Response to Reviewer #1 – *Geoscientific Model Development*
Manuscript EGUsphere-2024-4010
Authors: Jinhui Zheng, Le Yu *, Zhenrong Du, Liujun Xiao, and Xiaomeng Huang

Reviewer #1

Dear Reviewer Theodoros Mavromatis,

We sincerely thank you for your insightful comments and constructive suggestions, which have helped us improve the quality of our manuscript. Below, we provide detailed responses to each point. The original *reviewer comments are presented in italic*, while the authors' responses are provided in blue.

The line number is based on the clean version of the revised manuscript, not the track change version.

This is a significant contribution. Among others issues (noted on the attached manuscript that should be taken care of) a major one is related to the selection of the "extreme" years chosen in this study.

Response:

We sincerely thank the reviewer for the thoughtful and encouraging evaluation of our work, as well as for the valuable suggestions regarding the selection of extreme weather years. Your insightful comments have been instrumental in guiding us to improve the rigor and clarity of our analysis. In response to your recommendation, we have carefully re-evaluated our approach and made substantial revisions to the manuscript concerning the selection of representative years for extreme weather analysis. Drawing upon comprehensive survey data and official reports, we have now selected 2009, 2010, 2012, and 2018 as the representative extreme weather years used in this study. We believe these years more accurately and comprehensively reflect the diverse and significant impacts of extreme weather events on crop production in the study region. The updated rationale and supporting information can be found in lines 283–301 of the revised manuscript.

To briefly summarize:

According to the Ministry of Ecology and Environment of the People's Republic of China (www.mee.gov.cn), 2009 was marked by record-breaking high temperatures in the study area, with multiple locations exceeding historical maximums. In 2010, the frequency of meteorological disasters notably increased, and numerous extreme weather events were documented. In 2012, China experienced 38 severe rainfall events, including the devastating "7.21" event. The region also endured concurrent droughts and cold waves (Zhang et al., 2018; Zhao et al., 2019). In 2018, extreme low temperatures caused widespread frost damage, which had a pronounced impact on agricultural productivity (China Meteorological Administration, www.cma.gov.cn).

We believe that this more cautious and evidence-based selection of extreme weather years provides a stronger foundation for evaluating the WOFOST-EW model under diverse and complex extreme weather conditions. These revisions have enhanced the scientific robustness, representativeness, and overall credibility of our study. Relevant references and background information have been added to support these changes. Once again, we sincerely appreciate your constructive feedback, which has greatly contributed to the improvement of our manuscript.

Lines 20-25: Report it as RRMSE as well.

Response:

Thank you very much for your careful review of the model performance evaluation metrics. We fully agree with your observation that reporting the Relative Root Mean Square Error (RRMSE) provides a standardized measure of relative error, which is highly valuable for assessing model performance and facilitating comparisons with other studies. Accordingly, we have calculated the RRMSE values as per your suggestion and added them to the abstract (lines 20–25). We have also included the corresponding explanations in the results section of the manuscript (lines 251–257, 259–282, and 296–301). We greatly appreciate your constructive feedback, which has helped improve the clarity and scientific rigor of our work.

Lines 20-25: see my comments in the text for these specific years.

Response:

Thank you for your valuable suggestion. As mentioned above, we have made substantial revisions to the study, including the re-selection of representative years for extreme weather events. Please refer to lines 283–301 in the revised manuscript for the updated content.

Lines 30-35: Which one? There are two in the reference list.

Response:

Thank you very much for pointing out this oversight. In response to your suggestion, we have revised the reference list to distinguish multiple publications by authors with the same surname published in the same year. Specifically, we have added “a,” “b,” etc., after the publication year to ensure that each reference is clearly identified and cited appropriately. Please refer to lines 33–34 in the revised manuscript for the updated citations.

Lines 35-45: Rephrase or delete. In my opinions, the previous sentences do not result in this one.

