ISPRS Journal of Photogrammetry and Remote Sensing 239 (2026) 147–163

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

TMF: Temporal multi-modal fusion framework for estimating wheat yield from multi-source satellite and environmental data across the European Union

Zhixian Lin [a,b], Sheng Wang [a,b,c,*], Kaiyu Guan [a,b,d,e,*], Jiaying Zhang [a,b], Senthold Asseng [f], René Gislum [g], Jørgen Eivind Olesen [h,i], Klaus Butterbach-Bahl [c,j], Miroslav Trnka [i,k], Rui Zhou [a,e]

[a] Agroecosystem Sustainability Center, Institute for Sustainability, Energy, and Environment, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA
[b] Department of Natural Resources and Environmental Sciences, College of Agricultural, Consumer and Environmental Sciences, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA
[c] Pioneer Center Land-CRAFT, Department of Agroecology, Aarhus University, Aarhus 8000, Denmark
[d] National Center for Supercomputing Applications, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA
[e] Siebel School of Computing and Data Science, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA
[f] Department of Life Science Engineering, Chair of Digital Agriculture, HEF World Agricultural Systems Center, Technical University of Munich, Freising, Germany
[g] Department of Agroecology, Aarhus University, Forsøgsvej 1, 4200 Slagelse, Denmark
[h] Department of Agroecology, Aarhus University, Blichers Allé 20, Tjele 8830, Denmark
[i] Global Change Research Institute of the Czech Academy of Sciences, Brno 60300, Czech Republic
[j] Karlsruhe Institute of Technology, Institute for Meteorology and Climate Research, Atmospheric Environmental Research (IMK-IFU), Kreuzeckbahnstrasse 19, Garmisch-Partenkirchen 82467, Garmisch-Partenkirchen, Germany
[k] Department of Agrosystems and Bioclimatology, Faculty of AgriSciences, Mendel University in Brno, Brno, Czech Republic

ARTICLE INFO

Keywords:
Crop yield estimation
Multi-modal deep learning
Transformer
Data fusion
Remote sensing

ABSTRACT

Accurate large-scale crop yield estimation is critical for agricultural management and advance warning of potentially compromised food security. While recent advances in remote sensing provide diverse data sources for crop monitoring, effectively integrating these data sources for accurate yield estimation at the continental scale remains challenging due to complex crop-environment interactions and high spatial heterogeneity. In this study, we developed a Temporal Multi-modal Fusion (TMF) framework for end-of-season wheat yield estimation at the sub-national level across the European Union from 2001 to 2019. Our framework integrated time-series data across the entire growing season from climate variables, satellite-based remote sensing measurements including vegetation indices (e.g., Enhanced Vegetation Index) and productivity indicators (e.g., Solar-Induced Fluorescence, Gross Primary Productivity), and static soil properties. By employing parallel transformer encoders followed by a late-fusion strategy, our architecture preserves modality-specific temporal dynamics before explicitly anchoring them against static spatial constraints. Leave-one-year-out cross-validation demonstrated that the TMF framework achieved accurate yield estimation with an RMSE of 0.75 Mg·ha^{-1}, 7–38% lower than baseline models (LSTM, GBRT, RF, and ANN). The model demonstrated better spatial stability and captured severe yield anomalies during extreme climate events (e.g., the 2003 and 2018 droughts). The ablation study and SHAP-based interpretability analysis revealed that while fusing more input modalities consistently improved model performance, substantial information redundancy exists among multi-modal inputs. Although satellite-derived measurements effectively capture seasonal canopy dynamics and interannual yield fluctuations, static soil properties provide complementary spatially structured information for yield estimation. Their integration reduced RMSE by 28–33% compared to climate-only inputs and improved prediction stability across the study domain. Furthermore, the model's temporal contribution patterns aligned with wheat phenology, with feature importance peaking during reproductive and grain-filling stages. These findings highlight that the TMF framework provides a robust and scalable approach for continental-scale crop monitoring across highly diverse agroecosystems.

* Corresponding authors.
E-mail addresses: shengwang12@gmail.com (S. Wang), kaiyug@illinois.edu (K. Guan).

https://doi.org/10.1016/j.isprsjprs.2026.05.041
Received 25 August 2025; Received in revised form 30 April 2026; Accepted 22 May 2026
Available online 9 June 2026

1. Introduction

Agroecosystems face great challenges in maintaining productivity while ensuring sustainable food production, making accurate crop yield modeling essential for effective agricultural policy and research (Lobell et al., 2014; Rezaei et al., 2023). As one of the most critical staple crops in the European Union (EU), wheat accounts for approximately 35% of total arable land use and is critical for regional and global food security (Eurostat, 2021; Trnka et al., 2014). Accurate continental-scale wheat yield estimation provides critical support for ensuring agroecosystem sustainability and food security in response to climatic variability (Lipper et al., 2014; Moore and Lobell, 2015). Beyond socioeconomic benefits, reliable yield estimations also facilitate the scientific understanding of crop yield gaps and the response of crops to environmental stress (Dhakar et al., 2022; Guan et al., 2017). However, achieving accurate yield estimations at large spatial scales remains challenging due to complex crop-environment interactions and spatial heterogeneity.

Crop yield is the outcome of the complex interactions among genetic factors, environmental conditions, and management practices (Bailey-Serres et al., 2019). These interactions not only determine the final yield but also influence crop development and growth throughout the growing season. Environmental factors significantly influence crop development and growth through meteorological variables (temperature, precipitation, radiation) and soil properties (texture, organic matter, water-holding capacity). Wheat development follows distinct phenological stages with varying sensitivities of growth to environmental stress (Kahiluoto et al., 2019). For example, temperature extremes during critical developmental stages, particularly flowering and grain filling, can dramatically reduce yield potential (Asseng et al., 2015). This temporal complexity is further compounded by the spatial heterogeneity of environmental conditions across the EU agroecological zones (Metzger et al., 2005), creating diverse growing conditions that make large-scale yield estimation challenging but also critical. Although various efforts have been made for wheat yield modeling across Europe (van der Velde and Nisini, 2019), most studies have focused on specific regions (e.g., Danube basin) or countries (e.g., France, Germany, the Czech Republic, or Moldova) using localized datasets or narrow temporal windows (Bazzi et al., 2024; Bregaglio et al., 2023; Bueechi et al., 2023; Meitner et al., 2023; Paudel et al., 2022; Potopová et al., 2020), with limited consideration of the spatial heterogeneity and environmental variability of large-scale modeling. Additionally, increasing extreme weather events introduce non-linear crop responses that conventional models struggle to capture (Ben-Ari et al., 2018; de Nóia Júnior et al., 2023), presenting significant challenges for developing robust prediction models applicable across varied regions and seasons.

Satellite remote sensing has significantly advanced crop growth monitoring by providing consistent, spatially explicit, and timely observations. Vegetation indices (VIs) derived from multispectral satellite imagery, such as the Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI), serve as valuable indicators of canopy development and vigor throughout the growing season (Weiss et al., 2020). Time-series analysis of VIs enables the detection of critical phenological transitions and growth anomalies, offering early indicators of potential yield variations (Sakamoto et al., 2013; Zheng et al., 2016). Besides traditional VIs, advanced productivity and biophysical metrics like Solar-Induced Chlorophyll Fluorescence (SIF) and Gross Primary Productivity (GPP) offer deeper insights into crop physiological processes, providing a direct link to final yield by capturing photosynthetic activity and carbon assimilation processes (Guan et al., 2016; Peng et al., 2020). Previous studies have demonstrated the potential of these advanced metrics to improve yield estimation accuracy, particularly under stress conditions (Guan et al., 2017; Wang et al., 2023). However, remote sensing alone cannot capture the full range of environmental factors influencing crop growth, such as soil water retention and atmospheric dryness. Recent studies have shown that yield estimation based on diverse data generally outperforms models based on a single data source (Bueechi et al., 2023; Cai et al., 2019; Gámez et al., 2025; Li et al., 2022; Zhuang et al., 2024). Developing effective methods to integrate remote sensing-derived information with other environmental factors remains a key challenge in building more accurate and robust large-scale yield estimation models.

Crop yield modeling approaches have evolved from mechanistic process-based models to increasingly sophisticated data-driven frameworks (Jones et al., 2017; Liu et al., 2024; Rosenzweig et al., 2013; Zhou et al., 2023a). While process-based models provide physiological insights, their large-scale application is often hindered by heavy parameterization and calibration requirements (Martre et al., 2015; Wang et al., 2017). Data-driven machine learning approaches such as Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting can capture non-linear relationships without detailed process understanding, but often treat factors as independent features, overlooking temporal dependencies in crop growth (Jiang et al., 2020). Some recent studies have further incorporated crop-model outputs or domain knowledge into data-driven frameworks, including knowledge-guided machine learning (KGML) (Fang et al., 2026; Liu et al., 2024) and data assimilation approaches (Huang et al., 2019; Yang et al., 2023), to enhance robustness and generalization. Despite these developments, accurately capturing complex spatio-temporal interactions from multi-source satellite and environmental data at continental scales remains challenging, particularly under diverse agroecological conditions and extreme climate variability. With the increasing availability of diverse environmental and remote sensing data, there is a growing need for more sophisticated modeling approaches that can effectively capture complex spatio-temporal patterns and fully exploit the information embedded in multi-source datasets.

Deep learning methods have demonstrated strong performance in capturing complex temporal patterns from multi-source data in crop growth dynamics, offering new opportunities for yield estimation (LeCun et al., 2015; Reichstein et al., 2019). Given that crop growth is a complex dynamic process characterized by significant spatial heterogeneity from planting to maturity, deep neural networks utilize hierarchical structures to extract intricate spatiotemporal features from high-dimensional data. Recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM), have demonstrated success in modeling time-series data by maintaining internal memory states that capture temporal dependencies (Jiang et al., 2020; Lin et al., 2020; Schwalbert et al., 2020; Zhong et al., 2025). More recently, attention mechanisms and transformer architectures have shown promise in crop yield modeling by identifying critical growth periods and environmental conditions that disproportionately influence final yields (Guo et al., 2024; Tian et al., 2024; Xiong et al., 2024). Another significant advantage of deep learning is its ability to enable the effective fusion of heterogeneous data sources through specialized architectures and training strategies (Mena et al., 2025). Attention-based multi-modal fusion methods offer a more adaptive and effective approach for capturing and integrating the most relevant information for improved results in crop modeling (Maimaitijiang et al., 2020).

Despite these advancements, comprehensive multi-modal fusion approaches for pan-European wheat yield estimation remain underexplored. Existing transformer-based studies have largely been focused on regional or national scales (Guo et al., 2024; Zhang et al., 2025a), limiting their evaluation across diverse agroecological zones. Furthermore, many studies rely primarily on a narrow set of temporal inputs, typically vegetation indices and climate variables (Xiong et al., 2026; Zhang et al., 2025b), with relatively limited incorporation of crop physiological indicators or static soil constraints. Systematic assessments of multi-modal fusion strategies for handling the inherent heterogeneity of agricultural data remain scarce. Moreover, deep learning approaches often face criticism for their “black box” nature, offering limited transparency into their decision-making processes (Hu et al., 2023). To address this, Explainable AI (XAI) techniques, e.g., SHapley Additive exPlanations (SHAP), have been applied to decipher model

behavior in crop yield estimation (Barredo Arrieta et al., 2020; von Bloh et al., 2024). However, quantifying the specific contributions of different modalities across growth stages, particularly under extreme climate conditions, remains critical for validating model reliability and enabling broader agronomic applications.

In this study, we developed a Temporal Multi-modal Fusion (TMF) framework for accurately estimating soft wheat yields at sub-national scales across the EU's diverse agroecological zones from 2001 to 2019. The framework integrates time-series data from climate variables, satellite-based remote sensing measurements including vegetation indices and productivity indicators, and static soil properties to capture complex spatio-temporal patterns underlying wheat yield formation. Our approach employs transformer encoders to extract temporal features from time-series variables, which are then fused with soil features using a late fusion strategy. Specifically, we aim to answer three research questions: (1) To what extent can the TMF framework capture the complex spatio-temporal interactions between multi-source data for wheat yield estimation across the EU? (2) How do fusion strategies and modality combinations affect the synergistic integration of heterogeneous data for wheat yield estimation? (3) What are the relative contributions of different temporal features and data modalities in wheat yield estimation? We hypothesized that 1) multi-modal data integration is essential to resolve the high spatio-temporal heterogeneity inherent in continental-scale yield modeling; 2) static soil properties provide structural spatial yield patterns that complement temporal crop dynamics captured by satellite observations; and 3) the model's reliance on different modalities will dynamically align with key crop phenological stages and environmental stress responses.

2. Materials and methods

2.1. Study area and crop yield data

This study focused on soft wheat (predominantly winter wheat) yield estimation across the European Union at the sub-national level using the administrative classification of NUTS-2016 (Nomenclature of Territorial Units for Statistics in 2016) from 2001 to 2019 (Fig. 1). The yield data were obtained from the Harmonized European Union subnational crop statistics database (Ronchetti et al., 2024). Although spring and winter wheat varieties are included, winter wheat dominates EU production, accounting for more than 93% of total wheat cultivation (Eurostat, 2024; Le Gouis et al., 2020). After removing outliers using the interquartile range (IQR) method, the dataset comprised 14,749 NUTS-year records of 852 regions, predominantly at NUTS level 3 due to varying country sizes.

Substantial spatial heterogeneity exists in wheat yields across the EU, with the highest productivity in Western and Northern Europe (yields > 8 Mg·ha^{-1}), and lower yields in Eastern and Southern regions. While yields remained relatively stable during the study period, yield losses occurred in 2003 and 2018 due to severe drought and heatwave events (Clarke et al., 2021; Kahiluoto et al., 2019). The growth season was defined as 50 weeks from September 1st of the sowing year to August 15th of the harvest year, which ensures coverage of all critical growth stages across all European latitudes.

To understand the long-term dynamics of wheat production across the EU, we analyzed the temporal trends of crop yields across all NUTS regions from 2001 to 2019 (**Fig. S1**). The statistical analysis ($p < 0.05$) revealed that wheat yields in most EU regions have stagnated, with no significant increasing or decreasing trends observed over the 19-year period. This observation aligns with previous studies reporting a plateauing (stagnation) of wheat yields in most European regions due to constraints from climatic change and agronomic practices (Brisson et al., 2010; Le Gouis et al., 2020; Moore and Lobell, 2015; Ray et al., 2012). Spatially, significant positive trends were primarily concentrated in Eastern Europe (e.g., the Baltic countries), reflecting potential improvements in agricultural management and the closing of yield gaps in these regions (Schils et al., 2018). This prevailing yield stagnation highlights the challenge of yield estimation in the EU, as models must effectively capture high-frequency inter-annual variability driven by environmental fluctuations rather than relying on long-term linear trends.

