Agricultural and Forest Meteorology 389 (2026) 111404

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Winter wheat yield responses to extreme heat in the North China Plain: GGCMI-AgMIP model evaluation and a heat-stress phenology improvement in JULES-crop

Lirong You [a,*], Karina Bett-Williams [b,c], Stephen Sitch [a], Christoph Müller [h], Huiyi Yang [d,e], Jefferson Goncalves De Souza [a], Munir Hoffmann [i], Maria Carolina Duran Rojas [a], Beiyao Xu [f,g]

[a] *Faculty of Environment, Science and Economy, University of Exeter, Exeter, EX4 4QF, UK*
[b] *Global Systems Institute, University of Exeter, Exeter, EX4 4PY, UK*
[c] *UK Met Office, Fitzroy Road, Exeter, EX1 3PB, UK*
[d] *Natural Resources Institute, University of Greenwich, Chatham, SE10 9LS, UK*
[e] *School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK*
[f] *Institute for Climate and Atmospheric Science, University of Leeds, Leeds, LS2 9JT, UK*
[g] *School of Atmospheric Sciences, Nanjing University, Nanjing, 210023, China*
[h] *Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Potsdam, 14473, Germany*
[i] *Tropical Plant Production and Agricultural Systems Modelling (TROPAGS), University of Goettingen, Grisebachstr. 6, Goettingen, 37077, Germany*

ARTICLE INFO

Keywords:
Climate change
Crop model evaluation
Winter wheat
Heat stress
Phenology

ABSTRACT

Crop models are used within the GGCMI-AgMIP framework to assess climate impacts on crop yields, but their ability to reproduce observed yield responses to regional climatic variability and extremes remains insufficiently evaluated. To help fill this evaluation gap, we use the North China Plain (NCP), China, as a regional case study and combine prefecture-level winter wheat yield and climate records with GGCMI Phase 2 simulations. The analysis focuses on three growing-season climatic drivers: mean temperature, precipitation, and extreme heat days (EHD), defined as growing-season days with daily maximum temperature exceeding 34 °C. Rather than evaluating yield levels alone, we compare effective yield–climate response types for these climatic drivers, using observed responses as a benchmark for diagnosing model behaviour. We use targeted JULES-crop experiments to test whether a heat-stress phenology modification can reduce the diagnosed bias in simulated extreme-heat responses. Observations show a widespread significant negative linear yield response to EHD in the central NCP, whereas responses to growing-season mean temperature and precipitation are more spatially heterogeneous. Most GGCMI models fail to reproduce this observed response type to extreme heat, suggesting a systematic weakness in the representation of heat-stress impacts. The original JULES-crop configuration fails to capture significant negative linear yield-EHD relationships at the observed heat-sensitive sites. Incorporating heat-stress effects into phenological development and senescence improves this functional response in JULES-crop, increasing the number of sites with a consistent significant negative linear yield-EHD relationship from 0 to 7 out of 16 observed significant sites. Although this targeted modification does not eliminate all model-observation discrepancies, it demonstrates that phenological sensitivity to extreme heat is a plausible process-level source of model bias. This study extends GGCMI Phase 2 evaluation from broad-scale yield-level comparison towards regional, response-based diagnosis, and identifies heat-stress phenology as a practical pathway for improving winter wheat simulations under intensifying heat extremes.

1. Introduction

Wheat is a staple crop that sustains billions of people worldwide, with China being one of the largest producers and accounting for approximately 18% of global wheat production (Liu et al., 2021). The North China Plain (NCP) is one of the most important wheat-producing regions in China, accounting for approximately 71% of the country's wheat production (Sun et al., 2006). Winter wheat dominates wheat cultivation in China, contributing approximately 90% of total wheat yield and planting area (of Statistics of China, 2015). Climate

* Corresponding author.
E-mail address: gk25435@bristol.ac.uk (L. You).

https://doi.org/10.1016/j.agrformet.2026.111404
Received 14 November 2025; Received in revised form 31 July 2026; Accepted 1 August 2026
Available online 12 August 2026

change and climate variability are expected to affect wheat production through changes in temperature, precipitation, and the frequency of extreme climatic events (Kang et al., 2009; Lv et al., 2013; Yang et al., 2017). In the NCP, more frequent extreme high-temperature events and heavy precipitation have been projected, with potentially important consequences for wheat production (H. Wang et al., 2022; Bai et al., 2021). These conditions make the NCP a valuable regional case for assessing whether crop models can reproduce observed winter wheat yield responses to climatic variability and extremes.

Crop models are widely used to assess climate impacts on crop yields and to support regional to global yield projections (Jägermeyr et al., 2021; Leng, 2017; Zhang et al., 2018). However, substantial uncertainty remains in simulated wheat yield responses to climatic drivers (Müller et al., 2021; Wang et al., 2020, 2017; Martre et al., 2015). This uncertainty is not only related to the magnitude of simulated yields, but also to whether models reproduce the observed functional relationships between yield and climatic drivers (Corbeels et al., 2018; Wang et al., 2017). Agreement between simulated and observed yields over historical periods does not necessarily ensure that crop models reproduce yield responses under climate-change or extreme conditions (Easterling et al., 1996). Recent analyses from the Global Gridded Crop Model Intercomparison (GGCMI) have therefore emphasized the need to evaluate functional responses to individual drivers, rather than relying only on historical yield observations (Müller et al., 2024). Therefore, in addition to conventional yield-level comparison, response-based evaluation is needed to diagnose whether crop models capture the effective yield–climate relationships observed at regional scales.

GGCMI Phase 2, conducted under the Agricultural Model Intercomparison and Improvement Project (AgMIP), provides a multi-model framework for assessing crop yield responses to climate variability at the global scale. GGCMI models have been widely applied to investigate climate impacts on major crops across different regions worldwide and have demonstrated reasonable skill in capturing large-scale yield variability (Müller et al., 2017; Franke et al., 2020). However, previous evaluations have also shown substantial regional differences in model performance, including relatively low correlations between simulated and observed wheat yields in China (Franke et al., 2020). It therefore remains unclear how well GGCMI Phase 2 models reproduce observed yield–climate response types within specific high-production regions such as the NCP.

The key climatic drivers considered in this study are growing-season mean temperature, precipitation, and extreme heat days (EHD). Among these climatic drivers, extreme heat is of particular concern for winter wheat production. Many existing crop models do not fully represent the effects of extreme high-temperature events on crop yield (Webber et al., 2016; Rezaei et al., 2015b; B. Liu et al., 2016a), despite experimental and observational evidence that heat stress can strongly affect crop yield and quality (Akter and Rafiqul Islam, 2017; Hossain et al., 2013). In wheat, high temperatures during reproductive and grain-filling stages can accelerate development, shorten the grain-filling period, enhance senescence, and reduce final yield (Farooq et al., 2011; B. Liu et al., 2016b). Previous studies have further suggested that heat-stress effects on crop phenology are important but are often insufficiently represented in crop models (Rezaei et al., 2015a; Kumar et al., 2023; Mankin et al., 2024). This limitation is particularly relevant for winter wheat systems in the NCP, where regional phenological patterns, irrigation practices, and late-season heat events interact to shape yield responses (N. Wang et al., 2022; S. Wang et al., 2022; Yang et al., 2022). Thus, heat-stress effects on phenological development represent a plausible process-level source of bias in simulated winter wheat responses to extreme heat.

JULES-crop provides a suitable process-based model framework for testing this potential source of bias. JULES-crop is a crop modelling module embedded within the Joint UK Land Environment Simulator (JULES), designed to simulate crop growth and yield responses to climate and management practices at regional to global scales (Osborne et al., 2014; Williams et al., 2017). Although JULES-crop (JULES-crop-original) contributed to GGCMI Phase 2, winter wheat outputs from JULES-crop were not available in the GGCMI Phase 2 dataset analysed here. Nevertheless, its explicit representation of phenological development makes JULES-crop a suitable process-based framework for investigating heat-stress phenology as a potential source of bias in simulated winter wheat responses to extreme heat.

In this study, we combine prefecture-level winter wheat yield and climate records from the NCP with GGCMI Phase 2 simulations and targeted JULES-crop experiments. Specifically, we aim to: (1) characterize observed winter wheat yield response types to growing-season mean temperature, precipitation, and EHD; (2) evaluate whether GGCMI Phase 2 models and the original JULES-crop configuration reproduce these observed regional response types; and (3) test whether incorporating heat-stress effects into JULES-crop phenological development improves the simulated response of winter wheat yield to extreme heat. By linking observed yield–climate response benchmarks with multi-model evaluation and a targeted process-level model experiment, this study extends GGCMI Phase 2 evaluation from broad-scale yield-level comparison towards regional, response-based diagnosis. It also identifies heat-stress phenology as a plausible contributor to model bias and as a practical pathway for improving winter wheat simulations under intensifying heat extremes.

2. Data and methods

2.1. Winter wheat yield data and meteorological data

Wheat yield data at the prefecture level for four provinces in the NCP — Shandong, Henan, Hebei, and Anhui — were obtained from the provincial Statistical Yearbooks of the National Bureau of Statistics of China (of Statistics of China, 2020), covering the period from 1990 to 2019. Meteorological station data for the same period, including mean temperature, maximum temperature, and precipitation, were obtained from the National Meteorological Information Center of the China Meteorological Administration (CMA) (China Meteorological Data Service Center, 2024). Meteorological data were first linked to yield records at the prefecture level by matching records from the same prefecture. For prefectures with both yield records and local meteorological station data, observations from the station located within the same prefecture were used. If no meteorological station data were available for a given prefecture, daily meteorological observations were assigned from the geographically nearest prefecture with available station data, based on the shortest geographic distance. Each prefecture was thus associated with a single, continuous meteorological time series for subsequent analysis. Because the analysis focuses on detrended interannual variability and relative yield–climate relationships rather than absolute climate magnitudes, these scale mismatches are expected to have less influence on the response-based evaluation, although some uncertainty remains.

2.2. Model output data from GGCMI phase 2

We used GGCMI Phase 2 as a standardized multi-model benchmark to evaluate whether current global crop models reproduce observed regional responses of winter wheat yield to growing-season mean temperature, precipitation, and EHD. GGCMI Phase 2 provides historical yield simulations from 14 crop models for the period 1980–2010 under a range of experimental conditions, including different atmospheric CO_2 concentrations, temperature, water and nitrogen levels, and adaptation assumptions (Müller et al., 2017). To ensure consistency with observed yield data and to focus on baseline model performance under historical conditions, we selected the GGCMI simulations forced with a CO_2 concentration of 360 ppm (C360), no temperature offset (T0), no adaptation, full irrigation (Winf), and high nitrogen input (N200). The irrigation and nitrogen assumptions are broadly consistent with intensive winter wheat management in the North China Plain, where irrigation and fertilizer inputs are widely applied to avoid strong water or nutrient limitations (Koch et al., 2020). Under these selection criteria, seven crop models provided complete winter wheat yield outputs: APSIM-UGOE, EPIC-TAMU, GEPIC, LPJ-GUESS, LPJmL, pDSSAT, and PEPIC. Gridded yield data were matched to prefecture-level observations using the nearest grid cell to the prefecture centroid.

2.3. JULES-crop simulation

To further understand the performance of crop models in simulating winter wheat in the NCP and improve the simulation of yield under extreme high-temperature conditions, we conducted a study using the JULES-crop model. JULES-crop contributed to GGCMI Phase 2; however, it did not provide data on winter wheat. JULES is the land-surface component of the UK Met Office Unified Model and the next generation of UK Earth System Model (UKESM) (Best et al., 2011; Clark et al., 2011). JULES-crop is a crop-specific parameterization in JULES, which is developed for studying both the impact of climate change on crop yields and the impact of crops on the climate, especially in regions where the majority of the vegetation is crops. In JULES-crop, each crop is considered as an additional plant functional type (PFT) with extra parameters to be specified, which are different from natural PFTs (Osborne et al., 2014; Williams et al., 2017). For each crop PFT in each year, the JULES-crop parameterization sows, accumulates the effective temperature, partitions net primary productivity (NPP) into carbon pools during crop development, and finally harvests. JULES also includes options for irrigating PFTs. Thermal times for each crop PFT are provided in a spatially-varying ancillary, which allows a (very limited) representation of differences between crop varieties.

The meteorological data required for driving JULES-crop include air temperature, air pressure, specific humidity, precipitation, wind speed and downward fluxes of short-wave and long-wave radiation. We used ERA5, the fifth generation of global atmospheric reanalysis data provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). Because the driving data are not AgMERRA as used in GGCMI Phase 2 (for temperature, ERA5 and AgMERRA show a high correlation coefficient of about 0.98), our comparison does not assess absolute model skill but rather the emergent functional performance revealed through regression analysis, i.e., the extent to which each model's yield responds consistently to its driving meteorological variables. The input winter wheat ancillary data of sowing and harvest dates are from Liu et al. (2014), which are used to generate the thermal time ancillary. This ensures that the baseline growing season length reflects local observations before additional modifications are made to represent heat effects on phenology. Thus, our JULES-crop analysis is regarded as a region-specific evaluation rather than a strict protocol-compliant comparison with the GGCMI ensemble.

Two versions of the JULES-crop model were used in this study. The baseline version, hereafter referred to as *JULES-crop-original*, follows the standard JULES-crop configuration and does not explicitly account for the effects of extreme heat on crop phenology. The modified version, referred to as *JULES-crop-improved*, incorporates heat stress effects into phenological development by modifying the effective temperature calculation and introducing a heat-induced senescence acceleration, thereby allowing extreme high temperatures to accelerate crop development and maturity.