Response:

Thank you for your suggestion. As recommended, we have removed the sentence accordingly. Please refer to lines 45–47 in the revised manuscript.

Lines 65-70: Define this. What does it include?

Response:

Thank you very much for your valuable comment. We agree that the definition of "extreme climate" was not clearly articulated in the original manuscript, which may lead to confusion. In the revised manuscript, we have clarified this term explicitly (see lines 67–75).

In this study, "extreme climate" refers to abnormal temperature, precipitation, or drought events that occur during the entire growth season of winter wheat (from sowing to maturity). These events are characterized by their extremity, sudden onset, and damaging potential, which can significantly impact crop growth (Bai et al., 2024; Feng et al., 2019; Yu et al., 2025; Zheng and Zhang, 2025). Specifically, we used the following types of extreme climate indices as the basis for identification and simulation:

- Extreme temperature events: These include both heat and cold stress. Heat stress may cause premature senescence, poor grain filling, or even direct thermal damage, while cold stress can result in frost damage, delayed development, or yield loss. In our model, these impacts are quantified using indices such as high-temperature degree days (HDD) and low-temperature degree days (LDD) (Dong et al., 2023; Osman et al., 2020; Zhang et al., 2016; Zhang and Tao, 2019).

- Extreme precipitation events: These refer to unusually high-intensity rainfall over short periods or persistent abnormal precipitation patterns, which may lead to waterlogging, nutrient leaching, or increased disease pressure. We used indicators such as R95P (very wet days), R10 (number of days with precipitation >10 mm), and Rx1day (maximum 1-day precipitation) to characterize these events (Al-Sakkaf et al., 2024; Hong and Ying, 2018).

- Extreme drought conditions: These refer to prolonged periods of insufficient precipitation and/or high evapotranspiration that severely reduce soil moisture, causing water stress for crops. This can inhibit crop growth and photosynthesis and, in severe cases, lead to plant wilting or death. We used the Palmer Drought Severity Index (PDSI) and vapor pressure deficit (VPD) to represent drought conditions (Baydaroğlu et al., 2024; Kumar and Mahapatra, 2024; Oubaha et al., 2024; Peethani et al., 2024a; Pei et al., 2024a; SM et al., 2025; Yang et al., 2024; Zhang et al., 2025).

These events are defined as "extreme" because their intensity, duration, or frequency significantly exceed the normal historical range, with clear evidence of their adverse effects on wheat physiological processes and final yield. The WOFOST-EW model introduced in this study is designed to better simulate and assess the impact of these specific types of extreme climate conditions on wheat production. To eliminate any ambiguity, we have included this definition of “extreme climate” in lines 67–75 of the revised manuscript.

Lines 75-85: Elaborate on this. Mention a few dynamic changes in crop growth which are overlooked.

Response:

Thank you very much for your constructive suggestion. We have incorporated specific revisions in lines 80–87 of the revised manuscript to clarify these dynamic processes.

Specifically, in most previous studies, the outputs of crop models (such as biomass, leaf area index, or final yield) were often directly used as input variables for machine learning models. However, little attention has been paid to how extreme weather events nonlinearly disturb crop physiological and developmental processes at different growth stages. This form of static coupling overlooks the temporal sensitivity and continuity of extreme climate impacts. Phenological responses, for example, are often not fully considered—extreme heat or drought may lead to earlier heading or maturity in wheat (Hou et al., 2024; Liu et al., 2023), which can significantly affect dry matter accumulation. These nonlinear dynamics are frequently neglected in conventional machine learning frameworks.

Recognizing the omission of these critical dynamic responses, our study aims to enhance the phenology module of the WOFOST-EW model by integrating extreme weather indices with deep learning algorithms. This approach enables the model to more accurately and robustly capture the complex, stage-specific effects of extreme weather events on crop growth processes, thereby improving yield prediction accuracy under extreme climatic conditions.

Lines 75-85: Mention a few appropriate references for this statement.