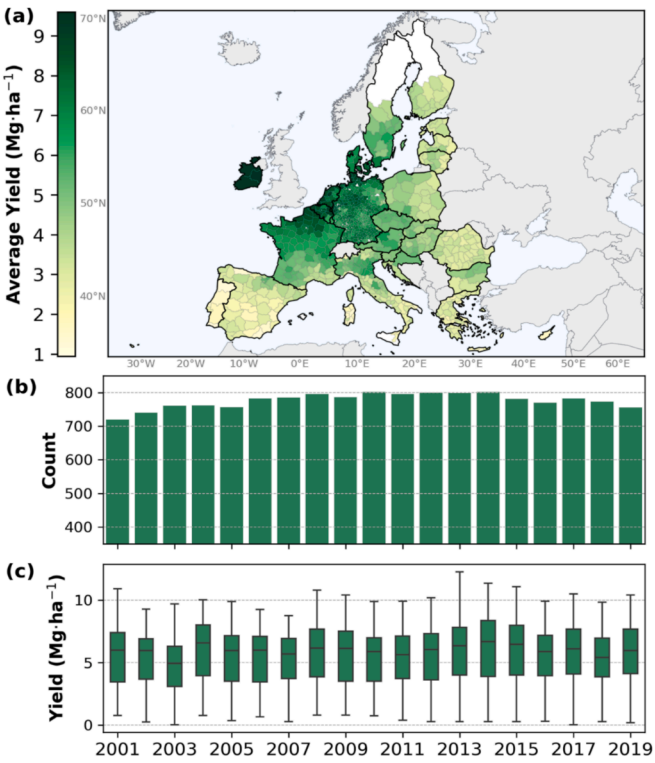

Fig. 1. Distribution of wheat yields in the EU from 2001 to 2019. (a) Average yield distribution of the study period across the EU, (b) annual count of yield observations, and (c) boxplot of crop yield. In (c), the line shows the median, box edges indicate the 25th and 75th percentiles, and whiskers represent 1.5 × interquartile range.
Source: Harmonized European Union subnational crop statistics database (Ronchetti et al., 2024)

2.2. Multi-source dataset and preprocessing

This study integrated multi-source satellite Earth observation and environmental data to capture the complex dynamics of wheat growth across the EU. The dataset combines climate data, satellite-based remote sensing measurements, including vegetation indices and productivity indicators, and static soil properties (Table 1). All data were consolidated at the NUTS regional level with cropland masks to ensure consistent coverage of the study area and wheat growth season (Fig. 2**a**).

2.2.1. Climate data

Time-series climate data were obtained from ERA5-Land daily reanalysis at 0.1° spatial resolution (Muñoz-Sabater et al., 2021). The variables used in this study included temperature (mean, maximum, and minimum air temperature), moisture (precipitation, evaporation, air relative humidity, vapor pressure deficit), and radiation components (incoming longwave and shortwave radiation). Relative humidity (RH) and vapor pressure deficit (VPD) were derived using the August-Roche-Magnus approximation from ERA5-Land temperature data.

2.2.2. Satellite-based measurements of vegetation

To comprehensively capture wheat growth dynamics and productivity, we used satellite-based measurements, including vegetation indices and productivity indicators. Vegetation indices reflect canopy

Table 1
Summary of the collected datasets for wheat yield estimation in the EU.

Category	Variables	Spatial Resolution	Temporal Resolution	Source
Crop yield	Sub-national yield statistics	NUTS-2/3	Year	Harmonized European Union subnational crop statistics (Ronchetti et al., 2024)
Climate data	Temperature (mean, maximum, and minimum), precipitation, evapotranspiration, RH, VPD, incoming radiation (shortwave, longwave)	0.1°	1-day	ERA5-Land Daily Aggregated (Muñoz-Sabater et al., 2021)
Satellite-based measurements (Vegetation indices)	NDVI EVI WDRVI	250 *m*	16-day	MODIS (MOD13Q1.061)
Satellite-based measurements (Productivity indicators)	SIF	0.05°	4-day	CSIF dataset (Zhang et al., 2018)
	fAPAR	1 km	10-day	CLMS (https://land.copernicus.eu/en/products/vegetation/fraction-of-absorbed-photosynthetically-active-radiation-v2-0–1 km)
	GPP	0.05°	8-day	GOSIF GPP (Li and Xiao, 2019)
Soil properties	Clay content, sand content, organic carbon, bulk density, field capacity, pH (0, 30 cm)	250 *m*	/	OpenLandMap Soil (Hengl, 2018)

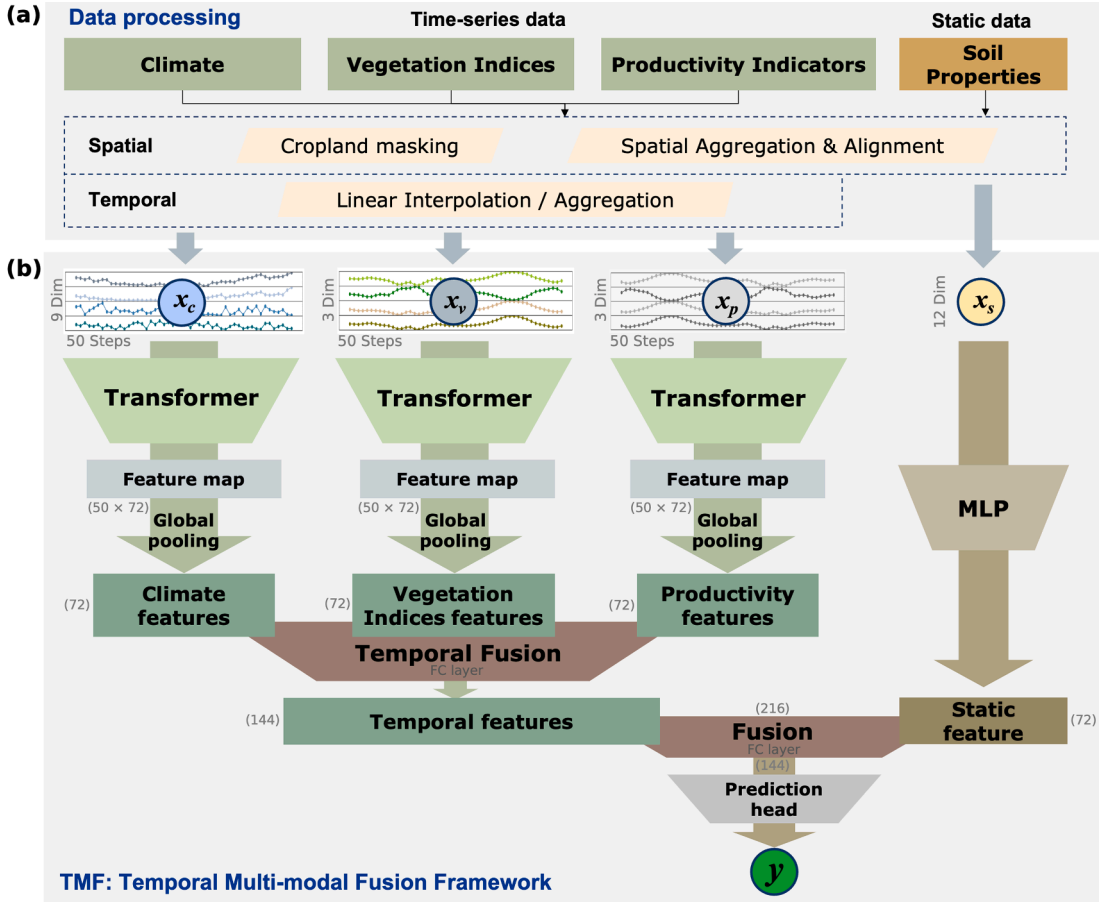

Fig. 2. (a) Multi-source data processing workflow, and (b) Temporal Multi-modal Fusion (TMF) framework for wheat yield estimation. The TMF employs parallel Transformer encoders for each temporal modality (climate, vegetation indices, and productivity indicators), followed by a two-stage hierarchical late fusion.

structure and greenness, while productivity indicators provide direct insights into photosynthetic activity and carbon uptake.

Vegetation indices derived from MODIS (MOD13Q1.061) 16-day composite products at 250 m resolution were used to monitor wheat growth dynamics. Specifically, the Normalized Difference Vegetation Index (NDVI) reflects overall greenness, while the Enhanced Vegetation Index (EVI) improves sensitivity in high biomass regions. Additionally, the Wide Dynamic Range Vegetation Index (WDRVI) was calculated from surface reflectance bands to better characterize vegetation during peak growing seasons (Gitelson, 2004).

Productivity indicators, including SIF, the fraction of absorbed photosynthetically active radiation (fAPAR), and GPP, were used for crop yield estimations. These productivity indicators can characterize crop photosynthetic activity and carbon fixation. Specifically, SIF is a direct proxy of photosynthetic activity, capturing chlorophyll's re-emission of absorbed solar radiation and showing strong correlations with productivity across various vegetation types (Guan et al., 2016). Here, we used the CSIF (Continuous Solar-Induced Fluorescence) dataset at 0.05° resolution and 4-day intervals (Zhang et al., 2018), which was trained using OCO-2 SIF observations and MODIS reflectance data. The

fAPAR quantifies the proportion of incoming solar radiation absorbed by vegetation for photosynthesis, serving as a key indicator of potential photosynthetic capacity (Baret et al., 2007). The fAPAR dataset was obtained from the Copernicus Land Monitoring Service (CLMS), which provides global coverage at 1 km resolution and 10-day intervals. GPP represents the total carbon dioxide uptake through photosynthesis, providing a direct measurement of vegetation productivity and showing strong links to crop yield formation (Jiang et al., 2021). GPP estimates in this study were derived from the GOSIF (Global OCO-2-based SIF) dataset (Li and Xiao, 2019), which provides data at 0.05° resolution and 8-day intervals by combining SIF observations with a light-use efficiency model.

2.2.3. Soil properties

Static soil properties were derived from OpenLandMap at 250 m to capture the spatial heterogeneity of soil conditions affecting wheat growth (Hengl, 2018). OpenLandMap provides information about soil properties at seven standardized depth points based on machine learning models trained on harmonized soil profile observations and environmental covariates. For this study, six key soil characteristics were selected: clay content, sand content, organic carbon content, bulk density, field capacity, and pH. These properties were extracted at the surface (0 cm) and subsurface (30 cm) depths to represent soil conditions that influence wheat growth and yield formation.

2.2.4. Data processing

The multi-source dataset underwent spatial and temporal processing to ensure consistency for large-scale yield estimation (Fig. 2**a**). The processing workflow consisted of two main steps: spatial and temporal harmonization. For spatial processing, we first applied a cropland mask using the Land Cover CCI Climate Research Data Package (CRDP) annual land cover product at 300 m resolution to retain only cropland pixels (Defourny et al., 2023). The masked datasets were then aggregated to the NUTS regional level through pixel averaging. Temporal harmonization was applied by interpolating time-series data with coarser-than-weekly resolution to weekly intervals using linear interpolation, whereas data with higher temporal frequency were aggregated to weekly means.

The data processing was performed using Google Earth Engine (GEE), which efficiently handled the large-scale processing of MODIS products, ERA5-Land climate data, and soil properties (Gorelick et al., 2017). Processing of the productivity indicators (SIF, fAPAR, and GPP) was conducted using Python geospatial libraries, specifically rasterio for raster operations, geopandas for vector processing, and xarray for multi-dimensional array manipulation.

2.3. Temporal multi-modal fusion framework

The proposed temporal multi-modal fusion (TMF) framework integrates both temporal and static data for sub-national crop yield estimation through a hierarchical architecture (Fig. 2**b**). The TMF consists of two main components: (1) parallel temporal feature extraction for time-series variables and static soil properties encoding, and (2) hierarchical multi-modal fusion for yield estimation. This architectural design reflects the heterogeneous controls on wheat yield formation, where yield outcomes are shaped by both cumulative seasonal dynamics and time-invariant environmental constraints. Temporal variables (e.g., climate and satellite-based measurements) capture stage-dependent growth responses and interannual variability, whereas static soil properties represent time-invariant environmental information related to regional yield variability. By employing modality-specific extractors followed by late fusion, the framework preserves the unique temporal signatures of diverse data sources while effectively capturing the non-linear interactions between dynamic growth processes and static environmental drivers.

Feature extraction: The framework processes two types of input features: temporal and static data. For dynamic temporal modalities, including climate variables $X_c \in R^{(T\times d_c)}$, vegetation indices $X_v \in R^{(T\times d_v)}$, and productivity indicators $X_p \in R^{(T\times d_p)}$ (where $T = 50$ represents sequence length and d represents feature dimensions), we employed parallel transformer encoders for each modality. These encoders first project the input to a common feature space using a linear layer, and then add sinusoidal positional encodings to retain temporal order. Each transformer encoder comprises three layers with four attention heads per layer, incorporating multi-head self-attention and feedforward sublayers with residual connections and layer normalization. This architecture is specifically leveraged to capture the non-linear temporal dependencies within each modality. By utilizing multi-head self-attention, the model can simultaneously focus on multiple critical phenological windows, such as the sensitive flowering stage and the grain-filling period (Rußwurm and Körner, 2020; Shaw et al., 2018). The static soil properties $X_s \in R^{d_s}$ are processed by a multilayer perceptron (MLP) branch with layer normalization to extract static patterns, which learns non-linear relationships among soil properties that are relevant to regional yield variability.

Multi-modal fusion: The framework adopts a hierarchical late fusion strategy in which modality-specific temporal representations are first fused across temporal modalities, and then integrated with static soil features in a second fusion stage. This design reflects the heterogeneous roles of different data sources in yield formation, enabling each modality to capture its own temporal dynamics before integrating their complementary information at a higher representation level. The fusion process follows a hierarchical structure: Temporal features from climate variables, vegetation indices, and productivity indicators are first extracted through global pooling operations after transformer encoding. These modality-specific temporal representations are then concatenated and processed through a fully connected temporal fusion layer to generate an integrated temporal representation. This step integrates the seasonal information separately encoded from climate variables, vegetation indices, and productivity indicators. The integrated temporal features are then combined with the encoded soil features through a dedicated fusion module based on a fully connected layer. The final yield estimation is computed through a two-layer estimation head with dropout regularization.

