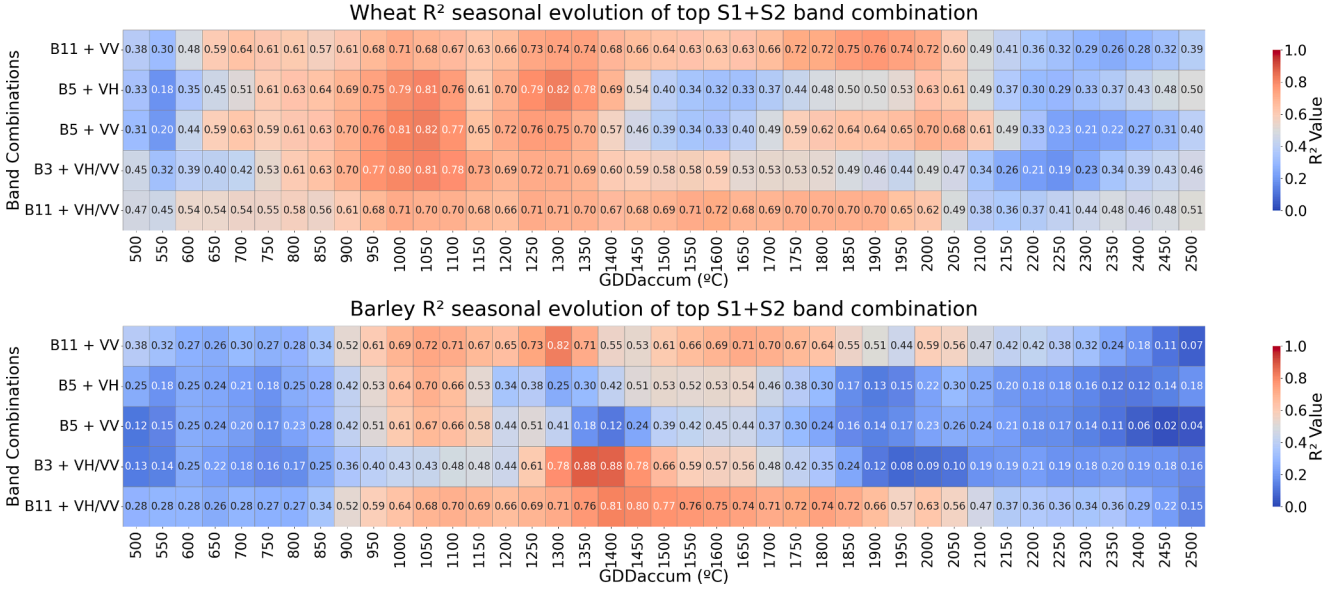



**Fig. A4.** Temporal evolution of the $R^2$ across the $GDD_{accum}$ considering one S1 parameter and one S2 spectral band against the yield of wheat (top) and barley (bottom) in the calibration area until the 20202021 season. Top combinations are selected based on the maximum seasonal $R^2$ achieved.