2.4. Heat stress index

The heat stress index is used to measure the impact of extreme high temperatures on winter wheat growth, primarily reflecting the severity of high-temperature conditions during the winter wheat growing season. Different heat stress indices focus on intensity, frequency, or duration when describing the severity of extreme heat events (B. Liu et al., 2016b). In this study, we employ the concept of extreme heat days (EHD) to investigate the impact of extreme high temperatures. This is commonly used in studies on winter wheat in the NCP (Shi and Shi, 2016). Extreme heat days are defined as days within the winter wheat growing season where the daily maximum temperature exceeds the threshold of 34 °C (Shi and Shi, 2016). In Eqs. (1) and (2), EHD represents the number of extreme heat days, n is the number of days in the growing season, d is the daily extreme heat day count factor, T_{max} is the daily maximum temperature, and T_{crit} is the temperature threshold, set here at 34 °C.

$$EHD = \sum_{t=1}^{n} d_{eh,t} \tag{1}$$

$$d_{eh,t} = \begin{cases} 0, \mathrm{T_{max}} \le \mathrm{T_{crit}} \\ 1, \mathrm{T_{max}} > \mathrm{T_{crit}} \end{cases} \tag{2}$$

2.5. Regression analysis

Univariate regression analysis is used to assess the response of NCP winter wheat yield to individual climatic drivers (mean temperature, precipitation, and EHD) during the growing season. As these factors are not independent, univariate regression does not isolate a pure causal effect but rather captures the effective or bulk relationship between yield and each factor. By this we mean the overall correlation that reflects both direct and indirect influences. Such effective relationships are useful for comparing the behaviour of different models against each other and against observations. However, they should not be interpreted as causal responses of yield to a single climatic variable, but only as emergent functional patterns suitable for evaluation purposes.

Following Xu et al. (2022), we characterized site-level yield responses to each climatic driver using single-factor regression analysis, with the primary objective of classifying response types rather than achieving optimal predictive fit. We considered linear and quadratic relationships because they capture the two dominant patterns relevant for model evaluation: (i) monotonic responses (linear) and (ii) non-monotonic responses with an optimum/threshold-like curvature (quadratic). More flexible forms (e.g., segmented or higher-order regressions) were not adopted because the available record length at each prefecture is limited, and additional parameters would increase the risk of overfitting and reduce comparability across sites and models.

For each prefecture, we fitted a linear model,

$$Y = a_1 + b_1 X, \tag{3}$$

where Y is the detrended winter wheat yield, X is the detrended climatic driver (mean temperature, precipitation, or EHD), a_1 is the intercept, and b_1 is the linear slope.

We also fitted a quadratic model,

$$Y = a_2 + b_2 X + c_2 X^2, \tag{4}$$

where a_2 is the intercept, b_2 is the linear-term coefficient, and c_2 is the quadratic-term coefficient. For negative (convex) quadratic relationships ($c_2 < 0$), the optimum X value is

$$X_o = -\frac{b_2}{2c_2}. \tag{5}$$

Response types were classified using coefficient significance tests ($p \le 0.05$). A site was classified as *linear* if b_1 was significant and c_2 was not significant; the sign of b_1 indicates positive or negative linear response. A site was classified as *quadratic* if c_2 was significant (regardless of the significance of b_2), with the sign of c_2 indicating concave or convex curvature; otherwise, the site was classified as *non-responsive*. To reduce non-climatic effects (e.g., technology trends), all yield and meteorological time series were detrended using the locally weighted regression model (LOWESS) (Cleveland, 1979; Cleveland and Devlin, 1988; Lu et al., 2017).

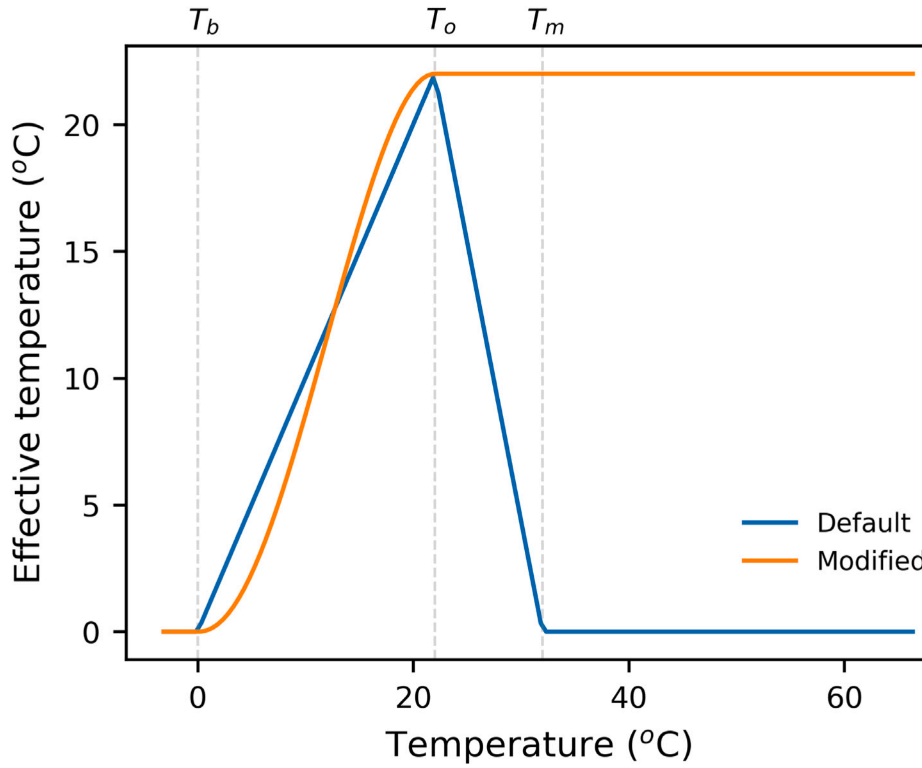

Fig. 1. Effective temperature calculations. The default calculation used in JULES-crop-original is shown as the blue line. The modified routine is displayed as the orange line. The base temperature (T_b, 0 °C), optimum temperature (T_o, 22 °C), and maximum temperature (T_m, 32 °C) are indicated by dashed lines and follow the standard JULES-crop configuration.

2.6. Heat-stress phenology modification in JULES-crop

Accurate representation of crop phenological stages under high temperatures is critical for modelling yield in wheat. To minimize uncertainties in crop yield simulations, it is important to improve temperature response in phenological development and senescence in models (Wang et al., 2017). Building on that motivation, our modifications to the JULES-crop-original model primarily focus on improving the effective temperature calculation that drives phenology, so that extremely high temperatures are treated more appropriately.

Thermal time is used to measure the accumulation of heat required at various growth stages during the plant growth process, and it is also a variable controlled by crop models to regulate crop phenological stages (Saiyed et al., 2009). In the standard version of JULES-crop (JULES-crop-original), the calculation of thermal time is related to effective temperature (Osborne et al., 2014). In Eq. (6), T_{eff} represents effective temperature, and T_b, T_o, and T_m are the base temperature, optimum temperature, and maximum temperature, respectively, and are specified for each crop type individually. The temperature thresholds (T_b, T_o, T_m) were held constant across all simulations to isolate the effects of phenological formulation. Cultivar-specific parameter tuning and sensitivity tests were not conducted, as the objective of this study is to evaluate relative differences between model formulations rather than to optimize cultivar-specific thresholds. The maximum temperature parameter (T_m = 32°C) used here defines the upper bound of the thermal response curve for phenological development and does not represent a physiological damage threshold. In contrast, the 34 °C threshold used previously to define EHD follows regional observational studies and serves as a statistical indicator of extreme heat events. These two temperature thresholds therefore serve different purposes and are not directly comparable. In this calculation, when the temperature exceeds T_o, T_{eff} decreases as temperature increases, indicating that the higher the temperature, the longer it takes for the crop to reach the next stage (Fig. 1).

$$T_{\text{eff}} = \begin{cases} 0, & T \le T_{\text{b}} \text{ or } T \ge T_{\text{m}}, \\ T - T_{\text{b}}, & T_{\text{b}} < T \le T_{\text{o}}, \\ (T_{\text{o}} - T_{\text{b}}) \left(1 - \dfrac{T - T_{\text{o}}}{T_{\text{m}} - T_{\text{o}}}\right), & T_{\text{o}} < T < T_{\text{m}}. \end{cases} \tag{6}$$

However, studies have shown that high temperature events during the reproductive stage accelerate development and lead to advanced maturity of wheat Farooq et al. (2011), B. Liu et al. (2016b), and the reproductive stage of NCP wheat occurs around May, making it susceptible to high temperatures. Therefore, we modified the calculation of T_{eff} to better suit the characteristics of winter wheat:

$$T_{\text{eff}} = \begin{cases} 0, T \le T_{\text{b}} \\ T_o \left(\dfrac{T_m - T}{T_m - T_o} \times \dfrac{T - T_b}{T_o - T_b} \right)^{\frac{T_o - T_b}{T_m - T_o}} & , \quad T_b < T \le T_o \\ T_o, T > T_{\text{o}} \end{cases} \tag{7}$$

We used the ‘Beta routine’ described by B. Liu et al. (2016b), who compared four temperature-response routines for wheat phenology and found that the Beta routine reproduced heat-stress effects more reliably than the alternatives. In this formulation, T_{eff} reaches its maximum value at T_o and remains constant above T_o, thereby accelerating development and maturity relative to the original calculation.

In addition, following B. Liu et al. (2016b), we implemented a senescence acceleration function (HT_{eff}) to further capture premature leaf senescence induced by heat stress.

While this does not explicitly simulate physiological senescence (e.g., leaf yellowing), it is consistent with the simplified phenology framework of JULES-crop-original, as it models a swifter decline in the green leaf LAI. NT_{eff} represents the effective temperature under conditions without heat stress (Eq. (7)). HT_{eff} is determined by the cumulative daily heat degree days (HDD), calculated as detailed in

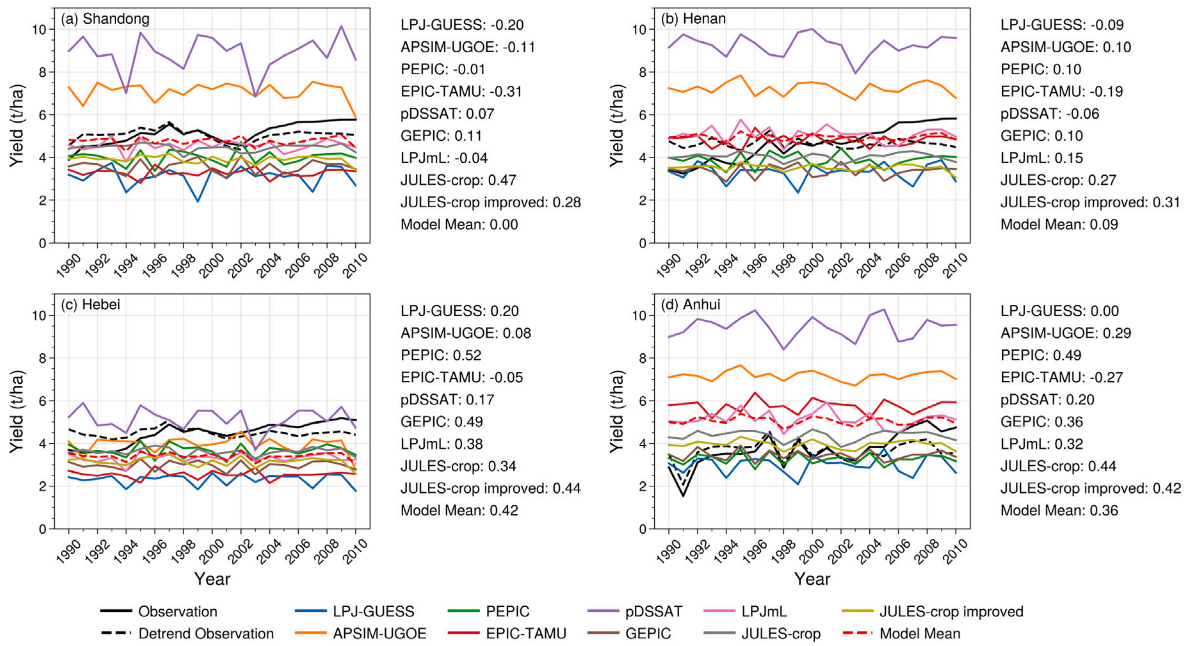

Fig. 2. Time series of winter wheat yields from observations and model outputs from 1990 to 2010. Four provinces in the NCP are listed, Shandong (a), Henan (b), Hebei (c), and Anhui (d). A set of correlation coefficients between detrended modelled and observed yields are displayed at the right of each panel.