Response:

Thank you for your suggestion. We have added relevant references to support this statement. Please refer to lines 80–87 in the revised manuscript.

Table 1: Move this Table in the Supplementary material. Presenting growing season temperature and precipitation for each station, would be more relevant.

Response:

Thank you for your valuable suggestion. We have moved the table to the Supplementary Materials as Table S1 and made the corresponding adjustments in the main text. Additionally, we have included information on the temperature and precipitation during the winter wheat growing season for each station. Please refer to Table S1 in the Supplementary Materials.

Lines 105-110: Mention them.

Response:

Thank you for your suggestion. We have added details about the main elements included in the weather dataset. Please refer to lines 105–109 in the revised manuscript.

Lines 105-110: How about available water capacity for each layer?

Response:

The soil data from the ISRIC global database include Available Water Capacity (mm/m). We have explicitly added this important information to the revised manuscript to provide readers with a more comprehensive understanding of the soil data details. Please refer to lines 118–121 in the revised manuscript.

Lines 120-125: In years 2008 and 2018.

Response:

Thank you for your suggestion. We have made the corresponding additions in the revised manuscript. Please refer to lines 130–132.

Lines 140-145: Why these two indices were preferred over the others included in the CHM_Drought database?

Response:

Thank you for your valuable question. We selected the Palmer Drought Severity Index (PDSI) and Vapor Pressure Deficit (VPD) from the CHM_Drought database for several key reasons:

First, PDSI is one of the most widely used drought indices (Oubaha et al., 2024; Yang et al., 2024; Zhang et al., 2025). It accounts for antecedent precipitation and water supply-demand balance, providing clear physical meaning and is particularly suitable for assessing agricultural drought. In this study, we focus on the winter wheat growing season, and PDSI effectively captures drought processes at this timescale. Numerous studies have demonstrated its significant correlation with crop yield (Baydaroğlu et al., 2024; Kumar and Mahapatra, 2024; Peethani et al., 2024a; Pei et al., 2024a; SM et al., 2025).

Second, VPD is a key variable measuring atmospheric dryness, directly affecting crop transpiration and water stress. Studies have reported a steady global increase in VPD from 2010 to 2019, severely hindering agricultural production (Koehler et al., 2023; Nesmith and Ritchie, 1992). Under extreme heat and low humidity, elevated VPD exacerbates crop transpiration and water loss, posing a direct threat to yields. Thus, VPD can partly reflect the stress intensity of short-term extreme heat and drought events (Yu et al., 2024).

Although the CHM_Drought database provides various indices, our literature review and preliminary analyses show that these two indices exhibit stronger applicability and higher historical validation reliability for drought monitoring and assessment in the North China Plain (Li et al., 2024; Luan et al., 2024; Wu et al., 2024). They better correspond with observed agricultural drought events and yield losses in the region.

We have further elaborated on the rationale for selecting these indices in lines 146–158 of the revised manuscript. We appreciate your insightful comment, which helped us clarify the methodological choices in our study.

Lines 190-195: Are these extreme indices are estimated every day and then be used an input to LSTM? How about PDSI and VPD? On which temporal basis are estimated? Why not the PDSI over the ScPDSI? Why R95P, R10 and Rxwday are estimated on annual and not on growing season basis? Elaborate on these.

Response:

Thank you very much for your detailed question regarding the processing of extreme weather indices. First, we would like to clarify that all extreme indices used in this study—including PDSI, VPD, R95P, R10, Rx1day, and others mentioned later—were calculated over the entire growing period of winter wheat. This approach ensures that the extreme weather events we analyze are closely linked to the actual growth and developmental stages of the crop, allowing for a more accurate assessment of their impacts on yield. Below is a detailed explanation based on the manuscript content:

> *Are these extreme indices are estimated every day and then be used an input to LSTM? How about PDSI and VPD? On which temporal basis are estimated?*

Response:

In our study, high-temperature degree days (HDD) and low-temperature degree days (LDD) are cumulative indices based on wheat growth stages. Specifically, the high-temperature thresholds were set at 25°C from sowing to heading and 30°C from heading to maturity; the low-temperature thresholds were −5.7°C and −2°C for the corresponding periods (Farooq et al., 2011; Liu et al., 2013; Porter and Gawith, 1999). These indices are calculated daily by summing temperature deviations and directly reflect the sustained impact of extreme temperatures during critical phenological stages. Their timing strictly corresponds to the wheat growth cycle, making them suitable as input features for the LSTM to characterize stage-specific climate stress intensity. Although extreme precipitation indices such as R95P, R10, and Rx1day are often reported on an annual scale in the literature, we recalculated them based on the wheat growing season to more accurately capture precipitation extremes encountered during crop growth. The Palmer Drought Severity Index (PDSI) reflects long-term soil moisture conditions by integrating precipitation, temperature, and potential evapotranspiration, and is typically calculated on a monthly or seasonal basis (Zhang and Miao, 2024). In this study, PDSI and vapor pressure deficit (VPD) were averaged or accumulated over the growing season to represent the sustained effect of soil moisture deficits. We acknowledge that the original manuscript's descriptions were incomplete; thus, we have revised the definitions in Table 1 for greater accuracy and supplemented the explanations in lines 135–162 of the revised manuscript to improve clarity.

Why not the PDSI over the ScPDSI?

Response:

The Self-Calibrating Palmer Drought Severity Index (ScPDSI) modifies the empirical constants used in the original PDSI calculation by dynamically adjusting them to local climate conditions, allowing it to automatically calibrate drought behavior at any location (Dai, 2011; Zhang and Miao, 2024). However, studies have shown that in most regions of China, PDSI exhibits a stronger correlation with normalized difference vegetation index (NDVI) anomalies, simulated soil moisture anomalies (SMA), and the land water storage deficit index (WSDI) compared to ScPDSI (Zhong et al., 2019). This suggests that PDSI is more representative for characterizing agricultural drought impacts on crops in the Chinese context. One main reason is that ScPDSI tends to capture meteorological droughts as less severe than PDSI, which may be attributed to two factors: 1) modifications in ScPDSI reduce sensitivity to different potential evapotranspiration (PET) estimation methods, leading to lower responsiveness under wet or dry conditions (van der Schrier et al., 2011); 2) adjustments to the self-calibrating persistence factors and climate characteristic parameter may increase ScPDSI's sensitivity to dataset-specific features, sometimes weakening its drought detection performance in certain regions. For example, Liu et al. (2016) confirmed this in the Yellow River Basin of northern China. Although the self-calibrating procedure improves spatial consistency and controls extreme event frequency (Dai, 2011b; Trenberth et al., 2014), considering our study's specific objective to evaluate WOFOST-EW model performance under extreme conditions, we argue that PDSI sufficiently and effectively reflects drought conditions impacting crop growth within the current data and regional context. Moreover, PDSI has been widely validated in numerous crop-related studies (Islam et al.; Peethani et al., 2024b; Pei et al., 2024b; Yan et al., 2016). To clarify this choice further, we added a detailed explanation in lines 146–162 of the revised manuscript.

Why R95P, R10 and Rxwday are estimated on annual and not on growing season basis? Elaborate on these.

Response:

Thank you very much for pointing out the potential ambiguity in our description. We would like to clarify that R95P, R10, and Rx1day in our study were recalculated specifically for the winter wheat growing season to more directly capture the extreme precipitation events encountered during the crop growth period. We have revised the definitions in Table 1 as well as the related content in lines 135–162 to avoid any confusion for readers. Once again, we sincerely appreciate your constructive comments and careful review.

Lines 200-205: between measured and estimated yield?