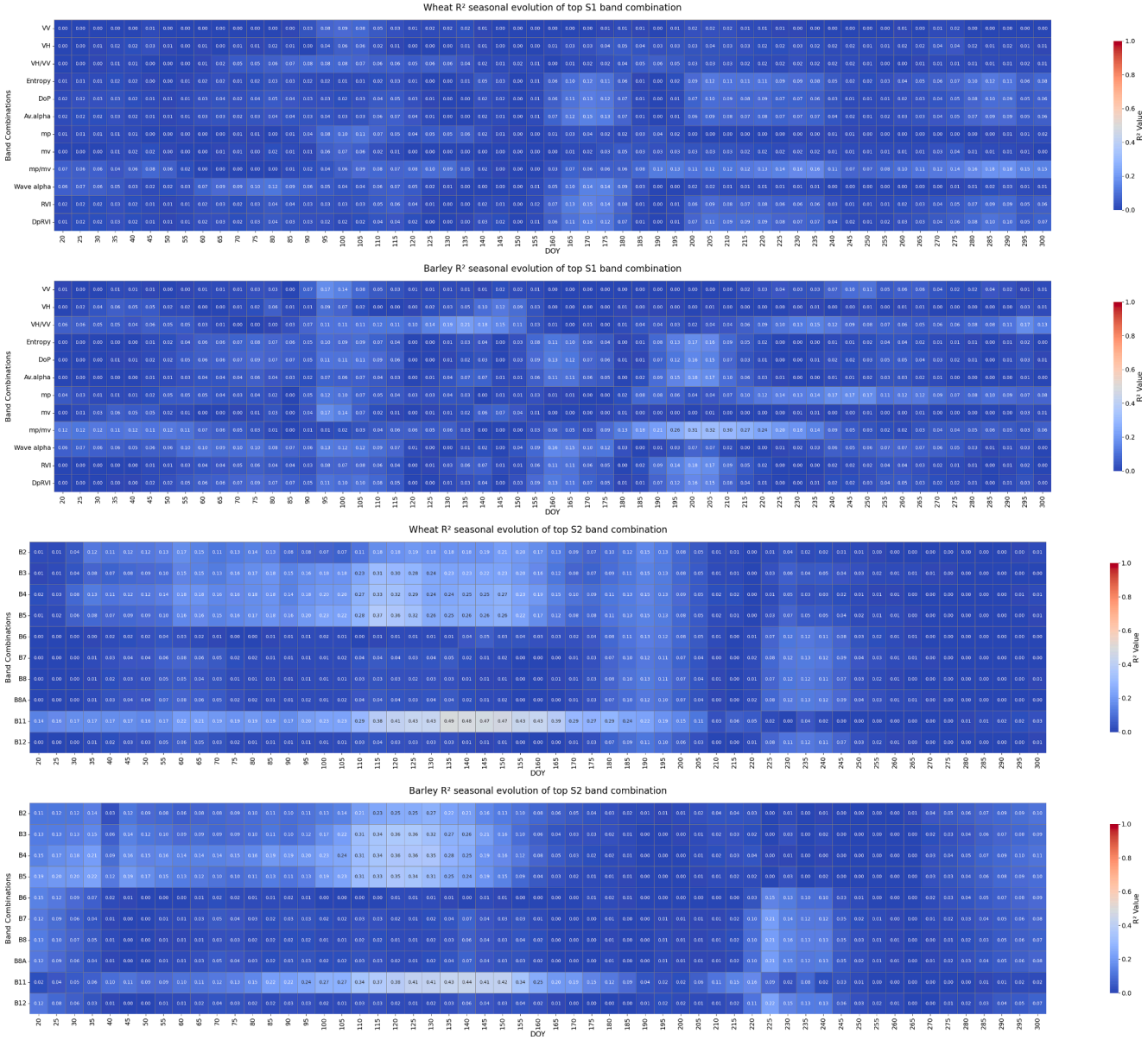



**Fig. A5.** Temporal evolution of the $R^2$ across the DOY considering one S1 (a, b) or S2 (c, d) spectral band against the yield of wheat (a, c) and barley (b, d) for all parcels.