The TMF uses the Adam optimizer to optimize the weights of the network. The mean squared error (MSE) loss function is employed for model optimization, directly measuring the discrepancy between estimated and observed yield values. We used the same set of hyperparameters across all folds in LOYO-CV. The model dimension (d_model) was particularly investigated through grid search over the range [32, 48, 72, 96, 128]. This hyperparameter search was conducted for both late-fusion and early-fusion TMF variants using an identical search space and evaluation protocol. In both cases, a model dimension of 72 was determined as the optimal value based on cross-validation performance. Other architectural parameters were set based on preliminary experiments, with the number of attention heads fixed at 4 and the number of transformer layers at 3. The implementation utilizes the PyTorch framework and is trained on a high-performance computing infrastructure comprising an AMD Milan Processor (2.45 GHz/64 Core), 256 GB RAM, and four NVIDIA A100 GPUs (40 GB VRAM).

2.4. Baseline models

To evaluate the effectiveness of our proposed TMF framework, we developed four baseline models for comparison, including LSTM (Hochreiter and Schmidhuber, 1997), Artificial Neural Network (ANN) (Agatonovic-Kustrin and Beresford, 2000), Gradient Boosting Regression Trees (GBRT) (Friedman, 2001), and RF (Breiman, 2001). All baseline models were trained and evaluated using identical data splits, preprocessing strategies, and evaluation metrics as the TMF framework to ensure fair and consistent comparison.

The bidirectional LSTM network serves as our primary temporal deep learning baseline. The model processes concatenated temporal features through two stacked LSTM layers, while static soil properties are processed through an MLP identical in structure to that used in the TMF framework. The temporal and static features are concatenated and pass through a fully connected fusion layer followed by an estimation head. The LSTM hidden size was determined through a grid search over the range [64, 128, 256, 512].

For non-sequential machine learning baselines, the ANN architecture was optimized through grid search over different hidden layer configurations, ranging from [32, 16] to [1024, 512], employing ReLU activation and dropout for regularization. The GBRT and RF models underwent randomized search over key hyperparameters, including the number of trees, maximum depth, and minimum sample requirements. These non-sequential models process the flattened temporal features alongside static soil properties, with 762 input variables (15 temporal variables $\times$ 50 timesteps + 12 soil properties).

For all baseline models (ANN, LSTM, RF, and GBRT), hyperparameters were optimized prior to model evaluation and then fixed across all leave-one-year-out cross-validation (LOYO-CV) folds to ensure fair and consistent comparison. ANN and LSTM architectures were selected via grid search over predefined configurations, while RF and GBRT hyperparameters were optimized using RandomizedSearchCV with 500 iterations. The final hyperparameter search spaces and selected configurations for all baseline models, as well as the proposed TMF framework, are summarized in **Table S1**.

2.5. Experimental design

To systematically evaluate the proposed TMF framework, we designed a structured experimental design comprising three components: (a) model performance assessment, (b) multi-modal fusion strategies evaluation, and (c) SHAP-based interpretability analysis (Fig. 3).

(a) Model performance assessment. The model performance was evaluated using leave-one-year-out cross-validation (LOYO-CV). In each fold, data from one of the 19 years (2001 to 2019) was held out as the test set, while the remaining years were used as the training set, enabling evaluation of the model's performance in estimating crop yield across different temporal conditions. This design prevents within-year information leakage and allows rigorous assessment of temporal generalization to unseen climatic conditions. Model performance was assessed using multiple metrics, including Root Mean Square Error (RMSE), coefficient of determination (R^2), and relative RMSE normalized by overall mean yield (rRMSE). We report both the pooled performance across all LOYO test folds and individual fold accuracies to provide detailed insights into model stability across years.

(b) Multi-modal fusion strategies evaluation. We conducted experiments from the perspectives of input modality combinations and fusion architecture to systematically evaluate the effectiveness of different modal integration approaches. The modality combination analysis examined 14 model variants through ablation studies, which evaluate modality-level structural contributions to predictive performance. The experiments were organized into two main groups: models with and without soil properties. For models incorporating soil properties, we evaluated combinations of three temporal modalities (climate, vegetation indices, and productivity indicators) with soil data (3 T + 1S, 2 T + 1S, 1 T + 1S). Similarly, we tested temporal modality combinations without soil properties (3 T, 2 T, 1 T) to quantify the impact of static soil information on yield estimation accuracy. The architectural analysis compared late and early fusion approaches for modal integration. While our proposed late fusion strategy preserves modality-specific temporal

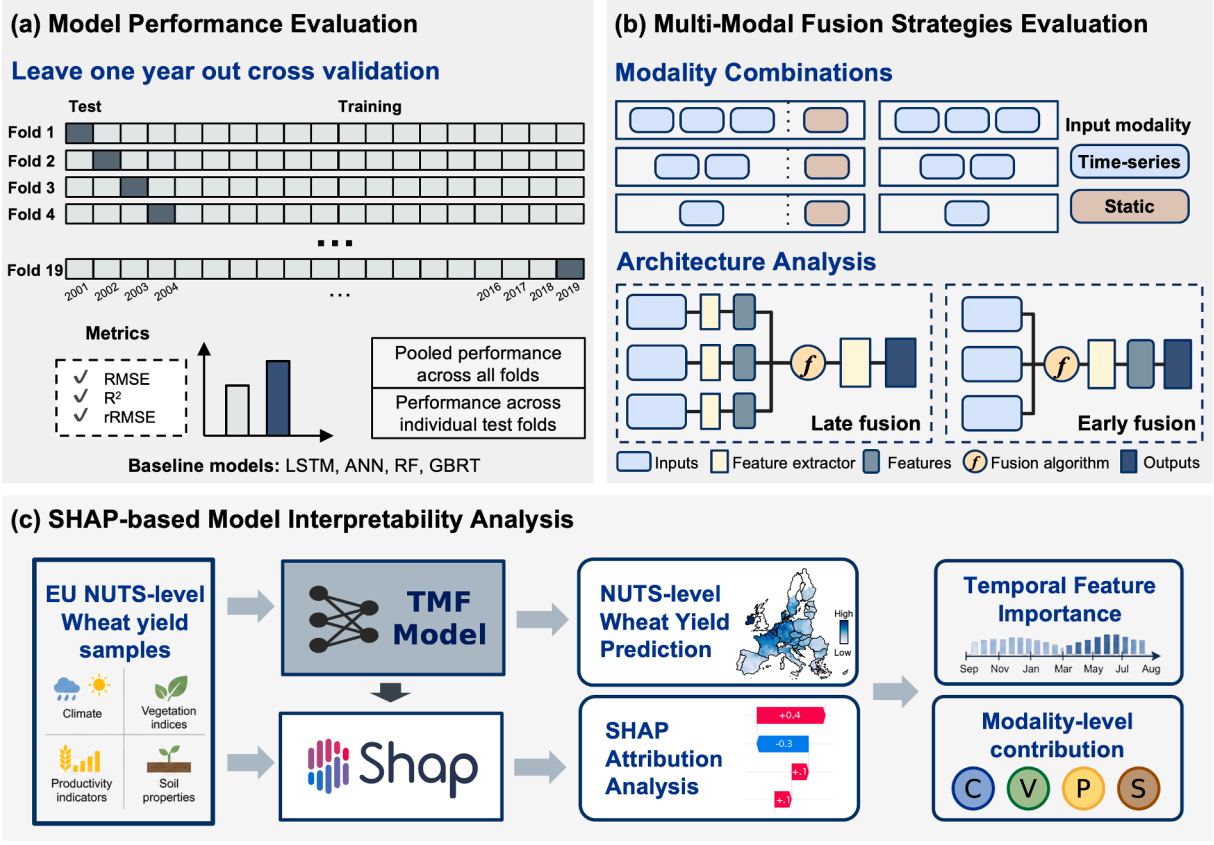

Fig. 3. Experimental design to evaluate and understand the proposed TMF framework for large-scale wheat yield estimation: (a) model performance evaluation using leave-one-year-out cross-validation (LOYO-CV), (b) multi-modal fusion strategies evaluation, and (c) SHAP-based model interpretability analysis.

patterns through independent processing streams, the early fusion variant concatenates all input features at the beginning of the network.

(c) SHAP-based interpretability analysis. Model interpretability was investigated through SHAP analysis, which provides individualized feature contribution assessments based on coalitional game theory (Lundberg et al., 2020). Specifically, SHAP quantifies the marginal contribution of each input variable within the fully integrated TMF framework, enabling feature-level interpretation of how static and temporal variables influence yield estimations. We calculated SHAP values for all features across temporal and static modalities. The temporal dimension analysis quantified the feature importance variations during key growth stages and extreme events, providing insights into the model's adaptation to different environmental conditions. The modality-level contribution analysis compared the relative importance of climate variables, vegetation indices, productivity indicators, and soil properties to yield estimation.

3. Results

3.1. Relationship among input variables

The Pearson correlation analysis revealed distinct patterns between wheat yield and growing season accumulated temporal input variables across the EU (Fig. 4). For the correlation analysis, weekly time-series variables were summed over the defined growing season to obtain accumulated metrics that characterize overall seasonal effects. Yield showed positive correlations with productivity indicators and VIs. Among them, SIF ($r = 0.55$) and fAPAR ($r = 0.54$) showed relatively strong linear relationships with yield. Conversely, negative correlations were observed with climate variables, specifically temperature-related variables, VPD, and incoming shortwave radiation. We also found strong intercorrelations within certain variable groups, for example, among the three VIs (NDVI, EVI, and WDRVI) ($r > 0.9$) and among the three productivity indicators (SIF, fAPAR, and GPP) ($r > 0.8$). Additionally, cross-group correlations between VIs and productivity indicators were also high ($r > 0.8$), suggesting potential multicollinearity among these variables. Similarly, strong correlations were observed between vegetation indices and productivity indicators, which can be attributed to their common derivation from remote sensing spectral bands.

Soil properties showed clear south-to-north latitudinal spatial patterns across the EU, indicating the spatial variation in soil and climate that is critical for yield estimation at continental scales (**Fig. S2**). Northern European regions were characterized by lower clay and bulk density but higher sand content, while southern regions exhibited inverse patterns. Soil organic carbon content and field capacity showed a clear north–south gradient, with higher values in northern regions and lower values in Mediterranean areas. Soil pH showed more acidic soils (pH 5.0–6.0) in northern Europe and alkaline soils (pH > 6.5) in southern regions. These spatial patterns were consistent between surface (0 cm) and subsurface (30 cm) depths, though with slight variations in magnitude. Correlation analysis between soil properties and yield (**Fig. S3**) showed weak relationships between soil properties and crop yield ($r < 0.5$), but strong correlations between surface and subsurface measurements of the same properties. This pronounced spatial variability highlights the importance of incorporating soil properties in large-scale yield estimation models to capture region-specific growing conditions and reduce bias across diverse agroecological zones.

3.2. Yield estimation performance

The proposed TMF framework achieved robust performance in large-scale wheat yield estimation, outperforming all baseline models using LOYO-CV (Fig. 5). The TMF framework with a late fusion strategy

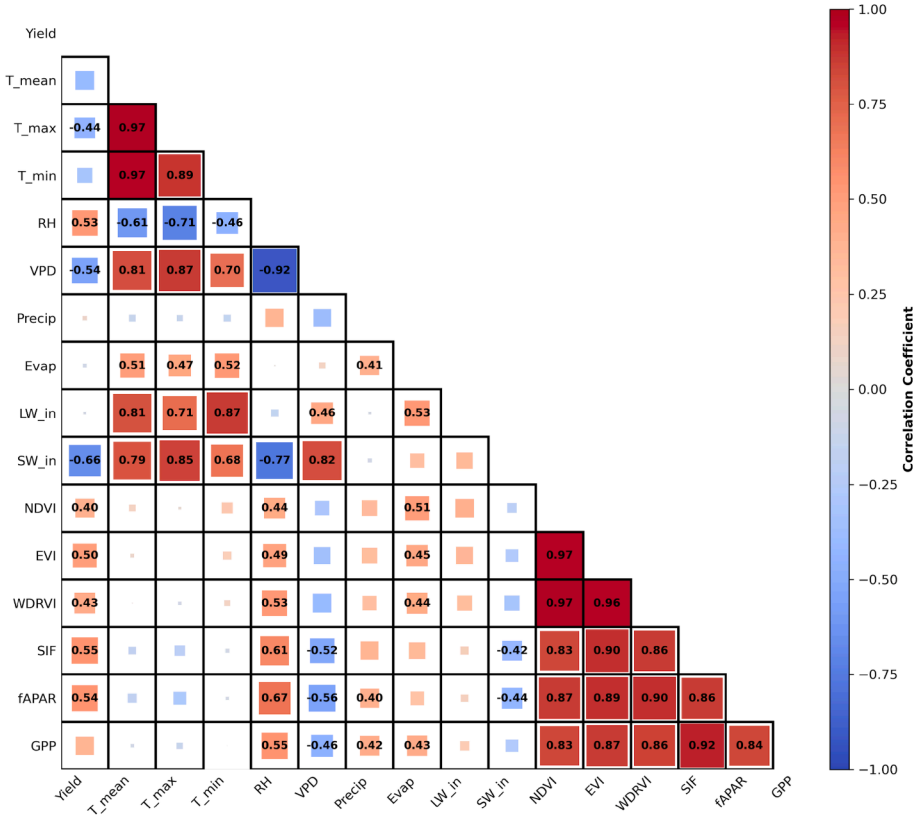

Fig. 4. Correlation heatmap of the sub-national level wheat yield and growing season accumulated input time-series variables. Weekly time-series data were summed over the defined growing season. Only correlation coefficients with $|r| > 0.5$ are labeled numerically. Black bold frames indicate statistically significant correlations at the $p < 0.05$ level.

demonstrated the best overall performance (RMSE = 0.75 Mg·ha^{-1}, R^2 = 0.88, and rRMSE = 13.4%), reducing RMSE by 0.06–0.46 Mg·ha^{-1} (7.0–37.8%) compared to baseline models. While the early fusion variant showed slightly degraded performance (RMSE = 0.79 Mg·ha^{-1} and R^2 = 0.87), it still outperformed all baseline models, suggesting the effectiveness of transformer-based architectures in capturing complex temporal dependencies in time-series data for crop yield estimation. We also observed that temporal deep learning models consistently outperformed traditional machine learning approaches in large-scale wheat yield estimation. Among the baseline models, LSTM achieved the strongest performance (RMSE = 0.81 Mg·ha^{-1}), followed by GBRT (RMSE = 0.98 Mg·ha^{-1}), RF (RMSE = 1.14 Mg·ha^{-1}), and ANN (RMSE = 1.21 Mg·ha^{-1}).

