ARTICLE IN PRESS

AIIA-00352; No of Pages 23

Artificial Intelligence in Agriculture xxx (xxxx) xxx

KeAi
CHINESE ROOTS
GLOBAL IMPACT

Contents lists available at ScienceDirect

Artificial Intelligence in Agriculture

journal homepage: http://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/

CropFusionNet: an interpretable deep learning framework for uncertainty-aware crop yield forecasting across Germany

Amit Kumar Srivastava [a,1], Krishnagopal Halder [a,*,1], Gina Lopez [b], Kaushik Muduchuru [a], Luis Alfredo Pires Barbosa [a], Kazi Jahidur Rahaman [a], Dominik Behrend [b], Liangxiu Han [c], Claas Nendel [a,d,e], Gang Zhao [f], Thomas Gaiser [b], Manmeet Singh [g], Karthikeyan Lanka [h], Jingye Han [i], Ioannis N. Athanasiadis [i], Michael Maerker [a], Wenzhi Zeng [j], Karam Alsafadi [k], Jaber Rahimi [l,m], Frank Ewert [a,b]

[a] *Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Str. 84, 15374 Müncheberg, Germany*
[b] *Institute of Crop Science and Resource Conservation, University of Bonn, Bonn 53111, Germany*
[c] *School of Computing, and Mathematics, Faculty of Science and Engineering, Manchester Metropolitan University, John Dalton Building, Chester Street, Manchester M1 5GD, UK*
[d] *Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany*
[e] *Global Change Research Institute of the Czech Academy of Sciences, Brno, Czech Republic*
[f] *State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Northwest A&F University, Yangling 712100, China*
[g] *Earth, Environmental and Atmospheric Sciences, Western Kentucky University, Bowling Green, KY, United States of America*
[h] *Centre of Studies in Resources Engineering, Indian Institute of Technology (IIT), Mumbai, India*
[i] *Wageningen University and Research, Wageningen, the Netherlands*
[j] *College of Agricultural Science and Engineering, Hohai University, No 8, West Focheng Road, Nanjing, Jiangsu Province, China*
[k] *Key Laboratory of Marine Environmental Science, College of the Environment and Ecology, Xiamen University, Xiamen 361102, China*
[l] *Pioneer Center Land-CRAFT, Department of Agroecology, Aarhus University, Aarhus, Denmark*
[m] *Institute of Meteorology and Climate Research, Atmospheric Environmental Research (IMK-IFU), Karlsruhe Institute of Technology (KIT), Garmisch-Partenkirchen, Germany*

ARTICLE INFO

Article history:
Received 17 April 2026
Received in revised form 1 July 2026
Accepted 24 August 2026
Available online xxxx

Keywords:
Crop yield forecasting
Explainable Artificial Intelligence (XAI)
Temporal Fusion Transformer
Climate extremes
Uncertainty-aware forecasting

ABSTRACT

Escalating climate fluctuations and the increasing frequency of compound extreme weather events pose severe threats to global food security. Current operational crop yield forecasting systems, which predominantly rely on process-based models or traditional statistical approaches, often underestimate yield losses during climatic extremes due to their inability to capture complex, non-linear climate-yield dynamics. While deep learning (DL) offers a transformative alternative, its widespread adoption remains limited by its "black-box" nature and the difficulty of processing long agro-meteorological time series without losing critical signals from climatic anomalies. To address these gaps, we propose CropFusionNet, a novel and interpretable architecture inspired by the Temporal Fusion Transformer (TFT), designed for district-level (NUTS-3) crop yield forecasting. CropFusionNet integrates daily time-varying climatic variables with static agro-environmental covariates and offers a more interpretable deep learning framework, enabling attribution of predictions to key environmental drivers, though not fully mechanistic. Applied to Germany's principal crop commodities, CropFusionNet consistently outperformed established deep learning baselines and operational frameworks such as MARS and ABSOLUT. It achieved high predictive accuracy for silage maize ($R^2 = 0.75$; MAPE = 8.66%), winter wheat ($R^2 = 0.60$; MAPE = 8.06%), and a moderate accuracy for winter barley ($R^2 = 0.45$; MAPE = 10%), successfully capturing inter-annual variability and accurately reproducing extreme negative yield anomalies during severe drought years (2003, 2018, and 2022). Lead-time analyses further demonstrated that reliable forecasts for winter cereals (wheat and barley) can be made 40–60 days before harvest, whereas spring crops (silage maize) can be predicted up to 70 days in advance. By disentangling key environmental drivers and effectively representing non-linear system dynamics, CropFusionNet represents a fundamental transformation towards scalable, accurate, uncertainty-aware, and interpretable agricultural forecasting, providing essential support for proactive food security management. The model implementation is available at https://github.com/geonextgis/CropFusionNet.

* Corresponding author.
E-mail address: Krishnagopal.Halder@zalf.de (K. Halder).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.aiia.2026.08.016

1. Introduction

Climate fluctuations account for approximately one-third of the variation in global crop yields, with projections indicating an increasing frequency and intensity of extreme weather events that threaten food system stability (Dhillon et al., 2025). The catastrophic summers of 2003 and 2018 exemplify these emerging threats: concurrent high temperatures and precipitation deficits resulted in severe yield losses across Europe for diverse crop species (Strer et al., 2018; Ostberg et al., 2018). For key crops grown across Germany, including winter wheat, winter barley, and maize, compound dry and hot events have emerged as dominant risk factors, often exceeding the cumulative impact of individual stressors (Strer et al., 2018). The 2016 European crop failures, caused by compound warm temperatures and excessive rainfall in some regions while others experienced severe drought, further illustrate the spatial and crop-specific heterogeneity of climate impacts (Webber et al., 2020). Therefore, developing robust forecasting systems capable of anticipating these climatic shocks for multiple crop species represents a critical prerequisite for regional food security, agricultural sustainability, and economic stability (Schmidt and Felsche, 2024).

Current operational crop yield forecasting systems in Europe rely predominantly on process-based crop models (PBCMs) or statistical regression approaches (Maestrini and Basso, 2018; Pagani et al., 2019; van der Velde and Nisini, 2019; Ziliani et al., 2022). The MARS Crop Yield Forecasting System (MCYFS) combines process-based simulations (primarily WOFOST and WARM), meteorological observations, and remote sensing indicators to produce regional forecasts (van der Velde et al., 2019; Baruth et al., 2007). While offering biophysical interpretability, PBCMs face substantial challenges in calibration and parameter transferability across diverse agroecological zones and management regimes (Leng and Hall, 2020; Xiong et al., 2024). More critically, recent assessments indicate that PBCMs systematically underestimate yield losses during extreme climatic events, such as Germany's 2018 drought, due to insufficient representation of compound and interacting stress factors (Becker et al., 2025). The district-level statistical model ABSOLUT v1.2 represents an alternative approach, evaluating time-aggregated meteorological predictors within multiple linear regression models (Conradt, 2022). While this approach is scalable and regionally specific, it requires predefined functional forms (e.g., quadratic temperature terms), which limits its ability to capture emergent nonlinear interactions between weather variables and crop-specific physiological processes. Furthermore, ABSOLUT does not systematically integrate high-resolution soil data, satellite-derived vegetation indicators, or modern nonlinear learning algorithms.

Machine learning approaches demonstrated significant improvements over traditional methods, with random forests and ensemble methods achieving R^2 values between 0.50 and 0.70 for major crops at regional scales (Lischeid et al., 2022; Jeong et al., 2016). Deep learning (DL) has emerged as particularly useful for modelling complex environmental interactions and crop-specific responses (You et al., 2017). Convolutional Neural Networks (CNNs) excel at spatial pattern recognition in satellite imagery, while Long Short-Term Memory (LSTM) networks capture cumulative physiological effects of weather variability across growing seasons (Kamilaris & Prenafeta-Boldú, 2018; Rashid et al., 2021; Zachow et al., 2025). However, recurrent architectures struggle with very long sequences, leading many studies to rely on temporal aggregation of daily time series. This approach risks smoothing the impact of short-term extreme events (e.g., heatwaves, drought spells) particularly when rapid weather fluctuations coincide with critical phenological stages, thereby disproportionately affecting final crop yield. Transformer-based architectures address these constraints by using self-attention mechanisms to model long-range dependencies without recursive state propagation, enabling efficient encoding of both local and global temporal interactions (Xiong et al., 2024; Vani and Guruprakash, 2024). Recent advances in multi-modal learning further enable integration of diverse data sources, including optical satellite imagery, weather data, soil properties, and digital elevation models, to improve prediction accuracy (Bach et al., 2015; Overweg et al., 2021).

Despite these advances, critical limitations persist. Most existing deep learning approaches treat temporal meteorological data and static environmental variables as separate input streams, concatenating or averaging them without preserving their structural relationships (Sun et al., 2019). This fails to capture complex interactions between dynamic weather patterns and static landscape characteristics that fundamentally drive crop growth processes. More fundamentally, the black box nature of deep learning limits adoption by agricultural stakeholders who require transparent predictions for decision-making (Hassija et al., 2024; Liakos et al., 2018). This challenge is compounded by the equivocality problem identified by Lischeid et al. (2022), which shows that multiple machine learning models can achieve nearly identical performance using entirely different feature sets and functional relationships. This ambiguity undermines confidence in predictions and highlights that "science cannot be delegated to artificial intelligence alone" and that "expert judgment is indispensable for selection of the most plausible and reliable model" (Lischeid et al., 2022). While recent advances in Explainable Artificial Intelligence (XAI), including SHAP, Regression Activation Mapping, and attention-based mechanisms, enable better understanding of model behaviour (Lundberg & Lee, 2017; Ribeiro et al., 2016; Filippi et al., 2024; Vani and Guruprakash, 2024), XAI applications for interpreting yield failures across Germany's diverse soil and climate zones remain largely unexplored. Moreover, existing studies rarely investigate optimal lead times for yield forecasting, often focusing on end-of-season predictions that offer limited practical value (Crane-Droesch, 2018). The critical early warning window, typically 1 to 4 months before harvest, remains poorly characterized despite its importance for crop insurance, market preparation, and food security planning. Comprehensive comparisons against established operational systems like MARS or a system "ABSOLUT" developed at PIK (Potsdam Institute for Climate Impact Research, Germany) also remain scarce (Fritz et al., 2019), limiting understanding of whether deep learning truly advances operational capabilities.

To address these limitations, we develop CropFusionNet, a novel Temporal Fusion Transformer (TFT)-inspired architecture designed for diverse agro-environmental inputs and interpretable crop yield forecasting (Lim et al., 2021). The model combines LSTM-based sequential encoding with multi-head self-attention to capture short-range and long-range temporal dependencies. Static and time-varying inputs are integrated through dedicated Variable Selection Networks (VSNs) and Gated Residual Networks (GRNs), which dynamically weight features and generate interpretable importance scores for key environmental drivers. This design provides transparency into how temporal climate dynamics and static characteristics influence crop-specific yield forecasts. We apply this architecture to district-level (NUTS-3) forecasting across Germany's principal arable crops: winter wheat (27% of total agricultural area), winter barley (10%), and silage maize (18%) (Blickensdörfer et al., 2022). These crops differ substantially in phenology, rooting depth, and physiological sensitivity to environmental stress. Maize is particularly sensitive to heat and drought stress during mid-summer grain filling, whereas winter cereals show critical vulnerability during spring stem elongation and grain development (Vanongeval and Gobin, 2023; Qi et al., 2022).

This research addresses four specific questions: (1) How does a deep learning architecture designed for complete integration of daily meteorological time series and static environmental data improve district-level yield forecast accuracy and generalizability compared to conventional fusion methods? (2) Does CropFusionNet demonstrate superior robustness and predictive skill compared to state-of-the-art deep learning baselines (LSTM, CNN, Transformer) and crop yield forecasting systems like MARS and ABSOLUT, across diverse climatic conditions and crop types? (3) To what extent do XAI mechanisms provide meaningful and agronomically plausible insights into feature contributions, particularly under extreme weather conditions? (4) What is the optimal forecast lead time, and how does this early warning window differ between winter and spring crops?

2. Materials and methods

2.1. Data

2.1.1. Crop yield data

District-level yield observations were obtained from the harmonized dataset developed by Duden et al. (2024), which provides annual crop yield records at the NUTS-3 administrative level (397 districts) for the period 1979–2021. These yield estimates are based on surveys conducted on representative farms within each district. The compiled dataset reports yield in tonnes per hectare (t ha^{-1}) and covers crops that together represent approximately 80% of Germany's arable land, corresponding to about 9.5 million hectares annually. To extend the temporal coverage, we incorporated additional observations from recent years using publicly available agricultural statistics available online at Regionalstatistik (https://www.regionalstatistik.de/genesis/online, **Table S1**), extending the dataset to 2023 for winter wheat, winter barley, and silage maize. However, due to the availability constraints of other data sources, particularly remote sensing products, the analysis in this study was restricted to the period 2001–2023.

2.1.2. Time-varying covariates

Daily meteorological and soil covariates (e.g., soil moisture and soil temperature) were obtained at a spatial resolution of 1 × 1 km from the German Weather Service (Deutscher Wetterdienst, DWD). The dataset includes standard meteorological variables such as daily precipitation sums (PREC), sunshine duration (SUND), solar radiation (RAD), and minimum (TMIN), mean (TMEAN), and maximum (TMAX) air temperatures (Table S1). In addition, the dataset contains modelled daily soil moisture (SM), expressed as the percentage of plant-available water, derived using the AMBAV model (Friesland and Löpmeier, 2007). Daily soil temperature (ST) at a depth of 5 cm was also included, estimated using the AMBETI model, which simulates the heat and water exchange between the atmosphere, vegetation canopy, and soil (Braden and Wetterdienst, 1995). Furthermore, reference evapotranspiration (ET0) was calculated using the standardized Penman-Monteith equation (Howell and Evett, 2004). To capture atmospheric and hydrological stress conditions, additional agro-climatic indicators were derived, including Vapor Pressure Deficit (VPD) and Climatic Water Balance (CWB) (**Appendix A**).

Satellite-derived biophysical and vegetation indices were extracted via Google Earth Engine (GEE) to capture crop development dynamics. All remote sensing features were derived from MODIS (Moderate Resolution Imaging Spectroradiometer) Terra products. Specifically, the Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) were extracted from the MOD13Q1.061 product, which provides global 16-day composites at a 250 m spatial resolution. The Fraction of Photosynthetically Active Radiation (FPAR) and Leaf Area Index (LAI) were sourced from the MOD15A2H.061 product, offering 8-day composites at a 500 m spatial resolution (see **Table S1**).

2.1.3. Static covariates

To account for inherent environmental and management characteristics across regions, a suite of static covariates was incorporated and aggregated to the district-level. These variables represent geophysical conditions and irrigation infrastructure that influence crop productivity and resilience across Germany.

Soil suitability for agriculture was characterized using the Müncheberg Soil Quality Rating (SQR) dataset (Mueller et al., 2013), which provides a comprehensive assessment of soil productivity potential on a scale from 0 (unsuitable for agriculture) to 100 (prime farmland). Topographic variables (e.g., elevation and slope) were derived from the Copernicus DEM GLO-30 (Trevisani et al., 2023) to represent landscape structure. To account for human interventions in the hydrological cycle, we calculated the irrigated area fraction using the European Crop-specific IRrigated Area (ECIRA) dataset (Zhu et al., 2025) (**Appendix B**). This dataset provides high-resolution (1-km) annual estimates of the irrigated area for 16 specific crop types, including winter wheat, winter barley, and silage maize.

2.1.4. Ancillary data

Crop phenological information was obtained from the German Meteorological Service (DWD) phenological observation network. The DWD phenological database contains long-term observations of key crop development stages recorded by trained observers across Germany. The observations include phenophases such as sowing, emergence, flowering, and maturity for major crops. For this study, phenology records corresponding to winter wheat, winter barley, and silage maize were extracted and processed from DWD stations. These data provide spatially distributed estimates of crop developmental timing and were used to support feature construction for yield forecasting.

High-resolution crop type maps for Germany (2017–2021) were used to delineate crop-specific areas. For each target crop (winter wheat, winter barley, and silage maize), binary masks were generated using standardized crop identifiers from the High-Resolution Layer (HRL) Cropland product of the Copernicus Land Monitoring Service (CLMS).

2.2. Data preprocessing

To develop a standardized dataset suitable for crop yield forecasting, a comprehensive data preprocessing and harmonization pipeline was implemented. All spatial datasets were aggregated (mean) to the NUTS-3 administrative district-level to match the spatial resolution of the crop yield observations and were temporally restricted to the 2001–2023 study period. A key component of the preprocessing strategy involved aligning time-varying covariates with biologically meaningful crop growth stages. Station-level observations from the phenology data were quality-controlled (using QUALITAETS_BYTE values QB = 1 and 2), converted to day-of-year (DOY) values, and aggregated to derive representative crop specific district-level phenological events. Based on these phenological records, the median sowing and harvest months were determined for each crop over the entire study period (2001−2023). To capture pre-season environmental conditions, the temporal window was extended by three months prior to the median sowing month. The first day of this extended start month and the last day of the median harvest month were then used to define the maximum temporal boundaries (the fixed sequence length) for extracting time-varying predictors for each district-year combination present in the crop yield dataset (refer Section 2.1.1). Additionally, to account for interannual variability and extreme climate years that may force early maturity, year-specific masking was applied to the training, validation, and test datasets. Whenever the reported harvest date for a specific district-year occurred earlier than the median harvest month, all subsequent daily timesteps were masked. This prevents the model from erroneously incorporating irrelevant post-harvest environmental noise (e.g., bare soil signals) into the crop growth representations. Furthermore, the pre-planting period was also masked for remote sensing variables, as vegetation signals from previous cropping cycles are unlikely to influence the current growing season (Liu et al., 2019).