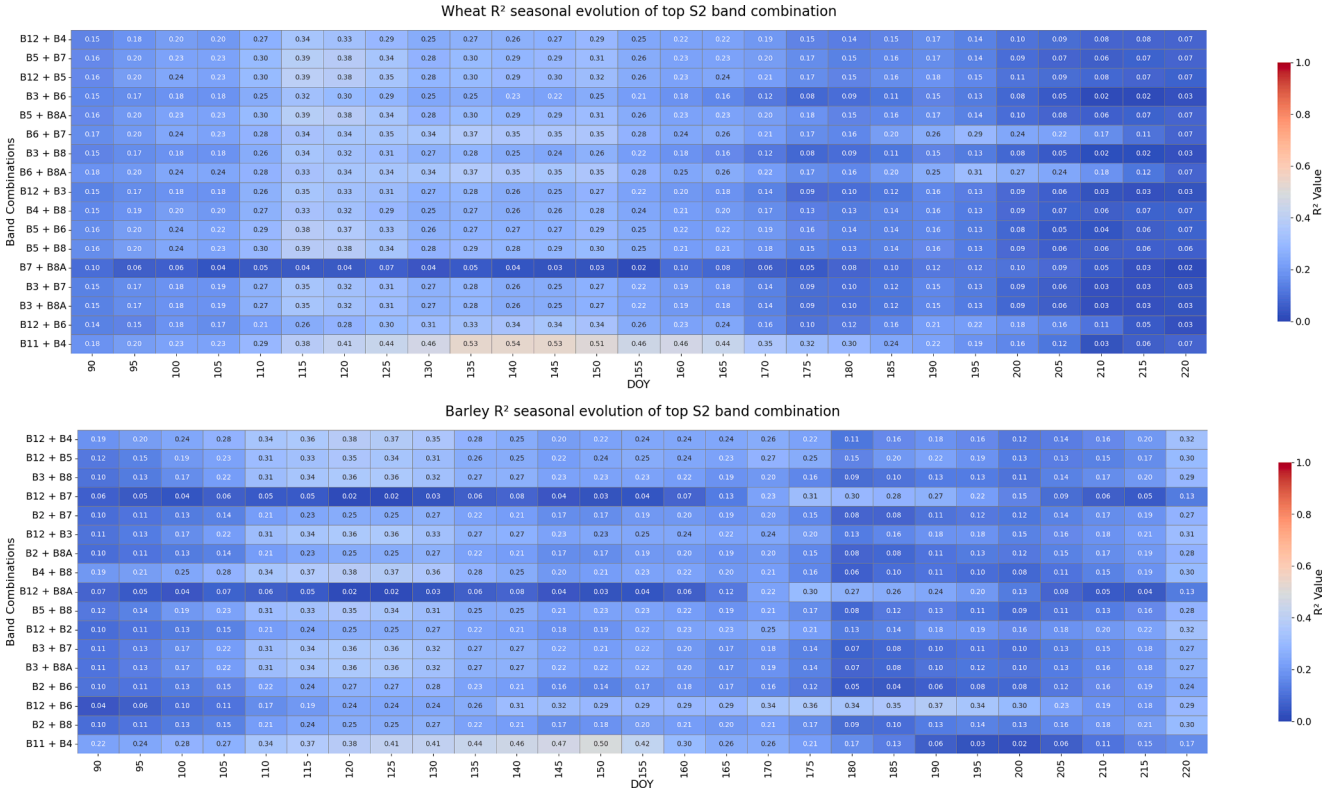



**Fig. A6.** Temporal evolution of the $R^2$ across the DOY considering the linear combination of two S2 spectral bands against the yield of wheat (top) and barley (bottom) for all parcels. Top combinations are selected based on the maximum seasonal $R^2$ achieved.