Eqs. (9) and (10), where $AHDD$ represents the extent of heat stress experienced by wheat from the post-heading stage to the end of the maturation period. HDD quantifies the degree to which hourly temperatures exceed a threshold temperature (T_h), as detailed in Eqs. (11) and (12). HTS denotes the sensitivity of wheat to high temperatures and is associated with the wheat variety. GDD_{RGP}, or growing degree days, is set at 520 °C d, following B. Liu et al. (2016b).

$$T_{eff} = NT_{eff} + HT_{eff} \tag{8}$$

$$HT_{eff} = \frac{HTS \times AHDD_i}{GDD_{RGP}} \tag{9}$$

$$AHDD_i = \sum_{j=1}^{i} HDD_j \tag{10}$$

$$HDD_j = \frac{1}{24}\sum_{t=1}^{24} HD_t \tag{11}$$

$$HD_t = \begin{cases} 0, T < T_h \\ T - T_h, T \geq T_h \end{cases} \tag{12}$$

3. Results

3.1. Model–data comparison of winter wheat yield

The time series of observed and modelled winter wheat yield from 1990 to 2010 across four provinces in the NCP (Shandong, Henan, Hebei, and Anhui) reveals notable discrepancies between model simulations and observed data (Fig. 2). The multi-model mean yield is closer to the observed values in terms of absolute yield levels in Shandong and Henan (approximately 4–6 t/ha), but it underestimates by about 1 t ha^{-1} in Hebei and overestimates by around 1 t ha^{-1} in Anhui. Among the models, pDSSAT and APSIM-UGOE consistently overestimate yield, particularly in Shandong, Henan, and Anhui, whereas LPJ-GUESS tends to underestimate yield across all regions in the NCP.

To evaluate interannual variability, correlation coefficients between detrended simulated and detrended observed yields are calculated for each province. Generally, the coefficients of multi-model mean remain relatively low in Shandong (0.00) and Henan (0.09) (Fig. 2a,b). But for Hebei and Anhui, there are higher correlation coefficients of 0.42 and 0.36, respectively. Compared to the other models, PEPIC, GEPIC, LPJmL, JULES-crop-original, and JULES-crop-improved demonstrate a better agreement with observed interannual variability in these two provinces. However, EPIC-TAMU, APSIM-UGOE, and LPJ-GUESS exhibit relatively poor performance, showing negative correlations in certain provinces. Overall, the correlation results show pronounced spatial heterogeneity in model–observation agreement across the four provinces.

The simulations of winter wheat yield in the NCP indicate substantial variability across different models, and there are also pronounced regional variations within each model, even at the provincial level. For example, in many prefectures of Anhui (9 out of 15 prefectures), JULES-crop (including its improved version) and APSIM-UGOE frequently achieve significant positive correlations, whereas PEPIC, GEPIC, and LPJmL capture significant positive correlations at 2–3 prefectures in Hebei (Fig. 3). This discrepancy reveals that these models tend to perform better at reproducing interannual yield variability in the southern and eastern parts of the NCP, whereas EPIC-TAMU and LPJ-GUESS exhibit negative correlations across much of the region, including 9 out of 17 prefectures in Shandong, suggesting notable biases (Fig. 3a,d). In several prefectures across the NCP, JULES-crop-original showed slightly higher model–observation yield correlation coefficients than JULES-crop-improved. Correlation coefficients also varied substantially within individual provinces; for example, some prefectures in Shandong and Henan showed significant positive correlations, whereas others showed weak or negative correlations.

3.2. Observed yield responses to climatic drivers

3.2.1. Observed response of winter wheat yield to climatic drivers

Before examining yield responses, we briefly assess the interdependence among mean temperature, precipitation, and EHD during the wheat growing season. EHD is positively correlated with mean temperature and negatively correlated with precipitation across much of the NCP, whereas mean temperature and precipitation show weak or insignificant correlations in most prefectures (Fig. A.1). This indicates that EHD captures information on extreme thermal stress that is not fully represented by seasonal mean temperature or precipitation

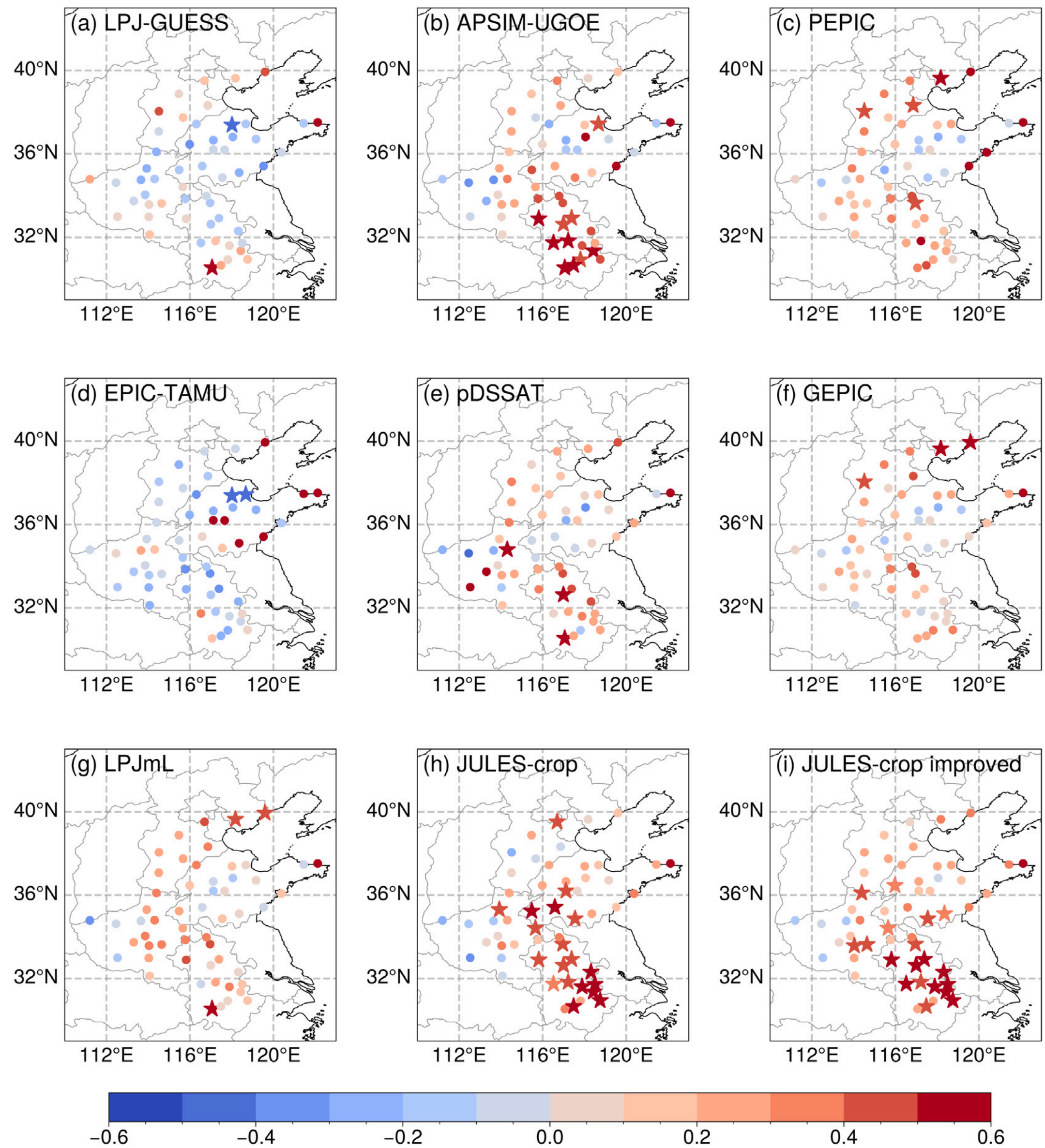

Fig. 3. Correlation coefficients between modelled and observed yields at prefectures in the NCP. Distributions of coefficients from LPJ-GUESS (a), APSIM-UGOE (b), PEPIC (c), EPIC-TAMU (d), pDSSAT (e), GEPIC (f), LPJmL (g), JULES-crop-original (h), and JULES-crop-improved (i) are shown in each panel. The stars indicate a significance level less than 0.05.

alone, supporting its inclusion as a separate explanatory variable in the subsequent yield response analysis.

To determine the response types of the NCP winter wheat yields to various climatic factors, univariate linear and quadratic regression analyses are conducted between each climatic factor and yield. Fig. 4 indicates that four prefectures in Henan and Anhui show a significant negative quadratic response to the average temperature during the growth period, with the optimum average temperature from 14.9 °C to 15.7 °C. The linear slopes are $-353.3\,\mathrm{kg\,ha^{-1}\,°C^{-1}}$ in Anhui, $-323.0\,\mathrm{kg\,ha^{-1}\,°C^{-1}}$ in Shandong, and $266.5\,\mathrm{kg\,ha^{-1}\,°C^{-1}}$ in Hebei. Here, the linear slope represents the regression coefficient from the univariate linear regression between detrended winter wheat yield and a given climatic driver at the prefecture level, quantifying the change in yield per unit change in the climatic variable. For EHD, many observation points in northern Shandong and western Henan suggest a negative linear response, with an average slope of $-87.1\,\mathrm{kg\,ha^{-1}\,d^{-1}}$, meaning that an increase in EHD would significantly reduce wheat yield in these prefectures. Significant negative responses to EHD occurred in more prefectures than significant responses to growing-season mean temperature.

The response of winter wheat yield to local variations in precipitation is statistically significant in 38% of the analysed prefectures. A positive linear relationship is primarily observed in Henan and Shandong, with an average slope of $552.1\,\mathrm{kg/ha/mm/d}$. In contrast, a negative linear response is evident in Anhui, Shandong, and Hebei, with an average slope of $-361.0\,\mathrm{kg/ha/mm/d}$. Despite these linear trends, the dominant yield–precipitation relationship follows a negative quadratic pattern. The spatial distribution of optimal precipitation values further reveals a decreasing trend from south to north, with maximum yield occurring at approximately 2.0 mm/day in the southern regions and 0.8 mm/day in the northern regions.

3.3. Modelled response of winter wheat yield to climatic drivers

Based on the observed wheat yield responses in the NCP to various climatic drivers (Section 3.2), modelled responses in this region from GGCMI simulations can be further evaluated.

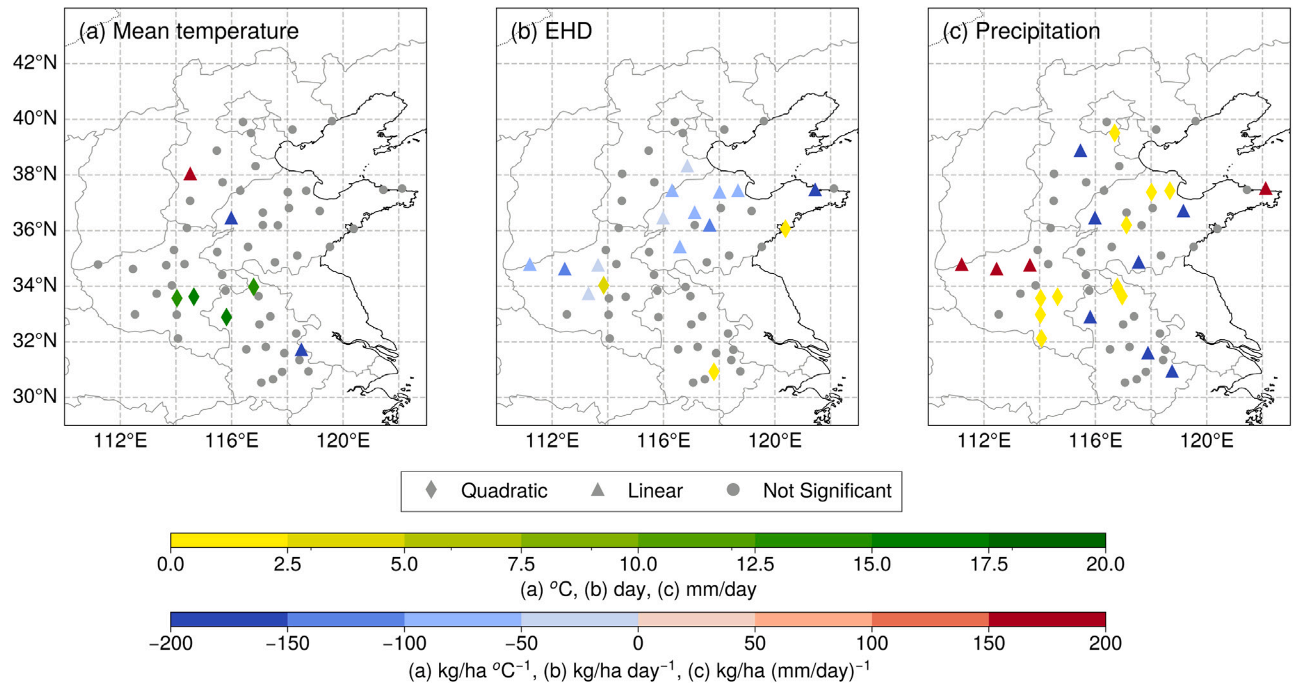

Fig. 4. Linear regression slopes (kg ha^{-1} per unit climatic variable) and quadratic regression optimum values from observations. (a) shows the regression between yields and mean temperature, (b) regression between yields and EHD, and (c) regression between yields and precipitation. Sites with linear slopes are displayed as triangles, while quadratic regressions are represented by diamonds. Grey dots indicate sites with no statistical significance (P-value > 0.05).