Response:

Thank you for pointing out this issue. The statement refers to the approach where the parameters corresponding to the minimum root mean square error (RMSE) between observed and simulated yield, as well as between observed and simulated phenological stages, are considered optimal. This is a standard method for assessing model fit during calibration and optimization (Chen and Tao, 2020; Zheng and Zhang, 2023). We have revised this sentence to improve clarity and accuracy. Please refer to lines 214–225 in the revised manuscript.

Lines 200-205: Which parameters of the Table S2 were calibrated and for which target (anthesis, maturity and yield)?

Response:

We sincerely thank the reviewer for the careful review and valuable comments. We have added supplementary explanations and clarifications in the revised manuscript; please refer to lines 214–225 in the main text as well as Text S1, S2 and Tables S3, S4 in the supplementary materials. Specifically, we included the sensitivity analysis results of the WOFOST model (Table S3) and identified the parameters to be adjusted based on this analysis (Table S4). We appreciate your important questions, which helped us better clarify the rationale and optimization process of the model parameter settings.

Lines 210-215: Estimate and show the respective results for MAE (mean absolute error). How about the results for r2?

Response:

We fully agree with your viewpoint that Mean Absolute Error (MAE) and the Coefficient of Determination (R^2) are important metrics for evaluating model prediction accuracy and explanatory power. Accordingly, we have calculated MAE and R^2 as per your suggestion and have detailed these evaluation results in lines 230–257 of the revised manuscript.

Lines 245-250: Express it and in the form or relative RMSE (within parentheses).

Response:

Thank you very much for your valuable suggestion. Following your advice, we have calculated the RRMSE and included it in parentheses in lines 259–273 of the revised manuscript to more clearly demonstrate the model's performance.

Figure 5: Which statistic is this? Has this been described in section 2.3.6?

Response:

Thank you very much for your question. The statistic shown in Figure 5c is the simulation bias, defined as the difference between simulated and observed yields (simulated value minus observed value). This metric is used to assess the model's systematic overestimation or underestimation of yield across different years. We have added a clear definition of this statistic in Section 2.3.5 of the revised manuscript, along with its corresponding formula (Equation 16).

Lines 255-260: Is this for the calibration or validation period?

Response:

Thank you very much for your question. The statistical results shown in Figure 5c indeed correspond to the validation period. To avoid any confusion, we have revised the figure caption in the updated manuscript to explicitly indicate the time period represented. Please refer to the updated Figure 5 in the revised manuscript.

Lines 260-265: This statistic should be defined in 2.3.6 section. Express it also in its relative version (relative MAD).

Response:

Thank you very much for your valuable suggestion. We have explicitly added the definition of the MAE statistic in Section 2.3.5 of the revised manuscript. Additionally, following your recommendation, we have further calculated and reported the Mean Relative Error (MRE) to more intuitively represent the relative proportion of simulation errors to observed values. This metric helps to better understand the model performance across different counties and years. The related results have been added in lines 270–275 of the revised manuscript.

Lines 265-270: Which statistic is this? Has this been described in section 2.3.6?

Response:

Thank you very much for your valuable suggestion. We have explicitly added the definition of the Mean Absolute Error (MAE) in Section 2.3.5 of the revised manuscript. Additionally, following your advice, we further calculated and reported the Mean Relative Error (MRE) to provide a more intuitive representation of the simulation errors relative to the observed values. This metric helps better understand the model's performance across different counties and years. The related results have been added in lines 269–273 of the revised manuscript.

Figure 7: 2001?

Response:

Thank you for pointing this out. The error has been corrected, and the entire manuscript has been carefully reviewed. Please refer to Figure 7 in the revised manuscript.

Lines 275-280: I am not sure that the selected years are the best options. The observed yield of 2018 (see my red rectangle) is one of the highest in the study period. That means that wheat in reality recovered from the extreme weather noted in 2018 (Table S1) at many stations as can be seen from Fig. 8c, d. 2008 yield is also one of the highest in the study period. Maybe different years should be selected.