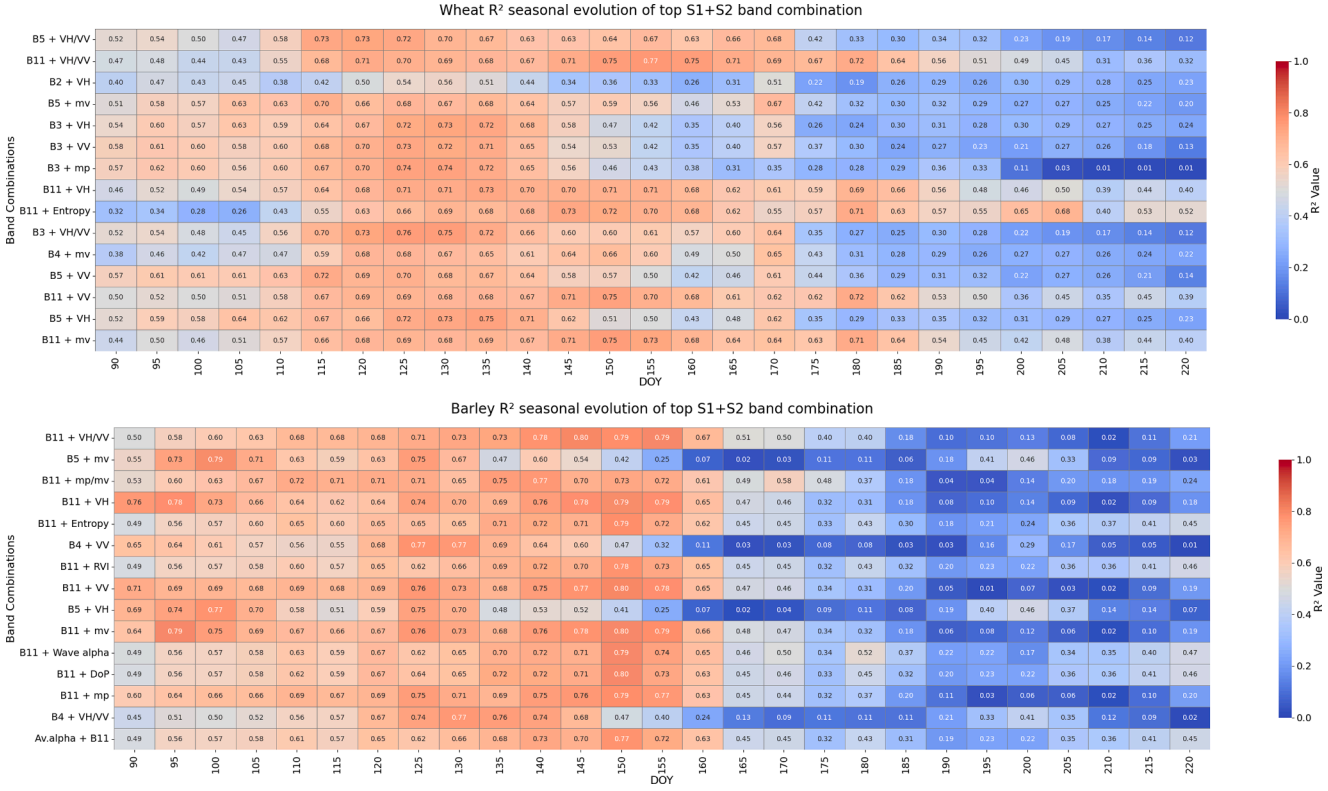



**Fig. A7.** Temporal evolution of the $R^2$ across the DOY considering one S1 parameter and one S2 spectral band against the yield of wheat (top) and barley (bottom) in the calibration area until the 20202021 season. Top combinations are selected based on the maximum seasonal $R^2$ achieved.