Regarding the temporal robustness of model performance, the TMF (LateFusion) maintained consistent results across years (mean RMSE = 0.75 ± 0.09 Mg·ha^{-1}) (**Fig. S4**). Such temporally consistent performance indicates high stability in yield estimation across diverse climatic conditions. In contrast, traditional machine learning approaches showed higher variability, with ANN exhibiting the greatest error (mean RMSE = 1.16 ± 0.37 Mg·ha^{-1}). The RMSE heatmaps further highlighted the year-to-year performance dynamics (Fig. 6), where TMF (LateFusion) maintained high accuracy in nearly all years and demonstrated strong performance during extreme drought years such as 2003 and 2018 (Ciais et al., 2005; Büntgen et al., 2021). These results demonstrate the robustness of the TMF framework in capturing year-to-year crop-environment variation, improving both average yield estimation accuracy and model temporal generalizability under varying and adverse environmental conditions in large-scale crop yield estimation. In addition, a spatial cross-validation experiment based on NUTS-1 regions (**Fig. S5**) was conducted to further assess model performance under spatially independent evaluation. The results showed that the TMF framework maintains strong predictive skill (RMSE = 0.89 Mg·ha^{-1}, R^2 = 0.84), although performance decreased compared to the LOYO-CV setting due to increased spatial distribution shifts.

The spatial distribution of estimation errors illustrates distinct patterns in model performance across regions and climate conditions (Fig. 7). Overall, all models tended to overestimate yield in Eastern Europe, reflecting the inherent spatial heterogeneity in environmental and management practices. Among all models, the TMF (LateFusion) showed higher spatial consistency in yield estimations, with smaller error magnitudes and more balanced error distributions across regions and years than baseline models. This spatial consistency was more evident in normal years (2008 and 2013), where TMF (LateFusion) maintained estimation errors largely within ± 1 Mg·ha^{-1} across most NUTS regions.

During years of extreme climatic events, all models showed more significant spatial variability in yield estimation errors compared to normal years. Spatial analysis of yield anomalies (**Fig. S6**) confirmed that drought-induced yield losses exhibit strong spatial heterogeneity. Regions with larger yield loss tend to correspond to areas of model overestimation, suggesting that abrupt and extreme yield reductions remain more challenging to fully capture. Temporal deep learning models (TMF (LateFusion), TMF (EarlyFusion), and LSTM) could better reproduce the yield loss under extreme climate stress than traditional machine learning approaches. The TMF framework showed the best overall performance compared with other models in both drought and normal years. TMF (EarlyFusion) slightly outperformed TMF (LateFusion) in 2003 (RMSE = 0.83 vs. 0.84 Mg·ha^{-1}), but TMF (LateFusion) achieved better accuracy in other years and more consistent spatial performance. In 2018, when severe drought and heat stress impacted central and northern Europe, most models overestimated yields in the affected regions. However, temporal deep learning models showed more

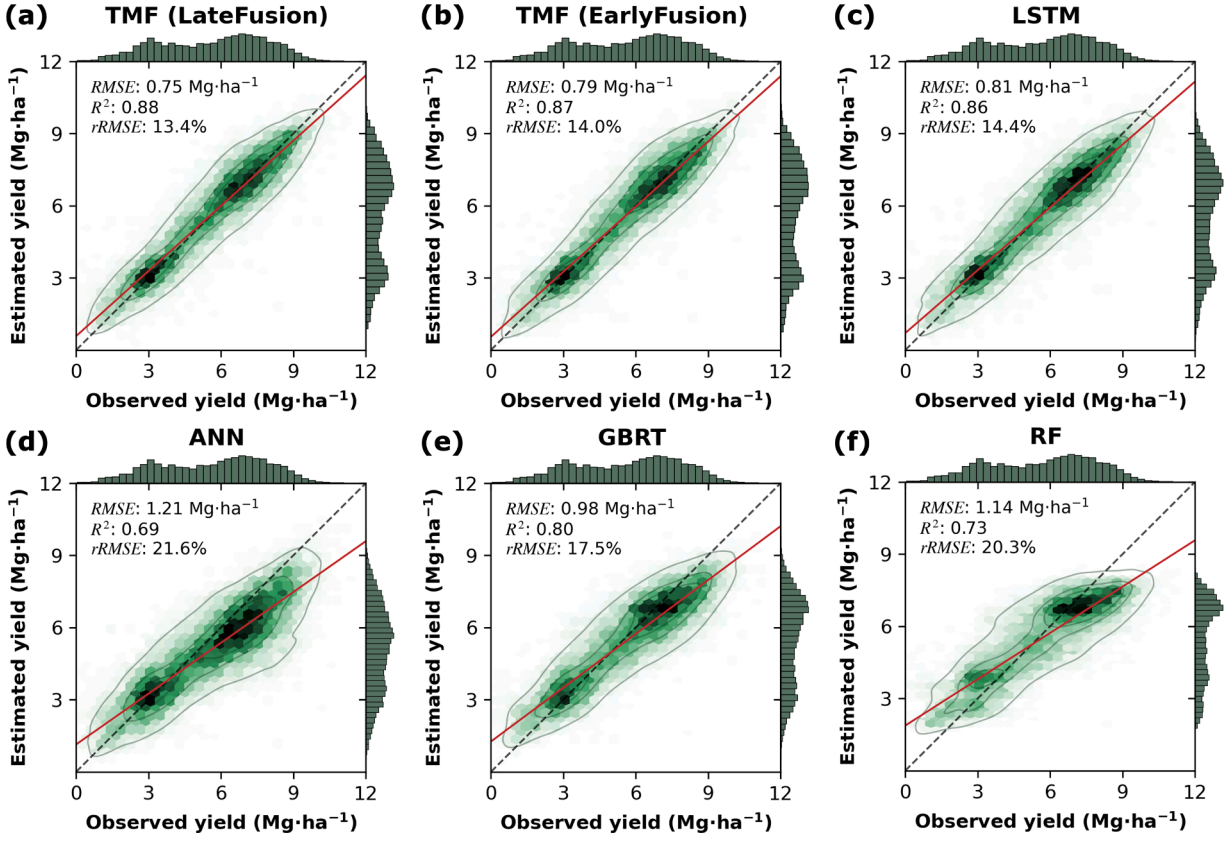

Fig. 5. Scatter plots comparing estimated and observed yields across six models using pooled test results from LOYO-CV: (a) TMF (LateFusion), (b) TMF (EarlyFusion), (c) LSTM, (d) ANN, (e) GBRT, and (f) RF. The black dashed line represents the 1:1 line, and the red solid line shows the fitted regression line. Green shading indicates point density, with darker colors representing higher density. Histograms on the axes show the distribution of actual (x-axis) and estimated (y-axis) yields.

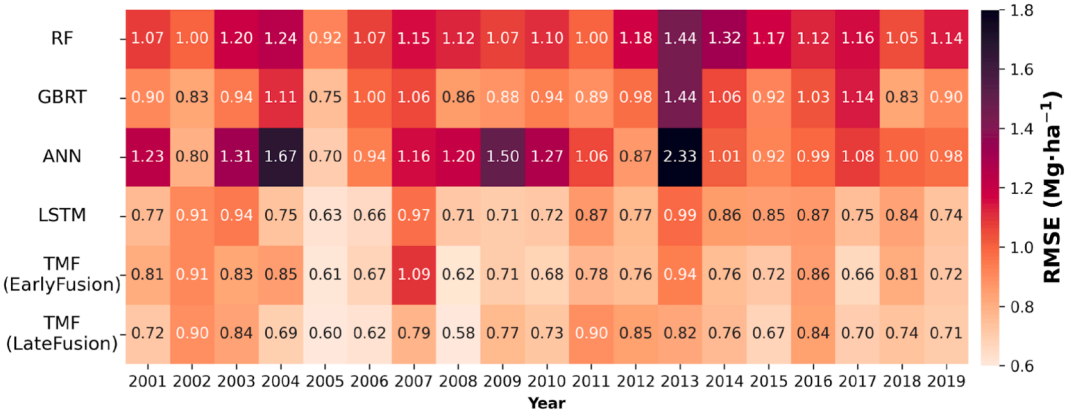

Fig. 6. Heatmap of year-wise RMSE for different models under LOYO-CV (2001–2019).

moderate overestimation (generally < 1.5 Mg·ha^{-1}) compared to traditional approaches, suggesting a better capability in capturing crop yield responses to extreme conditions.

3.3. Differences among input modality combinations for yield estimation

The ablation experiment revealed how different combinations of input modalities influence the performance of the TMF (LateFusion) model (Fig. 8). Yield estimation performance generally improved as more input modalities were fused. The combination of all four input modalities—climate variables, vegetation indices, productivity indicators, and soil properties (CVPS)—achieved the best performance, yielding the lowest RMSE of 0.75 Mg·ha^{-1}, the highest R^2 of 0.88 on the pooled test sets, and the most stable performance across years. The model performance gradually degraded as fewer modalities were used, with three-modality combinations showing RMSEs of 0.76–0.89 Mg·ha^{-1}, two-modality combinations around 0.76–0.96 Mg·ha^{-1}, and single-modality yield estimations resulting in the poorest performance (RMSEs of 0.98–1.13 Mg·ha^{-1}). Although partial redundancy exists among certain variables, the consistent performance gains suggest that the TMF framework is able to extract complementary information from heterogeneous modalities.

Soil properties notably improved yield estimation performance when incorporated into different modality combinations through the TMF framework, with RMSE decreasing by 13.2–26.7%. For instance, adding soil properties to the CVP combination reduced RMSE from 0.89 to 0.75 Mg·ha^{-1}. Moreover, integrating satellite-derived measurements and soil properties (CVS, CPS, and CVPS) reduced RMSE by 28.1–33.3% compared to climate-only inputs (C). The consistent improvement across different combinations highlights the complementary role of static soil properties in enhancing spatial differentiation when integrated with temporal climate and vegetation signals for large-scale crop yield estimation. For models using only a single time-series input (C, V, and P), productivity indicators showed superior performance (0.98 Mg·ha^{-1}) compared to climate and vegetation indices (1.13 and 1.06 Mg·ha^{-1}). This suggests that productivity indicators, which more directly reflect photosynthetic activity and canopy productivity, provide signals more closely linked to final yield.

To further assess the robustness of our TMF framework for large-scale yield modeling, we conducted a stability analysis using Monte Carlo dropout during inference. This approach enables quantification of yield estimation stability by generating multiple yield estimates for each sample during the testing phase, while randomly deactivating neurons according to the dropout probability (Goel and Chen, 2021). We analyzed the spatial distribution of the yield estimation variance across different input modality combinations and environmental conditions by generating 30 yield estimates (Fig. 9). Results revealed that models incorporating soil properties (CVPS and VPS) significantly reduced the yield estimation variance compared to their counterparts without soil. Wheat production areas with intensive agriculture (Western Europe) benefited most in model stability from soil data integration. These findings highlight that multi-modal fusion with static soil properties not only improves accuracy but also enhances model estimation stability across heterogeneous environmental conditions for large-scale yield estimation.

3.4. Feature importance of multi-modal inputs

The SHAP-based model interpretability analysis of the TMF (LateFusion) model quantifies how different input modality variables from different growth stages contributed to estimated wheat yields within the integrated TMF framework (Fig. 10). Feature importance generally peaked during the mid-to-late growth periods (April–July), corresponding to reproductive and grain-filling stages when wheat yield is most sensitive to environmental conditions (Asseng et al., 2015). This temporal concentration of importance suggests that the model captures biologically meaningful yield sensitivity windows rather than uniformly weighting the growing season. Vegetation indices (NDVI and EVI) emerged as the strongest contributors, with higher importance values than other modalities. Productivity indicators also exhibited moderate importance, peaking slightly earlier in the growth season compared to vegetation indices. Among climate variables, RH demonstrated higher contributions than temperature metrics or radiation parameters, indicating that moisture-related stress signals may play a more direct role in yield variability within the EU. These results demonstrate the TMF framework's capacity to integrate temporal patterns and leverage the informative signals across modalities, while also aligning well with crop phenology and sensitivity periods for yield. To further understand the model's behavior under extreme conditions, we compared the temporal feature importance between drought years (2003, 2018) and normal years (2008, 2013) (**Fig. S7**). While satellite-based measurements (EVI, NDVI) consistently dominated feature contributions across all years, their importance was notably intensified during the reproductive and grain-filling stages in drought years. This shift indicates the TMF framework's ability to adaptively capture stress-induced changes in crop physiology and adjust feature weighting accordingly for robust yield estimation under varying environmental conditions.

The accumulated SHAP values over the entire growth period illustrate the relative influence of different input modalities in wheat yield estimation (Fig. 11). Among temporal input features, satellite-based measurements (vegetation indices and productivity indicators) demonstrated higher importance than climate variables, with EVI showing the strongest contribution ($|\text{SHAP}| > 1.2$), followed by NDVI, fAPAR, and WDRVI. These results indicate that satellite-based measurements primarily drive year-to-year yield fluctuations by capturing phenological responses to climatic conditions. Despite being static features, soil

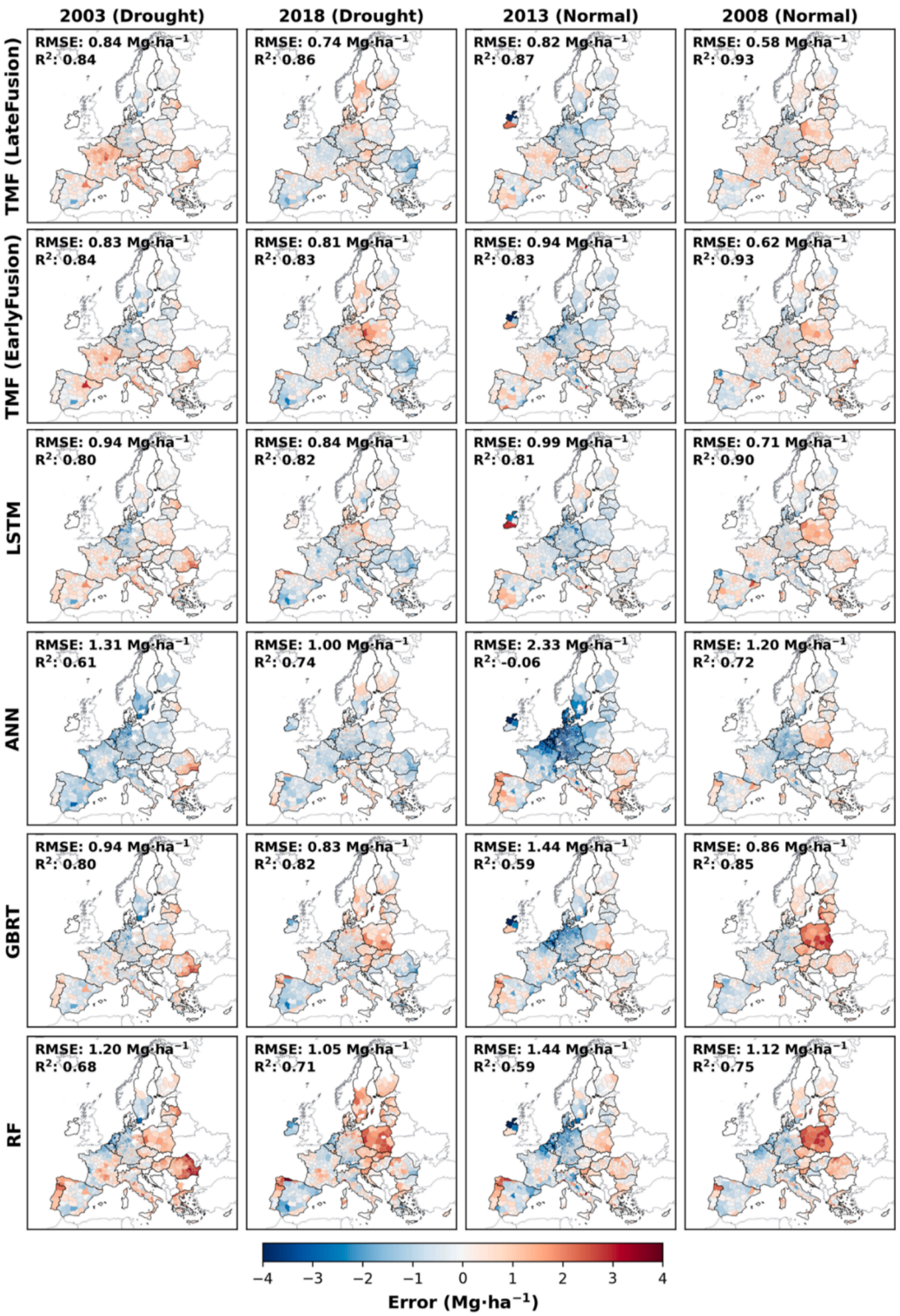

Fig. 7. Maps of the estimation error (estimated − observed yield) at the NUTS level for different models in representative drought (2003, 2018) and normal years (2008, 2013). The selected drought years correspond to documented large-scale European heatwave and drought events, while the normal years provide baseline comparisons. The RMSE and R^2 values shown in the upper-left corner of each panel indicate year-specific predictive accuracy.