Yield response to mean temperature during the growing season varies considerably among models, with some showing a higher proportion of linear or quadratic regressions, while others exhibit no significant regression in most prefectures in the NCP (Figs. A.5, A.3). For example, positive linear responses are found in the central NCP in GEPIC, and a small number of prefectures in APSIM-UGOE (5 prefectures), PEPIC (2 prefectures), and JULES-crop-original (1 prefecture) (Figs. A.5, 5d, Table A.2). A large number of negative linear responses are observed in the northern NCP for EPIC-TAMU and in the southern NCP for pDSSAT. Compared to the observed linear regression, GEPIC reproduces the increasing linear response between yield and mean temperature in 1 site in western Hebei, with a slope of $256\,\mathrm{kg\,ha^{-1}\,°C^{-1}}$ (observed: $233\,\mathrm{kg\,ha^{-1}\,°C^{-1}}$). Negative quadratic response types consistent with the observations were simulated mainly by pDSSAT and APSIM-UGOE, with the optimal average temperature ranging from 14.1 °C to 16.2 °C for pDSSAT and from 7.0 °C to 11.0 °C for APSIM-UGOE. However, both models exhibit a northern offset, with only one prefecture in pDSSAT aligning with observations.

Some GGCMI models simulate significant negative linear yield–EHD responses at only a limited number of sites in the NCP. In contrast, JULES-crop-improved reproduces this response at 7 of the 16 sites where observations show a significant negative yield–EHD relationship, compared with none for JULES-crop-original (Fig. 5b,e,h), attributed to an adjusted effective temperature calculation. As shown in Fig. A.6, the JULES-crop-improved consistently reaches maturity earlier than the default version across years, indicating that the negative yield response to EHD primarily arises from accelerated phenological progression. The average slopes range from $-172\,\mathrm{kg\,ha^{-1}\,d^{-1}}$ to $-55\,\mathrm{kg\,ha^{-1}\,d^{-1}}$, closely matching the observed range of −32 to $-110\,\mathrm{kg\,ha^{-1}\,d^{-1}}$.

Under irrigation conditions, pDSSAT (9 sites), LPJ-GUESS (5 sites) and APSIM-UGOE (2 sites) show some negative quadratic responses to precipitation in the NCP, with an optimum value of 1.8 to 3.3 mm/day, higher than the observed value. LPJmL and APSIM-UGOE also show numerous negative linear responses.

In summary, the performance of models' yield responses to mean temperature, the EHD, and precipitation in the NCP region remains generally poor. All models, except for the JULES-crop-improved, fail to exhibit a consistent negative linear response to EHD in the central NCP. Accurately capturing the impact of extreme high temperatures on yield remains a challenge for most crop models.

4. Discussion

4.1. Observed yield–climate responses in the NCP

Our observational analysis across four provinces in the NCP (Shandong, Henan, Hebei, and Anhui) reveals a detailed relationship between winter wheat yields and climatic drivers (Fig. 4). EHD provides a useful diagnostic variable for capturing yield responses to extreme heat stress. Although EHD is related to both growing-season mean temperature and precipitation, it captures short-duration high-temperature events that are not fully represented by seasonal mean temperature or precipitation alone.

The more spatially widespread significant negative yield response to EHD, compared with that to growing-season mean temperature (Fig. 4a,b), indicates that short-duration heat extremes are more consistently associated with winter wheat yield losses across the NCP, consistent with previous studies (Yang et al., 2017; Ullah et al., 2022). One possible reason that wheat in these northern areas suffers more pronounced heat stress is that its heading and maturity stages occur relatively late, meaning that the critical and most heat-sensitive phases of its growth cycle coincide with the hottest period of the year. This timing likely exacerbates the impact of heat stress (Liu et al., 2014).

Meanwhile, in most regions, the mean temperature during the growth period does not show a significant response type (Fig. 4a), which might be attributed to the varying impacts of average temperatures at different growth stages. As found by Warrington et al. (1977), lower temperatures during the ear development phase are conducive to a higher number of potentially fertile florets, whereas temperatures prior to anthesis do not significantly affect final ear weight. Furthermore, Mamrutha et al. (2020) observed that high night-time temperatures during the grain-filling period significantly reduce

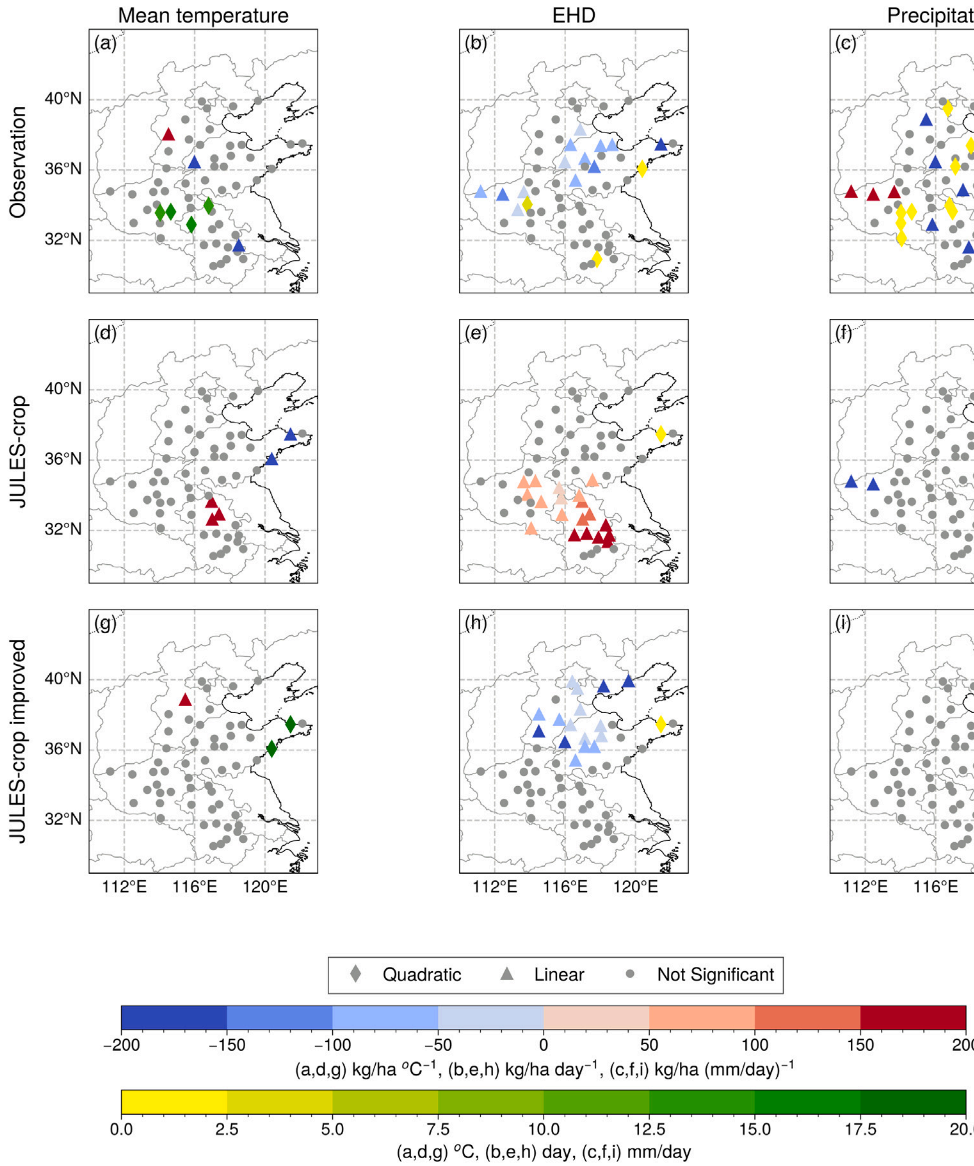

Fig. 5. Linear regression slopes and quadratic regression optimum values between yields and mean temperature (a, d, g), the EHD (b, e, h), and precipitation (c, f, i) from observation (a, b, c), JULES-crop-original (d, e, f) and JULES-crop-improved (g, h, i). Sites with linear slopes are displayed by triangles; negative quadratic responses are shown as diamonds; positive quadratic responses are indicated by purple squares. The black dots indicate sites where the significance level is not reached (P-value > 0.05).

yields, while high nighttime temperatures before tillering and during the booting stage do not significantly impact yield. Alternatively, it is also possible that the interannual variation in mean temperature is simply not large enough to exert a meaningful influence on yield in these regions during the observation period. Therefore, a more refined investigation into the role of temperature at distinct growth stages — for example, by combining stage-specific field observations with model experiments, or through controlled studies such as growth chamber or field manipulation experiments — might provide clearer insights into the role of mean temperature.

Regarding precipitation, we observed different response types. For regions with negative linear responses in Anhui province (Fig. 4c), abundant rainfall, often accompanied by continuous cloudy and rainy conditions with lower solar radiation, is a likely cause of decreased wheat yields (Thorne and Wood, 1987; Abbate et al., 1997). Additionally, excessive rainfall may damage crops through multiple pathways. Apart from increasing the risk of root diseases (Tao et al., 2008), extreme precipitation events can reduce nitrogen availability during key developmental stages, thereby limiting tillering, and may also disrupt pollination and reduce grain filling (Fu et al., 2023).

Observational records indicate an increasing trend in precipitation in these regions over recent decades, accompanied by more frequent flood events (Wang, 2007; S. Wang et al., 2015). Enhancing agricultural drainage systems may possibly address the negative impact of current and future increases in precipitation. Regions with positive linear responses to precipitation (western Henan) typically experience drier climates where precipitation has not yet reached an optimal level (Fig. 4c). In contrast, in regions with negative quadratic responses (across all provinces in NCP), precipitation below the optimal threshold boosts yields, but yields decline once rainfall exceeds that threshold. These varying response patterns may stem from differences in soil water-holding capacity (Lehane and Staple, 1965), irrigation conditions (Chen et al., 2014), and management practices. Moreover, as described in the results, the spatial distribution of optimal precipitation shows a decreasing trend from south to north, with higher optimal values observed in warmer southern regions. This pattern suggests that in warmer climates, wheat may require more precipitation to achieve optimal yields, likely due to increased evaporative demand under higher temperatures.

4.2. Modelled yield–climate responses and diagnosed biases

In addition to observed yield–climate relationships, this study provides insights into the ability of different crop models to reproduce regional winter wheat responses to climatic drivers in the NCP. Our results indicate notable uncertainties in how different models capture the effects of temperature, extreme heat, and precipitation (Fig. A.5). This is consistent with prior experience from GGCMI Phase 1 (Müller et al., 2019), where even under unified inputs and management assumptions, models exhibited discrepancies in reproducing yields and their spatiotemporal responses to climatic drivers (Müller et al., 2017). Such differences were further confirmed in the GGCMI Phase 2 CTWN (CO_2–Temperature–Water–Nitrogen) experiment: even under comparable forcing conditions and perturbation magnitudes, different models' simulations of single climate factors or their interactions still diverged substantially, leading to considerable uncertainty from regional to global scales (Franke et al., 2020). These inter-model variations in climate response can amplify uncertainty in future yield projections, as shown in Müller et al. (2021), where GGCMI models were forced with CMIP5 and CMIP6 projections. More recently, by analysing multiple models' sensitivities to the same driving forces, Müller et al. (2024) emphasized that significant differences in response patterns exist, calling for functionality-based evaluation and comparison to clarify the sources of these disparities.

The generally moderate model–observation correlations for detrended yields, often below 0.6, likely reflect a combination of model structural limitations and uncertainties arising from the comparison between prefecture-level yield statistics and gridded crop model simulations. The substantial variation in correlations among prefectures within the same province, such as Shandong and Henan, further suggests that local environmental conditions and management heterogeneity may influence model–observation agreement at sub-provincial scales. Prefecture-level yield statistics aggregate diverse local management practices, including cultivar choice, irrigation scheduling, and fertilizer application, which are not fully represented in the models. In addition, non-climatic factors such as pests, diseases, and policy-driven management changes may introduce variability in reported yields that is not represented in climate-driven crop models, thereby reducing model–observation correlations. Similar ranges of model–observation correlation have been reported in previous regional and multi-model crop evaluations, particularly for wheat in China (Franke et al., 2020). These limitations indicate that correlation-based evaluation alone is insufficient for diagnosing model behaviour, motivating the response-based comparison of yield–climate relationships used here.

The comparison between JULES-crop-original and JULES-crop-improved illustrates the distinction between aggregate yield correlation and targeted response-based evaluation. Across the NCP, the modified phenology does not uniformly improve model–observation yield correlations (Fig. 3h, i): JULES-crop-improved shows higher correlations than JULES-crop-original in 38 out of 54 prefectures with available data, and the mean correlation increases from 0.25 to 0.30. However, correlations do not improve everywhere, with JULES-crop-original showing higher correlations in 16 prefectures. This is not unexpected because the yield correlation coefficient is an aggregate metric calculated over the full time series and reflects multiple sources of interannual yield variability, including years and locations where extreme heat is not the dominant yield driver. By contrast, the heat-stress phenology modification is designed to improve the simulated response to extreme heat. Importantly, the modified configuration improves the targeted yield–EHD response metric. This distinction highlights the value of response-based evaluation.