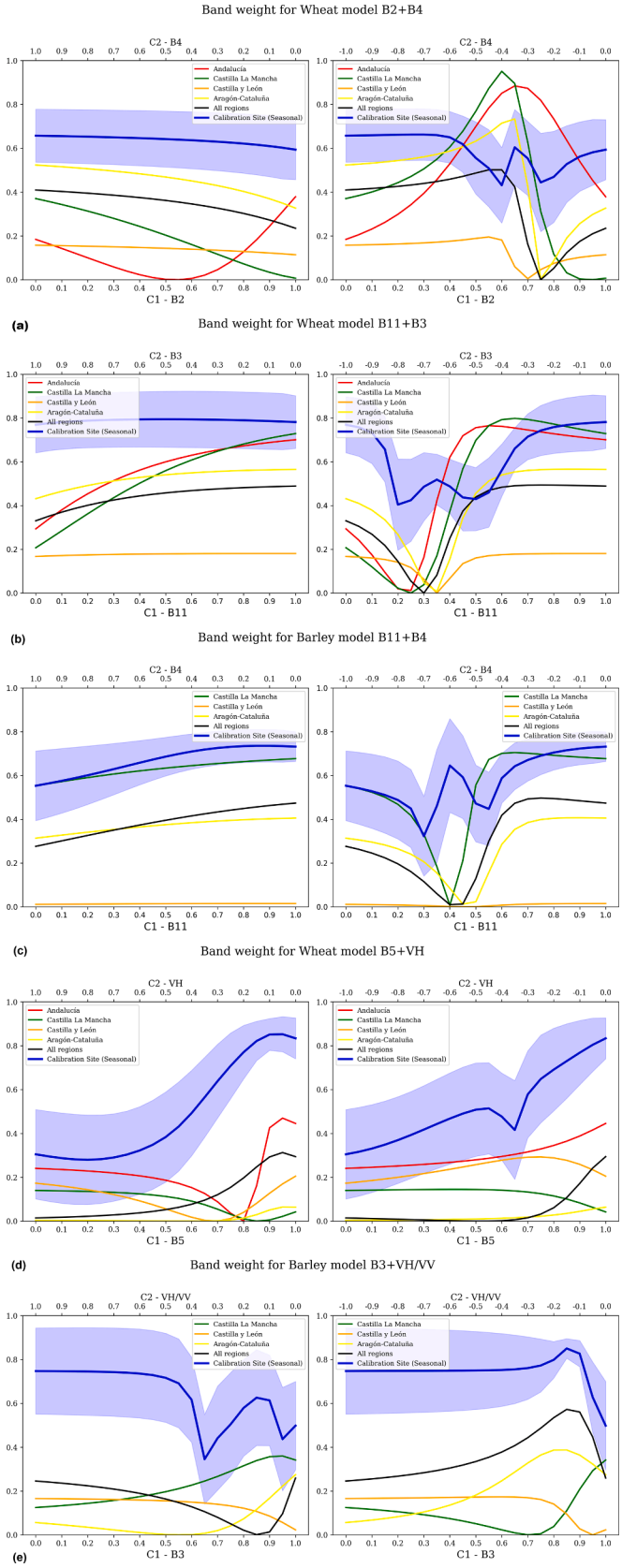
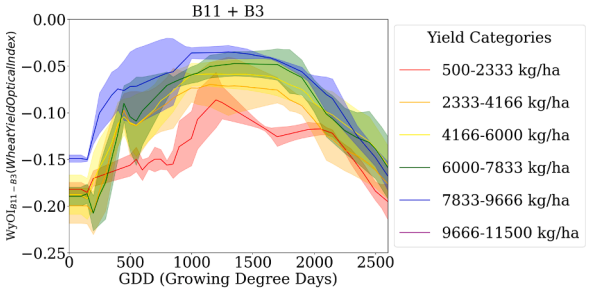
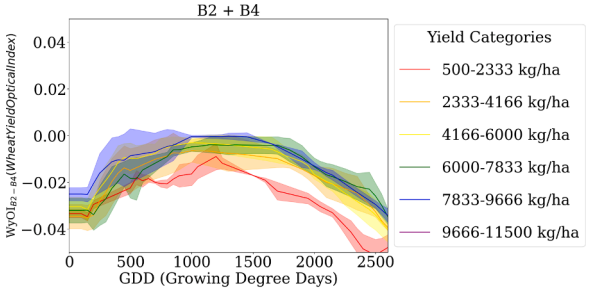


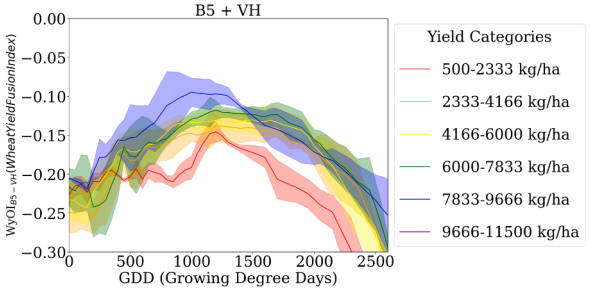

**Fig. A8.** $WyOI_{B2\text{-}B4}$ (a), $WyOI_{B11\text{-}B3}$ (b), $ByOI_{B11\text{-}B4}$ (c), $WyFI_{B5\text{-}VH}$ (d) and $ByFI_{B3\text{-}VH}$ (e) stability analysis by analyzing the $R^2$ variability when changing the relative weight of each S2 band considering both positive (left) or one negative and one positive (right) against the yield. The black line represents the average $R^2$ across all seasons in the calibration site and the shaded area represents its standard deviation.