properties also showed notable importance in yield estimation. Notably, the subsurface depth (30 cm) soil properties contributed more to yield estimation than those at the surface (0 cm). Overall, the SHAP analysis demonstrated that our TMF framework could effectively synthesize multi-source information through temporal learning and multi-modal fusion, capturing both crop growth dynamics and environmental conditions for improved yield estimation.

4. Discussion

This study demonstrated that integrating multi-modal data through a transformer-based fusion framework (i.e., TMF framework) can substantially improve the accuracy and robustness of large-scale wheat yield estimation across the EU. The TMF framework effectively captured both temporal dynamics and spatial heterogeneity through a late fusion

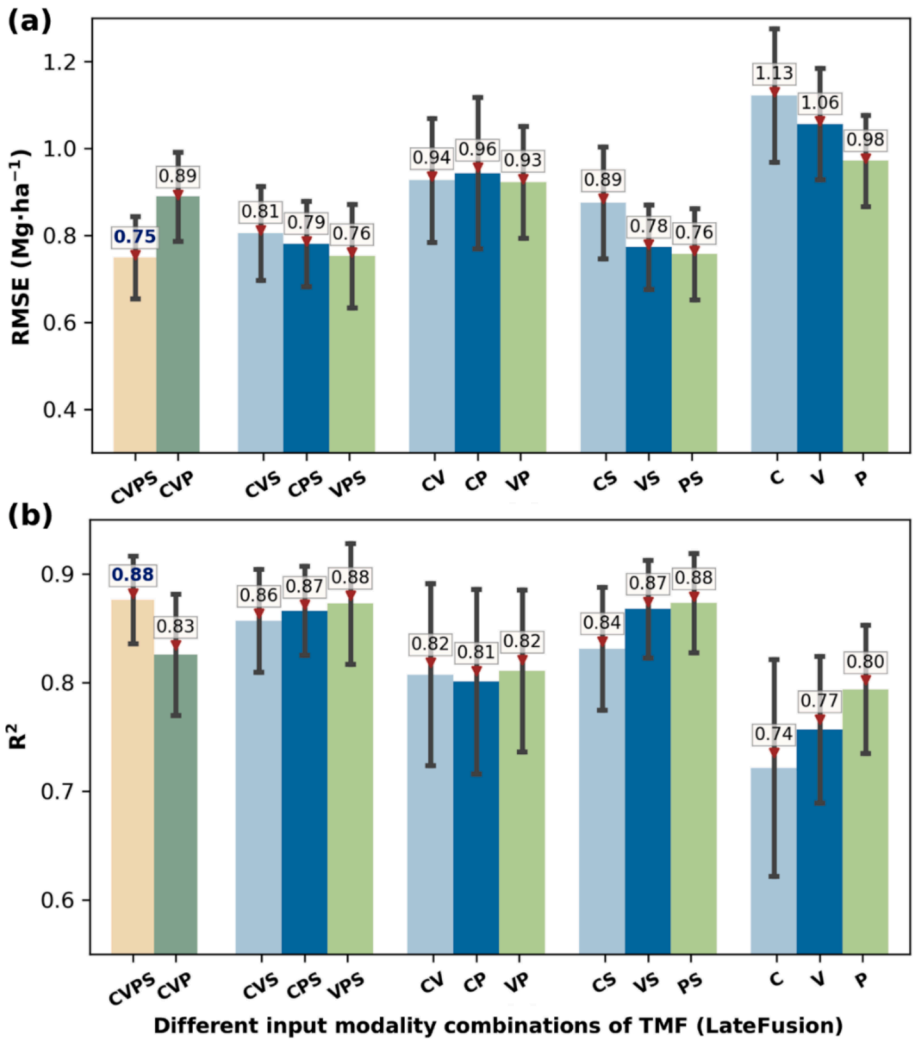

Fig. 8. Ablation study evaluating model performance under different input modality combinations for the TMF (LateFusion) using LOYO-CV from 2001 to 2019: (a) RMSE and (b) R^2. Bars and error bars indicate the means ± standard deviation across individual years (LOYO folds). Red markers and their associated values indicate metrics calculated on pooled test sets, with bold values highlighting optimal performance for each metric. Abbreviations: C = Climate variables, V = Vegetation indices, P = Productivity indicators, and S = Soil properties.

strategy that preserves modality-specific temporal patterns, with consistent performance across diverse agroecological zones and climatic conditions. The inclusion of static soil properties enhanced model performance and stability. Moreover, we also found that remote sensing variables (EVI, NDVI, and fAPAR), especially during critical growth stages (grain-filling), contributed most to yield estimations.

4.1. Advantages of the multi-modal fusion framework for large-scale yield estimations

The proposed TMF framework achieved improved wheat yield estimation across the EU, representing a significant improvement over baseline models by reducing RMSE by 7.0–37.8% (Fig. 5). Our work developed one of the first comprehensive, remote sensing-based frameworks for EU-wide wheat yield estimation at the NUTS-3 level. Previous remote sensing studies on crop yield estimation in Europe have focused on smaller regions, limited agroecological zones, or county-level analyses (Bregaglio et al., 2023; de Nóia-Júnior et al., 2025; Paudel et al., 2022; Ronchetti et al., 2023). For example, the EU Joint Research Center (JRC) has developed an agroecosystem model-based yield forecasting system at the NUTS-1 level (van der Velde et al., 2019a), which is limited to the national scale and failed during extreme events (e.g., in France in 2016). Developing fine-scale yield estimation models across large regions presents challenges due to the spatial heterogeneity in climate, soil conditions, and management practices (Feng et al., 2021; Schils et al., 2018). The robustness of the TMF framework is particularly vital in the context of the observed yield stagnation across much of the EU (Moore and Lobell, 2015). Our framework provides a viable option to address these challenges by effectively integrating high-resolution multi-source data to capture both common patterns and regional variations in crop-environment relationships.

Our results showed significant advancements in crop yield modeling through the integration of transformer-based temporal learning and multi-modal fusion strategies. The transformer encoder architecture with self-attention mechanisms showed improved performance compared with LSTM (reducing RMSE by 7.0%), suggesting its advantage in capturing complex temporal dependencies across wheat phenological stages, consistent with recent findings by Guo et al. (2024) and Rußwurm and Körner (2020). This is also supported by the early-fusion TMF variant, which uses concatenated temporal inputs similar to LSTM but still outperformed the LSTM baseline. Our late fusion approach, which preserves modality-specific temporal patterns before integration, demonstrated an additional improvement over early fusion (approximately 4% lower RMSE), aligning with results from computer vision research (Baltrušaitis et al., 2019; Huang et al., 2020). In addition to overall accuracy, the TMF framework exhibited high temporal

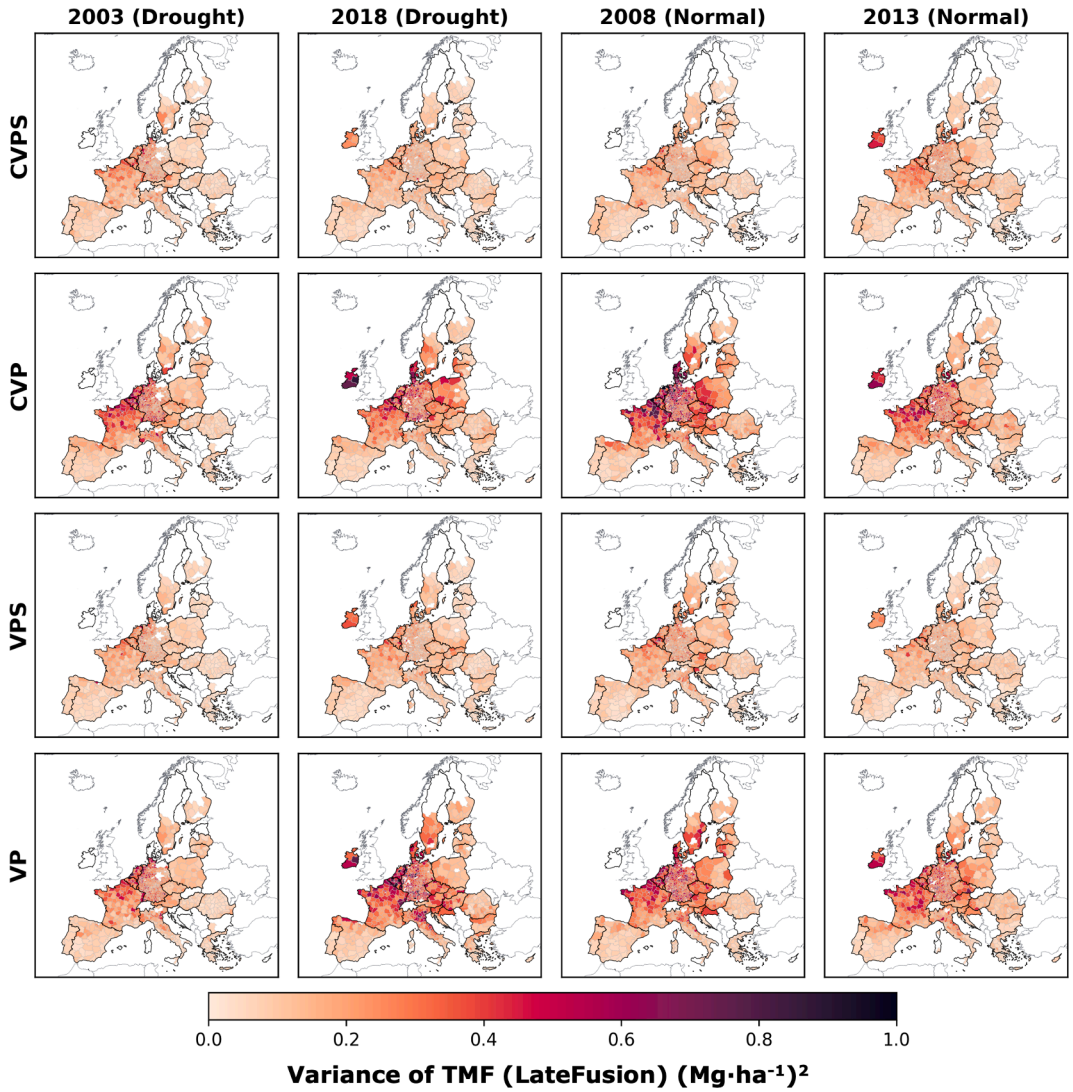

Fig. 9. Estimated wheat yield stability for two sets of models with and without soil input (CVPS vs. CVP and VPS vs. VP) based on Monte Carlo dropout in representative drought (2003, 2018) and normal years (2008, 2013). Abbreviations: C = Climate variables, V = Vegetation indices, P = Productivity indicators, and S = Soil properties.

robustness across years (Fig. 6), maintaining consistent performance under varying climatic conditions and outperforming baseline models in both normal and extreme years (Fig. 7).

The incorporation of static soil properties with temporal features enhanced accuracy and improved the yield estimation stability at large-scales, reducing RMSE by 13–27% across modality combinations (Fig. 8, Fig. 9). In European agricultural monitoring, process-based operational systems (e.g., the JRC MARS framework based on WOFOST) have long recognized soil properties (e.g., texture and water-holding capacity) as fundamental spatial constraints for simulating water stress and baseline yield potential (van der Velde et al., 2019b). While recent machine learning studies have also shown that accounting for soil-related spatial variability is essential for robust yield estimates (Cao et al., 2021; Kaur et al., 2023; Zhong et al., 2022), a systematic quantification of its complementary role within multi-modal deep learning at a pan-European scale remains limited. Our results demonstrate that soil properties contribute unique spatial predictive information that remote sensing and climate variables cannot fully substitute. This finding is particularly meaningful in the European context, where pronounced spatial gradients in soil properties (**Fig. S2**) suggest that soil health plays a critical role in regulating wheat yield patterns across the EU. These spatially structured soil signals provide additional information beyond temporal inputs, enabling the TMF framework to better account for both spatial and temporal variations in yield formation across heterogeneous European environments.