Satellite-derived vegetation indices were masked using crop-specific binary layers derived from the HRL Cropland product of the Copernicus Land Monitoring Service to ensure that the extracted signals represented the conditions of the target crops. Because long-term high-resolution crop type maps are not available for the entire study period, a maximum occurrence approach was applied across the available years (2017–2021) to generate a single master mask identifying pixels classified as the target crop at least once during this period. This mask was applied during the extraction of MODIS Terra products, together with standard quality assurance (QA) bitmasks to remove pixels affected by clouds or shadows. The filtered values of NDVI, EVI, FPAR, and LAI were subsequently spatially averaged to produce crop-specific remote sensing time series for each district. To harmonize these data with the daily temporal resolution of the meteorological variables, the remote sensing series were further processed to generate daily observations. First, a Savitzky-Golay smoothing filter was applied to reduce noise while preserving phenological patterns (Jönsson and Eklundh, 2004). The smoothed series were then resampled to daily resolution using cubic spline interpolation, producing continuous vegetation trajectories throughout the growing season. Remaining gaps at the temporal boundaries were filled using forward and backward filling to ensure complete daily records. The resulting NDVI, EVI, FPAR, and LAI time series were merged into a unified dataset for each district-year combination.

Data with incomplete observations after processing were excluded. In rare cases where satellite extraction failed entirely, typically in geographically small or highly heterogeneous districts with no valid MODIS pixels available for aggregation despite the existence of official yield records, or where phenological data were not recorded, data from the nearest neighbouring district were used as a spatial fallback to maintain dataset completeness. Daily meteorological, soil, and agro-climatic variables were spatially aggregated to the NUTS-3 district-level to match the spatial resolution of the crop yield observations. A similar post-harvest masking strategy was applied to these variables to exclude environmental conditions occurring after the crop growth period. Time-invariant static covariates were processed in a comparable manner to capture intra-district environmental heterogeneity. For the Müncheberg Soil Quality Rating (SQR) dataset and the Copernicus DEM GLO-30, both the mean and standard deviation were calculated for each district to represent average conditions and their spatial variability. In addition, annual irrigation estimates from the ECIRA dataset were aggregated to quantify the crop-specific fraction of irrigated land within each district.

Finally, both time-varying and static covariates were integrated into unified, crop-specific datasets. Time-varying covariates -including meteorological, soil, agro-climatic, and remote sensing variables -were aligned with phenology-defined growing seasons for each district-year combination to produce fixed-length sequential inputs suitable for deep learning models. Static covariates, such as soil quality, topographic features, and irrigated area fraction were appended to the corresponding district records to provide additional contextual information. A rigorous quality control procedure was then applied to identify and remove district-year samples with missing or inconsistent records across the different data sources, ensuring that only complete observations were retained.

For model development, the dataset was partitioned into training, validation, and test subsets. From the 23-year study period, approximately 80% of the years (2001–2018) were used as the training set, ~10% of the years (2019–2021) as the validation set, and the remaining 10% (2022—2023) as the test set. Importantly, standardization parameters were estimated exclusively from the training data and subsequently applied to the validation and test sets to prevent information leakage and ensure unbiased model evaluation (**Appendix C**).

2.3. CropFusionNet model

The proposed model, CropFusionNet, is an interpretable deep learning architecture designed for uncertainty-aware crop yield forecasting. The model is inspired by the Temporal Fusion Transformer (TFT) (Lim et al., 2021), which integrates attention mechanisms, gating structures, and variable selection for interpretable multi-horizon forecasting. CropFusionNet adapts and extends this framework to address agricultural forecasting scenarios, where crop productivity is driven by complex interactions between dynamic environmental conditions and site-specific static characteristics. Unlike the original TFT architecture, which is designed for multi-horizon time-series forecasting, CropFusionNet is reformulated as a regression model to predict end-of-season crop yield. Consequently, several architectural components are modified to better suit this single-target prediction task while retaining the interpretability advantages of the TFT framework. The model receives two distinct types of data as input: static covariates (e.g., topographic variables and soil quality) that describe time-invariant characteristics of the landscape, and daily time-varying covariates (e.g., meteorological variables and remotely sensed vegetation indices) that capture the temporal evolution of crop growth and environmental conditions throughout the growing season. The final output is an uncertainty-aware, quantile-based crop yield estimate, providing both the most likely estimate and the predictive uncertainty. Overall, CropFusionNet consists of the following main components: feature embedding, variable selection networks, temporal representation learning, temporal pooling, and uncertainty-aware quantile head for crop yield forecasting. A schematic overview of the proposed architecture is provided in Fig. 1.

2.3.1. Feature embedding

CropFusionNet transforms heterogeneous input variables into a unified latent representation through a feature embedding module. Since the model supports both static and temporal variables with different data types (numerical and categorical), embedding layers are used to project these inputs into a common feature space. For static variables, two types of embeddings are applied. Static categorical variables are mapped to dense vector representations using learnable embedding matrices or lookup tables, enabling the model to represent discrete categories within a continuous latent space. Static numerical variables are projected into the same embedding space using linear transformations. Each numerical variable is passed through a fully connected layer that maps the scalar input to a vector of dimension d_{model}, where d_{model} denotes the embedding dimension. The resulting embedded vectors are concatenated to form the static feature representation. Similarly, time-varying variables are embedded at each timestep to represent the temporal evolution of environmental conditions and crop development. Time-varying categorical variables are encoded using embedding layers applied independently at each timestep, while time-varying numerical variables are transformed using linear projection layers. To ensure consistent processing across temporal sequences, these transformations are applied in a time-distributed manner so that each timestep is embedded independently while preserving the sequential structure of the data. The embedded temporal features are subsequently concatenated across variables to produce a sequence of feature vectors representing the environmental and crop conditions throughout the growing season. By transforming all features into uniform d_{model}-dimensional vectors, the network can seamlessly integrate and propagate both static and time-varying covariates into the subsequent modules without dimensionality conflicts.

2.3.2. Variable selection networks

In agricultural modelling, the importance of meteorological and remote sensing variables can vary substantially across different crop phenological stages and site-specific characteristics (Zeng et al., 2025). Consequently, not all input variables contribute equally to crop yield prediction at every stage of the growing season. To adaptively suppress irrelevant noise while retaining the most informative features and corresponding timesteps, CropFusionNet incorporates a Variable Selection Network (VSN). This module is constructed using two key components: the Gated Linear Unit (GLU) and the Gated Residual Network (GRN).

2.3.2.1. Gated linear unit (GLU). GLU acts as an efficient suppression mechanism, allowing the model to dynamically control the flow of information. It determines which parts of the input feature representations are relevant and which should be filtered out. Given an input x, GLU applies two independent linear transformations: one acts as the

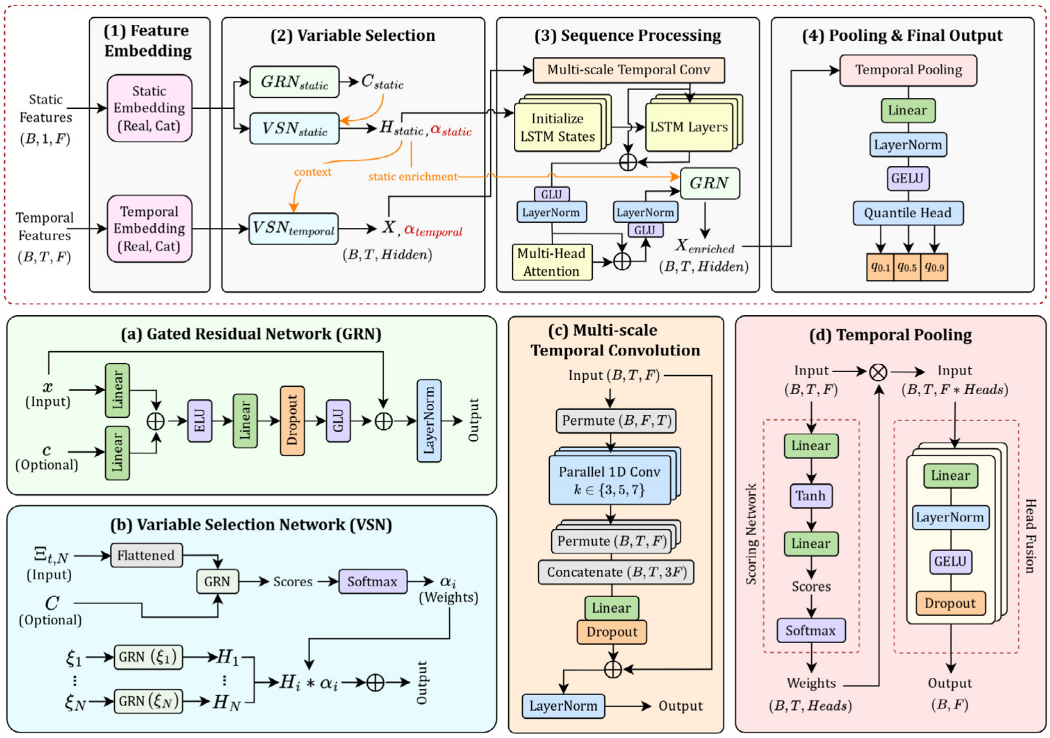

Fig. 1. Architecture of CropFusionNet for uncertainty-aware quantile-based crop yield forecasting. The model consists of four main stages: (1) Feature embedding, where static and temporal inputs are transformed into latent representations; (2) Variable selection, where Gated Residual Networks (GRNs) and Variable Selection Networks (VSNs) identify the most relevant static and time-varying features; (3) Sequence processing, which captures short- and long-term temporal dependencies using multi-scale temporal convolutions, LSTM layers initialized from static context, and multi-head self-attention; and (4) Pooling and final output, where attention-based pyramidal pooling aggregates the temporal sequence into a compact representation that is used to generate quantile-based crop yield predictions through a quantile prediction head. The lower panels illustrate the internal structures of the key modules: (a) GRN, (b) VSN, (c) multi-scale temporal convolution, and (d) the temporal pooling mechanism. Orange arrows indicate the flow of contextual information from static features into temporal processing, while α_{static} and $\alpha_{temporal}$ denote the interpretable variable selection weights learned for static and time-varying covariates, respectively. (Note: Real = Numerical, Cat = Categorical).

primary feature projection, while the other acts as a gating signal scaled between 0 and 1 via a Sigmoid activation (σ). This operation can be expressed as:

$$GLU(x) = \sigma(W_1 x + b_1) \odot (W_2 x + b_2)$$

where $\odot$ denotes element-wise multiplication. This ensures that irrelevant features are suppressed (multiplied by a value close to 0) while important features are passed through (multiplied by a value close to 1). Here, W denotes the learnable weight matrices associated with the linear transformations, while b represents their corresponding bias terms.

2.3.2.2. Gated residual network (GRN). Building upon the gating mechanism, the GRN serves as the primary feed-forward module within CropFusionNet. It is designed to provide complex, non-linear processing while ensuring stable gradient flow during training via residual connections. Operationally, the GRN first projects the input feature vector through a primary linear transformation. If contextual information is available–such as static embeddings–it is independently projected into the same latent space and added to the transformed input. This mechanism allows the network to condition temporal representations on static contextual features. The resulting representation is then passed through an Exponential Linear Unit (ELU) activation function (Shah et al., 2016) to introduce nonlinearity. Following the activation, a second linear projection is applied, and dropout is used for regularization to reduce overfitting. To dynamically filter the flow of information, the transformed output is passed through a GLU, which controls how much of the transformed signal should be propagated forward. Finally, a residual shortcut adds the original input back to the transformed representation–using a linear projection when the input and output dimensions differ–and the output of the block is stabilized using Layer Normalization. Mathematically, the overall transformation can be expressed as:

$$GRN(x, c) = LayerNorm\left(x_{skip} + GLU\left(Dropout\left(W_2\left(ELU\left(W_1 x + W_c c\right)\right)\right)\right)\right)$$

where x represents the primary input, c denotes the optional context vector, and x_{skip} represents the residual connection.

2.3.2.3. Variable Selection Network (VSN). Building upon these foundational components, the Variable Selection Network (VSN) employs GRNs within a dual-pathway architecture to simultaneously estimate the relative importance of input variables and enrich their feature representations. In CropFusionNet, this mechanism is implemented in two sequential stages. The first stage identifies the most informative static covariates, while the second stage filters the time-varying covariates conditioned on the learned static context.

Let the set of static variable embeddings be denoted as Ξ_{static}. To capture the overall static context, the concatenated static embeddings are first passed through a dedicated GRN to produce a foundational context vector C_{static}. The VSN is then applied to the static variable embeddings using C_{static} as its own conditioning context. Following the general VSN formulation, the network computes sparse selection weights α_{static}, which quantifies the relative importance of each static feature. In parallel, each static embedding is individually transformed through a dedicated GRN. The final static representation H_{static} is obtained by aggregating these transformed features using the learned selection weights, producing a compact vector that captures the most relevant static information for subsequent temporal processing. These operations can be expressed as:

$$C_{static} = GRN_{static}(\Xi_{static})$$

$$H_{static} = VSN_{static}(\Xi_{static}, C_{static})$$

Once the most salient static features are distilled into H_{static}, this vector serves as the conditioning context for the temporal sequence. At a given timestep t, let the set of time-varying input features be denoted as $\Xi_{t,N}$, which consists of N individual variable embeddings $\xi_1, \ldots, \xi_N$. To compute temporal variable selection weights, these embeddings are first concatenated into a single flattened vector $\Xi_{t,N}$. The static representation H_{static} is then used as a contextual input, allowing the network to integrate static information into the temporal selection process. Both the flattened temporal inputs and the static context are passed through a global scoring GRN to produce raw importance scores for each variable. To ensure that the model does not learn from invalid timesteps, a masking operation is applied to the raw scores by replacing the corresponding values with $-\infty$. When the scores are subsequently passed through a Softmax activation, this operation forces the weights of invalid timesteps to zero, while the valid timesteps produce normalized selection weights $\alpha_{t,i} \in [0, 1]$ that sum to 1 across all input variables:

$$\alpha_t = Softmax(Mask(GRN_{score}(\Xi_{t,N}, H_{static})))$$

Concurrently, in the second pathway, each individual variable embedding ξ_i is processed through its own independent GRN. This ensures that the network learns feature-specific, non-linear transformations without inference from other variables, resulting the transformed representations $H_{t,i}$. Finally, the temporal VSN aggregates the information by taking the weighted sum of these individually transformed features. Each transformed representation $H_{t,i}$ is scaled by its corresponding selection weight $\alpha_{t,i}$, and the results are summed across the feature dimension:

$$\text{Output}_t = \sum_{i=1}^{N} \alpha_{t,i} H_{t,i}$$

This architecture produces a single, consolidated hidden representation for each timestep. By dynamically scaling the feature representations before sequence processing, the VSN explicitly suppresses irrelevant noise and ensures the downstream recurrent and attention mechanisms focus exclusively on the most crop yield-limiting factors at any given point in the season. Furthermore, extracting both α_{static} and α_t provide inherent model interpretability, allowing stakeholders to quantify which variables drove the crop yield prediction at specific phenological stages.

2.3.3. Temporal representation learning

Following variable selection, CropFusionNet employs a hierarchical temporal processing module designed to capture both short and long-term temporal dependencies throughout the crop growing season. This module consists of multi-scale temporal convolutions, static-conditioned recurrent modelling, and self-attention mechanisms.

2.3.3.1. Multi-scale temporal convolution. Meteorological time-series data often exhibit short-term anomalies, such as short heatwaves or sudden precipitation spikes, which can profoundly impact final crop yield outcome (Vogel et al., 2019). To capture such localized temporal events, the hidden temporal representations produced by the variable selection networks are first processed using a lightweight multi-scale 1D convolution block. The module applies parallel 1D convolution with varying kernel sizes ($k \in \{3, 5, 7\}$) to capture different localized receptive fields. The outputs of these parallel convolutions are concatenated along the feature dimension and linearly projected back to the original input hidden dimension. Finally, dropout is applied, and the result is added to the original input via a residual connection and normalized. Mathematically, for an input sequence X, the operation is defined as:

$$C_k = Conv1D_k(X)$$

$$X_{conv} = LayerNorm(X + Dropout(W_{proj}[C_3, C_5, C_7]))$$

where $[C_3, C_5, C_7]$ denotes the concatenation of the multi-scale convolutional outputs.

2.3.3.2. Sequence processing with LSTM. To model the sequential dynamics of crop development, such as the cumulative influence of temperature, precipitation, or vegetation growth over time, CropFusionNet employs Long Short-Term Memory (LSTM) network layers (Sherstinsky, 2020). Unlike conventional recurrent architectures that initialize the hidden states with zeros, the proposed architecture conditions the temporal recurrence on the final static representation H_{static} obtained from VSN_{static}. This design allows the network to adapt its temporal dynamics according to site-specific static characteristics, such as soil quality, management practices, or topographic variability. The initial hidden state h_0 and cell state c_0 of the LSTM are generated via linear projections of the static representation vector H_{static}:

$$h_0 = W_h H_{static} + b_h$$

$$c_0 = W_c H_{static} + b_c$$

The temporally enriched sequence X_{conv}, is processed through the LSTM layers, and the output is refined using a GLU and skip connection to facilitate stable gradient flow:

$$X_{lstm} = LSTM(X_{conv}, (h_0, c_0))$$

$$X_{lstm} = LayerNorm(GLU(X_{lstm} + X_{conv}))$$

2.3.3.3. Self-attention and static enrichment. While LSTMs excel at modelling sequential data, their ability to capture long-range dependencies–such as the relationship between a drought during planting and heat stress during harvesting–can degrade over long sequences. To address this, CropFusionNet employs a Multi-Head Self-Attention mechanism (Voita et al., 2019). First, standard sinusoidal positional encodings (PE) are added to the sequence to ensure the attention mechanism retains the absolute chronological order of the timesteps (Vaswani et al., 2017). The sequence is then processed through multiple attention heads:

$$X_{attn} = MultiHeadAttention(X_{lstm} + PE)$$

Similar to the recurrent layer, the attention output is regularized using a GLU-based skip connection and Layer Normalization.

Finally, to ensure that the globally attended temporal representation remains heavily influenced by the underlying site-specific static covariates, the sequence undergoes Static Enrichment. The attention outputs are passed through a GRN, using the H_{static} vector as the conditioning input:

$$X_{enriched} = GRN(X_{attn}, H_{static})$$

This enriched sequence efficiently fuses dynamic temporal environmental conditions with static spatial traits, creating a highly information-rich representation ready for final pooling and crop yield prediction.