In our study, while a few models (e.g., pDSSAT, APSIM-UGOE, and GEPIC) reproduce the observed responses to temperature and precipitation in a limited number of sites, most fail to capture the widespread negative yield response to EHD, particularly in central Shandong and western Henan (Fig. A.5). The closer alignment of JULES-crop-improved with the observed EHD response is consistent with the incorporation of heat-stress effects into phenological development and senescence (Fig. 5), suggesting that better representation of heat stress processes may enhance model performance in this region. Moreover, recent studies comparing alternative approaches with the GGCMI framework further support the importance of explicitly representing high-temperature effects. For example, Lucia Martin et al. (2025) examined winter wheat suitability across France using climate storyline scenarios and showed that warm summers combined with mild winters substantially reduce the area suitable for winter wheat cultivation. Their results indicate that insufficient vernalization under warmer winters leads to a marked contraction of winter wheat area, with losses exceeding those estimated within the GGCMI framework. This comparison suggests that GGCMI-based assessments may underestimate the adverse impacts of high-temperature conditions when key physiological constraints, such as vernalization and heat sensitivity, are not explicitly represented.

4.3. Implications

Our findings highlight that extreme heat presents a pronounced threat to winter wheat yields in significant portions of the NCP, especially in Shandong and Henan, where observational evidence shows a clear negative linear impact of EHD on yield (Fig. 4b). Although high-temperature events often coincide with drought, and moderate increases in irrigation can help alleviate heat stress (Tack et al., 2017; Birthal et al., 2021) by maintaining adequate soil moisture and providing canopy cooling (Siebert et al., 2017), caution is warranted in certain regions. For example, in parts of central Shandong that display a negative linear response to EHD and a negative linear or quadratic response to precipitation, increasing irrigation should be approached carefully. Excess water can reduce water use efficiency (WUE), worsen soil aeration, and increase disease pressure (Tao et al., 2008), and may also cause nitrogen leaching that limits nutrient availability for crops (Fu et al., 2023), ultimately diminishing the benefits of mitigating heat stress. In such areas, balancing the timing and volume of irrigation is crucial to avoid oversupply while still protecting crops during high-temperature events. Additionally, other measures, such as applying exogenous protectants (e.g., salicylic acid or KCl) (Akter and Rafiqul Islam, 2017; Prasad et al., 2020) or using moderate amounts of nitrogen fertilizer (Ru et al., 2024), may be better suited to address heat stress without the drawbacks associated with excessive irrigation.

The diversity of yield responses to temperature, precipitation, and EHD across the GGCMI models underscores the importance of accurately representing wheat developmental processes (Fig. A.5). Our assessment indicates that many models in the GGCMI ensemble fail to capture the negative linear impact of EHD. Incorporating calculations related to heat stress into these models would potentially improve their performance in this region. First, phenological characteristics determine the timing of crop growth and development, directly influencing biomass accumulation and distribution, and thus yield simulations. Consequently, phenology is a critical process in crop models (Craufurd et al., 2013). Introducing heat stress into phenology calculations can enhance the accuracy of crop phenology predictions (Shi et al., 2015; B. Liu et al., 2016b). While phenology validation was not conducted in this study, the modified heat stress function was adopted from B. Liu et al. (2016b), who demonstrated improved agreement between simulated phenology and field observations. However, it should be noted that improved yield simulations may not always correspond to more accurate phenology representations (Zhu et al., 2019). Second, heat stress accelerates grain senescence and reduces photosynthetic

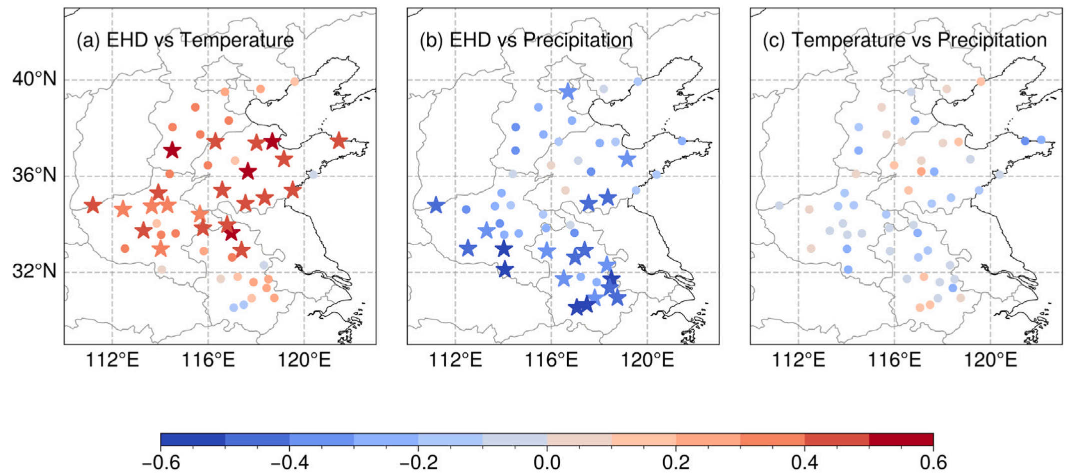

Fig. A.1. Correlations between EHD and mean temperature (a), EHD and mean precipitation (b), mean temperature and mean precipitation (c). The stars indicate the sites with P-value ≤ 0.05.

capacity, leading to a decrease in final grain weight and grain number (Shirdelmoghanloo et al., 2016; Mirosavljević et al., 2021). By integrating a process-based approach that includes such effects, Liu et al. (2020) demonstrated improved agreement between simulated and observed yields. Alternatively, empirical reduction functions for final yield or grain number are also commonly employed (Rezaei et al., 2015b). Similarly, where observations show negative yield responses to precipitation (Fig. 4c), crop models may require process-based or empirical representations of excess-rainfall damage that account for regional soil and management conditions (Li et al., 2019; Sereyrorth et al., 2020; Stephens and Lyons, 1998).

4.4. Uncertainties and limitations

This analysis provides valuable insight into model performance by assessing whether models can capture the overall or "effective" relationships between yield and climate variables. Evaluating how well these bulk correlations are reproduced helps identify key discrepancies across models. However, because temperature, EHD, and precipitation are interdependent in the observational record, this approach cannot disentangle the independent contributions of each driver, which is essential for guiding improvements in process-based models. A promising way to pinpoint the quantitative link between each environmental factor and yield loss is through the design of multi-factor field experiments (H.-l. Wang et al., 2015; Zhang et al., 2022), which capture yield responses under diverse heat and water conditions.

In addition, there are further limitations to note. First, the heat stress effect incorporated into the JULES-crop-improved phenology calculation acts directly on the duration of wheat development rather than on yield itself, following previous studies where model-simulated growth-phase lengths were compared with field experimental data (Shi et al., 2015; B. Liu et al., 2016b). In contrast to these studies, however, our evaluation compares modelled yields against observed yields. Related to this, it is important to note that the heat stress routine in the model, by accelerating wheat maturation (Fig. A.6), potentially enables NCP wheat to evade subsequent heat stress events. If harvest dates are not accurately captured, this may introduce errors especially when additional process-based methods representing heat stress effects are incorporated. Previous work has shown that the quantitative attribution of yield loss to high temperature is highly sensitive to how phenological timing is represented, underscoring the importance of accurate harvest date simulation (Zhu et al., 2019). Second, differences in climate forcing represent an important source of uncertainty across the model evaluations. GGCMI simulations are driven by AgMERRA, whereas the JULES-crop simulations use ERA5. Previous GGCMI analyses have shown that the choice of climatic forcing dataset can exert strong regional influences on simulated crop productivity and the representation of extreme events (Ruane et al., 2021). While AgMERRA temperature has been shown to be highly consistent with ERA5 in the study region (mean correlation ~0.98), differences between them may influence the magnitude of simulated yield variability and climate sensitivities. However, the objective of this study is to compare relative yield response patterns across models rather than to assess absolute model skill, which reduces sensitivity to systematic forcing biases. Third, the relatively coarse spatial resolution of the crop model outputs (0.5° for GGCMI) introduces additional uncertainty when matching gridded simulations to prefecture-level yield statistics. Representing each administrative unit using a single nearest grid cell may not fully capture local heterogeneity in climate and management conditions, potentially affecting site-level relationships. Regional crop-model simulations have been shown to be sensitive to both spatial input resolution and spatially varying management (Constantin et al., 2019). In addition, the GGCMI simulations analysed in this study are based on the Winf (water-sufficient) configuration, which represents a standardized irrigation scenario. Although winter wheat production in the North China Plain is predominantly irrigated (Koch et al., 2020), Winf should be interpreted as an upper-bound approximation of irrigation intensity rather than a perfect representation of field management. Due to the lack of spatially explicit irrigation amount data, more nuanced irrigation simulations were not feasible. Consequently, some precipitation-related yield variability in reality may not be fully reflected in the standardized water-sufficient simulations. However, because our objective is to compare relative yield response patterns under consistent management assumptions rather than to reproduce absolute yield skill under site-specific water constraints, this limitation does not alter the main conclusions of the study.

Finally, two types of fixed temperature thresholds are used in this study: the 34 °C threshold used to define EHD as a diagnostic index, and the T_b, T_o, T_m, and T_h parameters used in the JULES-crop phenology and heat-stress formulations. The use of a common EHD threshold facilitates consistent comparison across observations and models, but may not fully capture spatial variation in actual heat sensitivity arising from cultivar traits. Recent studies have shown substantial variation in heat tolerance among crop varieties and marked geographic heterogeneity in temperature thresholds for heat-induced yield loss, suggesting that the use of fixed thresholds may bias estimates of extreme-heat exposure and the associated yield losses (Sun et al., 2025; Zhao et al., 2026). Therefore, part of the model–observation mismatch in yield–EHD relationships may reflect the use of a common diagnostic threshold, in addition to differences in model structure and

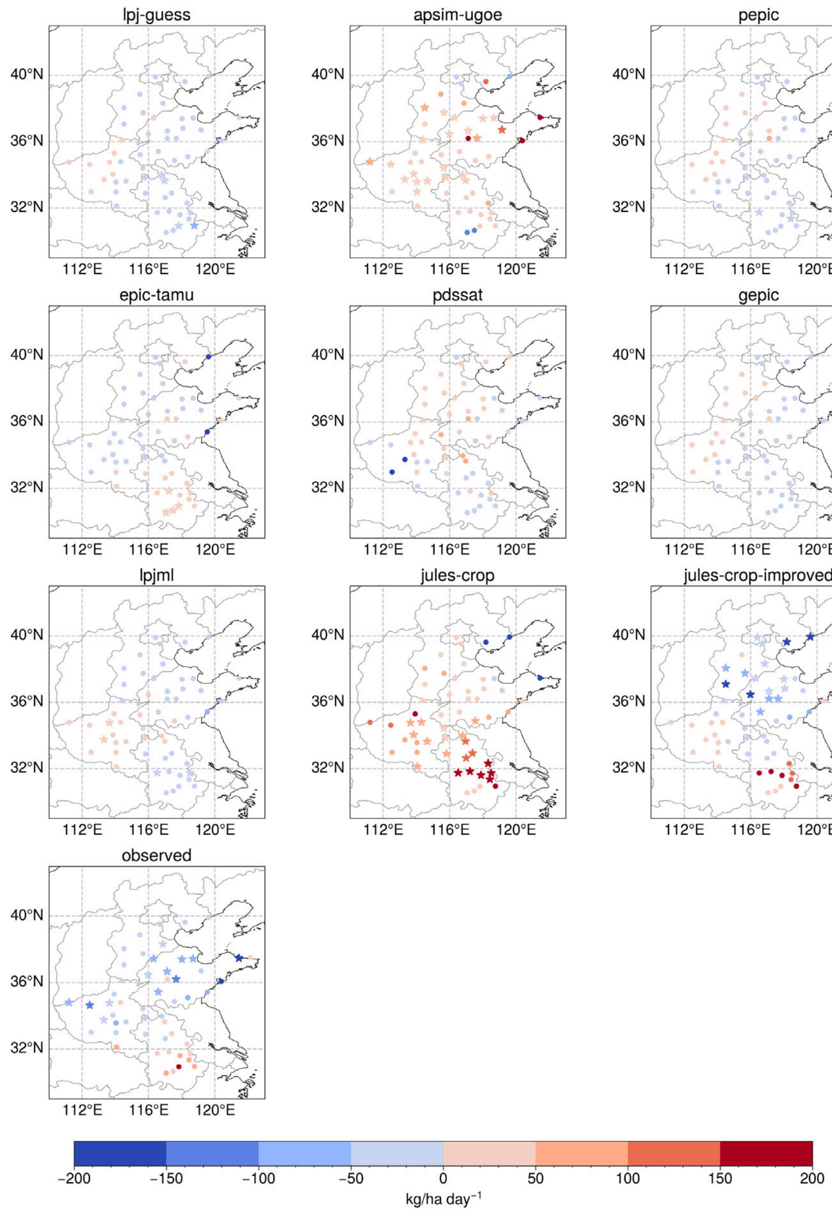

Fig. A.2. Linear regression slope (unit: kg/ha) between model output yields and EHD. The stars indicate sites where the significance level is reached (P-value < 0.05).

heat-stress representation. Future work should evaluate crop-, cultivar-, and region-specific thresholds and represent varietal heterogeneity more explicitly.

5. Conclusion

This study provides a systematic evaluation of winter wheat yield responses to climatic drivers in the North China Plain (NCP) by combining observations with multi-model simulations from the GGCMI framework and targeted JULES-crop experiments. Using a regression-based approach, we diagnose how well current crop models reproduce effective yield–climate relationships across a major wheat-producing region.

Observations reveal that EHD shows a widespread negative association with winter wheat yield in the middle of the NCP, with significant responses occurring in more prefectures than for growing-season mean temperature in many locations. In contrast, yield responses to mean temperature and precipitation are more spatially heterogeneous and often nonlinear, reflecting interactions among climate, soil, irrigation, and management conditions.