Response:

We sincerely appreciate the reviewer's careful evaluation and valuable suggestion regarding the selection of extreme weather years. Based on your insightful comments, we have re-examined the choice of representative years to better reflect the actual impacts of extreme weather on wheat yields in the study region. After a thorough review of official reports and observational data, we have revised our selection to include 2009, 2010, 2012, and 2018 as the key years representing extreme weather events (see lines 283–301 and Figs. 9 and 10 in the revised manuscript).

While it is true that the observed yield in 2018 was relatively high, this year was characterized by severe frost damage and other extreme low-temperature events documented by official sources (China Meteorological Administration). Our analysis also considers the spatial and temporal variability of impacts, as shown in Fig. 8, which captures differing regional responses to the extreme conditions (Ministry of Ecology and Environment of the People's Republic of China). The years 2009, 2010, and 2012 correspond to well-documented episodes of record-breaking heat, increased meteorological disasters, and extreme rainfall events, respectively (Zhao et al., 2019; Zheng et al., 2018).

We believe that this selection better balances the representation of diverse extreme weather types and their influence on crop production across the region. Detailed justification and supporting data have been added in lines 283–301 of the revised manuscript. We thank the reviewer again for the constructive advice, which has strengthened the rigor and clarity of our study.

Lines 280-285: No such counties in Table S1.

Response:

Thank you very much for your careful observation. The previous wording was indeed inaccurate. Shanxi and Hebei refer to provinces rather than county names. We have revised the relevant statements (lines 283–301 in the revised manuscript) to ensure clearer and more accurate expression, avoiding any potential confusion. We sincerely appreciate your meticulous review and valuable comments.

Figure 8: yield

Response:

Thank you for your suggestion. We have revised the figure caption accordingly. Please refer to Figure 9 in the revised manuscript.

Lines 310-315: Maybe estimating R95P, R10 and Rxwday on growing season basis rather on annual could help.

Response:

Thank you very much for your valuable comment. We fully understand your concern and would like to clarify that all extreme climate indices used in this study—including R95P, R10mm, and Rx1day—are calculated specifically for the winter wheat growing season (from sowing to maturity), rather than on an annual basis. This approach ensures that the identified extreme events are closely aligned with the actual crop growth period, allowing for a more accurate assessment of their potential impact on yield. We have added detailed explanations in lines 135–162 and updated Table 1 in the revised manuscript to prevent any possible misunderstanding.

Lines 320-325: I am not sure I understand this statement. Elaborate and make it more clearly. An appropriate reference would help.

Response:

Thank you for pointing out this issue. What we intended to convey is that the increasing frequency of extreme weather events leads to greater variability and unpredictability in meteorological observations such as temperature and precipitation. This variability complicates the derivation of stable and representative input parameters for crop models—such as thermal time and stress thresholds—thereby introducing uncertainty into model simulations (Gao et al., 2020; Gao et al., 2021). Such instability may cause deviations in model outputs and ultimately affect the accuracy of crop growth predictions. We have revised the original unclear statements and added relevant references accordingly. Please refer to lines 370–383 in the revised manuscript.

Lines 325-330: see my previous comment.

Response:

Thank you for raising this point. Here, we intended to express that in the North China Plain, frequent extreme high and low temperatures disrupt the consistency of daily weather inputs used in the model (Gu et al., 2024). This inconsistency affects the reliability of key model parameters, such as effective temperature accumulation and phenological thresholds, ultimately reducing the accuracy of crop growth and yield simulations under these extreme conditions (Bai et al., 2024). We have revised the previously unclear statements and supplemented relevant references accordingly. Please refer to lines 375–380 in the revised manuscript.

Lines 325-330: Rephrase.

Response:

Thank you for your suggestion. This sentence has been appropriately revised. Please refer to line 385 in the revised manuscript.

Lines 340-345: This is an overstatement. Only anthesis, maturity and final yield were checked. Rephrase.

Response:

Thank you for your suggestion. This sentence has been appropriately revised. Please refer to lines 390–395 in the revised manuscript.