2.3.4. Temporal pooling

After modelling both short and long-range temporal dependencies, the resulting temporal representations must be aggregated into a fixed-length vector suitable for final crop yield prediction. Conventional pooling techniques, such as mean or max pooling across the time dimension, may dilute the impact of short-term extreme events (Ding et al., 2019). To address this limitation, CropFusionNet learns to selectively aggregate information from the most informative periods of the growing season.

Let the final temporal representation after the attention and static enrichment stages be denoted as $X_{enriched} \in R^{T \times F}$, where T represents the number of timesteps and F denotes the hidden feature dimension. To determine the relative importance of each timestep, a lightweight scoring network computes attention scores for all timesteps across P parallel pooling heads simultaneously. This network consists of two linear transformations separated by a nonlinear activation function. The resulting attention score tensor is given by

$$S = W_{s2} \cdot tanh\ (W_{s_1} \cdot X_{enriched} + b_{s1}) + b_{s2} \quad where\ S \in R^{T \times P}$$

where W_{s1} and W_{s2} are learnable weight matrices, and b_{s1} and b_{s2} denote their corresponding bias terms. The scoring network thus produces temporal importance scores for each timestep across all P pooling heads. Before applying the Softmax normalization, a masking strategy–like the one used in the Variable Selection Network (Section 2.3.2.3)–is applied to ensure that invalid timesteps do not influence the aggregation. As a result, for each head, the weights are constrained to the interval $[0, 1]$ and sum to one across all timesteps.

$$A = Softmax(Mask(S))$$

The attention weights are then used to compute weighted temporal summaries of the sequence. For each head $p \in \{1, \dots, P\}$, a weighted sum of the temporal representations is calculated, resulting in multiple pooled representations $H_p \in R^F$ that capture different perspectives of the temporal dynamics:

$$H_p = \sum_{t=1}^{T} A_{t,p} \cdot X_t$$

These head-specific representations are subsequently concatenated into a single high-dimensional vector $Z \in R^{P \cdot F}$:

$$Z = [H_1, H_2, \dots, H_P]$$

The final pooling module progressively compresses the large, concatenated head representations through a sequence of linear transformations. Each intermediate stage is followed by layer normalization, GELU nonlinear activation, and dropout, which can be mathematically expressed as:

$$Z^{(l)} = Dropout\left(GELU\left(LayerNorm\left(W^{(l)} Z^{(l-1)} + b^{(l)}\right)\right)\right)$$

where $W^{(l)}$ and $b^{(l)}$ represent the learnable weights and biases for that specific layer. The pyramidal structure gradually reduces the dimensionality of the concatenated vector, enabling the network to retain the most informative components of the aggregated temporal representation while filtering redundant information.

2.3.5. Quantile head

Crop yield forecasting is inherently uncertain; therefore, deterministic point estimates often fail to capture the full range of variability required for effective agricultural risk management and informed decision-making. To address this limitation, CropFusionNet is formulated as a quantile-based predictive model that predicts multiple quantiles of the target crop yield distribution.

Following the temporal pooling, the final temporal representation $H_{final} \in R^F$ is passed through an output projection block. This block refines the aggregated feature representation before the final prediction by applying a linear transformation, followed by Layer Normalization and a GELU activation function:

$$H_{proj} = GELU(LayerNorm(W_{proj} H_{final} + b_{proj}))$$

where W_{proj} and b_{proj} represent the learnable weights and biases of the projection layer.

The enriched representation is then passed to a linear prediction head that produces Q target quantiles (e.g., the 10th, 50th, and 90th percentiles of the crop yield distribution). This mapping can be written as:

$$\widehat{Y} = W_q H_{proj} + b_q$$

where $W_q \in R^{Q \times F}$ and $b_q \in R^Q$ are the learnable parameters of the final prediction head. The resulting output $\widehat{Y} \in R^Q$ contains the predicted crop yield values for each specified quantile, providing an uncertainty-aware estimate of the expected crop yield distribution.

2.4. Baseline models

To benchmark the performance and architectural innovation of CropFusionNet, we compared it against three established deep learning baselines: a Vanilla Long Short-Term Memory (LSTM) network, a Simple Transformer, and a 1D Residual Convolutional Neural Network (ResCNN). To ensure a fair comparison, all baseline models share the identical feature embedding strategy as CropFusionNet–mapping categorical and numerical variables into a unified latent space and concatenating static covariates to every temporal step prior to sequence processing.

- **Vanilla LSTM:** Models temporal sequences recursively using standard LSTM cells, utilizing the final hidden state of the sequence for quantile prediction.
- **Simple Transformer:** Captures global temporal dependencies using standard multi-head self-attention layers (with positional encodings), applying global average pooling across the sequence length to generate the final quantile prediction.
- **ResCNN:** Extracts local temporal patterns using a stack of 1D convolutional residual blocks, followed by global average pooling.

Detailed architectural configurations for these baseline models are provided in **Appendix D**.

2.5. Experimental setting

All experiments were conducted on a high-performance computing cluster equipped with NVIDIA Tesla V100 GPUs (16 GB memory) using PyTorch 2.9.1 framework. To identify suitable model configurations for CropFusionNet, a random search hyperparameter optimization strategy was employed. In total, 100 independent training trials were executed across the three studied crops. The search space consisted of eight key hyperparameters: learning rate (10^{-3}, 10^{-4}, 10^{-5}), batch size (32, 64, 128), LSTM hidden dimension (128, 256, 512), LSTM layers (1,2,3), attention heads (4, 8, 16), pooling heads (4, 8, 16), embedding dimension (4, 8, 16), and dropout rate (0.3, 0.4, 0.5). These specific ranges were selected based on the optimization ranges used in Lim et al. (2021) and on preliminary exploratory runs to ensure stable model convergence without over-parameterization. However, the upper bounds for computationally expensive parameters, specifically the batch size (up to 128) and LSTM hidden dimensions (up to 512), were constrained by the 16 GB memory capacity of the available NVIDIA Tesla V100 GPUs to prevent out-of-memory errors during training. Each training trial randomly sampled one configuration from this predefined parameter space.

Model training was performed for a maximum of 500 epochs using the Adam optimizer with weight decay (10^{-5}) to regularize the model parameters and reduce overfitting. An early stopping strategy with a patience of 10 epochs was applied based on the validation loss, allowing training to terminate automatically when no further improvement was observed. The model was trained using quantile loss (pinball loss) to enable uncertainty-aware quantile-based crop yield prediction for the 0.1, 0.5, and 0.9 quantiles.

To better understand the relationship between sampled hyperparameter configurations and model performance, a Random Forest surrogate model was trained to approximate the mapping between hyperparameter settings and validation loss. This approach provides a model-agnostic way to estimate the relative importance of different hyperparameters within the explored search space. In addition, a Kruskal-Wallis non-parametric statistical test was applied to examine whether the distributions of validation losses differed across the categorical levels of each hyperparameter. These analyses were performed to support a systematic interpretation of the hyperparameter search process and to ensure that the final model configuration was selected in a transparent and reproducible manner. The final CropFusionNet configurations used in subsequent experiments were selected from the best-performing trials within the explored hyperparameter search space for each crop (**Appendix E**).

In addition, a lead-time forecasting evaluation was conducted to assess model performance under varying data availability conditions prior to harvest. For this analysis, input features were progressively truncated to simulate different forecasting horizons, ensuring that only information available up to each prediction date was used. Specifically, meteorological and remote sensing variables were restricted to observations preceding the forecast time, thereby preventing the use of future information and avoiding information leakage. This setup reflects a realistic operational forecasting scenario, where predictions must rely solely on in-season data available at the time of prediction. Furthermore, to quantify the statistical robustness of the reported point estimates given the relatively limited number of test years (2022–2023), we employed a bootstrap resampling approach. We generated 1000 independent bootstrap samples with replacement from the held-out test predictions. For each iteration, we computed the coefficient of determination (R^2) and the Root Mean Square Error (RMSE) to derive the 95% confidence intervals (defined by the 2.5th and 97.5th percentiles of the bootstrap distributions). This ensures the evaluated accuracy is stable across inter-annual and spatial sample variations.

Finally, to evaluate the model for agricultural early warning, continuous yield predictions were recast into a three-class categorical classification problem based on terciles of the observed yield distribution. For each crop, the 33rd and 67th percentile thresholds were derived from the observed yields of the combined held-out validation and test years (2019–2023). These observation-derived thresholds were applied to both the true observed yields and the model's predictions to assign each sample into Low, Normal, and High yield classes. Categorical predictive skill was then quantified using overall accuracy and macro-F1 scores.

3. Results

3.1. Model performance and benchmark comparisons

3.1.1. Inter-annual variability and temporal robustness

To obtain a clear and comprehensive assessment of model performance, the initial analysis focused on evaluating how well CropFusionNet captures the temporal fluctuations in crop yield over time. This provides a direct indication of the model's sensitivity to climate-driven variability, which strongly influences crop yield dynamics. This evaluation was conducted by comparing the model's median ($q_{0.5}$) predicted crop yields with observed district-level crop yields across the entire 2001–2023 period. As shown in Fig. 2, temporal distribution boxplots indicate that CropFusionNet effectively replicates both inter-annual variability and long-term trends for winter wheat, winter barley, and silage maize. The model demonstrates consistent generalization across time, maintaining consistent performance throughout the training (2001–2018), validation (2019–2021), and testing (2022–2023) periods. Importantly, the model successfully captured extreme drought events in 2003 and 2018 by predicting crop yield distributions lower than the average for all crops, and it continued to align closely with observed trends in the validation and test periods for example, slightly underestimating winter wheat yields and slightly overestimating silage maize yields in 2021, consistent with the actual observed distributions.

3.1.2. Comparison with deep learning baselines

To assess the architectural advantages of CropFusionNet, its predictive performance was benchmarked against three established deep learning models: a Vanilla LSTM, a Simple Transformer, and a 1D ResCNN. The results presented in Fig. 3 and Table 1 are based on the combined validation and test period (2019–2023), enabling a comprehensive comparison of model performance across unseen years. CropFusionNet consistently outperformed the baseline models across the majority of statistical metrics (**Appendix F**). For winter wheat, it achieved the highest predictive accuracy, with R^2 of 0.60 and a Pearson correlation coefficient (r) of 0.78, while substantially reducing prediction errors (RMSE = 0.74 t ha^{-1}, MAPE = 8.06%). In contrast, the Vanilla LSTM exhibited the lowest performance, with a higher MAPE (10.47%) and lower R^2 (0.32). Although the Simple Transformer and ResCNN demonstrated comparable performance, they consistently lagged behind CropFusionNet across all metrics. Predicting winter barley yields was more challenging for all models; nonetheless, CropFusionNet achieved the lowest errors (RMSE = 0.89 t ha^{-1}, MAPE = 10.1%) and the highest explained variance (R^2 = 0.44). While the ResCNN attained a slightly higher correlation (r = 0.70 compared to CropFusionNet's 0.68), its overall accuracy was on the lower side (R^2 = 0.34, RMSE = 0.97 t ha^{-1}). For silage maize, CropFusionNet again showed strong performance, reaching an R^2 of 0.75 and r of 0.87, with the lowest RMSE (4.46 t ha^{-1}) and MAPE (8.66%). The baseline models exhibited higher error margins, with the ResCNN producing the highest RMSE (5.77 t ha^{-1}) and MAPE (11.87%). To further evaluate the model's generalizability, we assessed CropFusionNet exclusively on the held-out test set (2022–2023) (**Fig. S7**). For winter wheat, the model achieved an R^2 of 0.61 (95% CI: 0.55–0.67), an RMSE of 0.76 t ha^{-1} (95% CI: 0.70–0.82), and a MAPE of 8.20%. While predicting winter barley yields proved more challenging across all models, CropFusionNet still yielded an R^2 of 0.20 (95% CI: 0.08–0.32), an RMSE of 0.98 t ha^{-1} (95% CI: 0.91–1.05), and a MAPE of 10.47%. For silage maize, the model demonstrated strong and highly stable performance, reaching an R^2 of 0.70 (95% CI: 0.64–0.76), with an RMSE of 4.65 t ha^{-1} (95% CI: 4.25–5.06) and a MAPE of 9.83%. The relatively narrow confidence intervals derived from bootstrap resampling confirm that the model's predictive performance is statistically robust and not unduly influenced by specific district-year anomalies within the test period.

Beyond point-prediction accuracy, CropFusionNet demonstrated its significant strength in capturing the underlying distribution of yield observations (**Fig. S1**). Kernel density estimates (KDEs) reveal that CropFusionNet closely reproduces the shape, central tendency, and variability of the observed yield distributions across all three crops. For winter wheat, the predicted density curve aligns almost perfectly with the observed data, producing a negligible bias (+0.05 t ha^{-1}) and a low Kolmogorov-Smirnov (KS) statistic (0.05), indicating no statistically significant difference between the predicted and observed distributions ($p \approx 0.06$). In contrast, the baseline models show noticeable deviations, often failing to accurately capture the peak density. For winter barley, which proved challenging for all architectures, CropFusionNet still maintained the closest distributional alignment

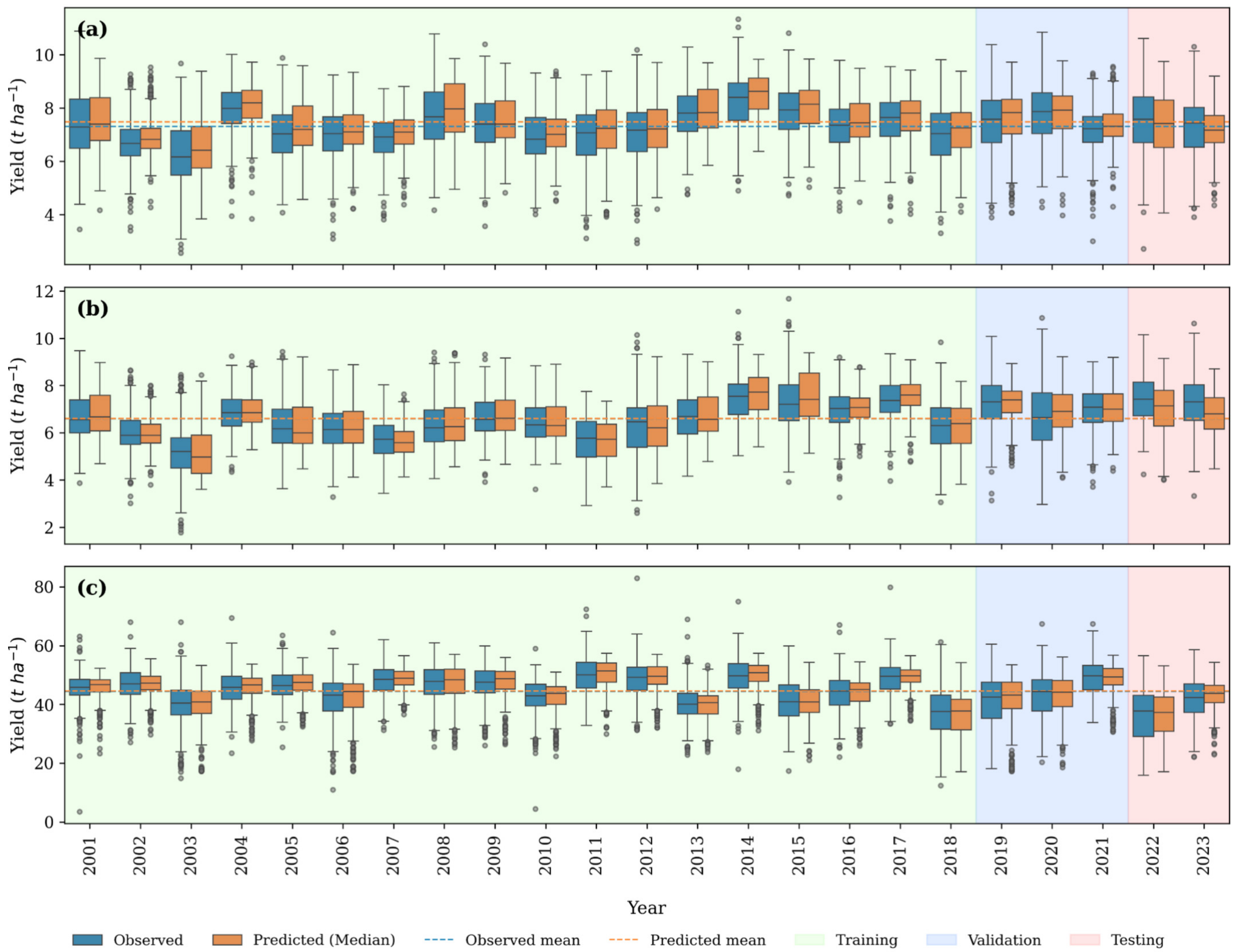

Fig. 2. Temporal distribution boxplots comparing observed district-level yields across Germany with CropFusionNet median (q0.5) predictions for (a) winter wheat, (b) winter barley, and (c) silage maize from 2001 to 2023. Background shading denotes the data partitioning into training (2001–2018), validation (2019–2021), and testing (2022–2023) periods.

(bias = -0.13 t ha^{-1}). The baseline models, particularly the Simple Transformer and ResCNN, exhibited a large negative bias, shifting their predictive distributions far to the left of the true observed yields. Similarly, for silage maize, CropFusionNet accurately matched the observed peak and spread with minimal bias (-0.06 t ha^{-1}), avoiding the clear rightward shift (overestimation) seen in the distributions of the Simple Transformer and ResCNN models. This strong agreement between predicted and observed distributions is further supported by the Quantile-Quantile (Q-Q) plots (**Fig. S2**). While the baseline models show clear deviations, particularly in the tails of the distribution, CropFusionNet closely follows the 1:1 reference line across all quantiles. This indicates that the model reliably captures not only typical yield values but also extreme events. To further evaluate the reliability of the predicted uncertainty bounds, we assessed the Prediction Interval Coverage Probability (PICP) for the nominal 80% prediction interval (**Appendix G**). The empirical coverage rates were 71.5% for winter wheat, 71.1% for silage maize, and 63.2% for winter barley (**Table S2**). While winter wheat and silage maize were reasonably close to the 80% prediction interval, indicating sufficient calibration, the lower coverage indicates that the model's prediction intervals are generally too narrow. Narrow and overconfident prediction intervals when forecasting out-of-distribution climate events are a known phenomenon in deep learning (Amini et al., 2020), highlighting the structural challenge of relying on deterministic single-model architectures to capture total predictive variance during extreme years.