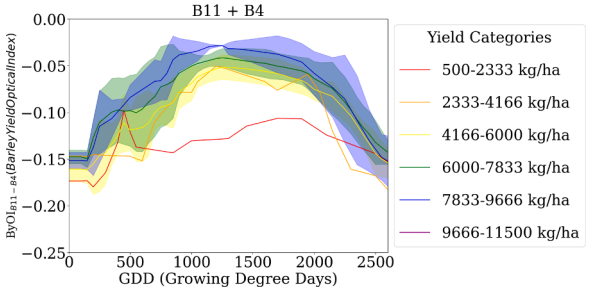

**(a)** WyOI B11-B3 (1400 - 1950) = -0.76 * B11 + 0.24 * B3



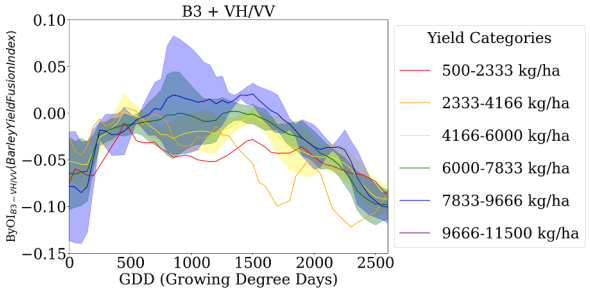

**(b)** WyOI B2-B4 (1300 - 1700) = 0.57 * B2 - 0.43 * B4




**(c)** WyFI B5-VH (900 - 1300) = -0.91 * B5 - 0.09 * VH




**(d)** ByOI B11-B4 (1400 - 1700) = -0.73 * B11 + 0.27 * B4




**(e)** ByFI B3-VH/VV (1300 - 1500) = -0.89 * B3 + 0.11 * VH/VV

**Fig. A9.** Temporal profiles of $WyOI_{B11-B3}$ (a), $WyOI_{B2-B4}$ (b), $WyFI_{B5-VH}$ (c), $ByOI_{B11-B4}$ (d) and $ByFI_{B3-ratio}$ (e) across yield categories.

## Data availability

The authors do not have permission to share data.

## References


Amin, E., Pipia, L., Belda, S., et al., 2024. In-season forecasting of within-field grain yield from Sentinel-2 time series data. Int. J. Appl. Earth Obs. Geoinf. 126.

Aslan, M.F., Sabanci, K., Aslan, B., 2024. Artificial intelligence techniques in crop yield estimation based on Sentinel-2 data: a comprehensive survey. Sustainability 16 (18), 8277.

Bailey, N., Hochman, Z., Mao, Y., Silvapulle, M. J., & Silvapulle, P. (2020). Statistical modelling and forecast evaluation of the impact of extreme temperatures on wheat crops in North Western Victoria.

Breiman, L., 2001. Random Forests. Machine Learning 45 (1), 5–32.

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci. Data 5 (1), 1–12.

Becker-Reshef, I., Franch, B., Sahajpal, R., & Sadeh, Y. (2026). Crop yield forecasting and estimation. En S. Liang (Ed.), Comprehensive remote sensing (2nd ed., pp. 159–181). Elsevier. https://doi.org/10.1016/B978-0-443-13220-9.00095-0.

Cammarano, D., Ceccarelli, S., Grando, S., et al., 2019. The impact of climate change on barley yield in the Mediterranean basin. Eur. J. Agron. 106, 1–11.

Cavalaris, C., Megoudi, S., Maxouri, M., et al. (2021). Modeling of Durum Wheat Yield Based on Sentinel-2 Imagery. *Agronomy 2021, Vol. 11, Page 1486*, *11*(8), 1486.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Chen, J., Jönsson, P., Tamura, M., et al., 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. Remote Sens. Environ. 91 (3–4), 332–344.

Cloude, S. (2009). Polarisation: Applications in Remote Sensing. *Polarisation: Applications in Remote Sensing*.

Del Pozo, A., Yáñez, A., Matus, I. A., et al. (2016). Physiological traits associated with wheat yield potential and performance under water-stress in a mediterranean environment. *Frontiers in Plant Science*, *7*(JULY2016), 210392.