4.2. Contributions of climate, remote sensing, and soil inputs for yield estimations

The interpretability analysis through SHAP provided insights into how different input modalities contributed to wheat yield estimation. Temporal contributions from all modalities generally peaked during the later part of the growing period (Fig. 10), which corresponds to critical reproductive and grain-filling stages for wheat when environmental conditions most significantly influence final yield formation (Asseng et al., 2011; Zachow et al., 2024). The concentration of importance during these stages is consistent with known agronomic significance of reproductive and grain-filling stages in determining final wheat yield. The key stages highlighted by the SHAP analysis are consistent with existing literature in both process-based and empirical models for yield estimation and agronomic studies (Lobell et al., 2012; Mäkinen et al., 2018; Zhou et al., 2023b). Beyond peak importance periods, we also observed that vegetation indices (NDVI, EVI) displayed meaningful signals during early growth stages, suggesting that initial canopy

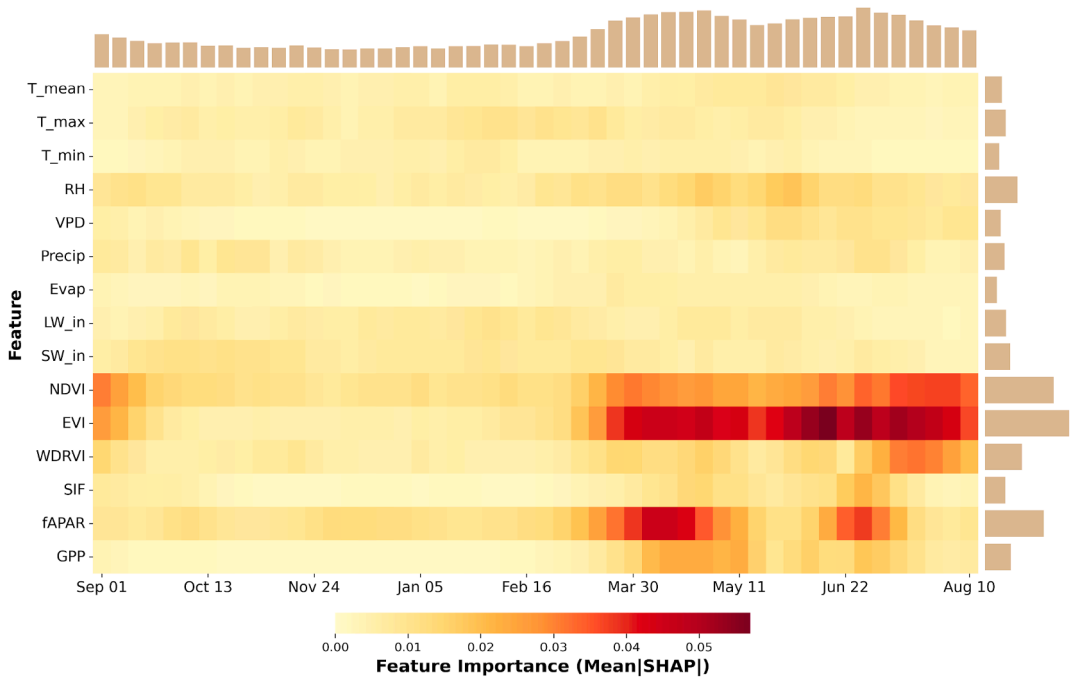

Fig. 10. SHAP-based contribution analysis of temporal input variables in the TMF (LateFusion) using LOYO-CV. Heatmap shows the multi-year average temporal contribution patterns of input variables for estimating wheat yields, where each row represents a feature variable, and columns indicate weekly contributions throughout the growing season. Marginal bar plots on the right demonstrate the overall variable importance, while those at the top show the temporal importance across the growing season.

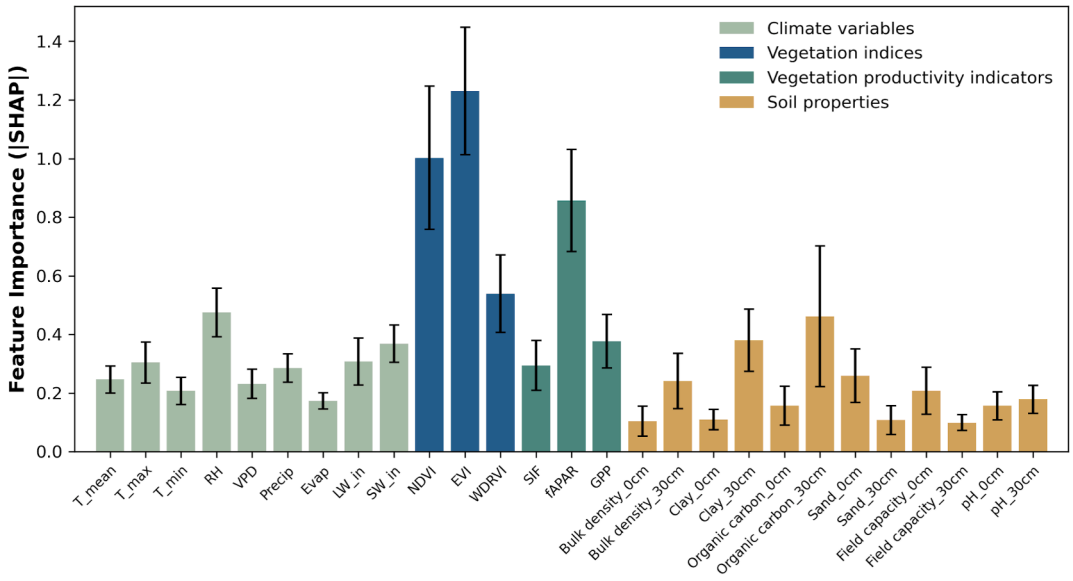

Fig. 11. Accumulated feature importance of temporal and static input variables for estimating wheat yield based on SHAP analysis from 2001 to 2019. For temporal features, absolute SHAP values were first accumulated across the time dimension for each sample and averaged across samples; for static features, the mean absolute SHAP values were directly computed across samples. Error bars represent the inter-annual standard deviation across years.

development captured by VIs provides valuable information for determining final yield (Skakun et al., 2017).

Remote sensing variables demonstrated strong contributions to yield estimation across our study region, with satellite-based remote sensing measurements (vegetation indices and productivity indicators) exhibiting higher contributions than climate variables (Fig. 11). Our ablation studies further confirmed these findings, as the combinations of VS and PS (vegetation indices + soil and productivity indicators + soil) achieved higher performance among dual-modality models (RMSE = 0.76–0.78 $Mg{\cdot}ha^{-1}$). There is growing interest in integrating productivity and biophysical variables into crop models, as they provide unique insights into photosynthetic efficiency compared to traditional VIs, especially during grain-filling stages (Cai et al., 2019; Chang et al., 2025; Guan et al., 2016; Wang et al., 2025). Despite their importance, we observed substantial information redundancy among different remote sensing modalities through correlation analysis (Fig. 4). Many productivity indicators and VIs showed high inter-correlations ($r > 0.8$), largely because they are derived from similar spectral bands and satellite platforms, potentially introducing a bias in feature importance attribution (Qiu et al., 2022). Although remote sensing variables effectively

capture temporal canopy dynamics and interannual variability, redundancy among vegetation indices and productivity indicators may limit additional gains. In contrast, static soil properties provide spatially structured predictive signals not captured by remote sensing inputs. Their integration reduced RMSE by 28.1–33.3% compared to climate-only inputs, indicating that soil information provides non-redundant signals that cannot be fully substituted by dynamic remote sensing features under our experimental setting. Notably, adding climate variables (CPS vs. PS) sometimes failed to improve performance, suggesting that increasing model complexity through additional data modalities can result in diminishing or even negative returns. This suggests that beyond simply adding more data sources, effective fusion strategies must balance information gain against model complexity to avoid overfitting (Zhang et al., 2021).

4.3. Potential improvement in data fusion for crop yield estimations

The study highlights the potential of multi-modal fusion and temporal deep learning approaches for large-scale crop yield estimation through integrating diverse satellite and environmental data. While our framework achieved significant improvements over conventional approaches, several refinements could further enhance its applicability. Since our study focused on NUTS-3 level yield estimation over 19 years, the satellite data used in this research have relatively coarse spatial resolution (250 m for VIs, 0.05° for SIF and GPP). Recent satellites with enhanced spatial and spectral capabilities, such as Sentinel-2 and TROPOMI SIF (Zhang et al., 2023), could improve feature extraction while maintaining regional consistency. Additionally, implementing crop-specific masks rather than general cropland identification could improve signal quality by reducing noise from non-target vegetation (Al-Shammari et al., 2021; Seguini et al., 2024). However, aggregating high-resolution inputs to the NUTS scale inevitably masks within-region variability due to the spatial constraints of official yield statistics. This limitation highlights the need for future research on cross-scale model transfer (Ma et al., 2024; Zhang et al., 2025a). Leveraging transfer learning or domain adaptation techniques, models trained on abundant regional-level data could be adapted to field-level yield estimation, thereby capturing fine-grained spatial heterogeneity even when high-resolution ground truth data remain limited. Such approaches may also facilitate model transfer across regions with distinct agroecological conditions, offering a promising pathway for broader geographic applications.

From a methodological perspective, exploring advanced fusion mechanisms like cross-attention (Li and Wu, 2024) or knowledge-guided neural networks (Liu et al., 2024) is likely to improve the capture of complex interactions between modalities. For example, process-based constraints (e.g., phenological stages, radiation-use efficiency principles, or water balance relationships) could be embedded into the multi-modal fusion architecture to reduce redundant correlations and improve extrapolation under extreme climate conditions. Exploring multi-task learning to simultaneously estimate yield in different agroecological zones could improve the model's ability to address regional variations in crop-environment relationships (Lin et al., 2022; Zhong et al., 2022). The flexible architecture of our fusion framework presents opportunities for future research to integrate additional data sources and extend applications to diverse crop types and geographical regions.

5. Conclusion

This study developed a Temporal Multi-modal Fusion (TMF) framework for large-scale wheat yield estimation across the European Union from 2001 to 2019. The TMF framework with a late fusion strategy achieved an RMSE of 0.75 $Mg{\cdot}ha^{-1}$ and an R^2 of 0.88 by effectively integrating climate variables, satellite-based remote sensing measurements including vegetation indices and productivity indicators, and soil properties. The transformer-based temporal learning with a late fusion strategy outperformed conventional approaches by preserving modality-specific temporal patterns (reduced RMSE by 7.0–37.8% compared to baseline models). Meanwhile, the new framework also effectively integrated heterogeneous information across modalities and demonstrated stable yield estimation performance across diverse years and regions. Ablation studies confirmed that model performance generally improved as more input modalities were fused, with the combination of all four input modalities yielding the best performance. Integrating satellite-derived measurements and static soil properties reduced RMSE by 28.1–33.3% compared to climate-only inputs, highlighting that soil heterogeneity provides spatially-structured predictive signals that complement temporally dynamic information captured by remote sensing inputs. This complementary representation also enhanced model stability and robustness during extreme climate years (e.g., 2003 and 2018 droughts). Interpretability analysis revealed that remote sensing variables contributed more than climate variables, with feature importance peaking during reproductive and grain-filling stages. Overall, our study demonstrates the TMF framework's potential for operational large-scale crop yield monitoring systems that can provide spatially explicit yield estimations across heterogeneous environments, offering potential support for agricultural policy, food security planning, and climate adaptation strategies.

CRediT authorship contribution statement

Zhixian Lin: Writing – original draft, Visualization, Validation, Software, Methodology, Data curation, Conceptualization. **Sheng Wang:** Writing – original draft, Supervision, Resources, Project administration, Methodology, Data curation, Conceptualization. **Kaiyu Guan:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Investigation, Funding acquisition. **Jiaying Zhang:** Writing – review & editing, Validation, Methodology. **Senthold Asseng:** Writing – review & editing, Conceptualization. **René Gislum:** Writing – review & editing, Conceptualization. **Jørgen Eivind Olesen:** Writing – review & editing, Conceptualization. **Klaus Butterbach-Bahl:** Writing – review & editing, Conceptualization. **Miroslav Trnka:** Writing – review & editing, Conceptualization. **Rui Zhou:** Validation.

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.

Acknowledgement

We acknowledge the National Science Foundation (NSF) CAREER Award by the Environmental Sustainability Program (1847334) and the NASA ECOSTRESS Program (80NSSC23K0308). This work is also supported by the USDA NIFA Artificial Intelligence for Future Agricultural Resilience, Management, and Sustainability (AI-FARMS) project. S.W. acknowledges the support from the NASA Early Career Investigator Program in Earth Science (80NSSC24K1057). This work was also supported by the SCALE project (Food&Bio Cluster Denmark, Innovation Fund Denmark), the VILLUM Young Investigator 2024 project (00072051), and the Novo Nordisk Starting Grant (NNF23OC0087612). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and should not be construed to represent any official USDA or U.S. Government determination or policy.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.isprsjprs.2026.05.041.

References

Agatonovic-Kustrin, S., Beresford, R., 2000. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J. Pharm. Biomed. Anal. 22, 717–727. https://doi.org/10.1016/S0731-7085(99)00272-1.

Al-Shammari, D., Whelan, B.M., Wang, C., Bramley, R.G.V., Fajardo, M., Bishop, T.F.A., 2021. Impact of spatial resolution on the quality of crop yield predictions for site-specific crop management. Agric. For. Meteorol. 310, 108622. https://doi.org/10.1016/j.agrformet.2021.108622.

Asseng, S., Ewert, F., Martre, P., Rötter, R.P., Lobell, D.B., Cammarano, D., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Prasad, P.V. V., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Challinor, A.J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.-K., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P.J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z., Zhu, Y., 2015. Rising temperatures reduce global wheat production. Nat. Clim. Chang. 5, 143–147. https://doi.org/10.1038/nclimate2470.

Asseng, S., Foster, I., Turner, N.C., 2011. The impact of temperature variability on wheat yields. Glob. Chang. Biol. 17, 997–1012. https://doi.org/10.1111/j.1365-2486.2010.02262.x.

Bailey-Serres, J., Parker, J.E., Ainsworth, E.A., Oldroyd, G.E.D., Schroeder, J.I., 2019. Genetic strategies for improving crop yields. Nature 575, 109–118. https://doi.org/10.1038/s41586-019-1679-0.

Baltrušaitis, T., Ahuja, C., Morency, L.-P., 2019. Multimodal machine learning: a survey and taxonomy. IEEE Trans. Pattern Anal. Mach. Intell. 41, 423–443. https://doi.org/10.1109/TPAMI.2018.2798607.

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Niño, F., Weiss, M., Samain, O., Roujean, J.L., Leroy, M., 2007. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: part 1: principles of the algorithm. Remote Sens. Environ. 110, 275–286. https://doi.org/10.1016/j.rse.2007.02.018.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F., 2020. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012.

Bazzi, H., Ciais, P., Makowski, D., Baghdadi, N., 2024. Advancing winter wheat yield anomaly prediction with high-resolution satellite-based gross primary production. One Earth S2590332224005451. https://doi.org/10.1016/j.oneear.2024.10.021.

Ben-Ari, T., Boé, J., Ciais, P., Lecerf, R., Van der Velde, M., Makowski, D., 2018. Causes and implications of the unforeseen 2016 extreme yield loss in the breadbasket of France. Nat. Commun. 9, 1627. https://doi.org/10.1038/s41467-018-04087-x.