As shown in Fig. 4, CropFusionNet consistently achieved the lowest normalized root mean square error (NRMSE) across winter wheat (9.90%), winter barley (12.49%), and silage maize (10.53%), demonstrating a systematic reduction in relative predictive error compared to sequence-based and convolutional baselines. For winter wheat, its NRMSE was substantially lower than the Vanilla LSTM (12.86%), while for silage maize, CropFusionNet (10.53%) outperformed baselines ranging from 12.53% (Vanilla LSTM) to 13.61% (ResCNN). Although winter barley was more challenging, CropFusionNet remained the most robust (12.49%) compared to the Simple Transformer, which peaked at 16.23%. This trend is reflected in the MAPE distributions, where CropFusionNet maintained the lowest errors across all crops (8.06%, 10.11%, and 8.66%), further underscoring its superior strength in estimating accurate crop yield.

3.1.3. Benchmarking against operational forecasting systems

To evaluate its operational viability, CropFusionNet was also benchmarked against two crop yield forecasting systems using published results reported from the literature–the ABSOLUT model and the MARS system–with relative prediction errors calculated for the period 2018–2021 (Table 2). It should be noted that these comparisons

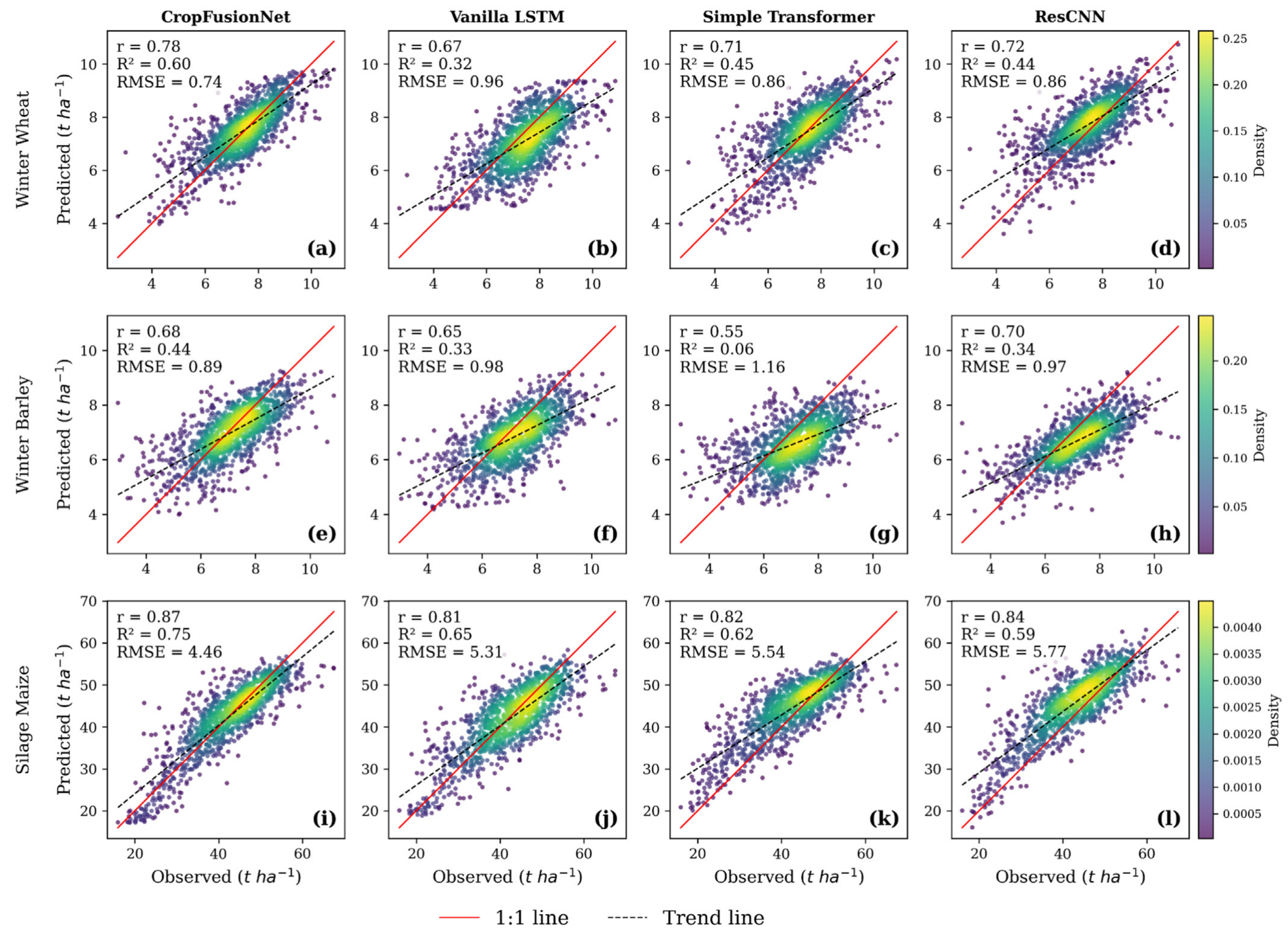

Fig. 3. Scatter density plots illustrating the predictive performance of CropFusionNet compared to baseline architectures (Vanilla LSTM, Simple Transformer, and ResCNN) for (a–d) winter wheat, (e–h) winter barley, and (i–l) silage maize. Performance is evaluated on the combined validation and test sets (2019–2023). The solid red line represents the ideal 1:1 relationship; the dashed black line indicates the actual prediction trend. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Quantitative performance evaluation of CropFusionNet and baseline deep learning models (Vanilla LSTM, Simple Transformer, and ResCNN) for forecasting the yields of winter wheat, winter barley, and silage maize. The input time series for these predictions were taken until July 31st for the winter crops (e.g., winter wheat and winter barley) and September 30th for the spring crop (e.g., silage maize). The metrics are derived from the combined validation and test sets (2019–2023) and include the Pearson correlation coefficient (r), coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

Crop	Model	r	R^2	RMSE ($t\ ha^{-1}$)	MAE ($t\ ha^{-1}$)	MAPE (%)
Winter Wheat	CropFusionNet	**0.78**	**0.60**	**0.74**	**0.56**	**8.06**
	Vanilla LSTM	0.67	0.32	0.96	0.76	10.47
	Simple Transformer	0.71	0.45	0.86	0.66	9.49
	ResCNN	0.72	0.44	0.86	0.66	9.79
Winter Barley	CropFusionNet	0.68	**0.44**	**0.89**	**0.67**	**10.11**
	Vanilla LSTM	0.65	0.33	0.98	0.77	11.16
	Simple Transformer	0.55	0.06	1.16	0.95	13.69
	ResCNN	**0.70**	0.34	0.97	0.76	10.81
Silage Maize	CropFusionNet	**0.87**	**0.75**	**4.46**	**3.41**	**8.66**
	Vanilla LSTM	0.81	0.65	5.31	4.12	10.34
	Simple Transformer	0.82	0.62	5.54	4.31	11.5
	ResCNN	0.84	0.59	5.77	4.48	11.87

rely on published results obtained using different modelling frameworks, spatial resolutions, and input data pipelines (detailed in **Appendix H**). Therefore, rather than representing strictly controlled comparisons, CropFusionNet's performance reflects the combined benefits of a more interpretable model architecture and a more comprehensive, higher-fidelity dataset.

Overall, CropFusionNet demonstrated highly competitive performance and frequently outperformed these established approaches or systems. This was particularly evident during the extreme climate anomaly of 2018. For winter wheat, CropFusionNet achieved a relative error of only 2.64%, substantially outperforming ABSOLUT (5.17%) and MARS (5.76%). A similar pattern was observed for winter barley, where CropFusionNet produced an almost perfect estimate (0.43%), while ABSOLUT underestimated yields (−2.97%) and MARS overestimated them (4.79%). For silage maize, CropFusionNet recorded a small relative error of −2.33%, compared with considerably larger deviations from ABSOLUT (9.92%) and a comparable error from the MARS system (2.3%). Across the subsequent years (2019–2021), CropFusionNet maintained consistently low error margins often below 3.7%, whereas the operational systems showed larger fluctuations, such as ABSOLUT's 11.89% error for winter barley in 2020 and MARS's 7.89% error for winter wheat in 2021.

Despite the methodological differences between the systems, this benchmark remains scientifically meaningful. Because all systems were evaluated using the same target metric (national relative error)

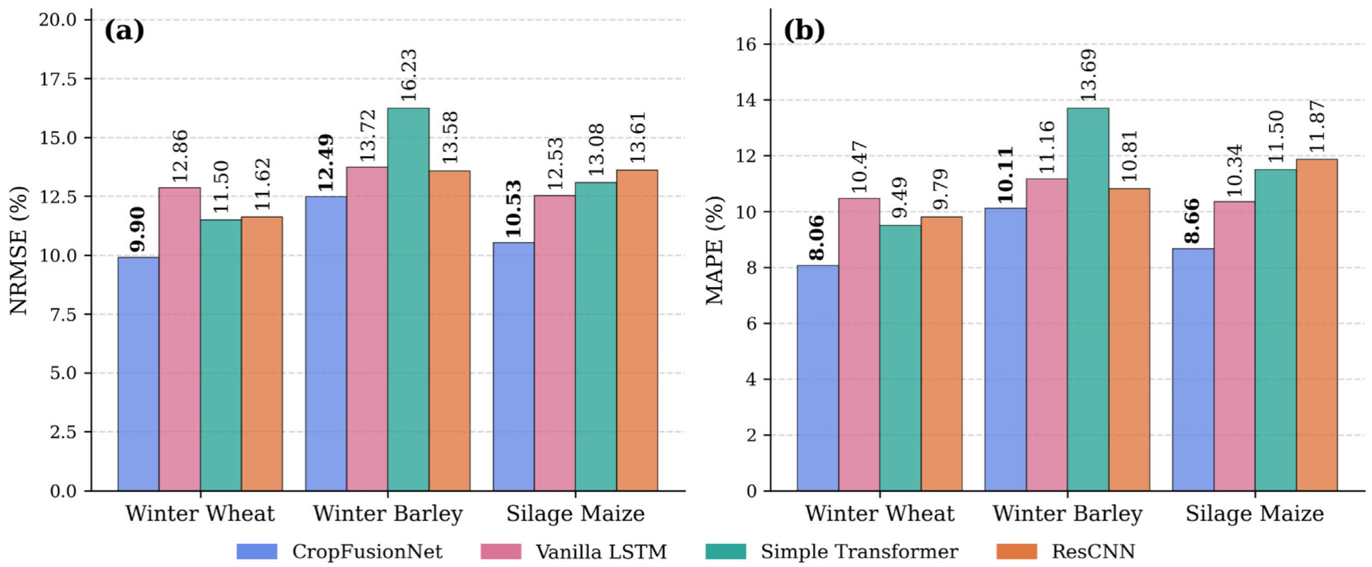

Fig. 4. Comparison of relative predictive errors across deep learning models for winter wheat, winter barley, and silage maize on the combined validation and test sets (2019–2023). Subplot (a) displays the Normalized Root Mean Square Error (NRMSE), and subplot (b) shows the Mean Absolute Percentage Error (MAPE).

and overlapping timeframe (2018–2021), CropFusionNet's sustained accuracy, particularly its ability to minimize errors during the 2018 compound hot and dry event, indicates that its capacity to process daily-scale inputs and learn nonlinear stress responses provides a genuine and operationally consequential advantage.

3.2. Spatial distribution of prediction errors and uncertainty

The robustness of CropFusionNet was further assessed by analyzing the spatial distribution of model performance at the district-level across Germany. Fig. 5 summarizes two complementary metrics: the relative Root Mean Square Error (rRMSE), which quantifies the magnitude of prediction errors relative to observed yields, and the Pearson correlation coefficient (r), which measures how well the model captures inter-annual yield variability over time within each district during the validation and test years. The rRMSE maps (top row) indicate consistently low prediction errors for winter wheat and silage maize across most districts. Winter wheat exhibits a median rRMSE of 7.8%, with most regions showing strong predictive accuracy. Silage maize shows similarly stable performance with a median rRMSE of 8.7%. In contrast, winter barley displays greater spatial heterogeneity and higher overall error levels, with a median rRMSE of 10.5% and more frequent localized deviations, particularly in the eastern region. The Pearson correlation maps (bottom row) evaluate the model's ability to reproduce year-to-year yield dynamics within each district. Higher correlation values indicate that the model successfully tracks inter-annual fluctuations in yield, even if absolute errors vary. Silage maize demonstrates the strongest temporal consistency, with a median r of 0.9 and uniformly high correlations across most districts. Winter wheat also shows robust performance with a median r of 0.7, although with slightly greater regional variability. Winter barley remains the most challenging crop, with a substantially lower median correlation of 0.2 and several regions–particularly in the east and northeast–demonstrating weak or even negative correlations. This pattern indicates that winter barley yield variability may be influenced more strongly by localized factors or regional heterogeneity that are not fully represented in the current model inputs.

To complement the evaluation of point prediction errors, we analyzed the spatial distribution of predictive uncertainty defined as the range between the 10th and 90th prediction percentiles (q0.9 - q0.1, measured in t ha^{-1}) across the training, validation, and testing sets. The uncertainty maps reveal highly consistent spatial patterns across all three data splits for each crop (**Fig. S6a–i**). The northeastern part of Germany (e.g., Brandenburg and Mecklenburg-Vorpommern) shows very high uncertainty (~1.6–2.0 t ha^{-1}) for winter crops compared to other regions. However, for silage maize, the spatial pattern is distinctly different; the most unpredictable regions are concentrated in the central-western part of Germany (e.g., Nordrhein-Westfalen, Hessen, and Rheinland-Pfalz). These findings identify specific regions that require special attention to understand complex, localized yield drivers that might currently be neither captured by the input data nor represented by the model itself.

Table 2
Comparison of near-harvest national average yield predictions for winter wheat, winter barley, and silage maize expressed as relative prediction errors (%) for the period 2018–2021. The performance of CropFusionNet is benchmarked against published results from two established national-scale operational forecasting systems: the ABSOLUT model and the European Commission's MARS system (Conradt, 2022).

Model / System	Crop	Forecast period	Relative Error (%)			
			2018	2019	2020	2021
CropFusionNet	Winter Wheat	31st Jul	**2.64**	**2.42**	**−0.09**	**3.44**
ABSOLUT		07–09 Jul	5.17	6.85	5.71	5.44
MARS		22–27 Jul	5.76	3.49	−3.55	7.89
CropFusionNet	Winter Barley	31st Jul	**0.43**	**0.20**	3.69	**0.51**
ABSOLUT		07–09 Jul	−2.97	1.66	11.89	0.98
MARS		24–29 Aug	4.79	−1.25	**2.23**	−1.54
CropFusionNet	Silage Maize	30th Sep	−2.33	**0.96**	**−0.05**	**−1.45**
ABSOLUT		07–09 Sep	9.92	−1.77	2.69	−5.61
MARS		14–20 Sep	**2.3**	1.03	−5.64	−3.24

3.3. Forecast skills at different lead times

We evaluated the model's forecasting skill using combined testing and validation datasets across multiple lead times to assess how far in advance of harvest reliable yield predictions can be generated for each crop. Forecast performance was analyzed at 10-day intervals over the final five months preceding harvest. The lead-time analysis (Fig. 6) shows a clear and consistent improvement in forecasting skill as the harvest date approaches across all three crops. For winter wheat, prediction errors remain relatively high at longer lead times

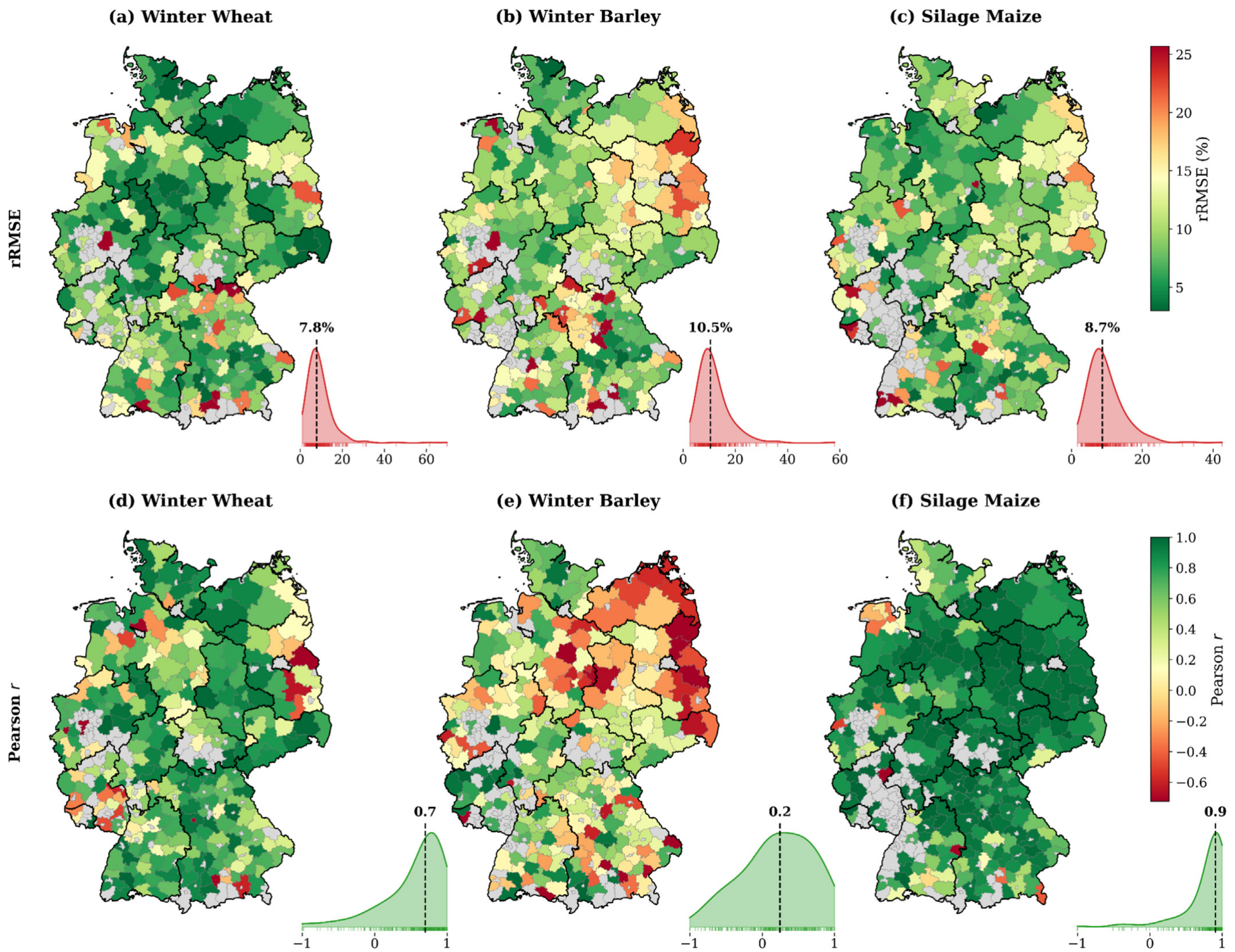

Fig. 5. Spatial distribution of CropFusionNet prediction performance at the district-level across Germany during the validation and test years (2019–2023). Panels (a-c) show the relative Root Mean Square Error (rRMSE) for winter wheat, winter barley, and silage maize, respectively, illustrating the magnitude of prediction errors relative to observed yields. Panels (d-f) present the Pearson correlation coefficient (r), indicating the model's ability to capture inter-annual yield variability within each district.