Doxani, G., Vermote, E., Roger, J.C., Gascon, F., Adriaensen, S., Frantz, D., Vanhellemont, Q., 2018. Atmospheric correction inter-comparison exercise. Remote Sens. (Basel) 10 (2), 352.

Falloon, P., Betts, R., 2010. Climate impacts on European agriculture and water management in the context of adaptation and mitigation—The importance of an integrated approach. Sci. Total Environ. 408 (23), 5667–5687.

Franch, B., Vermote, E., Skakun, S., et al., 2021. The ARYA crop yield forecasting algorithm: Application to the main wheat exporting countries. Int. J. Appl. Earth Obs. Geoinf. 104, 102552.

Franch, B., Vermote, E.F., Becker-Reshef, I., et al., 2015. Improving the timeliness of winter wheat production forecast in the United States of America, Ukraine and China using MODIS data and NCAR growing Degree Day information. Remote Sens. Environ. 161, 131–148.

Gaona, J., Benito-Verdugo, P., Martínez-Fernández, J., et al. (2022). Soil Moisture Outweighs Climatic Factors in Critical Periods for Rainfed Cereal Yields: An Analysis in Spain. *Agriculture 2022, Vol. 12, Page 533*, *12*(4), 533.

Grassini, P., van Bussel, L.G., Van Wart, J., Wolf, J., Claessens, L., Yang, H., Cassman, K. G., 2015. How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. Field Crop Res 177, 49–63.

Gorelick, N., Hancher, M., Dixon, M., et al., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Hosseini, M., McNairn, H., 2017. Using multi-polarization C- and L-band synthetic aperture radar to estimate biomass and soil moisture of wheat fields. Int. J. Appl. Earth Obs. Geoinf. 58, 50–64.

Hunt, M. L., Blackburn, G. A., Carrasco, L., et al. (2019). High resolution wheat yield mapping using Sentinel-2. *Remote Sensing of Environment*, *233*.

Iranzo, C., Montorio, R., García-Martín, A., 2022. Estimation of barley yield from Sentinel-1 and Sentinel-2 imagery and climatic variables. Revista De Teledeteccion 2022 (59), 61–72.

Kalecinski, N.I., Skakun, S., Torbick, N., et al., 2024. Crop yield estimation at different growing stages using a synergy of SAR and optical remote sensing data. Sci. Remote Sens. 10.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30*.

Kelly, P. (2019). *BRIEFING EPRS. European Parliamentary Research Service. Members' Research Service PE.* https://www.europarl.europa.eu/at-your-service/en/stay-informed/research-and-analysis.

Khosravi, I., Safari, A., Homayouni, S., 2018. MSMD: maximum separability and minimum dependency feature selection for cropland classification from optical and radar data. Int. J. Remote Sens. 39 (8), 2159–2176.

Li, X., Jin, H., Eklundh, L., Houssaine Bouras, E., et al., 2024. Estimation of district-level spring barley yield in southern Sweden using multi-source satellite data and random forest approach. Int. J. Appl. Earth Obs. Geoinf. 134, 104183.

Mandal, D., Bhattacharya, A., & Rao, Y. S. (2021). *Radar remote sensing for crop biophysical parameter estimation*. 252.

Mandal, D., Kumar, V., Ratha, D., et al., 2020. Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data. Remote Sens. Environ. 247, 111954.

Mascolo, L., Cloude, S.R., Lopez-Sanchez, J.M., 2022. Model-based Decomposition of Dual-Pol SAR Data: Application to Sentinel-1. IEEE Trans. Geosci. Remote Sens. 60.

Mascolo, L., Martinez-Marin, T., Lopez-Sanchez, J.M., et al., 2025. A Novel Dynamical Framework for Crop Phenology Estimation with Remote Sensing. IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing 18, 2208–2225.

McNairn, H., Hochheim, K., Rabe, N., 2004. Applying polarimetric radar imagery for mapping the productivity of wheat crops. Can. J. Remote. Sens. 30 (3), 517–524.