Bregaglio, S., Ginaldi, F., Raparelli, E., Fila, G., Bajocco, S., 2023. Improving crop yield prediction accuracy by embedding phenological heterogeneity into model parameter sets. Agr. Syst. 209, 103666. https://doi.org/10.1016/j.agsy.2023.103666.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brisson, N., Gate, P., Gouache, D., Charmet, G., Oury, F.-X., Huard, F., 2010. Why are wheat yields stagnating in Europe? a comprehensive data analysis for France. Field Crop Res 119, 201–212. https://doi.org/10.1016/j.fcr.2010.07.012.

Bueechi, E., Fischer, M., Crocetti, L., Trnka, M., Grlj, A., Zappa, L., Dorigo, W., 2023. Crop yield anomaly forecasting in the pannonian basin using gradient boosting and its performance in years of severe drought. Agric. For. Meteorol. 340, 109596. https://doi.org/10.1016/j.agrformet.2023.109596.

Büntgen, U., Urban, O., Krusic, P.J., Rybníček, M., Kolář, T., Kyncl, T., Ač, A., Koňasová, E., Čáslavský, J., Esper, J., Wagner, S., Saurer, M., Tegel, W., Dobrovolný, P., Cherubini, P., Reinig, F., Trnka, M., 2021. Recent european drought extremes beyond common era background variability. Nat. Geosci. 14, 191–196. https://doi.org/10.1038/s41561-021-00698-0.

Cai, Y., Guan, K., Lobell, D., Potgieter, A.B., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L., Peng, B., 2019. Integrating satellite and climate data to predict wheat yield in australia using machine learning approaches. Agric. For. Meteorol. 274, 144–159. https://doi.org/10.1016/j.agrformet.2019.03.010.

Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Zhang, J., Han, J., Xie, J., 2021. Integrating multi-source data for rice yield prediction across China using machine learning and deep learning approaches. Agric. For. Meteorol. 297, 108275. https://doi.org/10.1016/j.agrformet.2020.108275.

Chang, C.Y., Barnaby, J.Y., Maul, J.E., 2025. Sensitivity of sun-induced chlorophyll fluorescence (SIF) and hyperspectral reflectance to drought response in soybean genotypes with contrasting affinities for arbuscular mycorrhizal fungi. Remote Sens. Environ. 323, 114722. https://doi.org/10.1016/j.rse.2025.114722.

Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., Chevallier, F., De Noblet, N., Friend, A.D., Friedlingstein, P., Grünwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J.M., Papale, D., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J.F., Sanz, M.J., Schulze, E.D., Vesala, T., Valentini, R., 2005. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. Nature 437, 529–533. https://doi.org/10.1038/nature03972.

Clarke, D., Hess, T.M., Haro-Monteagudo, D., Semenov, M.A., Knox, J.W., 2021. Assessing future drought risks and wheat yield losses in England. Agric. For. Meteorol. 297, 108248. https://doi.org/10.1016/j.agrformet.2020.108248.

Defourny, P., Lamarche, C., Brockmann, C., Boettcher, M., Bontemps, S., De Maet, T., Duveiller, G.L., Harper, K., Hartley, A., Kirches, G., Moreau, I., Peylin, P., Ottlé, C., Radoux, J., Van Bogaert, E., Ramoino, F., Albergel, C., Arino, O., 2023. Observed annual global land-use change from 1992 to 2020 three times more dynamic than reported by inventory-based statistics. Prep.

Dhakar, R., Sehgal, V.K., Chakraborty, D., Sahoo, R.N., Mukherjee, J., Ines, A.V.M., Kumar, S.N., Shirsath, P.B., Roy, S.B., 2022. Field scale spatial wheat yield forecasting system under limited field data availability by integrating crop simulation model with weather forecast and satellite remote sensing. Agr. Syst. 195, 103299. https://doi.org/10.1016/j.agsy.2021.103299.

Eurostat, 2024. Cereals statistics. Available at: https://agriculture.ec.europa.eu/data-and-analysis/markets/overviews/market-observatories/crops/cereals-statistics_en (Accessed: 2025/3/20).

Eurostat, 2021. Cereals by NUTS 2 region. Available at: https://ec.europa.eu/eurostat/databrowser/product/page/ef_lac_cereals (Accessed: 2025/3/20).

Fang, R., Yu, L., Xie, N., Wang, K., Rao, K., Zhong, L., Chen, J., Wu, Z., 2026. A knowledge-informed fusion network for process-aware crop yield prediction. Int. J. Appl. Earth Obs. Geoinf. 146, 105154. https://doi.org/10.1016/j.jag.2026.105154.

Feng, L., Wang, Y., Zhang, Z., Du, Q., 2021. Geographically and temporally weighted neural network for winter wheat yield prediction. Remote Sens. Environ. 262, 112514. https://doi.org/10.1016/j.rse.2021.112514.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.

Gámez, A.L., Segarra, J., Vatter, T., Santesteban, L.G., Araus, J.L., Aranjuelo, I., 2025. Alfalfa yield estimation using the combination of sentinel-2 and meteorological data. Field Crops Res. 326, 109857. https://doi.org/10.1016/j.fcr.2025.109857.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. Big Remotely Sensed Data: Tools, Applications and Experiences 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Guan, K., Berry, J.A., Zhang, Y., Joiner, J., Guanter, L., Badgley, G., Lobell, D.B., 2016. Improving the monitoring of crop productivity using spaceborne solar-induced fluorescence. Glob. Chang. Biol. 22, 716–726. https://doi.org/10.1111/gcb.13136.

Guan, K., Wu, J., Kimball, J.S., Anderson, M.C., Frolking, S., Li, B., Hain, C.R., Lobell, D. B., 2017. The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. Remote Sens. Environ. 199, 333–349. https://doi.org/10.1016/j.rse.2017.06.043.

Guo, F., Wang, P., Tansey, K., Zhang, Y., Li, M., Liu, J., Zhang, S., 2024. A novel transformer-based neural network under model interpretability for improving wheat yield estimation using remotely sensed multi-variables. Comput. Electron. Agric. 223, 109111. https://doi.org/10.1016/j.compag.2024.109111.

Hengl, T., 2018. Clay content in % (kg / kg) at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (Version v02). Doi: 10.5281/zenodo.1476854.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780.

Hu, T., Zhang, X., Bohrer, G., Liu, Y., Zhou, Y., Martin, J., Li, Y., Zhao, K., 2023. Crop yield prediction via explainable AI and interpretable machine learning: dangers of black box models for evaluating climate change impacts on crop yield. Agric. For. Meteorol. 336, 109458. https://doi.org/10.1016/j.agrformet.2023.109458.

Huang, J., Gómez-Dans, J.L., Huang, H., Ma, H., Wu, Q., Lewis, P.E., Liang, S., Chen, Z., Xue, J.-H., Wu, Y., Zhao, F., Wang, J., Xie, X., 2019. Assimilation of remote sensing into crop growth models: current status and perspectives. Agric. For. Meteorol. 276–277, 107609. https://doi.org/10.1016/j.agrformet.2019.06.008.

Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I., Lungren, M.P., 2020. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. Npj Digit. Med. 3, 1–9. https://doi.org/10.1038/s41746-020-00341-z.

Jiang, C., Guan, K., Wu, G., Peng, B., Wang, S., 2021. A daily, 250m and real-time gross primary productivity product (2000–present) covering the contiguous united states. Earth Syst. Sci. Data 13, 281–298. https://doi.org/10.5194/essd-13-281-2021.

Jiang, H., Hu, H., Zhong, R., Xu, J., Xu, J., Huang, J., Wang, S., Ying, Y., Lin, T., 2020. A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: a case study of the US corn belt at the county level. Glob. Chang. Biol. 26, 1754–1766. https://doi.org/10.1111/gcb.14885.

Jones, J.W., Antle, J.M., Basso, B., Boote, K.J., Conant, R.T., Foster, I., Godfray, H.C.J., Herrero, M., Howitt, R.E., Janssen, S., Keating, B.A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., Wheeler, T.R., 2017. Brief history of agricultural systems modeling. Agr. Syst. 155, 240–254. https://doi.org/10.1016/j.agsy.2016.05.014.

Kahiluoto, H., Kaseva, J., Balek, J., Olesen, J.E., Ruiz-Ramos, M., Gobin, A., Kersebaum, K.C., Takáč, J., Ruget, F., Ferrise, R., Bezak, P., Capellades, G., Dibari, C., Mäkinen, H., Nendel, C., Ventrella, D., Rodríguez, A., Bindi, M., Trnka, M., 2019. Decline in climate resilience of european wheat. Proc. Natl. Acad. Sci. 116, 123–128. https://doi.org/10.1073/pnas.1804387115.

Kaur, A., Goyal, P., Rajhans, R., Agarwal, L., Goyal, N., 2023. Fusion of multivariate time series meteorological and static soil data for multistage crop yield prediction using multi-head self attention network. Expert Syst. Appl. 226, 120098. https://doi.org/10.1016/j.eswa.2023.120098.

Le Gouis, J., Oury, F.-X., Charmet, G., 2020. How changes in climate and agricultural practices influenced wheat production in western europe. J. Cereal Sci. 93, 102960. https://doi.org/10.1016/j.jcs.2020.102960.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Li, H., Wu, X.-J., 2024. CrossFuse: a novel cross attention mechanism based infrared and visible image fusion approach. Inf. Fusion 103, 102147. https://doi.org/10.1016/j.inffus.2023.102147.

Li, J., Hong, D., Gao, L., Yao, J., Zheng, K., Zhang, B., Chanussot, J., 2022. Deep learning in multimodal remote sensing data fusion: a comprehensive review. ArXiv220501380 Cs Eess.

Li, X., Xiao, J., 2019. Mapping photosynthesis solely from solar-induced chlorophyll fluorescence: a global, fine-resolution dataset of gross primary production derived from OCO-2. Remote Sens. 11, 2563. https://doi.org/10.3390/rs11212563.

Lin, T., Zhong, R., Wang, Y., Xu, J., Jiang, H., Xu, J., Ying, Y., Rodriguez, L., Ting, K.C., Li, H., 2020. DeepCropNet: a deep spatial-temporal learning framework for county-level corn yield estimation. Environ. Res. Lett. 15, 34016. https://doi.org/10.1088/1748-9326/ab66cb.

Lin, Z., Zhong, R., Xiong, X., Guo, C., Xu, J., Zhu, Y., Xu, J., Ying, Y., Ting, K.C., Huang, J., Lin, T., 2022. Large-scale rice mapping using multi-task spatiotemporal deep learning and sentinel-1 SAR time series. Remote Sens. 14, 699. https://doi.org/10.3390/rs14030699.

Lipper, L., Thornton, P., Campbell, B.M., Baedeker, T., Braimoh, A., Bwalya, M., Caron, P., Cattaneo, A., Garrity, D., Henry, K., Hottle, R., Jackson, L., Jarvis, A., Kossam, F., Mann, W., McCarthy, N., Meybeck, A., Neufeldt, H., Remington, T., Sen, P.T., Sessa, R., Shula, R., Tibu, A., Torquebiau, E.F., 2014. Climate-smart agriculture for food security. Nat. Clim. Chang. 4, 1068–1072. https://doi.org/10.1038/nclimate2437.

Liu, L., Zhou, W., Guan, K., Peng, B., Xu, S., Tang, J., Zhu, Q., Till, J., Jia, X., Jiang, C., Wang, S., Qin, Z., Kong, H., Grant, R., Mezbahuddin, S., Kumar, V., Jin, Z., 2024. Knowledge-guided machine learning can improve carbon cycle quantification in agroecosystems. Nat. Commun. 15, 357. https://doi.org/10.1038/s41467-023-43860-5.

Lobell, D.B., Roberts, M.J., Schlenker, W., Braun, N., Little, B.B., Rejesus, R.M., Hammer, G.L., 2014. Greater sensitivity to drought accompanies maize yield increase in the U.S. midwest. Science 344, 516–519. https://doi.org/10.1126/science.1251423.

Lobell, D.B., Sibley, A., Ivan Ortiz-Monasterio, J., 2012. Extreme heat effects on wheat senescence in India. Nat. Clim. Chang. 2, 186–189. https://doi.org/10.1038/nclimate1356.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67.

Ma, Y., Liang, S.-Z., Myers, D.B., Swatantran, A., Lobell, D.B., 2024. Subfield-level crop yield mapping without ground truth data: a scale transfer framework. Remote Sens. Environ. 315, 114427. https://doi.org/10.1016/j.rse.2024.114427.

Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., Fritschi, F.B., 2020. Soybean yield prediction from UAV using multimodal data fusion and deep learning. Remote Sens. Environ. 237, 111599. https://doi.org/10.1016/j.rse.2019.111599.

Mäkinen, H., Kaseva, J., Trnka, M., Balek, J., Kersebaum, K.C., Nendel, C., Gobin, A., Olesen, J.E., Bindi, M., Ferrise, R., Moriondo, M., Rodríguez, A., Ruiz-Ramos, M., Takáč, J., Bezák, P., Ventrella, D., Ruget, F., Capellades, G., Kahiluoto, H., 2018. Sensitivity of european wheat to extreme weather. Field Crops Res. 222, 209–217. https://doi.org/10.1016/j.fcr.2017.11.008.

Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J.W., Rötter, R.P., Boote, K.J., Ruane, A.C., Thorburn, P.J., Cammarano, D., Hatfield, J.L., Rosenzweig, C., Aggarwal, P.K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R.F., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Kumar, S.N., Nendel, C., O'leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J.W., Wolf, J., 2015. Multimodel ensembles of wheat growth: many models are better than one. Glob. Change Biol. 21, 911–925. Doi: 10.1111/gcb.12768.

Meitner, J., Balek, J., Bláhová, M., Semerádová, D., Hlavinka, P., Lukas, V., Jurečka, F., Žalud, Z., Klem, K., Anderson, M.C., Dorigo, W., Fischer, M., Trnka, M., 2023. Estimating drought-induced crop yield losses at the cadastral area level in the czech republic. Agronomy 13, 1669. https://doi.org/10.3390/agronomy13071669.

Mena, F., Pathak, D., Najjar, H., Sanchez, C., Helber, P., Bischke, B., Habelitz, P., Miranda, M., Siddamsetty, J., Nuske, M., Charfuelan, M., Arenas, D., Vollmer, M., Dengel, A., 2025. Adaptive fusion of multi-modal remote sensing data for optimal sub-field crop yield prediction. Remote Sens. Environ. 318, 114547. https://doi.org/10.1016/j.rse.2024.114547.

Metzger, M.J., Bunce, R.G.H., Jongman, R.H.G., Mücher, C.A., Watkins, J.W., 2005. A climatic stratification of the environment of europe. Glob. Ecol. Biogeogr. 14, 549–563. https://doi.org/10.1111/j.1466-822X.2005.00190.x.