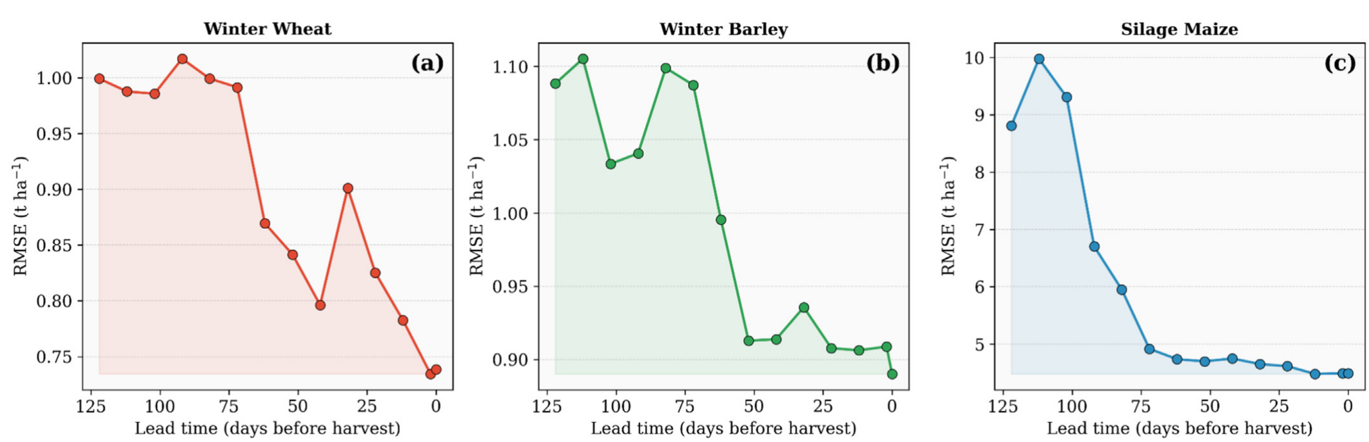

Fig. 6. Lead-time evaluation of CropFusionNet forecasting performance for winter wheat (a), winter barley (b), and silage maize (c). Root Mean Square Error (RMSE; t ha^{-1}) is shown as a function of lead time (days before harvest), with predictions generated at 10-day intervals over the final five months of the growing season.

(approximately 120 to 70 days before harvest), with RMSE values close to 1.0 t ha^{-1}. A significant improvement is observed after about 60 days before harvest, where the error declines steadily to around 0.73–0.75 t ha^{-1} near harvest. In the case of winter barley, the model exhibits comparatively weaker improvement over time. RMSE remains relatively high (around 1.05–1.10 t ha^{-1}) at longer lead times and decreases more gradually after approximately 70 days before harvest, eventually stabilizing near 0.90 t ha^{-1}. For silage maize, the model demonstrates the strongest lead-time dependency and the most substantial improvement in forecasting skill. RMSE is highest at longer lead times (approximately 9–10 t ha^{-1} at ~100 days before harvest) but declines sharply, reaching ~4.5 t ha^{-1} near harvest. Notably, errors fall below 5 t ha^{-1} even before 70 days prior to harvest, suggesting that reliable forecasts can be obtained relatively early for this spring crop as informative in-season signals become available. Minor non-monotonic fluctuations in RMSE are observed at intermediate lead times. These variations likely correspond to phenological transition phases and short-term environmental variability, during which yield formation becomes highly sensitive to climatic inputs. Such behaviour is consistent with previous crop forecasting studies (Zachow et al., 2025), where predictive uncertainty temporarily increases during critical growth stages before stabilizing closer to harvest.

3.4. Model interpretability and feature attribution

A key advantage of the CropFusionNet architecture lies in its inherent interpretability, driven by the Variable Selection Networks (VSNs). By extracting the softmax-normalized weights (α_{static} and $\alpha_{temporal}$), the model estimates the relative importance of both static and time-varying covariates in driving yield predictions. It is important to note that because these weights are generated through independent Softmax normalizations, they represent relative feature importance strictly within their respective domains. Furthermore, because the temporal VSN is mathematically conditioned on the static context (H_{static}), static landscape characteristics (e.g., soil, elevation, and slope) implicitly guide the model's attention to temporal events. Consequently, the absolute numerical values of α_{static} and $\alpha_{temporal}$ are on different scales and cannot be directly aggregated into a single, unified ranking; they must be interpreted as distinct spatial and temporal diagnostic metrics.

3.4.1. Global importance of static covariates

To identify the primary static drivers of crop yield, we aggregated static variable selection weights across all spatial units for winter wheat, winter barley, and silage maize (Fig. 7). Soil properties and topography emerged as the dominant determinants. Mean soil quality consistently ranked as the most important covariate, accounting for 0.30–0.38 of the total attribution weight across crops, highlighting the importance of inherent fertility and water-holding capacity. Elevation was the second most influential factor, reflecting altitude-driven microclimatic effects, while within-district variability of soil quality also played a meaningful role. The variability (standard deviation) of soil and topographic properties generally contributed less than their corresponding mean values but remained relevant for explaining yield variability, with the exception of slope variability (SLP sd) for silage maize. In contrast, the irrigated fraction had minimal influence at the large scale, consistent with Germany's predominantly rainfed winter cereal production. Fig. 8 maps the spatial distribution of static feature importance across Germany. Mean soil quality (SQR mean) is driving the predictive performance for winter crops in the northern districts and few districts in the south where soil quality is moderate (~50–60) (Fig. 8a, b). In central German districts with high soil quality, soil properties become non-limiting factors, as a result the topographical features contribute more to the yield variability. Elevation exhibits a clear spatial gradient, with high importance assigned primarily in central and southern regions, corresponding to the Central Uplands and Alpine Foreland, where higher altitudes impose distinct microclimates and shorter growing seasons. Irrigation remains near-zero everywhere, reflecting its non-limiting role and very small spatial coverage. These spatial patterns in VSN-derived weights suggest that the model differentiates between dominant environmental controls across regions, assigning higher importance to the most influential static factors under varying agro-ecological conditions. The model dynamically adjusts attention, emphasizing elevation in southern uplands and soil quality in northern plains, capturing the complex, region-aware interactions between crops and their environment while filtering out irrelevant geographic noise.

3.4.2. Global importance of time-varying covariates

The temporal evolution of feature importance for time-varying covariates at a daily scale across the crop growing periods is illustrated in Fig. 9. The heatmaps represent the mean temporal attention weights ($\alpha_{temporal}$) derived from the temporal VSN module, highlighting both the most important variables and their critical time windows. For winter wheat, temperature-related variables, particularly TMIN, dominate the importance throughout the main growing season, especially from autumn establishment through late winter and early spring. Vegetation

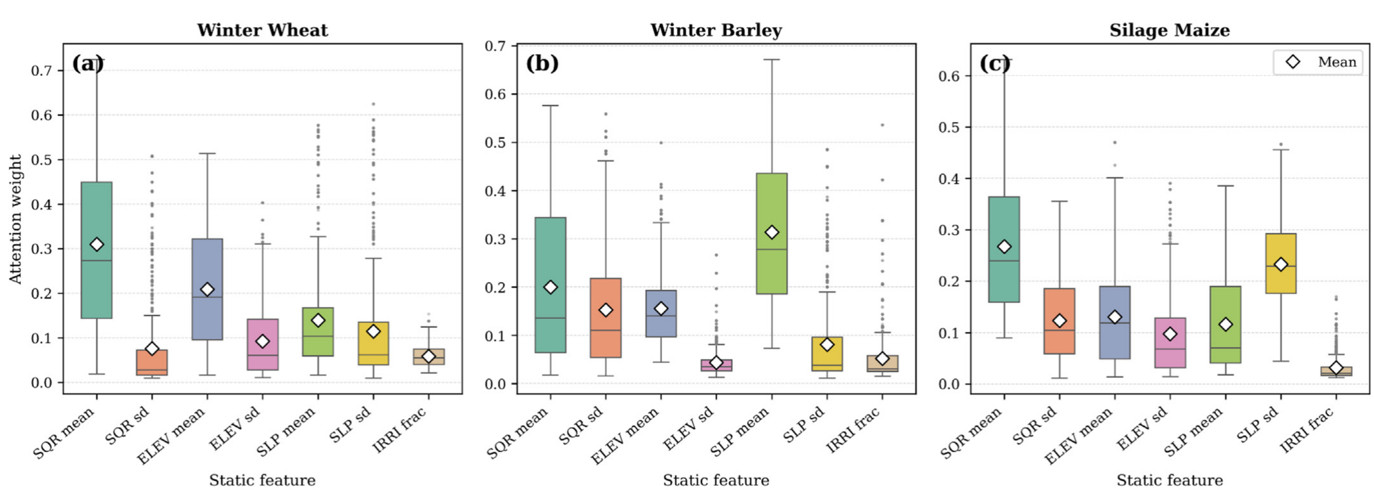

Fig. 7. Global importance of static covariates for winter wheat (a), winter barley (b), and silage maize (c). The values represent the mean softmax-normalized weights (α_{static}) extracted directly from the Static Variable Selection Network (VSN), averaged across all districts. Higher weights indicate a stronger influence on prediction. Static covariates include Soil Quality Rating (SQR mean, SQR sd), Elevation (ELEV mean, ELEV sd), Slope (SLP mean, SLP sd), and Irrigated area fraction (IRRI frac). (Note: mean = average; sd = standard deviation).

Fig. 8. Spatial distribution of static feature importance across German districts for (a) winter wheat, (b) winter barley, and (c) silage maize. Darker shades indicate regions where the model assigns greater weight to a given variable, thereby enriching the temporal representation and its contribution to final yield prediction. (Note: Full names of the variable abbreviations are provided in Table S1).

indices such as NDVI and EVI, proxies for canopy development and biomass accumulation, become increasingly important towards the later phenological stages, heading, flowering, and maturity. Interestingly, VPD shows high importance during the pre-planting phase. High VPD indicates dry air and greater evaporative demand, which can lead to soil moisture depletion prior to sowing. As a result, higher VPD negatively conditions the initial soil water availability at planting, influencing germination and delayed crop establishment. The model appears to capture this signal to encode pre-season environmental stress, which subsequently affects yield potential. A similar pattern is observed for winter barley. TMIN remains the most influential variable across much of the season, while VPD and ST contribute notably during early development. Vegetation-related features (e.g., FPAR, NDVI) become more relevant during the reproductive and grain-filling phases. In contrast, silage maize, being a spring crop, displays a distinctly different temporal pattern aligned with its summer growth cycle. Early-stage importance is driven by temperature (TMEAN) and water-related variables such as CWB and VPD, emphasizing the role of heat and water availability during crop establishment. As the season progresses, vegetation indices particularly EVI and FPAR become dominant, capturing peak biomass development during mid-to-late summer. Compared to winter crops, the importance is more concentrated within a shorter, well-defined growing window. Across all crops, temperature and vegetation indices consistently emerge as the primary drivers. These results demonstrate that the model effectively learns crop-specific phenological dynamics, assigning temporal importance in alignment with known growth stages. Figs. S3-S5 reveal greater spatial heterogeneity in temporal feature importance across Germany, reflecting underlying soil–atmosphere–plant continuum processes that shape yield variability. In the arid eastern German districts, silage maize yield is largely constrained by water availability, with PREC, ST, ET0 emerging as the dominant drivers. In contrast, Southern Germany shows a shift towards heat-related controls, where VPD, TMIN and SUND playing the main driving role. This east-south contrast highlights the transition from water-limited to energy-limited regimes. A similar spatial distinctiveness is evident in winter crops, underscoring the robustness of these regional mechanisms. These patterns indicate that the model adapts to region-specific conditions by assigning varying importance to temporal features, enabling representation of complex and non-linear interactions within the soil–atmosphere–plant system.

3.5. Sensitivity to extreme climate events

A common limitation of deep learning models in agricultural forecasting is their inherent tendency to regress towards the mean; which often leads to a diminished capacity to accurately represent the impacts of rare or unprecedented climate extremes (Olivetti and Messori, 2024; Qiao et al., 2025). This behaviour raises important questions about whether such models truly capture the underlying physiological responses of crops to environmental stress, or merely approximate average conditions. To address this concern, we investigated whether CropFusionNet effectively encodes the physiological signatures associated with environmental variability and stress. Specifically, we examined the model's internal latent representations by extracting the final high-dimensional temporal embedding (H_{final}) for all district-year combinations, immediately prior to its input into the uncertainty-aware quantile prediction head. This representation is expected to encapsulate the cumulative temporal dynamics of the growing season as learned by the model. To facilitate interpretation, we applied Principal Component Analysis (PCA) to project these high-dimensional embeddings into a lower-dimensional space (Fig. 10). This dimensionality reduction enabled us to visualize and assess how the model differentiates between

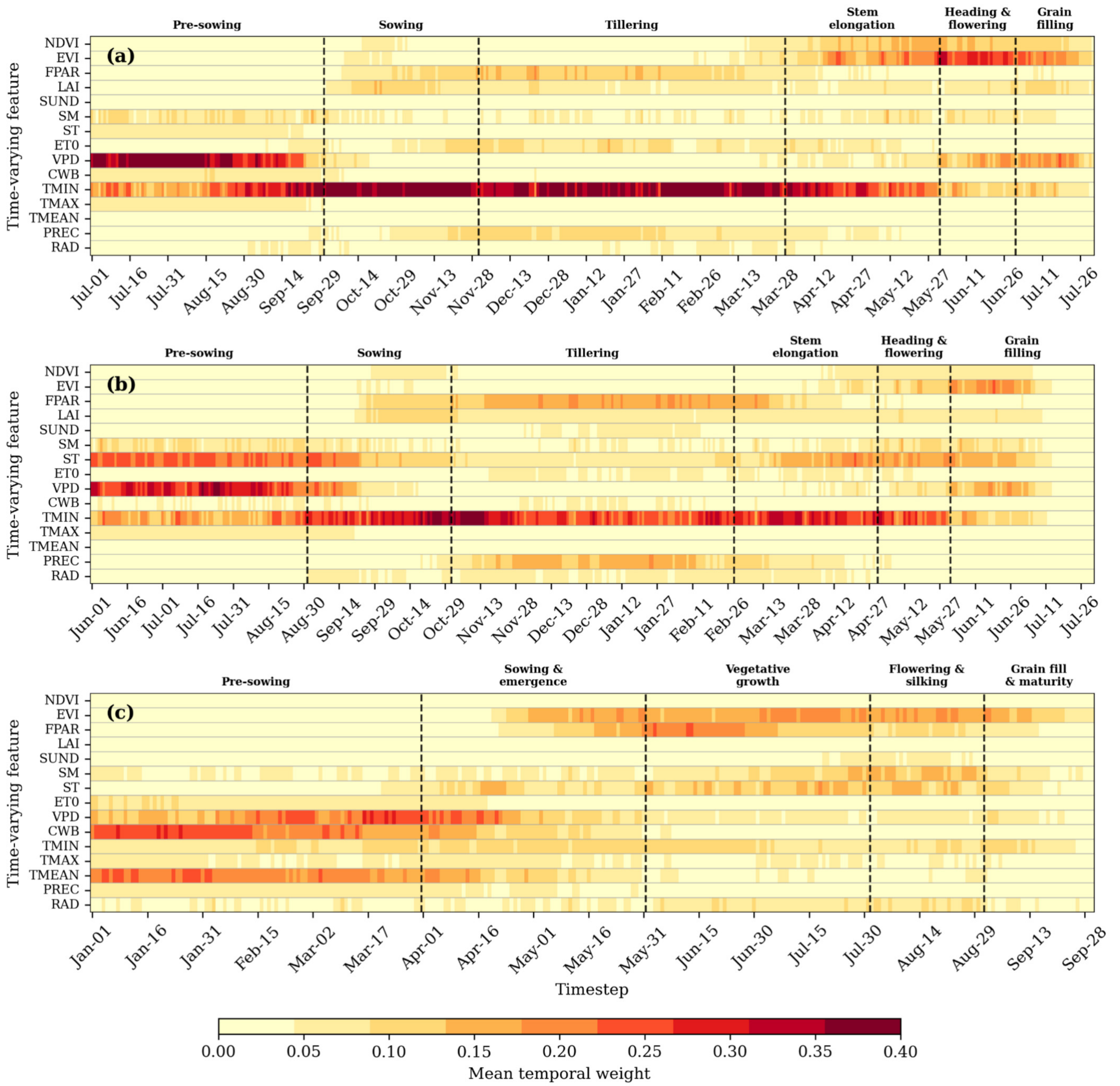

Fig. 9. Temporal evolution of feature importance for time-varying covariates at daily resolution for (a) winter wheat, (b) winter barley, and (c) silage maize. Colours represent the mean temporal attention weights ($\alpha_{temporal}$) derived from the temporal Variable Selection Network (VSN), with darker shades indicating greater contribution to yield prediction at specific timesteps. (Note: Full names of the variable abbreviations are provided in Table S1).

distinct types of growing seasons, including positive yield extremes, negative yield extremes, and near-normal conditions.