Ministerio de Agricultura Pesca y Alimentación (MAPA). (n.d.). *Visor SigPac V 4.18*. Retrieved July 9, 2025, from https://sigpac.mapama.gob.es/fega/visor/.

Moletto-Lobos, I.G., Franch, B., Guillem-Valls, A., et al., 2025. Enhancing Worldcereal Crop Calendars with Land Surface Phenology and Machine Learning. SSRN.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., et al., 2021. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth Syst. Sci. Data 13 (9), 4349–4383.

Perich, G., Turkoglu, M.O., Graf, L.V., et al., 2023. Pixel-based yield mapping and prediction from Sentinel-2 using spectral indices and neural networks. Field Crop Res 292, 108824.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems*, *31*.

Sagan, V., Maimaitijiang, M., Bhadra, S., et al., 2021. Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning. ISPRS J. Photogramm. Remote Sens. 174, 265–281.

Savitzky, A., Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem. 36 (8), 1627–1639.

Segarra, J., Buchaillot, M.L., Araus, J.L., Kefauver, S.C., 2020. Remote sensing for precision agriculture: Sentinel-2 improved features and applications. Agronomy 10 (5), 641.

Sharifi, A., 2021. Yield prediction with machine learning algorithms and satellite images. J. Sci. Food Agric. 101 (3), 891–896.

Singh, D., Singh, B., 2022. Feature wise normalization: an effective way of normalizing data. Pattern Recogn. 122, 108307.

Skakun, S., Franch, B., Vermote, E., Roger, J.C., Becker-Reshef, I., Justice, C., Kussul, N., 2017. Early season large-area winter crop mapping using MODIS NDVI data, growing degree days information and a Gaussian mixture model. Remote Sens. Environ. 195, 244–258.

Smith, M., Eddy, P.R., Bugden-Storie, J., et al., 2006. Multipolarized radar for delineating within-field variability in corn and wheat. Can. J. Remote. Sens. 32 (4), 300–313.

Steele-Dunne, S.C., McNairn, H., Monsivais-Huertero, A., et al., 2017. Radar Remote Sensing of Agricultural Canopies: a Review. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (5), 2249–2273.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram.*J. Geophys. Res. 106, D7, 71837192*.

Taud, H., & Mas, J. F. (2017). Multilayer perceptron (MLP). In *Geomatic approaches for modeling land change scenarios* (pp. 451-455). Cham: Springer International Publishing.

Uribeetxebarria, A., Castellón, A., & Aizpurua, A. (2023). Optimizing Wheat Yield Prediction Integrating Data from Sentinel-1 and Sentinel-2 with CatBoost Algorithm. *Remote Sensing 2023, Vol. 15, Page 1640*, *15*(6), 1640.

Veci, L., Lu, J., Foumelis, M., et al., 2017. ESA's Multi-mission Sentinel-1 Toolbox. EGUGA 19, 19398.

Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sens. Environ. 185, 46–56.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: a meta-review. Remote Sens. Environ. 236, 111402.

Wiseman, G., McNairn, H., Homayouni, S., Shang, J., 2014. RADARSAT-2 Polarimetric SAR response to crop biomass for agricultural production monitoring. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7 (11), 4461–4471.

Xiao, G., Zhang, X., Niu, Q., et al., 2024. Winter wheat yield estimation at the field scale using sentinel-2 data and deep learning. Comput. Electron. Agric. 216, 108555.

Zhao, Y., Potgieter, A. B., Zhang, M., et al. (2020). Predicting Wheat Yield at the Field Scale by Combining High-Resolution Sentinel-2 Satellite Imagery and Crop Modelling. *Remote Sensing 2020, Vol. 12, Page 1024*, *12*(6), 1024.

Zhuo, W., Huang, J., Li, L., et al. (2019). Assimilating Soil Moisture Retrieved from Sentinel-1 and Sentinel-2 Data into WOFOST Model to Improve Winter Wheat Yield Estimation. *Remote Sensing 2019, Vol. 11, Page 1618*, *11*(13), 1618.

Zupanc, A. (2017, December 19). *Improving Cloud Detection with Machine Learning*. Planet Stories - Medium. https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13.