Moore, F.C., Lobell, D.B., 2015. The fingerprint of climate trends on european crop yields. Proc. Natl. Acad. Sci. 112, 2670–2675. https://doi.org/10.1073/pnas.1409606112.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., et al., 2021. ERA5-land: a state-of-the-art global reanalysis dataset for land applications. Earth Syst. Sci. Data 13, 4349–4383.

Nóia Júnior, R. de S., Deswarte, J.-C., Cohan, J.-P., Martre, P., van der Velde, M., Lecerf, R., Webber, H., Ewert, F., Ruane, A.C., Slafer, G.A., Asseng, S., 2023. The extreme 2016 wheat yield failure in France. Glob. Change Biol. 29, 3130–3146. Doi: 10.1111/gcb.16662.

Nóia-Júnior, R. de S., Martre, P., Deswarte, J.-C., Cohan, J.-P., Van der Velde, M., Webber, H., Ewert, F., Ruane, A.C., Ben-Ari, T., Asseng, S., 2025. Past and future wheat yield losses in france's breadbasket. Field Crops Res. 322, 109703. Doi: 10.1016/j.fcr.2024.109703.

Paudel, D., Boogaard, H., De Wit, A., Van Der Velde, M., Claverie, M., Nisini, L., Janssen, S., Osinga, S., Athanasiadis, I.N., 2022. Machine learning for regional crop yield forecasting in Europe. Field Crops Res. 276, 108377. https://doi.org/10.1016/j.fcr.2021.108377.

Peng, B., Guan, K., Zhou, W., Jiang, C., Frankenberg, C., Sun, Y., He, L., Köhler, P., 2020. Assessing the benefit of satellite-based solar-induced chlorophyll fluorescence in crop yield prediction. Int. J. Appl. Earth Obs. Geoinformation 90, 102126. https://doi.org/10.1016/j.jag.2020.102126.

Potopová, V., Trnka, M., Hamouz, P., Soukup, J., Castraveț, T., 2020. Statistical modelling of drought-related yield losses using soil moisture-vegetation remote sensing and multiscalar indices in the south-eastern europe. Agric Water Manag 236, 106168. https://doi.org/10.1016/j.agwat.2020.106168.

Qiu, R., Li, X., Han, G., Xiao, J., Ma, X., Gong, W., 2022. Monitoring drought impacts on crop productivity of the U.S. midwest with solar-induced fluorescence: GOSIF outperforms GOME-2 SIF and MODIS NDVI, EVI, and NIRv. Agric. For. Meteorol. 323, 109038. https://doi.org/10.1016/j.agrformet.2022.109038.

Ray, D.K., Ramankutty, N., Mueller, N.D., West, P.C., Foley, J.A., 2012. Recent patterns of crop yield growth and stagnation. Nat. Commun. 3, 1293. https://doi.org/10.1038/ncomms2296.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., 2019. Deep learning and process understanding for data-driven earth system science. Nature 566, 195–204.

Rezaei, E.E., Webber, H., Asseng, S., Boote, K., Durand, J.L., Ewert, F., Martre, P., MacCarthy, D.S., 2023. Climate change impacts on crop yields. Nat. Rev. Earth Environ. 4, 831–846. https://doi.org/10.1038/s43017-023-00491-0.

Ronchetti, G., Manfron, G., Weissteiner, C.J., Seguini, L., Nisini Scacchiafichi, L., Panarello, L., Baruth, B., 2023. Remote sensing crop group-specific indicators to support regional yield forecasting in europe. Comput. Electron. Agric. 205, 107633. https://doi.org/10.1016/j.compag.2023.107633.

Ronchetti, G., Nisini Scacchiafichi, L., Seguini, L., Cerrani, I., van der Velde, M., 2024. Harmonized european union subnational crop statistics can reveal climate impacts and crop cultivation shifts. Earth Syst. Sci. Data 16, 1623–1649. https://doi.org/10.5194/essd-16-1623-2024.

Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., Winter, J.M., 2013. The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. Agric. for. Meteorol. Agricultural Prediction Using Climate Model Ensembles 170, 166–182. https://doi.org/10.1016/j.agrformet.2012.09.011.

Rußwurm, M., Körner, M., 2020. Self-attention for raw optical satellite time series classification. ISPRS J. Photogramm. Remote Sens. 169, 421–435. https://doi.org/10.1016/j.isprsjprs.2020.06.006.

Sakamoto, T., Gitelson, A.A., Arkebauer, T.J., 2013. MODIS-based corn grain yield estimation model incorporating crop phenology information. Remote Sens. Environ. 131, 215–231. https://doi.org/10.1016/j.rse.2012.12.017.

Schils, R., Olesen, J.E., Kersebaum, K.-C., Rijk, B., Oberforster, M., Kalyada, V., Khitrykau, M., Gobin, A., Kirchev, H., Manolova, V., Manolov, I., Trnka, M., Hlavinka, P., Palosuo, T., Peltonen-Sainio, P., Jauhiainen, L., Lorgeou, J., Marrou, H., Danalatos, N., Archontoulis, S., Fodor, N., Spink, J., Roggero, P.P., Bassu, S., Pulina, A., Seehusen, T., Uhlen, A.K., Żyłowska, K., Nieróbca, A., Kozyra, J., Silva, J.V., Maçãs, B.M., Coutinho, J., Ion, V., Takáč, J., Mínguez, M.I., Eckersten, H., Levy, L., Herrera, J.M., Hiltbrunner, J., Kryvobok, O., Kryvoshein, O., Sylvester-Bradley, R., Kindred, D., Topp, C.F.E., Boogaard, H., de Groot, H., Lesschen, J.P., van Bussel, L., Wolf, J., Zijlstra, M., van Loon, M.P., van Ittersum, M. K., 2018. Cereal yield gaps across europe. Eur. J. Agron. 101, 109–120. https://doi.org/10.1016/j.eja.2018.09.003.

Schwalbert, R.A., Amado, T., Corassa, G., Pott, L.P., Prasad, P.V.V., Ciampitti, I.A., 2020. Satellite-based soybean yield forecast: integrating machine learning and weather data for improving crop yield prediction in southern Brazil. Agric. For. Meteorol. 284, 107886. https://doi.org/10.1016/j.agrformet.2019.107886.

Seguini, L., Vrieling, A., Meroni, M., Nelson, A., 2024. Annual winter crop distribution from MODIS NDVI timeseries to improve yield forecasts for europe. Int. J. Appl. Earth Obs. Geoinformation 130, 103898. https://doi.org/10.1016/j.jag.2024.103898.

Shaw, P., Uszkoreit, J., Vaswani, A., 2018. Self-attention with relative position representations. Doi: 10.48550/arXiv.1803.02155.

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. https://doi.org/10.1016/j.rse.2017.04.026.

Tian, H., Wang, P., Tansey, K., Wang, J., Quan, W., Liu, J., 2024. Attention mechanism-based deep learning approach for wheat yield estimation and uncertainty analysis from remotely sensed variables. Agric. For. Meteorol. 356, 110183. https://doi.org/10.1016/j.agrformet.2024.110183.

Trnka, M., Rötter, R.P., Ruiz-Ramos, M., Kersebaum, K.C., Olesen, J.E., Žalud, Z., Semenov, M.A., 2014. Adverse weather conditions for european wheat production will become more frequent with climate change. Nat. Clim. Chang. 4, 637–643. https://doi.org/10.1038/nclimate2242.

van der Velde, M., Biavetti, I., El-Aydam, M., Niemeyer, S., Santini, F., van den Berg, M., 2019a. Use and relevance of european union crop monitoring and yield forecasts. Agr. Syst. 168, 224–230. https://doi.org/10.1016/j.agsy.2018.05.001.

van der Velde, M., van Diepen, C.A., Baruth, B., 2019b. The European crop monitoring and yield forecasting system: celebrating 25 years of JRC MARS bulletins. Agr. Syst. 168, 56–57. https://doi.org/10.1016/j.agsy.2018.10.003.

van der Velde, M., Nisini, L., 2019. Performance of the Mars-crop yield forecasting system for the european union: assessing accuracy, in-season, and year-to-year improvements from 1993 to 2015. Agr. Syst. 168, 203–212. https://doi.org/10.1016/j.agsy.2018.06.009.

von Bloh, M., Seiler, B., Van Der Smagt, P., Asseng, S., 2024. Explaining decision structures and data value for neural networks in crop yield prediction. Environ. Res. Lett. https://doi.org/10.1088/1748-9326/ad959f.

Wang, E., Martre, P., Zhao, Z., Ewert, F., Maiorano, A., Rötter, R.P., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Aggarwal, P. K., Anothai, J., Basso, B., Biernath, C., Cammarano, D., Challinor, A.J., De Sanctis, G., Doltra, J., Dumont, B., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.-K., Liu, L., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ripoche, D., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K., Wallach, D., Wang, Z., Wolf, J., Zhu, Y., Asseng, S., 2017. The uncertainty of crop yield projections is reduced by improved temperature response functions. Nat. Plants 3, 1–13. https://doi.org/10.1038/nplants.2017.102.

Wang, Y., Yu, Q., Liu, Z., Ren, W., Lu, X., 2025. A practical SIF-based crop model for predicting crop yields by quantifying the fraction of open PSII reaction centers (qL). Remote Sens. Environ. 320, 114658. https://doi.org/10.1016/j.rse.2025.114658.

Wang, Y.-Q., Leng, P., Shang, G.-F., Zhang, X., Li, Z.-L., 2023. Sun-induced chlorophyll fluorescence is superior to satellite vegetation indices for predicting summer maize yield under drought conditions. Comput. Electron. Agric. 205, 107615. https://doi.org/10.1016/j.compag.2023.107615.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: a meta-review. Remote Sens. Environ. 236, 111402. https://doi.org/10.1016/j.rse.2019.111402.

Xiong, X., Zhong, R., Tian, Q., Huang, J., Zhu, L., Yang, Y., Lin, T., 2024. Daily DeepCropNet: a hierarchical deep learning approach with daily time series of vegetation indices and climatic variables for corn yield estimation. ISPRS J. Photogramm. Remote Sens. 209, 249–264. https://doi.org/10.1016/j.isprsjprs.2024.02.008.

Xiong, X., Zhong, R., Jiang, H., Athanasiadis, I., Yang, Y., Zhu, L., Lin, T., 2026. Corn yield estimation under extreme climate stress with knowledge-encoded deep learning. ISPRS J. Photogramm. Remote Sens. 231, 101–118. https://doi.org/10.1016/j.isprsjprs.2025.10.020.

Zachow, M., Kunstmann, H., Miralles, D.J., Asseng, S., 2024. Multi-model ensembles for regional and national wheat yield forecasts in argentina. Environ. Res. Lett. 19, 84037. https://doi.org/10.1088/1748-9326/ad627c.

Yang, Q., Liu, L., Zhou, J., Ghosh, R., Peng, B., Guan, K., Tang, J., Zhou, W., Kumar, V., Jin, Z., 2023. A flexible and efficient knowledge-guided machine learning data assimilation (KGML-DA) framework for agroecosystem prediction in the US midwest. Remote Sens. Environ. 299, 113880. https://doi.org/10.1016/j.rse.2023.113880.

Zhang, H., Xu, H., Tian, X., Jiang, J., Ma, J., 2021. Image fusion meets deep learning: a survey and perspective. Inf. Fusion 76, 323–336. https://doi.org/10.1016/j.inffus.2021.06.008.

Zhang, L., Li, C., Zhang, G., Wu, X., Zhou, L., Chen, L., Jiao, Y., Liu, G., Hei, W., 2025a. Winter wheat yield estimation based on multisource remote sensing data: a dual-branch TCN-transformer model and analysis of growth-stage feature transition mechanisms. Comput. Electron. Agric. 239, 111014. https://doi.org/10.1016/j.compag.2025.111014.

Zhang, J., Guan, K., Chen, Z., Huang, Y., Zhao, K., Peng, B., Wang, S., Wu, X., Wang, S., Banerjee, A., Vergopolan, N., Fu, R., Zhao, S., Colussi, J., 2025b. Transfer learning for improved crop yield predictions in a cross-scale pathway: a case study for brazilian national soybean. Int. J. Appl. Earth Obs. Geoinf. 145, 104981. https://doi.org/10.1016/j.jag.2025.104981.

Zhang, Y., Joiner, J., Alemohammad, S.H., Zhou, S., Gentine, P., 2018. A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks. Biogeosciences 15, 5779–5800. https://doi.org/10.5194/bg-15-5779-2018.

Zhang, Z., Zhang, Y., Zhang, Y., 2023. Generating high-resolution total canopy SIF emission from TROPOMI data: algorithm and application. Remote Sens. Environ. 295, 113699. https://doi.org/10.1016/j.rse.2023.113699.

Zheng, H., Cheng, T., Yao, X., Deng, X., Tian, Y., Cao, W., Zhu, Y., 2016. Detection of rice phenology through time series analysis of ground-based spectral index data. Field Crops Res. 198, 131–139. https://doi.org/10.1016/j.fcr.2016.08.027.

Zhong, R., Zhu, Y., Wang, X., Li, H., Wang, B., You, F., Rodríguez, L.F., Huang, J., Ting, K. C., Ying, Y., Lin, T., 2022. Detect and attribute the extreme maize yield losses based on spatio-temporal deep learning. Fundam. Res. Doi: 10.1016/j.fmre.2022.05.006.

Zhong, R., Xiong, X., Tian, Q., Huang, J., Zhu, L., Yang, Y., Lin, T., 2025. PSNet: a deep learning framework following hierarchical yield level concept for crop yield estimation. Comput. Electron. Agric. 239, 110917. https://doi.org/10.1016/j.compag.2025.110917.

Zhou, J., Wang, D., Nezhad kheirollah, S., Maroufpoor, S., Band, S.S., 2023a. Sensitivity analysis of wheat yield based on growing degree days in different growth stages: application of machine learning approach enhanced by grey systems theory. Comput. Electron. Agric. 210, 107876. https://doi.org/10.1016/j.compag.2023.107876.

Zhou, W., Guan, K., Peng, B., Margenot, A., Lee, D., Tang, J., Jin, Z., Grant, R., DeLucia, E., Qin, Z., Wander, M.M., Wang, S., 2023b. How does uncertainty of soil organic carbon stock affect the calculation of carbon budgets and soil carbon credits for croplands in the U.S. midwest? Geoderma 429, 116254. https://doi.org/10.1016/j.geoderma.2022.116254.

Zhuang, H., Zhang, Z., Cheng, F., Han, J., Luo, Y., Zhang, L., Cao, J., Zhang, J., He, B., Xu, J., Tao, F., 2024. Integrating data assimilation, crop model, and machine learning for winter wheat yield forecasting in the north China plain. Agric. For. Meteorol. 347, 109909. https://doi.org/10.1016/j.agrformet.2024.109909.