Fig. 10 demonstrates that the first two principal components (PC1 and PC2) capture a substantial proportion of variance in the latent space, accounting for 74.6% for winter wheat, 50.0% for winter barley, and 58.1% for silage maize. The corresponding 2D projection (Fig. 11) reveals a structured organization of district-year samples according to their observed yield values. Notably, "negative extreme" data points exhibit a clear divergence from "normal" and "positive extreme" conditions, with the primary separation occurring along the PC1 axis. Analysis of centroid shifts relative to normal conditions (Fig. 12) further indicates an asymmetric response to climatic extremes. For silage maize, the centroid of negative extreme data points shifts substantially by −12.32 along PC1, whereas the positive extreme centroid shifts by only +6.97. Winter barley shows a similar, though less pronounced, asymmetric shift (−6.32 for negative vs. +4.66 for positive along PC1). In contrast, winter wheat exhibits a more symmetric, linear response to climate deviations (−7.31 vs. +7.14 along PC1).

Mapping specific historical years onto the latent space (Fig. 13) reveals that major drought and heatwave events are consistently located at the extreme negative end of the PC1 axis. Across all three crops, the severe compound drought years of 2003 and 2018 occupy the most negative positions in terms of mean yield anomalies. For silage maize, the extreme summer drought of 2022 clusters closely with these years,

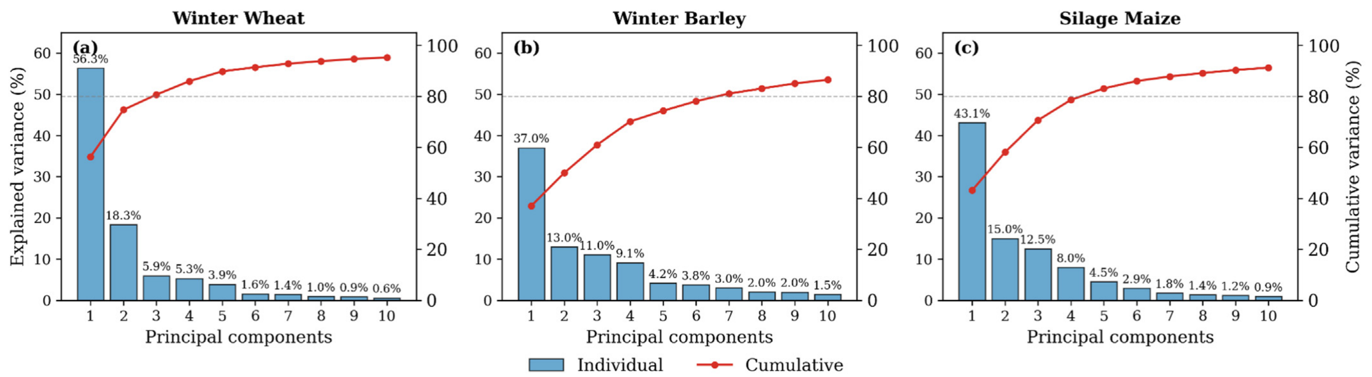

Fig. 10. Scree plots of principal component analysis (PCA) applied to CropFusionNet latent representations. The plots illustrate the variance explained by the first ten principal components (PCs) derived from the model's temporal representation (H_{final}) prior to the uncertainty-aware quantile-based output layer for (a) winter wheat, (b) winter barley, and (c) silage maize. Blue bars indicate the individual variance explained by each component, while the red line represents the cumulative explained variance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

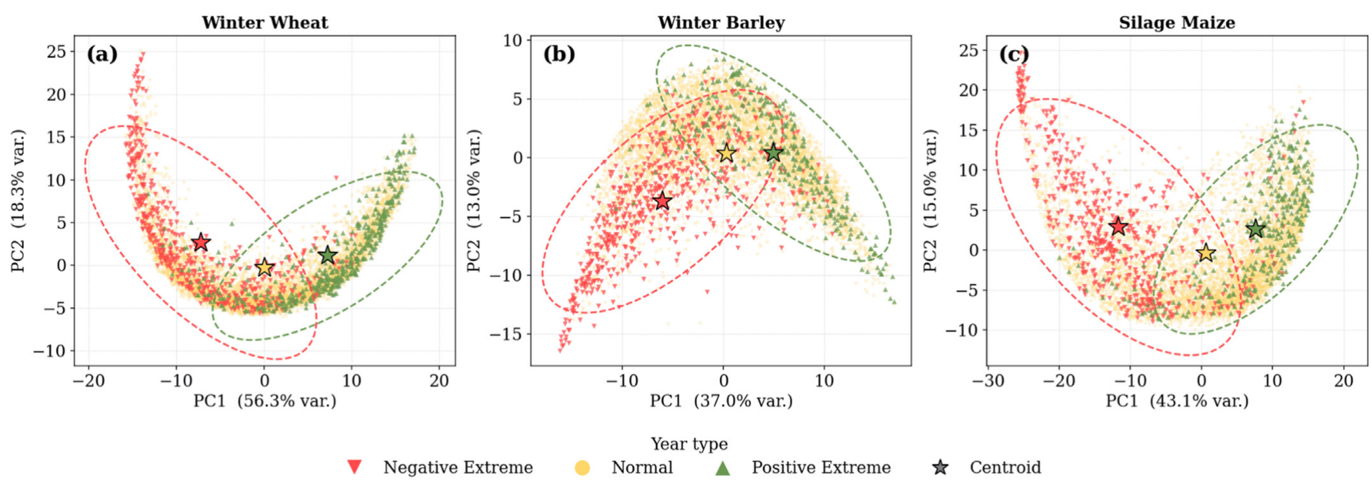

Fig. 11. Two-dimensional latent space representations (PC1 vs. PC2) categorized by normal and extreme years. Scatter plots displaying the distribution of latent representations for (a) winter wheat, (b) winter barley, and (c) silage maize. Each point represents a specific district-year combination. The samples are classified and coloured based on their detrended yield z-anomalies: values exceeding ± 1.5 standard deviations are defined as Negative Extremes (red triangles) and Positive Extremes (green triangles), while all intermediate values are classified as Normal years (yellow circles). Star markers denote the centroid of each distribution, and the dashed lines represent the 95% confidence ellipses for the extreme classes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

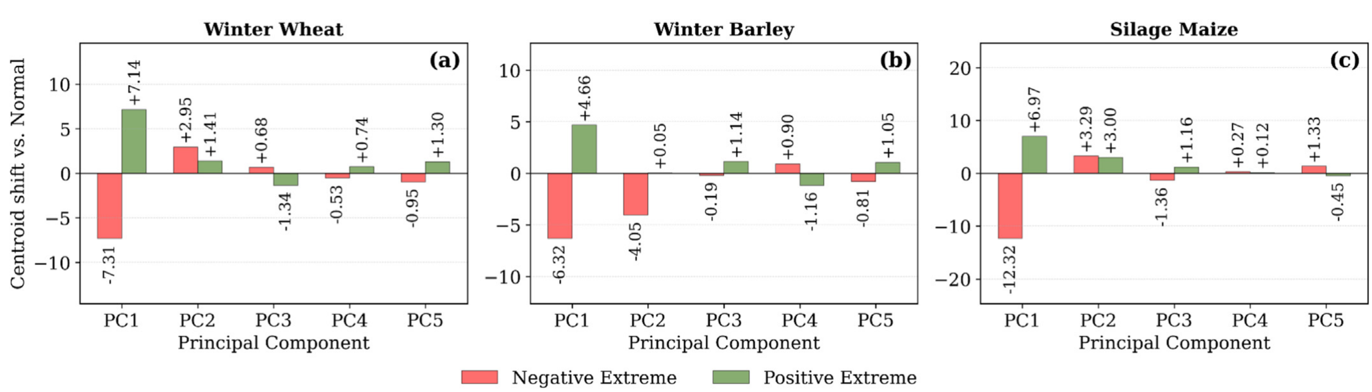

Fig. 12. Centroid shifts of extreme cases relative to normal conditions (on y-axis). Bars illustrate the directional displacement of negative extreme (red) and positive extreme (green) centroids with respect to the normal baseline across the first five principal components (on x-axis) for (a) winter wheat, (b) winter barley, and (c) silage maize. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

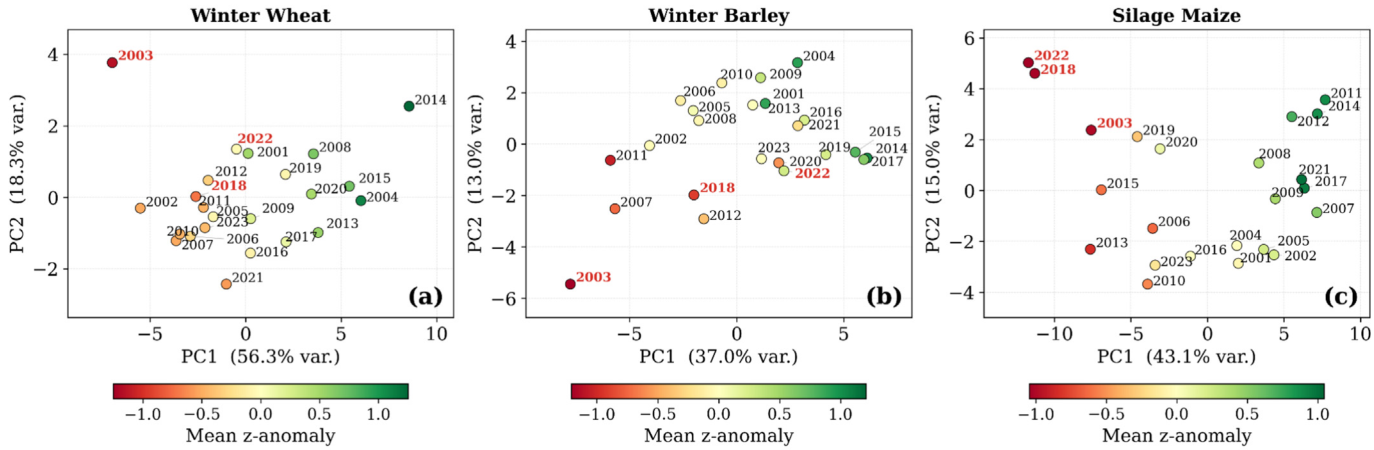

Fig. 13. Historical year-wise mean trajectories in the PC1-PC2 latent subspace. Aggregated year-wise representations for (a) winter wheat, (b) winter barley, and (c) silage maize from 2001 to 2023. Each point represents a specific year, coloured according to its mean yield z-anomaly (where deeper red indicates severe crop yield losses and deeper green indicates above average crop yields). Notably, historical catastrophic drought and heatwave years (2003, 2018, and 2022) are annotated in bold red text. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

highlighting the crop's pronounced sensitivity to mid-season heat and moisture stress. In contrast, highly productive years, such as 2014, are positioned towards the positive extreme of the distribution. These trajectories confirm that CropFusionNet's internal architecture explicitly encodes the biophysical stressors that drive major regional crop failures.

Recasting the continuous yield forecasts into a tercile-based classification problem demonstrated robust categorical skill (**Fig. S8**). CropFusionNet achieved overall accuracies of 0.64, 0.59, and 0.71 for winter wheat, winter barley, and silage maize, respectively, with corresponding macro-F1 scores of 0.65, 0.59, and 0.71, substantially exceeding the random baseline of 0.33. Performance was consistently higher for the Low and High yield classes than for the Normal class, reflecting the greater susceptibility of the middle tercile to boundary-crossing errors. Importantly, most misclassifications occurred between adjacent classes (Low↔Normal and Normal↔High), whereas severe misclassifications between the Low and High yield classes were rare (1.6%, 3.8%, and 1.0% for winter wheat, winter barley, and silage maize, respectively). These results indicate that CropFusionNet reliably distinguishes yield tiers while preserving the underlying yield gradient.

3.6. Ablation study of model components

CropFusionNet is a complex deep learning architecture; therefore, an ablation study was conducted to quantify the contribution of its individual components. Key modules including the Variable Selection Network (VSN), multi-scale temporal convolution, LSTM layers, multi-head self-attention, static enrichment, and attention-based temporal pooling were systematically removed one-by-one, and the resulting changes in coefficient of determination (ΔR^2) and root mean square error (ΔRMSE) were evaluated. All ablation experiments were conducted using the same training protocol and hyperparameter settings as the full model.

Fig. 14 highlights a clear crop-specific dependency on the architectural complexity. The VSN emerges as the only consistently indispensable component across all crops, as its removal leads to a systematic degradation in predictive performance, underscoring the importance of dynamically filtering noisy agro-meteorological inputs prior to sequence modelling. Silage maize benefits most from the full architecture, with the removal of the LSTM layers causing the largest performance decline ($\Delta R^2 = -0.070$; $\Delta \mathrm{RMSE} = +0.592$ t ha^{-1}), followed closely by the VSN ($\Delta R^2 = -0.063$; $\Delta \mathrm{RMSE} = +0.533$ t ha^{-1}). This indicates that capturing the cumulative effects of summer heat and moisture stress requires strong temporal memory. Winter barley exhibits mixed dependencies, with notable dependence on temporal pooling ($\Delta R^2 = -0.098$) and static enrichment ($\Delta R^2 = -0.094$). Interestingly, performance slightly improves when the multi-scale temporal convolution was removed ($\Delta R^2 = +0.051$; $\Delta \mathrm{RMSE} = -0.042$ t ha^{-1}), suggesting potential redundancy in this component. In contrast, winter wheat shows signs of architectural over-parameterization; aside from the VSN, the removal of deeper temporal processing modules leads to marginal performance gains, for example, excluding temporal convolution results in a $+0.040$ increase in R^2 and a -0.036 t ha^{-1} reduction in RMSE. Overall, the ablation results indicate that while the full CropFusionNet architecture is highly effective for dynamic spring crops such as silage maize, simpler model configurations may be sufficient and occasionally advantageous for winter cereals once the VSN has identified the most relevant input features.

Furthermore, while traditional yield forecasting models frequently rely on dekadal or monthly meteorological aggregates, CropFusionNet was designed to ingest daily time-varying input data to capture short-term climate fluctuations. To explicitly quantify the added value of this daily-scale information, we conducted an additional benchmark experiment to evaluate the model's sensitivity to temporally aggregated inputs. The results confirmed that aggregating daily-scale input data into 8-day, 16-day, or monthly intervals generally led to a clear degradation in predictive performance across all crops. These findings confirm that excessive temporal aggregation obscures high-frequency meteorological anomalies that are critical for accurate yield estimation. A detailed discussion of this experiment and the complete performance metrics are provided in **Appendix I** and Supplementary Table S3.