Most GGCMI models struggle to capture the observed negative linear yield response to EHD, highlighting a systematic weakness in representing heat stress processes. By explicitly incorporating heat stress effects into phenological calculations, the modified JULES-crop configuration demonstrates improved alignment with observed EHD–yield relationships in key regions. This result underscores the importance of representing heat stress mechanisms, particularly phenological sensitivity to high temperatures, for improving model realism under climate change conditions.

Overall, our findings highlight both the strengths and limitations of the current GGCMI ensemble for simulating winter wheat yields in the NCP. They suggest that improving crop model performance will require not only refinements to heat stress representation, but also better treatment of excess rainfall effects, growth-stage-specific temperature sensitivity, and regional management heterogeneity. These insights provide a clearer basis for interpreting model-based yield projections

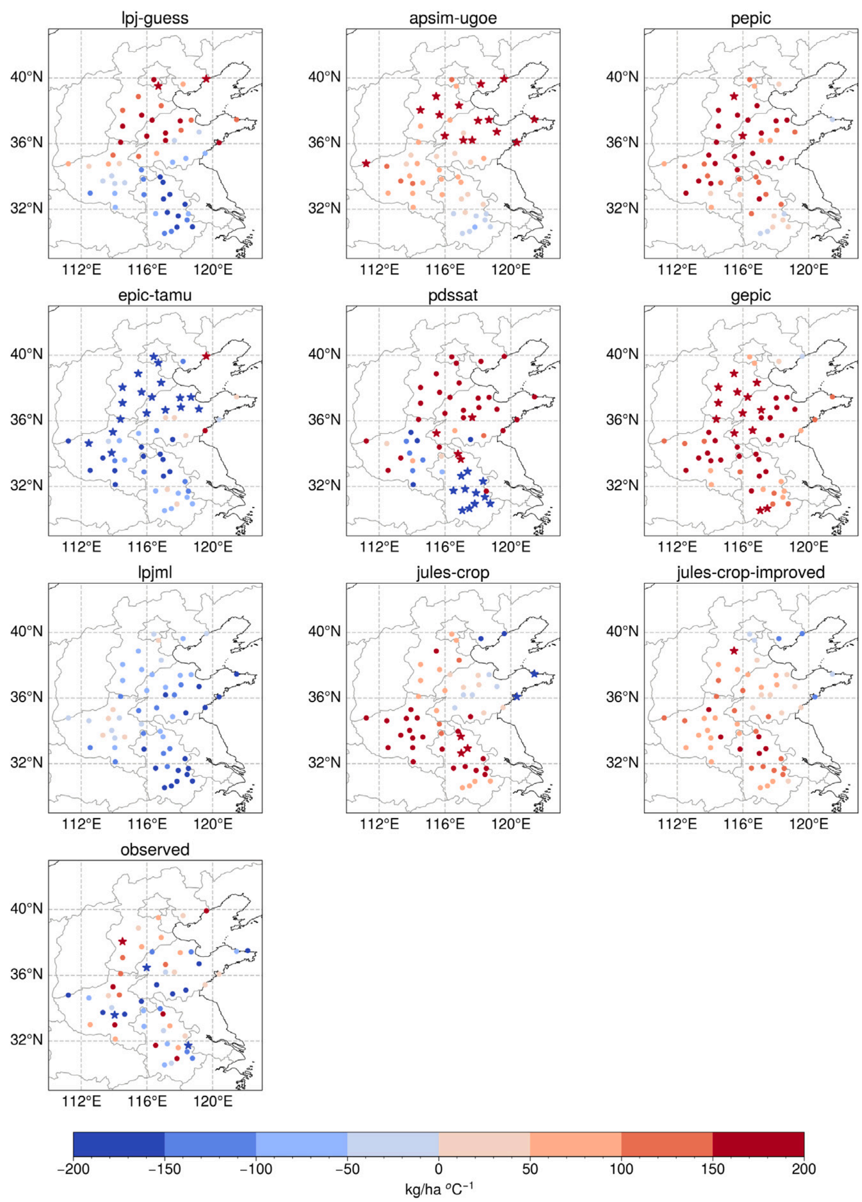

Fig. A.3. Linear regression slope (unit: kg/ha) between model output yields and mean temperature. The stars indicate sites where the significance level is reached (P-value < 0.05).

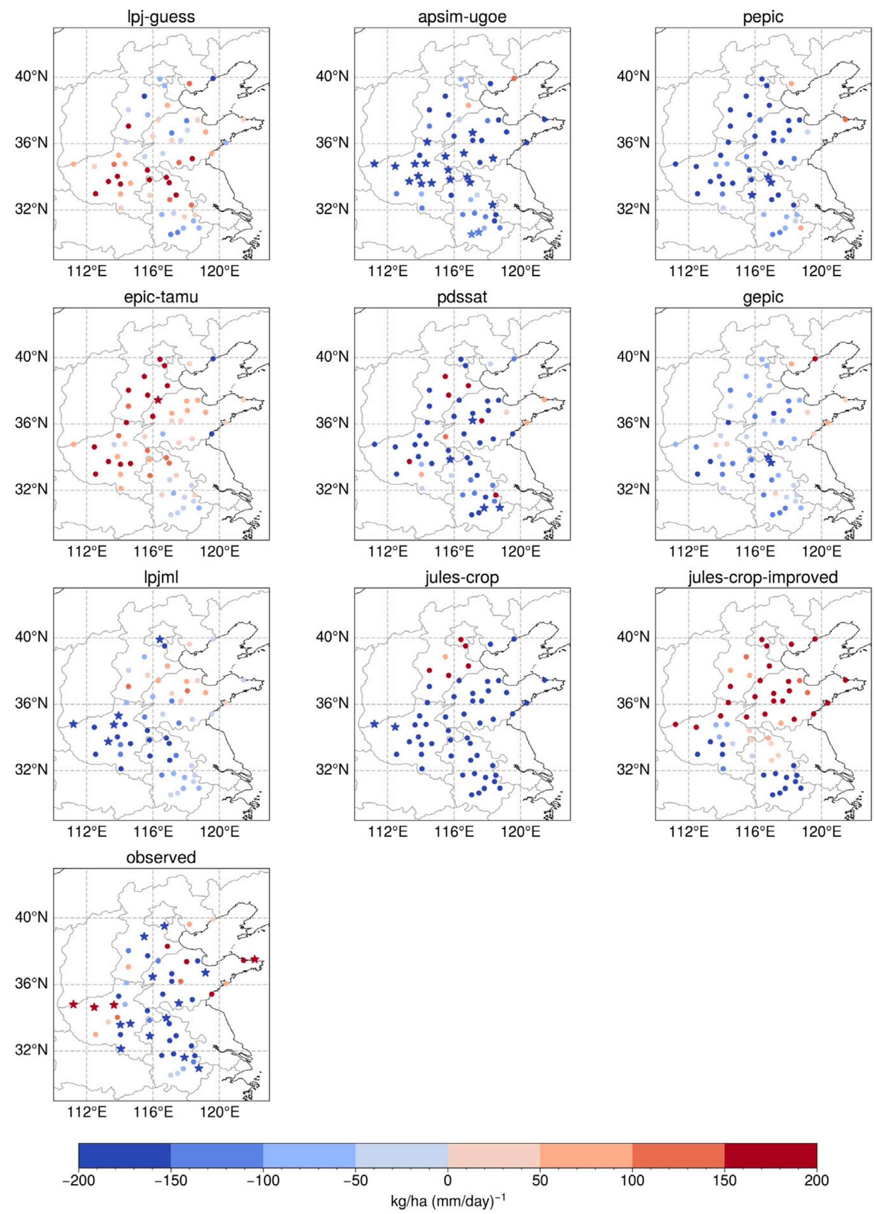

Fig. A.4. Linear regression slope (unit: kg/ha) between model output yields and precipitation. The stars indicate sites where the significance level is reached (P-value < 0.05).

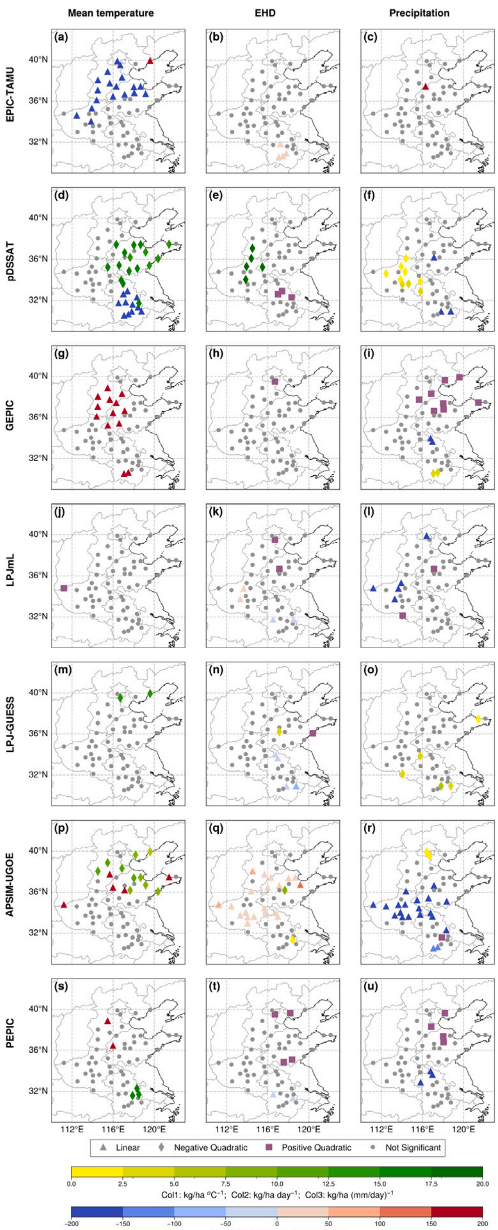

Fig. A.5. Linear regression slopes and quadratic regression optimum values between yields and mean temperature, EHD, mean precipitation from GGCMI Phase 2 models. Sites with linear responses are represented by triangles; negative quadratic responses are shown as diamonds; positive quadratic response patterns are indicated by purple squares. Each column is the regression of mean temperature, EHD, and precipitation. Rows are the results from different GGCMI models. The grey dots indicate sites where the significance level is not reached (P-value > 0.05).

and for guiding future model development in heat- and water-stressed agricultural systems.

CRediT authorship contribution statement

Lirong You: Writing – review & editing, Writing – original draft, Conceptualization, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Karina Bett-Williams:** Writing – review & editing, Supervision, Software, Methodology, Formal analysis, Conceptualization. **Stephen Sitch:** Writing – review & editing, Supervision, Software, Formal analysis. **Christoph Müller:** Writing – review & editing, Formal analysis. **Huiyi Yang:** Writing – review & editing, Supervision, Formal analysis. **Jefferson Goncalves De Souza:** Writing – review & editing, Software. **Munir Hoffmann:** Writing – review & editing, Software. **Maria Carolina Duran Rojas:** Writing – review & editing, Software. **Beiyao Xu:** Writing – review & editing.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT in order to improve the readability and language. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Funding sources

This work was supported by the China Scholarship Council [Grant number 202104910080]; the European Union's Horizon 2020 research and innovation programme, 4C project [grant number 821003]; UKRI-BBSRC, NetZeroPlus (NZ+) project [grant number BB/V011588/1]; and UKRI 'AI for Net Zero' Programme, ADD-TREES project [grant number EP/Y005597/1].

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 article.

Acknowledgements

We are grateful to Mike O'Sullivan from the Faculty of Environment, Science and Economy, University of Exeter, for his guidance on model simulations. We also thank Huiling Zou from the British Antarctic Survey for her assistance with formatting the manuscript. We thank Hao Yu from the Department of Mathematics at the University of Exeter for his assistance in collecting data. Their support was crucial to the completion of this work. We also acknowledge JASMIN, the UK data analysis facility, for providing the computational platform used for data processing and model simulations. We also gratefully acknowledge the use of the Monsoon2 system, a collaborative facility supplied under the Joint Weather and Climate Research Programme, a strategic partnership between the Met Office and the Natural Environment Research Council, for initial JULES training and testing.

Appendix. Supplementary materials

See Figs. A.1–A.6 and Tables A.1 and A.2

Data availability

Data will be made available on request.

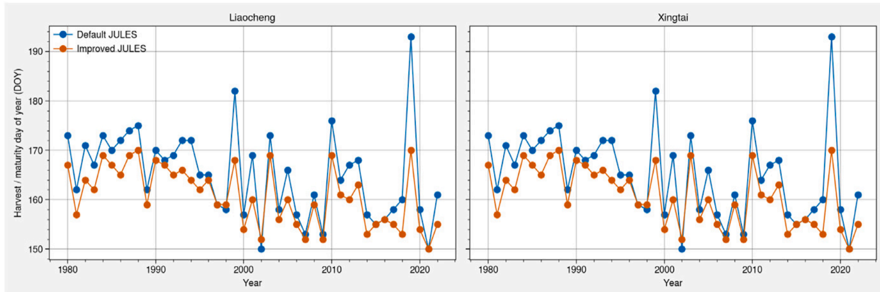

Fig. A.6. Interannual variation in simulated wheat harvest (maturity) dates derived from daily crop developmental index (DVI, cpft=0) for two representative sites (Liaocheng and Xingtai), where model yields exhibit a significant negative linear relationship with EHD. Harvest date is defined as the first day when $DVI \geq 1.99$. The improved JULES-crop configuration systematically reaches maturity earlier than the original configuration across years, indicating accelerated phenological development and a shortened grain-filling period under heat stress.