Lines 340-345: No comparison on long-term trends between WOFOST and WOFOST-EW was made. Rephrase.

Response:

Thank you for your suggestion. This sentence has been appropriately revised. Please see lines 390–395 in the revised manuscript.

Lines 345-350: see a previous comment for these years.

Response:

Thank you very much for your valuable suggestion. In response to your comment, we have carefully re-examined the criteria for selecting extreme years and made substantial revisions accordingly. The related content has been thoroughly updated in lines 284–301 of the revised manuscript, along with corresponding updates to the figures and supplementary materials. Please refer to our detailed response provided earlier.

Lines 350-355: Does the LSTM algorithm estimates one F(EW) value for each site? That means that this value is site- and study period- specific.

Response:

Thank you very much for raising this important point. We would like to provide further clarification as follows:

In this study, the LSTM model is trained separately for each county to fully capture the spatial heterogeneity of extreme weather impacts on crop yield. The model inputs include multiple extreme climate indices during the winter wheat growing season (e.g., HDD, LDD, R10mm, PDSI, etc.), and the output is the extreme weather impact factor, F(EW), calculated for each year and county. This factor quantifies the integrated effect of extreme climate conditions on that year's yield, making F(EW) a spatiotemporally dynamic output variable.

While the WOFOST-EW framework can, in principle, be extended to other crops and regions, there are some inherent limitations that we have discussed in lines 399–438 of the revised manuscript. First, although WOFOST is a generic crop growth model, its parameters and certain modules are more specifically tailored to particular crop types such as cereals. For structurally different crops—such as root/tuber crops, oilseed crops, or perennials—adjustments to WOFOST's internal parameters would be necessary. Second, crop growth characteristics vary across regions due to differences in local climate, soil conditions, and management practices. Even if the WOFOST model structure is applicable, detailed parameter localization and calibration would be required when applying it to new crop types and target areas.

Furthermore, crop growth is influenced not only by climatic factors but also strongly affected by soil fertility, pest and disease pressures, irrigation, fertilization, and other complex environmental and management variables. Currently, the extreme weather functions within WOFOST-EW primarily emphasize meteorological factors. Effectively integrating non-meteorological extreme stresses—such as sudden pest outbreaks or severe nutrient deficiencies—and their interactions with extreme weather remains an important direction for future research.

We have clarified the model structure and the interpretation of F(EW) in lines 164–178 of the revised manuscript, and have emphasized the model's regional adaptability and limitations accordingly.

Lines 360-365: Did the LSTM algorithm only aimed at minimizing the error between observed and estimated yield?

Response:

Thank you for your question. The role of the LSTM algorithm in this study is not simply to directly minimize the error between observed and simulated yields. Specifically, the core objective of the LSTM is to learn and estimate a spatiotemporally dynamic extreme weather function F(EW). This F(EW) variable is designed to capture how various extreme weather indicators—such as extreme temperatures, precipitation, and drought—complexly affect crop growth processes.

The LSTM learns the nonlinear dynamic relationships between extreme weather events and crop physiological responses, thereby generating a correction factor that adjusts the internal growth processes within the WOFOST model in real time. In other words, the training goal of the LSTM can be more accurately described as optimizing the estimation of F(EW), enabling the coupled WOFOST-EW model system to simulate crop growth dynamics and final outcomes—including phenological stages (e.g., heading, maturity) and final yield—more accurately under extreme climate conditions.

Thus, by learning the modulation effects of extreme weather on crop growth, the LSTM indirectly helps minimize the discrepancy between observed and simulated yields, while more finely capturing crop responses under extreme conditions.

We have elaborated on the specific role of the LSTM algorithm, its integration with the WOFOST model, and its training objective in the revised manuscript (see lines 164–178 and 209–211). Thank you again for your insightful question, which prompted us to clarify the internal logic of our model more clearly.

Lines 390-395: Report it as RRMSE as well.

Response:

Thank you for your suggestion. We have added the RRMSE values accordingly. Please refer to lines 440–447 in the revised manuscript.