4. Discussion

4.1. Architectural innovation in CropFusionNet

Accurately representing climate extremes remains a major challenge in agricultural system modelling (Feng et al., 2022; Richetti et al., 2025; Zheng et al., 2025). The proposed deep learning architecture, CropFusionNet, addresses this limitation by explicitly capturing the impact of extreme climatic conditions on end-of-season yield predictions (Figs. 11, 13), while simultaneously providing inherent model interpretability (Figs. 8, 9). Despite substantial progress in data-driven approaches over the past few decades, agricultural systems have traditionally been modelled using mechanistic (process-based) and statistical frameworks (Holzworth et al., 2015; Stock et al., 2024). Moreover, while these approaches have provided valuable insights, they often

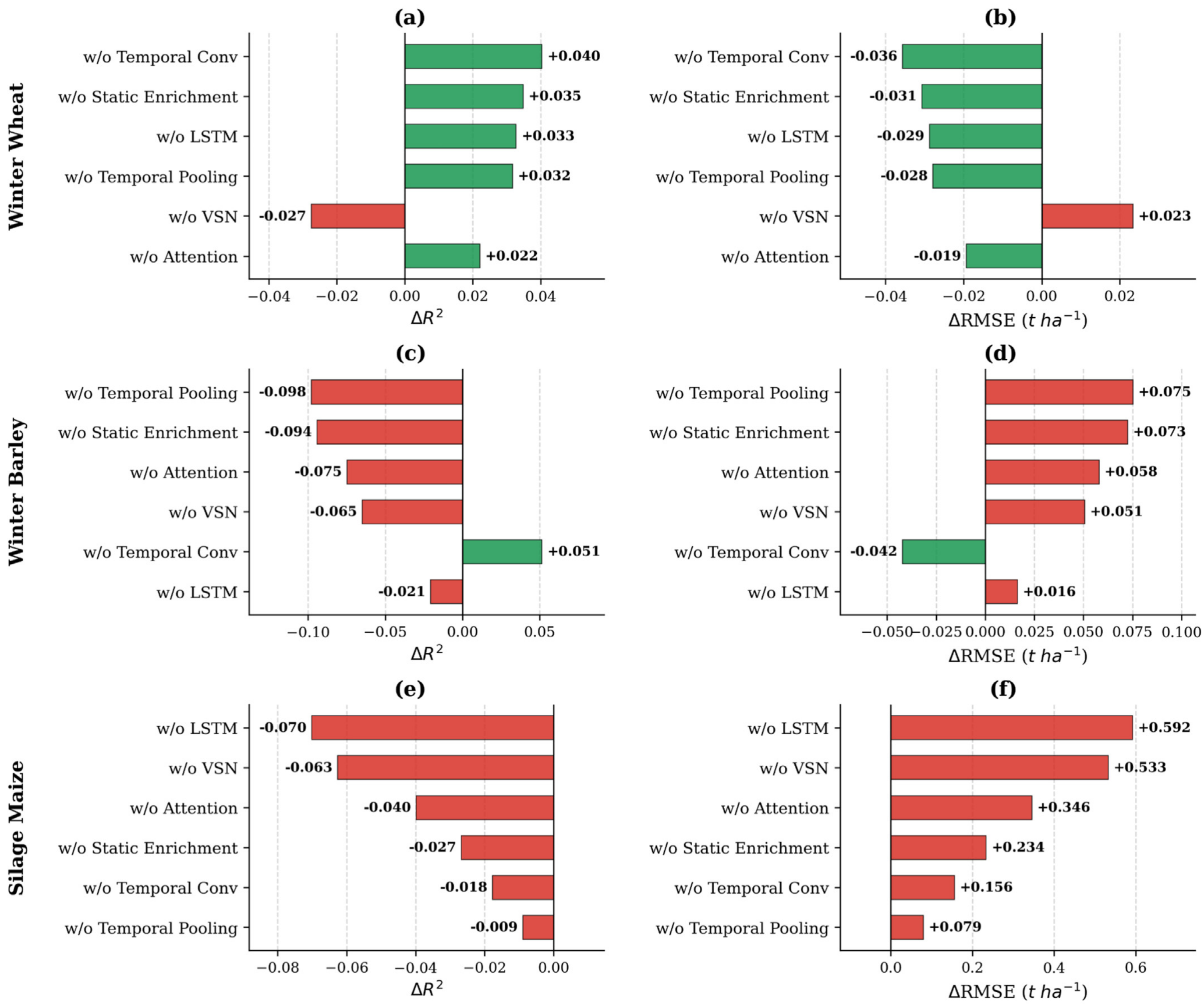

Fig. 14. Ablation study evaluating the contribution of individual architectural components in CropFusionNet for (a, b) winter wheat, (c, d) winter barley, and (e, f) silage maize. The left panels (a, c, e) display the change in the coefficient of determination (ΔR^2), while the right panels (b, d, f) show the change in Root Mean Square Error (ΔRMSE in t ha^{-1}) relative to the full model. Red bars indicate a degradation in model performance upon component removal, whereas green bars signify a performance improvement. (Note: w/o = without). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

struggle to accurately capture extreme events due to the inherent uncertainty and complexity of environmental processes (Feng et al., 2022; Filippi et al., 2025). Process-based models (e.g., WOFOST) are constrained by their reliance on predefined mathematical formulations of environmental processes, which limits their ability to adapt complex and dynamic stress responses under extreme conditions (Zheng et al., 2025). This limitation becomes further exacerbated when scaling to large and heterogeneous regions, where a single set of crop parameters is insufficient to represent diverse agro-ecological zones (Kallenberg et al., 2023). Consequently, model calibration becomes highly challenging, requiring extensive domain knowledge and region-specific experimentation, and often resulting in systematic under- or overestimation of yield extremes (Zampieri et al., 2018). In contrast, CropFusionNet, as a deep learning model, automatically updates its parameters via a user-defined loss function without requiring explicit intervention in model optimization. By fusing static environmental characteristics with dynamic meteorological inputs, it can efficiently generalize across large heterogeneous regions within the scope of the training data.

Conversely, statistical and traditional machine learning approaches typically rely on engineered inputs, such as temporally aggregated climate variables (e.g., weekly or monthly averages), which can obscure short-term yet high-impact climate shocks present in daily observations (Vandal et al., 2019). As a result, these methods fail to fully exploit the information richness of daily raw data. CropFusionNet, in contrast, operates on daily-scale meteorological inputs, preserving the temporal granularity and capturing the full spectrum of variability. Through its VSN and static enrichment modules, the model effectively conditions temporal inputs on static features, such as soil quality and topography. Moreover, CropFusionNet supports both numerical and categorical inputs for static and daily time-varying features, making it highly suitable for hybrid modelling, where outputs from crop simulation models (typically at daily resolution) can be integrated with input features to enhance predictions. By accommodating multiple data modalities and retaining the native structure of daily time series, CropFusionNet maximizes the information content available for accurate and flexible yield forecasting.

While process-based crop models offer interpretability, they come with significant limitations mentioned above. Recent deep learning approaches have improved predictive performance but are often treated as "black boxes," providing little interpretability for stakeholders. Consequently, interpretability in these models is typically addressed as post-hoc techniques such as SHAP or gradient-based importance scores, rather than being inherently integrated into the model architecture (Marcinkevičs and Vogt, 2023; Nieradzik et al., 2024). CropFusionNet, in contrast, generates normalized importance scores for both static and temporal inputs through its VSN module, which offers substantial practical and diagnostic value for end-users such as agronomists, agricultural advisors, and policymakers. For example, the spatial mapping of static drivers (α_{static}) allows policymakers to identify regions where inherent landscape constraints, such as low soil quality, disproportionately amplify climate vulnerability (Shook et al., 2021). This insight can guide targeted policy interventions, such as allocating soil conservation subsidies, optimizing land-use planning, or structuring region-specific crop insurance premiums. Similarly, the temporal attention weights ($\alpha_{temporal}$) provide time-dependent diagnostics during the growing season. If the model predicts a yield deficit, agronomists can trace the prediction back to its contributing environmental stressors (e.g., an anomalous spike in VPD during the flowering stage) (Isik et al., 2025). This transparency not only enhances stakeholder trust in data-driven forecasts but also enables better decision-making through timely mitigation strategies, such as adjusting regional irrigation schedules or prioritizing drought-resilient cultivars in identified stress hotspots (Filippi et al., 2025). Moreover, the importance of the VSN module is highlighted in the model component ablation study (Fig. 14), where removing it consistently degrades predictive performance across all crops, indicating that filtering inputs before temporal processing is a good choice for reducing noise present in the raw data. Such an architectural feature is absent in traditional baseline models, including LSTM, Transformer, and CNN architectures. This design not only improves predictive accuracy but also allows CropFusionNet to effectively model complex, nonlinear interactions in environmental data, making it both more interpretable and robust compared to conventional approaches.

Finally, CropFusionNet is designed to produce uncertainty-aware quantile-based forecasts, predicting crop yield across multiple quantiles rather than a single deterministic estimate. In contrast, most previous studies provide only point estimates, overlooking the inherent uncertainty in crop yield outcomes (Ye et al., 2021). By framing quantile-based yield forecasting, the model captures not only the most likely crop yield but also the best- and worst-case scenarios, offering stakeholders actionable insights for risk-aware decision-making (Lacasa et al., 2023). With these architectural innovations, CropFusionNet represents a promising model for next-generation crop yield forecasting tasks.

4.2. Role of key agro-climatic drivers in crop yield formation

A notable outcome of this study is the explicit ranking of key agro-climatic drivers shaping crop yield variability. Many yield prediction studies rely predominantly on basic meteorological inputs (e.g., precipitation and temperature) (Iizumi et al., 2021; Chen et al., 2023; Park et al., 2023), while process-based crop models integrate detailed physiological and soil processes to provide mechanistic insights into crop responses (Nendel et al., 2023). In contrast, our framework does not aim to replace such mechanistic representations but instead offers a data-driven perspective by incorporating physically meaningful stress indicators and vegetation proxies. This enables improved interpretability of yield variability and driver importance at scale, particularly in data-rich settings where full process-based modelling may not be feasible.

Although temperature fundamentally governs the phenological development, it is insufficient to explain yield variability in isolation particularly under compound climate extremes. CropFusionNet's temporal VSN module potentially captured this complexity, highlighting distinct crop- and phenological development stage-specific sensitivities. For winter cereals, TMIN dominated from establishment through early spring, aligning with known biological sensitivities to cold stress, vernalization, and early growth conditions. However, VPD is equally essential because it captures the effective water stress experienced by the crop, simultaneously accounting for atmospheric evaporative demand and soil moisture dynamics (Zhang et al., 2021). This explains why approaches relying solely on raw temperature and precipitation inputs often fail to capture extreme yield losses during compound drought-heat events (Ribeiro et al., 2020; Feng et al., 2022). Notably, our framework assigned greater importance to VPD in the pre-planting stage, plausibly encoding antecedent moisture demand and the conditioning of initial soil water availability for germination. Furthermore, as the season advances towards the growing phenological stages, vegetation indices (e.g., NDVI, EVI, LAI, and FPAR) become critical integrative proxies. Unlike daily weather variables that represent external forcing, these vegetation indices encode the cumulative effects of environmental conditions and management practices, directly reflecting the realized biophysical state of the plant. Similarly, for silage maize, mid-season vegetation, TMEAN, VPD, and CWB emerged as highly influential predictors, consistent with spring crop physiology and its greater sensitivity to heat and moisture stress during critical reproductive stages.

Apart from temporal dynamics, the spatial variability in the weights (**Fig. S3-S5**) highlights that crop yield responses are not governed by the same set of drivers as seen in Fig. 9 but rather emerge from region specific soil-plant-atmosphere interactions. The observed regional differences like the silage maize transition from water-limited systems in eastern Germany to more energy- and heat-limited regimes in the south. Such spatial differentiation also provides insight into why models relying on globally uniform predictor importance often underperform, particularly under extreme or compound events. For instance, while precipitation deficits may be the primary constraint in eastern regions, atmospheric demand indicators such as VPD become more critical in the south, where heat stress and evaporative demand dominate crop responses. We want to emphasize that the same climatic anomalies may translate into fundamentally different yield impacts depending on the regional context, reinforcing the CropFusionNet's ability to interpret agro-climatic drivers within the coupled environmental setting.

Accurately modelling crop state requires integrating dynamic meteorological drivers with static environmental constraints. Figs. 7 and 8 demonstrate that soil quality is a key determinant of yield potential, regulating water retention, nutrient availability, and resilience to climatic variability (Chen et al., 2024; Dhillon et al., 2025). Regions with higher soil quality can buffer transient water deficits, thereby mitigating the effects of short-term atmospheric stress. In contrast, lower-quality soils amplify the impact of the same climatic anomalies, leading to disproportionately larger yield reductions (Wang et al., 2024). Elevation, serving as a proxy for microclimatic gradients, further shapes these interactions by influencing temperature regimes, growing season length, and solar radiation exposure.

These outcomes imply that improving crop yield forecasting requires not only meaningful data sources but also frameworks capable of capturing cross-domain interactions between climate, vegetation, and land surface properties. While process-based crop models explicitly represent these interactions through predefined biophysical equations, their applicability at large spatial scales is often constrained by the need for site-specific parameterization and the difficulty of representing heterogeneous management conditions and stochastic disturbances. In contrast, DL approaches do not impose fixed functional relationships but instead learn implicit mappings between environmental drivers and yield outcomes directly from data. This allows DL models to capture complex, nonlinear interactions and to implicitly account for factors that are typically not represented in process-based frameworks, such as localized extreme weather events, pest and disease impacts, and non-agronomic influences on yield variability. As a result, DL models can provide more flexible and scalable representations of yield formation under heterogeneous and data-rich conditions. However, this flexibility comes at the cost of reduced interpretability, as DL models do not explicitly resolve the underlying physiological processes driving yield anomalies. Therefore, rather than replacing process-based models, DL approaches should be viewed as complementary tools that enhance predictive performance.

4.3. Contrasting forecast horizons for winter and spring crops

An important operational metric for any early warning in an agricultural system is the maximum lead time at which reliable forecasts can be generated. The evaluation of predictive skill across different temporal horizons (Fig. 6) reveals fundamentally divergent predictability behaviour between winter cereals (wheat, barley) and spring crops (silage maize). A general pattern observed is that forecast accuracy improves progressively as the growing season advances, which is expected as more in-season information becomes available (van der Velde et al., 2018). However, the rate of improvement and the stability of prediction errors differ notably between the crops.

Winter crops, such as wheat and barley, exhibit a more gradual and delayed improvement in forecast skill, with only modest gains observed until the later stages of the growing season. This behaviour can be attributed to their extended growing cycles (often exceeding 300 days) and exposure to a broad range of environmental conditions, including critical overwintering phases. Furthermore, yield formation in these crops is strongly aligned with vegetation dynamics during the flowering and grain-filling stages (Fig. 9). Therefore, truncating the prediction horizon several weeks or months before harvest limits the model's access to these highly informative vegetation signals, reducing its ability to accurately capture final yield outcomes. Consequently, winter wheat shows a significant improvement in predictive skill starting approximately 60 days prior to its late-July harvest (i.e., late May), where the RMSE drops sharply (by ~0.12 t ha^{-1}) and stabilizes near a final RMSE of ~0.74 t ha^{-1} at the end of the season. Winter barley exhibits a similarly delayed but more gradual improvement, stabilizing at an RMSE of ~0.91 t ha^{-1} roughly 40 to 50 days before harvest (early to late June).

In contrast, spring crops, such as silage maize, display a more stable and gradual decline in prediction errors as harvest approaches. This reflects the fact that maize growth occurs within a shorter and more climatically consistent period, where yield formation is continuously shaped by cumulative early (including pre-sowing) and mid-season conditions. Because the temporal Variable Selection Network (VSN) module assigns high weights to pre-sowing and early vegetative phases, the model rapidly gains predictive skill once these informative signals become available. As a result, forecast errors decrease in a much smoother trajectory compared to those of winter crops. For silage maize, reliable predictions can be extracted significantly earlier in the crop's relative life cycle. The prediction errors dipped below 5.0 t ha^{-1} as early as 72 days before the late-September harvest (mid-July) and continued falling sharply to a near-harvest RMSE of ~4.49 t ha^{-1}.

Ultimately, these dynamics emphasize that while early-warning forecasts for spring crops can be operationalized reliably by mid-summer, winter cereals require frequent updates extending through their highly sensitive late-spring phenological stages to achieve actionable risk mitigation.

4.4. Limitations of the current study

Achieving perfect predictive accuracy remains an inherent challenge in agricultural modelling due to the complex interactions among climate, environmental conditions, and management practices. Although CropFusionNet demonstrated strong predictive performance, several limitations should be acknowledged to guide future research and operational implementation.

1. **Data representation constraints:** The aggregation of input variables to the NUTS-3 district level inevitably obscures sub-district heterogeneity in soils, microclimate, vegetation dynamics, and topographic characteristics. In addition, the use of static crop masks (2017–2021) does not fully capture interannual variability in crop rotations and crop development. Furthermore, the framework does not explicitly incorporate dynamic management factors, such as fertilization, irrigation, pesticide application, and cultivar selection, which can substantially influence crop yields.
2. **Preprocessing assumptions:** Missing remote sensing time series were occasionally imputed using data from the nearest neighbouring district, affecting only a small proportion of the training samples (4.2% for winter wheat, 3.4% for winter barley, and 5.5% for silage maize). Although the overall proportion of imputed samples remains low, this spatial fallback may obscure regional crop dynamics in districts with distinctive agro-ecological conditions, potentially introducing noise into the training process and contributing to localized prediction errors.
3. **Distribution shift:** As a data-driven framework, CropFusionNet remains susceptible to distribution shifts under climatic conditions outside the training range. The limited size of the training dataset may also restrict direct transferability to regions with substantially different environmental conditions, making additional regional fine-tuning necessary for out-of-distribution applications.
4. **Operational deployment bottlenecks:** Although inference in CropFusionNet is highly efficient, with real-time yield prediction and attention-weight extraction for all districts across Germany completed within seconds on standard hardware, model training requires high-performance computing resources (e.g., NVIDIA GPUs) and approximately 2–3 h of computation due to the complexity of the model architecture. While our model training phase actively masks post-harvest data using historical phenology observations, in real-time predictive scenarios, exact harvest dates are unknown, forcing reliance on historical median phenology. During extreme climate anomalies, such as severe droughts that accelerate senescence, this reliance may temporarily introduce post-harvest noise into the predictive sequence before the season concludes. Operational deployment is also constrained by the latency of remote sensing products, such as MODIS LAI and FPAR, which are typically available only after an 8–14-day processing delay. Therefore, operational applications of CropFusionNet must account for this observation lag by using the most recently available composites or integrating short-term forecasting approaches for remote sensing inputs.

5. Conclusion

In this study, we introduce CropFusionNet, an interpretable deep learning framework for uncertainty- aware quantile-based crop yield forecasting that effectively integrates static features with daily time-varying agro-environmental inputs. Collectively, our findings highlight the need for a paradigm shift in agricultural modelling: rather than narrowly prioritizing predictive accuracy, deep learning approaches should be designed to capture and reflect the underlying biophysical system dynamics. By capturing complex, nonlinear climate-yield interactions, CropFusionNet overcomes key structural limitations that currently constrain operational forecasting frameworks. To further advance this system-oriented approach, future research should pursue three primary directions:

1. Expanding training data across broader spatio-temporal scales to improve generalization, particularly under shifting and unprecedented climatic regimes.
2. Integrating high-resolution Sentinel-1 and Sentinel-2 time series, field-level annual crop masks, and explicit farm management practices to reduce input uncertainty and improve signal fidelity.

3. Exploring hybrid approaches by coupling CropFusionNet with outputs from process-based models, allowing the architecture to leverage established biophysical priors. Concurrently, applying causal discovery and counterfactual analysis will be essential for distinguishing true physiological drivers from spurious statistical correlates.
4. CropFusionNet was developed and evaluated only for Germany's major crops under Central European conditions; its applicability to other regions, crops, or climates was not assessed. As a data-driven model, its learned relationships are conditioned on the training distribution, limiting direct transfer without fine-tuning. The modular architecture supports region-agnostic use and extending it to additional European regions and crops is identified as a priority for future work.
5. The uncertainty estimates in CropFusionNet primarily capture data-related (aleatoric) uncertainty, while the current deterministic architecture has limited ability to represent model-related (epistemic) uncertainty under out-of-distribution conditions. Future research should focus on disentangling aleatoric and epistemic uncertainty using approaches such as Deep Ensembles or Bayesian neural networks to improve the calibration and reliability of predictive uncertainty estimates.