Table A.1
Overview of crop models used in this study.

Model	Model category	Explicit heat stress	Primary pathway of temperature impact
APSIM (Keating et al., 2003; Holzworth et al., 2014)	Process-based farming systems crop model	No	Temperature controls phenological development via thermal time and regulates crop growth through general physiological response functions.
pDSSAT (Jones et al., 2003; Elliott et al., 2014)	Process-based system crop model	No	Temperature regulates photosynthesis, maintenance respiration and phenology, and directly influences reproductive processes including pod/seed set, kernel number formation and grain filling.
EPIC (Izaurralde et al., 2006)	Process-based system crop model	No	Temperature reduces potential biomass growth through a general temperature stress index in the radiation-use-efficiency (RUE) based growth framework.
GEPIC (Liu et al., 2007; Folberth et al., 2012)	Process-based system crop model (GIS-based EPIC)	No	Same as EPIC: temperature acts via a general stress factor that down-regulates potential biomass growth within the RUE framework.
PEPIC (W. Liu et al., 2016b,a)	Process-based system crop model (parallel EPIC implementation)	No	Same as EPIC: temperature influences crop growth through a general temperature stress index in the RUE-based growth formulation.
LPJ-GUESS (Lindeskog et al., 2013; Olin et al., 2015)	Process-based vegetation model (DGVM with crop PFTs)	No	Temperature regulates process-based photosynthesis and respiration rates and controls phenology via heat-unit accumulation.
LPJmL (von Bloh et al., 2018)	Process-based vegetation model (DGVM with crop functional types)	No	Temperature regulates process-based photosynthesis and respiration and controls crop development via heat-unit accumulation.
JULES-crop-original (Osborne et al., 2015; Williams et al., 2017)	Process-based land surface model with crop representation (optional dynamic vegetation module available)	No	Temperature regulates process-based photosynthesis and respiration and controls phenology via heat-unit accumulation.
JULES-crop-improved (Osborne et al., 2015; Williams et al., 2017; B. Liu et al., 2016b)	Process-based land surface model with crop representation (modified JULES-crop configuration)	Yes	Temperature regulates process-based photosynthesis and respiration. Explicit heat-stress effects are represented through a modified effective-temperature response that accelerates phenological development above the optimum temperature.

Table A.2
Summary of model skill in reproducing observed yield–climate response types at site level in NCP. Each cell reports *match /both_sig /obs_sig*: **match** is the number of sites where both the model and observation show a significant relationship and the response *type* matches (positive/negative linear, or quadratic where applicable); **both_sig** is the number of sites where both the model and observation are significant; **obs_sig** is the number of sites with a significant relationship in observations.

Model	Mean temperature	EHD	Precipitation
EPIC-TAMU	1/2/7	0/1/16	0/0/20
pDSSAT	1/2/7	1/1/16	3/6/20
GEPIC	1/2/7	0/0/16	0/3/20
LPJmL	0/0/7	0/3/16	0/3/20
LPJ-GUESS	0/0/7	0/2/16	1/2/20
APSIM-UGOE	0/2/7	0/10/16	1/9/20
PEPIC	0/2/7	0/0/16	1/4/20
JULES-crop-original	0/0/7	0/3/16	0/2/20
JULES-crop-improved	0/0/7	7/8/16	0/0/20

References

Abbate, P., Andrade, F., Culot, J., Bindraban, P., 1997. Grain yield in wheat: effects of radiation during spike growth period. Field Crop. Res. 54 (2–3), 245–257.

Akter, N., Rafiqul Islam, M., 2017. Heat stress effects and management in wheat. a review. Agron. Sustain. Dev. 37, 1–17.

Bai, H., Xiao, D., Wang, B., Liu, D.L., Feng, P., Tang, J., 2021. Multi-model ensemble of CMIP6 projections for future extreme climate stress on wheat in the north China plain. Int. J. Climatol. 41, E171–E186.

Best, M.J., Pryor, M., Clark, D., Rooney, G.G., Essery, R., Ménard, C., Edwards, J., Hendry, M., Porson, A., Gedney, N., et al., 2011. The joint UK land environment simulator (JULES), model description–part 1: energy and water fluxes. Geosci. Model. Dev. 4 (3), 677–699.

Birthal, P.S., Hazrana, J., Negi, D.S., Pandey, G., 2021. Benefits of irrigation against heat stress in agriculture: evidence from wheat crop in India. Agricult. Water. Manag. 255, 106950.

von Bloh, W., Schaphoff, S., Müller, C., Rolinski, S., Waha, K., Zaehle, S., 2018. Implementing the nitrogen cycle into lpjml (version 5.0). Geosci. Model. Dev. 11, 2789–2812.

Chen, S., Sun, H., Shao, L., Zhang, X., 2014. Performance of winter wheat under different irrigation regimers associated with weather conditions in the north China plain. Aust. J. Crop. Sci. 8 (4), 550–557.

of Statistics of China, N.B., 2015. China Statistical Yearbook. China Statistics Press, Beijing, Various years.

of Statistics of China, N.B., 2020. China Statistical Yearbook. China Statistics Press, Beijing.

China Meteorological Data Service Center, 2024. China meteorological data service center. Data repository of CMA surface and other meteorological observations. URL https://data.cma.cn/en.

Clark, D., Mercado, L., Sitch, S., Jones, C., Gedney, N., Best, M., Pryor, M., Rooney, G., Essery, R., Blyth, E., et al., 2011. The joint UK land environment simulator (JULES), model description–part 2: carbon fluxes and vegetation dynamics. Geosci. Model. Dev. 4 (3), 701–722.

Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Amer. Statist. Assoc. 74 (368), 829–836.

Cleveland, W.S., Devlin, S.J., 1988. Locally weighted regression: an approach to regression analysis by local fitting. J. Amer. Statist. Assoc. 83 (403), 596–610.

Constantin, J., Raynal, H., Casellas, E., Hoffmann, H., Bindi, M., Doro, L., Eckersten, H., Gaiser, T., Grosz, B., Haas, E., et al., 2019. Management and spatial resolution effects on yield and water balance at regional scale in crop models. Agricult. Forest. Meterol. 275, 184–195.

Corbeels, M., Berre, D., Rusinamhodzi, L., Lopez-Ridaura, S., 2018. Can we use crop modelling for identifying climate change adaptation options? Agricult. Forest. Meterol. 256, 46–52.

Craufurd, P.Q., Vadez, V., Jagadish, S.K., Prasad, P.V., Zaman-Allah, M., 2013. Crop science experiments designed to inform crop modeling. Agricult. Forest. Meterol. 170, 8–18.

Easterling, W.E., Chen, X., Hays, C., Brandle, J.R., Zhang, H., 1996. Improving the validation of model-simulated crop yield response to climate change: an application to the EPIC model. Clim. Res. 6, 263–273.

Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K., Best, N., Wilde, M., Foster, I., 2014. The parallel system for integrating impact models and sectors (pSIMS). Environ. Model. Softw. 62, 509–516.

Farooq, M., Bramley, H., Palta, J.A., Siddique, K.H., 2011. Heat stress in wheat during reproductive and grain-filling phases. Crit. Rev. Plant Sci. 30 (6), 491–507.

Folberth, C., Gaiser, T., Abbaspour, K.C., Schulin, R., Yang, H., 2012. Regionalization of a large-scale crop growth model for sub-saharan africa. Agric. Ecosyst. Environ. 151, 21–33.

Franke, J.A., Müller, C., Elliott, J., Ruane, A.C., Jägermeyr, J., Balkovic, J., Ciais, P., Dury, M., Falloon, P.D., Folberth, C., et al., 2020. The GGCMI phase 2 experiment: global gridded crop model simulations under uniform changes in CO 2, temperature, water, and nitrogen levels (protocol version 1.0). Geosci. Model. Dev. 13 (5), 2315–2336.

Fu, J., Jian, Y., Wang, X., Li, L., Ciais, P., Zscheischler, J., Wang, Y., Tang, Y., Müller, C., Webber, H., et al., 2023. Extreme rainfall reduces one-twelfth of China's rice yield over the last two decades. Nat. Food 4 (5), 416–426.

Holzworth, D.P., Huth, N.I., deVoil, P.G., Zurcher, E.J., Herrmann, N.I., McLean, G., Chenu, K., van Oosterom, E.J., Snow, V., Murphy, C., Moore, A.D., Brown, H., Whish, J.P.M., Verrall, S., Fainges, J., Bell, L.W., Peake, A.S., Poulton, P.L., Hochman, Z., Thorburn, P.J., Gaydon, D.S., Dalgliesh, N.P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp, J., Cichota, R., Vogeler, I., Li, F.Y., Wang, E., Hammer, G.L., Robertson, M.J., Dimes, J.P., Whitbread, A.M., Hunt, J., van Rees, H., McClelland, T., Carberry, P.S., Hargreaves, J.N., MacLeod, N., McDonald, C., Harsdorf, J., Wedgwood, S., Keating, B.A., 2014. APSIM – evolution towards a new generation of agricultural systems simulation. Environ. Model. Softw. 62, 327–350. http://dx.doi.org/10.1016/j.envsoft.2014.07.009.

Hossain, A., Sarker, M., Saifuzzaman, M., Teixeira da Silva, J., Lozovskaya, M., Akhter, M., 2013. Evaluation of growth, yield, relative performance and heat susceptibility of eight wheat (triticum aestivum l.) genotypes grown under heat stress. Int. J. Plant Prod. 7 (3), 615–636.

Izaurralde, R.C., Williams, J.R., McGill, W.B., Rosenberg, N.J., Quiroga Jakas, M.C., 2006. Simulating soil c dynamics with EPIC: Model description and testing against long-term data. Ecol. Model. 192, 362–384. http://dx.doi.org/10.1016/j.ecolmodel.2005.07.010.

Jägermeyr, J., Müller, C., Ruane, A.C., Elliott, J., Balkovic, J., Castillo, O., Faye, B., Foster, I., Folberth, C., Franke, J.A., et al., 2021. Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. Nat. Food 2 (11), 873–885.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model. Eur. J. Agron. 18, 235–265.

Kang, Y., Khan, S., Ma, X., 2009. Climate change impacts on crop yield, crop water productivity and food security–a review. Prog. Nat. Sci. 19 (12), 1665–1674.

Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D.P., Huth, N.I., Hargreaves, J.N.G., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J.P., Silburn, M., Wang, E., Brown, S., Bristow, K.L., Asseng, S., Chapman, S.C., McCown, R.L., Freebairn, D.M., Smith, C.J., 2003. An overview of APSIM, a model designed for farming systems simulation. Eur. J. Agron. 18, 267–288. http://dx.doi.org/10.1016/S1161-0301(02)00108-9.

Koch, J., Zhang, W., Martinsen, G., He, X., Stisen, S., 2020. Estimating net irrigation across the north China plain through dual modeling of evapotranspiration. Water Resour. Res. 56 (12), e2020WR027413.

Kumar, H., Chugh, V., Kumar, M., Gupta, V., Prasad, S., Kumar, S., Singh, C.M., Kumar, R., Singh, B.K., Panwar, G., et al., 2023. Investigating the impact of terminal heat stress on contrasting wheat cultivars: a comprehensive analysis of phenological, physiological, and biochemical traits. Front. Plant Sci. 14, 1189005.

Lehane, J., Staple, W., 1965. Influence of soil texture, depth of soil moisture storage, and rainfall distribution on wheat yields in southwestern saskatchewan. Can. J. Soil Sci. 45 (2), 207–219.

Leng, G., 2017. Recent changes in county-level corn yield variability in the united states from observations and crop models. Sci. Total Environ. 607, 683–690.

Li, Y., Guan, K., Schnitkey, G.D., DeLucia, E., Peng, B., 2019. Excessive rainfall leads to maize yield loss of a comparable magnitude to extreme drought in the United States. Global Change Biol. 25 (7), 2325–2337.

Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., Smith, B., 2013. Implications of accounting for land use in simulations of ecosystem carbon cycling in africa. Earth Syst. Dyn. 4, 385–407.

Liu, B., Asseng, S., Liu, L., Tang, L., Cao, W., Zhu, Y., 2016a. Testing the responses of four wheat crop models to heat stress at anthesis and grain filling. Global Change Biol. 22 (5), 1890–1903.

Liu, B., Gu, W., Yang, Y., Lu, B., Wang, F., Zhang, B., Bi, J., 2021. Promoting potato as staple food can reduce the carbon–land–water impacts of crops in China. Nat. Food 2 (8), 570–577.

Liu, B., Liu, L., Asseng, S., Zhang, D., Ma, W., Tang, L., Cao, W., Zhu, Y., 2020. Modelling the effects of post-heading heat stress on biomass partitioning, and grain number and weight of wheat. J. Exp. Bot. 71 (19), 6015–6031.

Liu, B., Liu, L., Asseng, S., Zou, X., Li, J., Cao, W., Zhu, Y., 2016b. Modelling the effects of heat stress on post-heading durations in wheat: A comparison of temperature response routines. Agricult. Forest. Meterol. 222, 45–58.

Liu, B., Liu, L., Tian, L., Cao, W., Zhu, Y., Asseng, S., 2014. Post-heading heat stress and yield impact in winter wheat of China. Global Change Biol. 20 (2), 372–381.

Liu, J., Williams, J.R., Zehnder, A.J.B., Yang, H., 2007. GEPIC–modelling wheat yield and crop water productivity with high resolution on a global scale. Agricult. Sys. 94, 478–493.

Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., Schulin, R., 2016a. Global investigation of impacts of PET methods on simulating crop-water relations for maize. Agricult. Forest. Meterol. 221, 164–175.

Liu, W., Yang, H., Liu, J., Azevedo, L.B., Wang, X., Xu, Z., Abbaspour, K.C., Schulin, R., 2016b. Global assessment of nitrogen losses and trade-offs with yields from major crop cultivations. Sci. Total Environ. 572, 526–537.

Lu, J., Carbone, G.J., Gao, P., 2017. Detrending crop yield data for spatial visualization of drought impacts in the United States, 1895–2014. Agricult. Forest. Meterol. 237, 196–208.

Lucia Martin, L., Smerald, A., Kiese, R., Klimiuk, T., Ludwig, P., Sanchéz-Benítez, A., Goessling, H., Scheer, C., 2025. The vulnerability of European agricultural areas to anthesis heat stress increases with climate change. Environ. Res.: Food Syst. 2 (2), 025002.

Lv, Z., Liu, X., Cao, W., Zhu, Y., 2013. Climate change impacts on regional winter wheat production in main wheat production regions of China. Agricult. Forest. Meterol. 171, 234–248.

Mamrutha, H., Rinki, K., Venkatesh, K., Gopalareddy, K., Khan, H., Mishra, C., Kumar, S., Kumar, Y., Singh, G., Singh, G., 2020. Impact of high night temperature stress on different growth stages of wheat. Plant Physiol. Rep. 25, 707–715.

Mankin, K.R., Edmunds, D.A., McMaster, G.S., Fox, F., Wagner, L.E., Green, T.R., 2024. Winter wheat crop models improve growth simulation by including phenological response to water-deficit stress. Environ. Model. Assess. 29 (2), 235–248.

Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J.W., Rötter, R.P., Boote, K.J., Ruane, A.C., Thorburn, P.J., Cammarano, D., et al., 2015. Multimodel ensembles of wheat growth: many models are better than one. Global Change Biol. 21 (2), 911–925.

Mirosavljević, M., Mikić, S., Kondić-Špika, A., Župunski, V., Zhou, R., Abdelhakim, L., Ottosen, C.-O., 2021. The effect of heat stress on some main spike traits in 12 wheat cultivars at anthesis and mid-grain filling stage. Plant Soil Environ. 67 (2), 71–76.

Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., et al., 2017. Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications. Geosci. Model. Dev. 10 (4), 1403–1422.

Müller, C., Elliott, J., Kelly, D., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Hoek, S., Izaurralde, R.C., et al., 2019. The global gridded crop model intercomparison phase 1 simulation dataset. Sci. Data 6 (1), 50.

Müller, C., Franke, J., Jägermeyr, J., Ruane, A.C., Elliott, J., Moyer, E., Heinke, J., Falloon, P.D., Folberth, C., Francois, L., et al., 2021. Exploring uncertainties in global crop yield projections in a large ensemble of crop models and CMIP5 and CMIP6 climate scenarios. Environ. Res. Lett. 16 (3), 034040.

Müller, C., Jägermeyr, J., Franke, J.A., Ruane, A.C., Balkovic, J., Ciais, P., Dury, M., Falloon, P., Folberth, C., Hank, T., et al., 2024. Substantial differences in crop yield sensitivities between models call for functionality-based model evaluation. Earth's Futur. 12 (3), e2023EF003773.

Olin, S., Schurgers, G., Lindeskog, M., Wårlind, D., Smith, B., Bodin, P., Holmér, J., Arneth, A., 2015. Modelling the response of yields and tissue c:N to changes in atmospheric CO2 and n management. Biogeosciences 12, 2489–2515.

Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., Wheeler, T., 2014. JULES-crop: a parametrisation of crops in the joint UK land environment simulator. Geosci. Model. Dev. Discuss. 7 (5), 6773–6809.

Osborne, T.M., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., Betts, R., Wheeler, T., 2015. JULES-crop: a parametrisation of crops in the joint UK land environment simulator. Geosci. Model. Dev. 8, 1139–1155. http://dx.doi.org/10.5194/gmd-8-1139-2015.

Prasad, R.B., Joshi, M.A., Basu, S., Gaikwad, K.B., 2020. Development of suitable mitigation strategy to counter the adverse effect of heat stress in wheat varieties (triticum aestivum l.). Int J Curr Microbiol App Sci 9, 646–653.

Rezaei, E.E., Siebert, S., Ewert, F., 2015a. Intensity of heat stress in winter wheat—phenology compensates for the adverse effect of global warming. Environ. Res. Lett. 10 (2), 024012.

Rezaei, E.E., Webber, H., Gaiser, T., Naab, J., Ewert, F., 2015b. Heat stress in cereals: Mechanisms and modelling. Eur. J. Agron. 64, 98–113.

Ru, C., Hu, X., Wang, W., 2024. Nitrogen mitigates the negative effects of combined heat and drought stress on winter wheat by improving physiological characteristics. Physiol. Plant. 176 (2), e14236.

Ruane, A.C., Phillips, M., Müller, C., Elliott, J., Jägermeyr, J., Arneth, A., Balkovic, J., Deryng, D., Folberth, C., Iizumi, T., et al., 2021. Strong regional influence of climatic forcing datasets on global crop model ensembles. Agricult. Forest. Meterol. 300, 108313.

Saiyed, I.M., Bullock, P.R., Sapirstein, H.D., Finlay, G.J., Jarvis, C.K., 2009. Thermal time models for estimating wheat phenological development and weather-based relationships to wheat quality. Can. J. Plant Sci. 89 (3), 429–439.

Sereyrorth, O., Yan, B., Chunpanha, K., Lybun, P., Linvolak, P., 2020. Study on the effects of extreme precipitation for seven growth stages of winter wheat in northern weihe loess plateau, China. J. Water Resour. Prot. 12 (4), 358–380.

Shi, X.-l., Shi, W.-j., 2016. Impacts of extreme high temperature on winter wheat yield in the huang-huai-hai plain. J. Ecol. Rural. Environ. 32 (2), 259–269. http://dx.doi.org/10.11934/j.issn.1673-4831.2016.02.014.

Shi, P., Tang, L., Lin, C., Liu, L., Wang, H., Cao, W., Zhu, Y., 2015. Modeling the effects of post-anthesis heat stress on rice phenology. Field Crop. Res. 177, 26–36.

Shirdelmoghanloo, H., Cozzolino, D., Lohraseb, I., Collins, N.C., 2016. Truncation of grain filling in wheat (triticum aestivum) triggered by brief heat stress during early grain filling: association with senescence responses and reductions in stem reserves. Funct. Plant Biology 43 (10), 919–930.

Siebert, S., Webber, H., Zhao, G., Ewert, F., 2017. Heat stress is overestimated in climate impact studies for irrigated agriculture. Environ. Res. Lett. 12 (5), 054023.

Stephens, D.J., Lyons, T., 1998. Rainfall-yield relationships across the Australian wheatbelt. Aust. J. Agric. Res. 49 (2), 211–224.

Sun, H.Y., Liu, C.-M., Zhang, X.Y., Shen, Y.J., Zhang, Y.Q., 2006. Effects of irrigation on water balance, yield and WUE of winter wheat in the north China plain. Agricult. Water. Manag. 85 (1–2), 211–218.

Sun, T., Miao, T., Hu, J., Zhou, H., Li, M., Zhang, J., Pan, Y., Bi, R., Ye, S., Wei, M., et al., 2025. Quantitative assessment of heat tolerance in new rice varieties through field survey during natural extreme heat events. Eur. J. Agron. 170, 127749.

Tack, J., Barkley, A., Hendricks, N., 2017. Irrigation offsets wheat yield reductions from warming temperatures. Environ. Res. Lett. 12 (11), 114027.

Tao, F., Yokozawa, M., Liu, J., Zhang, Z., 2008. Climate–crop yield relationships at provincial scales in China and the impacts of recent climate trends. Clim. Res. 38 (1), 83–94.

Thorne, G., Wood, D., 1987. Effects of radiation and temperature on tiller survival, grain number and grain yield in winter wheat. Ann. Botany 59 (4), 413–426.

Ullah, A., Nadeem, F., Nawaz, A., Siddique, K.H., Farooq, M., 2022. Heat stress effects on the reproductive physiology and yield of wheat. J. Agron. Crop Sci. 208 (1), 1–17.

Wang, H., 2007. Analysis of precipitation variation characteristics in fuyang city over the past four decades. Anhui Agric. Sci. 35 (6), 1883–1884.

Wang, S., Chen, J., Shen, M., Shi, T., Liu, L., Zhang, L., Dong, Q., Wang, C., 2022. Characterizing spatiotemporal patterns of winter wheat phenology from 1981 to 2016 in north China by improving phenology estimation. Remote. Sens. 14 (19), 4930.

Wang, S., Cheng, X., Yang, W., 2015. Precipitation characteristics in the chaohu lake basin over the past 56 years and their impact on drought and flood disasters. Res. Soil Water Conserv. 22 (3), 289–294.

Wang, E., Martre, P., Zhao, Z., Ewert, F., Maiorano, A., Rötter, R.P., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., et al., 2017. The uncertainty of crop yield projections is reduced by improved temperature response functions. Nat. Plants 3 (8), 1–13.

Wang, H., Wang, L., Yan, G., Bai, H., Zhao, Y., Ju, M., Xu, X., Yan, J., Xiao, D., Chen, L., 2022. Assessment and prediction of extreme temperature indices in the north China plain by CMIP6 climate model. Appl. Sci. 12 (14), 7201.

Wang, N., Wu, J., Gu, Y., Jiang, K., Ma, X., 2022. Factors influencing the spatiotemporal variability in the irrigation requirements of winter wheat in the north China plain under climate change. Agronomy 12 (9), 1987.

Wang, H.-l., Zhang, Q., Wang, R.-y., Gan, Y.-t., Niu, J.-y., Zhang, K., Zhao, F.-n., Zhao, H., 2015. Effects of air temperature increase and precipitation change on grain yield and quality of spring wheat in semiarid area of northwest China. Yingyong Shengtai Xuebao 26 (1), 67–75.

Wang, X., Zhao, C., Müller, C., Wang, C., Ciais, P., Janssens, I., Peñuelas, J., Asseng, S., Li, T., Elliott, J., et al., 2020. Emergent constraint on crop yield response to warmer temperature from field experiments. Nat. Sustain. 3 (11), 908–916.

Warrington, I., Dunstone, R., Green, L., 1977. Temperature effects at three development stages on the yield of the wheat ear. Aust. J. Agric. Res. 28 (1), 11–27.

Webber, H., Ewert, F., Kimball, B., Siebert, S., White, J.W., Wall, G., Ottman, M., Trawally, D., Gaiser, T., 2016. Simulating canopy temperature for modelling heat stress in cereals. Environ. Model. Softw. 77, 143–155.

Williams, K., Gornall, J., Harper, A., Wiltshire, A., Hemming, D., Quaife, T., Arkebauer, T., Scoby, D., 2017. Evaluation of JULES-crop performance against site observations of irrigated maize from mead, nebraska. Geosci. Model. Dev. 10 (3), 1291–1320. http://dx.doi.org/10.5194/gmd-10-1291-2017.

Xu, R., Li, Y., Guan, K., Zhao, L., Peng, B., Miao, C., Fu, B., 2022. Divergent responses of maize yield to precipitation in the united states. Environ. Res. Lett. 17 (1), 014016.

Yang, R., Dai, P., Wang, B., Jin, T., Liu, K., Fahad, S., Harrison, M.T., Man, J., Shang, J., Meinke, H., et al., 2022. Over-optimistic projected future wheat yield potential in the north China plain: the role of future climate extremes. Agronomy 12 (1), 145.

Yang, X., Tian, Z., Sun, L., Chen, B., Tubiello, F.N., Xu, Y., 2017. The impacts of increased heat stress events on wheat yield under climate change in China. Clim. Change 140, 605–620.

Zhang, J., Balkovič, J., Azevedo, L.B., Skalský, R., Bouwman, A.F., Xu, G., Wang, J., Xu, M., Yu, C., 2018. Analyzing and modelling the effect of long-term fertilizer management on crop yield and soil organic carbon in China. Sci. Total Environ. 627, 361–372.

Zhang, K., Nie, Z., Wang, J., Li, G., 2022. Simulation and analysis of the effects of precipitation and nitrogen application on the yield and biomass of spring wheat in dryland under elevated temperature. Arid. Zone Res. 39 (6), 1966–1975. http://dx.doi.org/10.13866/j.azr.2022.06.26.

Zhao, Q., Wang, C., Wang, X., Rezaei, E.E., Müller, C., Wang, E., Webber, H., Zhang, L., Li, X., Sang, Y., et al., 2026. Temperature thresholds of extreme heat-induced yield loss in maize and soybean reveal geographic heterogeneity across the northern hemisphere. Nat. Food 7 (2), 194–205.

Zhu, P., Zhuang, Q., Archontoulis, S.V., Bernacchi, C., Müller, C., 2019. Dissecting the nonlinear response of maize yield to high temperature stress with model-data integration. Global Change Biol. 25 (7), 2470–2484.