CRediT authorship contribution statement

Amit Kumar Srivastava: Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Krishnagopal Halder:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Gina Lopez:** Writing – review & editing, Writing – original draft, Methodology. **Kaushik Muduchuru:** Writing – review & editing, Writing – original draft, Validation. **Luis Alfredo Pires Barbosa:** Writing – review & editing. **Kazi Jahidur Rahaman:** Writing – review & editing. **Dominik Behrend:** Writing – review & editing. **Liangxiu Han:** Writing – review & editing. **Claas Nendel:** Writing – review & editing. **Gang Zhao:** Writing – review & editing. **Thomas Gaiser:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Manmeet Singh:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Karthikeyan Lanka:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jingye Han:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Ioannis N. Athanasiadis:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Michael Maerker:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Wenzhi Zeng:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Karam Alsafadi:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jaber Rahimi:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Frank Ewert:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Code availability

The code for this study, including the implementation of CropFusionNet, baseline models, and associated scripts for data preprocessing, evaluation, and figure generation, is openly available on GitHub at: https://github.com/geonextgis/CropFusionNet

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.

Acknowledgments

This research is supported by the Leibniz Association under LL-SYSTAIN (Grant Labs-2024-IOR), BBSRC (BB/S020969, BB/Y513763/1). We gratefully acknowledge the financial support of the German Research Foundation (DFG) - "FAIRagro" project number 501899475. We would also like to acknowledge the project entitled "Supporting Agroecological Transitions of Agroforestry Landscapes using an Innovative Transdisciplinary Assessment Framework" (AGRECO4CAST). The project is Co-funded by the European Union: "AGROECOLOGY Call 1-AGRECO4CAST (ID132)". Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority (funding ID for Germany: 031B1610A) can be held responsible for them. We also acknowledge PARAM-VC project (funding ID for Germany: 01DQ25001A), under Indo-German Science & Technology Centre (IGSTC) 2+2 call, funded jointly by the Department of Science and Technology (DST) under the Government of India and the Federal Ministry of Research, Technology and Space of Germany (BMFTR), "SUSTAINCrop" (HORIZON-EUROPE, funding ID: 0672EMSM), and Pioneer Center for Research in Sustainable Agricultural Futures (Land-CRAFT), DNRF grant number P2, Aarhus University, Denmark.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aiia.2026.08.016.

References

Amini, A., Schwarting, W., Soleimany, A., Rus, D., 2020. Deep evidential regression. Adv. Neural Inf. Proces. Syst. 33, 14927–14937.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.R., Samek, W., 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One 10 (7), e0130140.

BARUTH, B., GENOVESE, G., LEO, O., 2007. CGMS version 9.2-user manual and technical documentation.

Becker, R., Schauberger, B., Merz, R., Schulz, S., Gornott, C., 2025. The vulnerability of winter wheat in Germany to air temperature, precipitation or compound extremes is shaped by soil-climate zones. Agric. For. Meteorol. 361, 110322.

Blickensdörfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasmi, S., Hostert, P., 2022. Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany. Remote Sens. Environ. 269, 112831.

Braden, H., Wetterdienst, D., 1995. The Model AMBETI. (A Detailed Description).

Chen, P., Li, Y., Liu, X., Tian, Y., Zhu, Y., Cao, W., Cao, Q., 2023. Improving yield prediction based on spatio-temporal deep learning approaches for winter wheat: a case study in Jiangsu Province, China. Comput. Electron. Agric. 213, 108201.

Chen, F., Xu, X., Chen, S., Wang, Z., Wang, B., Zhang, Y., et al., 2024. Soil buffering capacity enhances maize yield resilience amidst climate perturbations. Agric. Syst. 215, 103870.

Conradt, T., 2022. Choosing multiple linear regressions for weather-based crop yield prediction with ABSOLUT v1. 2 applied to the districts of Germany. Int. J. Biometeorol. 66 (11), 2287–2300.

Dhillon, M.S., Koellner, T., Asam, S., Bogenreuther, J., Dech, S., Gessner, U., et al., 2025. Landscape structure, climate variability, and soil quality shape crop biomass patterns in agricultural ecosystems of Bavaria. Front. Plant Sci. 16, 1630087.
Ding, D., Zhang, M., Pan, X., Yang, M., He, X., 2019. Modeling extreme events in time series prediction. . In proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining (pp.pp. 1114–1122), July.
Duden, C., Nacke, C., Offermann, F., 2024. German yield and area data for 11 crops from 1979 to 2021 at a harmonized spatial resolution of 397 districts. Sci. Data 11 (1), 95.
Feng, P., Wang, B., Liu, D.L., Yu, Q., Hu, K., 2022. Coupling machine learning with APSIM model improves the evaluation of climate extremes impact on wheat yield in Australia. Modeling processes and their interactions in cropping systems: challenges for the 21st century, pp. 251–275.
Filippi, P., Whelan, B.M., Bishop, T.F., 2024. Explainable machine learning to map the impact of weather and soil on wheat yield and revenue across the eastern Australian grain belt. Agriculture 14 (12), 2318.
Filippi, P., Han, S.Y., Bishop, T.F., 2025. On crop yield modelling, predicting, and forecasting and addressing the common issues in published studies. Precis. Agric. 26 (1), 8.
Friesland, H., Löpmeier, F.J., 2007. The performance of the model AMBAV for evapotranspiration and soil moisture on Müncheberg data. Modelling Water and Nutrient Dynamics in Soil–Crop Systems: Proceedings of the Workshop on “Modelling Water and Nutrient Dynamics in Soil–Crop Systems” Held on 14–16 June 2004 in Müncheberg, Germany. Springer Netherlands, Dordrecht, pp. 19–26 , May.
Fritz, S., See, L., Bayas, J.C.L., Waldner, F., Jacques, D., Becker-Reshef, I., ... McCallum, I., 2019. A comparison of global agricultural monitoring systems and current gaps. Agric. Syst. 168, 258–272.
Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., et al., 2024. Interpreting black-box models: a review on explainable artificial intelligence. Cogn. Comput. 16 (1), 45–74.
Holzworth, D.P., Snow, V., Janssen, S., Athanasiadis, I.N., Donatelli, M., Hoogenboom, G., et al., 2015. Agricultural production systems modelling and software: current status and future prospects. Environ. Model Softw. 72, 276–286.
Howell, T.A., Evett, S.R., 2004. The Penman-Monteith Method. 14. USDA-Agricultural Research Service, Conservation & Production Research Laboratory, Washington, DC.
Iizumi, T., Shin, Y., Choi, J., Van der Velde, M., Nisini, L., Kim, W., Kim, K.H., 2021. Evaluating the 2019 NARO-APCC joint crop forecasting service yield forecasts for northern hemisphere countries. Weather Forecast. 36 (3), 879–891.
Isik, M.S., Ozturk, O., Celik, M.F., 2025. Kolmogorov–Arnold networks for interpretable crop yield prediction across the US corn belt. Remote Sens. 17 (14), 2500.
Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., ... Kim, S.H., 2016. Random forests for global and regional crop yield predictions. PloS one 11 (6), e0156571.
Jönsson, P., Eklundh, L., 2004. TIMESAT—a program for analyzing time-series of satellite sensor data. Comput. Geosci. 30 (8), 833–845.
Kallenberg, M.G., Maestrini, B., van Bree, R., Ravensbergen, P., Pylianidis, C., van Evert, F., Athanasiadis, I.N., 2023. Integrating Processed-based Models and Machine Learning for Crop Yield Prediction arXiv preprint arXiv:2307.13466.
Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. Comput. Electron. Agric. 147, 70–90.
Lacasa, J., Messina, C.D., Ciampitti, I.A., 2023. A probabilistic framework for forecasting maize yield response to agricultural inputs with sub-seasonal climate predictions. Environ. Res. Lett. 18 (7), 074042.
Leng, G., Hall, J.W., 2020. Predicting spatial and temporal variability in crop yields: an inter-comparison of machine learning, regression and process-based models. Environ. Res. Lett. 15 (4), 044027.
Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: a review. Sensors 18 (8), 2674.
Lim, B., Arık, S.Ö., Loeff, N., Pfister, T., 2021. Temporal fusion transformers for interpretable multi-horizon time series forecasting. Int. J. Forecast. 37 (4), 1748–1764.
Lischeid, G., Webber, H., Sommer, M., Nendel, C., Ewert, F., 2022. Machine learning in crop yield modelling: a powerful tool, but no surrogate for science. Agric. For. Meteorol. 312, 108698.
Liu, J., Shang, J., Qian, B., Huffman, T., Zhang, Y., Dong, T., et al., 2019. Crop yield estimation using time-series MODIS data and the effects of cropland masks in Ontario, Canada. Remote Sens. 11 (20), 2419.
Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems 30.
Maestrini, B., Basso, B., 2018. Predicting spatial patterns of within-field crop yield variability. Field Crop Res. 219, 106–112.
Marcinkevičs, R., Vogt, J.E., 2023. Interpretable and explainable machine learning: a methods-centric overview with concrete examples. Wiley Interdiscip. Rev. Data Min. Knowledge Discove. 13 (3), e1493.
Mueller, L., Schindler, U., Shepherd, T.G., Ball, B.C., Smolentseva, E., Pachikin, K., et al., 2013. The Muencheberg soil quality rating for assessing the quality of global farmland. Novel Measurement and Assessment Tools for Monitoring and Management of Land and Water Resources in Agricultural Landscapes of Central Asia. Springer International Publishing, Cham, pp. 235–248.
Nendel, C., Reckling, M., Debaeke, P., Schulz, S., Berg-Mohnicke, M., Constantin, J., et al., 2023. Future area expansion outweighs increasing drought risk for soybean in Europe. Glob. Chang. Biol. 29 (5), 1340–1358.
Nieradzik, L., Stephani, H., Sieburg-Rockel, J., Helmling, S., Olbrich, A., Keuper, J., 2024. Challenging the Black Box: A Comprehensive Evaluation of Attribution Maps of CNN Applications in Agriculture and Forestry. In Proceedings of the 19th International Joint Conference on Computer Vision. Imaging and Computer Graphics Theory and Applications 2, 483–492. https://doi.org/10.5220/0012363400003660 VISAPP; ISBN 978-989-758-679-8, SciTePress.
Olivetti, L., Messori, G., 2024. Advances and prospects of deep learning for medium-range extreme weather forecasting. Geosci. Model Dev. 17 (6), 2347–2358.
Ostberg, S., Schewe, J., Childers, K., Frieler, K., 2018. Changes in crop yields and their variability at different levels of global warming. Earth Syst. Dynam. 9 (2), 479–496.
Overweg, H., Berghuijs, H.N., Athanasiadis, I.N., 2021. Cropgym: a reinforcement learning environment for crop management. arXiv preprint, arXiv:2104.04326.
Pagani, V., Guarneri, T., Busetto, L., Ranghetti, L., Boschetti, M., Movedi, E., Campos-Taberner, M., Garcia-Haro, F.J., Katsantonis, D., Ricciardelli, E., Romano, F., Holecz, F., Collivignarelli, F., Granell, C., Casteleyn, S., Confalonieri, R., 2019. A high-resolution, integrated system for rice yield forecasting at district level. Agric. Syst. 181–190.
Park, Y., Li, B., Li, Y., 2023. Crop yield prediction using bayesian spatially varying coefficient models with functional predictors. J. Am. Stat. Assoc. 118 (541), 70–83.
Qi, Y., et al., 2022. Effects of high temperature and drought stresses on growth and yield of summer maize during grain filling in North China. Agriculture https://doi.org/10.3390/agriculture12111948.
Qiao, D., Dang, Q., Ruan, J., 2025. Deep Learning Struggles to Forecast Yields in Food-insecure, Low-production Regions. https://doi.org/10.21203/rs.3.rs-6800608/v1.
Rashid, M., Bari, B.S., Yusup, Y., Kamaruddin, M.A., Khan, N., 2021. A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. IEEE access 9, 63406–63439.
Ribeiro, A.F.S., Russo, A., Gouveia, C.M., Páscoa, P., Zscheischler, J., 2020. Risk of crop failure due to compound dry and hot extremes estimated with nested copulas. Biogeosci. Discuss. 2020, 1–21.
Ribeiro, M.T., Singh, S., Guestrin, C., 2016 August. “Why should i trust you?” Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144.
Richetti, J., Sadras, V.O., He, D., Leske, B., Hu, P., Beletse, Y., et al., 2025. Challenges in modelling the impact of frost and heat stress on the yield of cool-season annual grain crops. Front. Plant Sci. 16, 1613432.
Schmidt, M., Felsche, E., 2024. The effect of climate change on crop yield anomaly in Europe. Clim. Resilience Sustain. 3 (1), e61.
Shah, A., Kadam, E., Shah, H., Shinde, S., Shingade, S., 2016. Deep residual networks with exponential linear unit. . In proceedings of the third international symposium on computer vision and the internet (pp.pp. 59–65 , September.
Sherstinsky, A., 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Phys. d Nonlinear Phenomena 404, 132306.
Shook, J., Gangopadhyay, T., Wu, L., Ganapathysubramanian, B., Sarkar, S., Singh, A.K., 2021. Crop yield prediction integrating genotype and weather variables using deep learning. PLoS One 16 (6), e0252402.
Stock, M., Pieters, O., De Swaef, T., Wyffels, F., 2024. Plant science in the age of simulation intelligence. Front. Plant Sci. 14, 1299208.
Strer, M., Svoboda, N., Herrmann, A., 2018. Abundance of adverse environmental conditions during critical stages of crop production in northern Germany. Environ. Sci. Eur. 30 (1), 10.
Sun, J., Di, L., Sun, Z., Shen, Y., Lai, Z., 2019. County-level soybean yield prediction using deep CNN-LSTM model. Sensors 19 (20), 4363.
Trevisani, S., Skrypitsyna, T.N., Florinsky, I.V., 2023. Global digital elevation models for terrain morphology analysis in mountain environments: insights on Copernicus GLO-30 and ALOS AW3D30 for a large Alpine area. Environ. Earth Sci. 82 (9), 198.
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. Agric. Syst. 168, 203–212.
van der Velde, M., Baruth, B., Bussay, A., Ceglar, A., Garcia Condado, S., Karetsos, S., et al., 2018. In-season performance of European Union wheat forecasts during extreme impacts. Sci. Rep. 8 (1), 15420.
van der Velde, M., van Diepen, C.A., Baruth, B., 2019. The European crop monitoring and yield forecasting system: celebrating 25 years of JRC MARS bulletins. Agric. Syst. 168, 56–57.
Vandal, T., Kodra, E., Ganguly, A.R., 2019. Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation. Theor. Appl. Climatol. 137 (1), 557–570.
Vani, B.V., Guruprakash, C.D., 2024. Ia-Mstm: Improved adaptive moment estimation enabled mixed attention-based long short-term memory for crop yield prediction. 2024 IEEE North Karnataka Subsection Flagship International Conference (NKCon)IEEE, pp. 1–6 , September.
Vanongeval, F., Gobin, A., 2023. Adverse weather impacts on winter wheat, maize and potato yield gaps in northern Belgium. Agronomy https://doi.org/10.3390/agronomy13041104.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., et al., 2017. Attention is all you need. Adv. Neural Inf. Proces. Syst. 30.
Vogel, E., Donat, M.G., Alexander, L.V., Meinshausen, M., Ray, D.K., Karoly, D., et al., 2019. The effects of climate extremes on global agricultural yields. Environ. Res. Lett. 14 (5), 054010.
Voita, E., Talbot, D., Moiseev, F., Sennrich, R., Titov, I., 2019. Analyzing multi-head self-attention: specialized heads do the heavy lifting, the rest can be pruned. proceedings of the 57th annual meeting of the association for computational linguisticspp. 5797–5808, July.
Wang, X., Cheng, L., Xiong, C., Whalley, W.R., Miller, A.J., Rengel, Z., et al., 2024. Understanding plant–soil interactions underpins enhanced sustainability of crop production. Trends Plant Sci. 29 (11), 1181–1190.
Webber, H., Lischeid, G., Sommer, M., Finger, R., Nendel, C., Gaiser, T., Ewert, F., 2020. No perfect storm for crop yield failure in Germany. Environ. Res. Lett. 15, 104012. https://doi.org/10.1088/1748-9326/aba2a4.
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.

Ye, H., Nicholas, R.E., Roth, S., Keller, K., 2021. Considering uncertainties expands the lower tail of maize yield projections. PLoS One 16 (11), e0259180.

You, J., Li, H., Low, M., Lobell, D., Ermon, S., 2017. Deep Gaussian Processes for Crop Yield Prediction . KDD.

Zachow, M., Ofori-Ampofo, S., Kunstmann, H., Kuzu, R.S., Zhu, X.X., Asseng, S., 2025. Wheat yield forecasts with seasonal climate models and long short-term memory networks. Comput. Electron. Agric. 239, 110965.

Zampieri, M., Ceglar, A., Dentener, F., Toreti, A., 2018. Understanding and reproducing regional diversity of climate impacts on wheat yields: current approaches, challenges and data driven limitations. Environ. Res. Lett. 13 (2), 021001.

Zeng, X., Han, D., Tansey, K., Wang, P., Pei, M., Li, Y., et al., 2025. An interpretable wheat yield estimation model using time series remote sensing data and considering meteorological and soil influences. Remote Sens. 17 (18), 3192.

Zhang, J., Guan, K., Peng, B., Pan, M., Zhou, W., Jiang, C., et al., 2021. Sustainable irrigation based on co-regulation of soil water supply and atmospheric evaporative demand. Nat. Commun. 12 (1), 5549.

Zheng, J., Yu, L., Du, Z., Xiao, L., Huang, X., 2025. Modeling wheat development under extreme weather with WOFOST-EW v1. Geosci. Model Dev. 18 (21), 8379–8400.

Zhu, W., Baumert, J., Storm, H., Heckelei, T., Siebert, S., 2025. ECIRA-European crop-specific irrigated area at 1 km resolution annually from 2010 to 2020. Sci. Data 12 (1), 1349.

Ziliani, M.G., Altaf, M.U., Aragon, B., Houborg, R., Franz, T.E., Lu, Y., et al., 2022. Early season prediction of within-field crop yield variability by assimilating CubeSat data into a crop model. Agric. For. Meteorol. 313, 108736.