References

Al-Sakkaf, A. S., Zhang, J., Yao, F., Hamed, M. M., Simbi, C. H., Ahmed, A., and Shahid, S.: Assessing Exposure to Climate Extremes Over the Arabian Peninsula Using Era5 Reanalysis Data: Spatial Distribution and Temporal Trends, Atmos. Res., 300, 107224, https://doi.org/10.1016/j.atmosres.2024.107224, 2024.

Bai, H., Xiao, D., Tang, J., and De Li Liu: Evaluation of Wheat Yield in North China Plain Under Extreme Climate by Coupling Crop Model with Machine Learning, Comput. Electron. Agric., 217, 108651, https://doi.org/10.1016/j.compag.2024.108651, 2024.

Baydaroğlu, Ö., Yeşilköy, S., and Demir, I.: A Causality Perspective on the Impact of Hydroclimatic Extremes on Crop Yields, Copernicus Meetings, 2024.

Chen, Y. and Tao, F.: Improving the Practicability of Remote Sensing Data-Assimilation-Based Crop Yield Estimations Over a Large Area Using a Spatial Assimilation Algorithm and Ensemble Assimilation Strategies, Agric. For. Meteorol., 291, 108082, 2020.

Dai, A.: Drought Under Global Warming: A Review, Wiley Interdiscip. Rev. Clim. Change, 2, 45-65, 2011.

Dong, X., Zhang, T., Yang, X., Li, T., and Li, X.: Rice Yield Benefits from Historical Climate Warming to be Negated by Extreme Heat in Northeast China, Int. J. Biometeorol., 67, 835-846, https://doi.org/10.1007/s00484-023-02458-8, 2023.

Farooq, M., Bramley, H., Palta, J. A., and Siddique, K. H. M.: Heat Stress in Wheat During Reproductive and Grain-Filling Phases, Crit. Rev. Plant Sci., 30, 491-507, https://doi.org/10.1080/07352689.2011.615687, 2011.

Feng, P., Wang, B., De Li Liu, Waters, C., and Yu, Q.: Incorporating Machine Learning with Biophysical Model Can Improve the Evaluation of Climate Extremes Impacts on Wheat Yield in South-Eastern Australia, Agric. For. Meteorol., 275, 100-113, https://doi.org/10.1016/j.agrformet.2019.05.018, 2019.

Gao, Y., Wallach, D., Hasegawa, T., Tang, L., Zhang, R., Asseng, S., Kahveci, T., Liu, L., He, J., and Hoogenboom, G.: Evaluation of Crop Model Prediction and Uncertainty Using Bayesian Parameter Estimation and Bayesian Model Averaging, Agric. For. Meteorol., 311, 108686, https://doi.org/10.1016/j.agrformet.2021.108686, 2021.

Gao, Y., Wallach, D., Liu, B., Dingkuhn, M., Boote, K. J., Singh, U., Asseng, S., Kahveci, T., He, J., and Zhang, R.: Comparison of Three Calibration Methods for Modeling Rice Phenology, Agric. For. Meteorol., 280, 107785, 2020.

Gu, W., Ma, G., Wang, R., Scherer, L., He, P., Xia, L., Zhu, Y., Bi, J., and Liu, B.: Climate Adaptation through Crop Migration Requires a Nexus Perspective for Environmental Sustainability in the North China Plain, Nat. Food, 5, 569-580, 2024.

Hong, Y. and Ying, S.: Characteristics of Extreme Temperature and Precipitation in China in 2017 Based on ETCCDI Indices, Adv. Clim. Chang. Res., 9, 218-226, https://doi.org/10.1016/j.accre.2019.01.001, 2018.

Hou, M., Li, Y., Biswas, A., Chen, X., Xie, L., Liu, D., Li, L., Feng, H., Wu, S., and Satoh, Y.: Concurrent Drought Threaten Wheat and Maize Production and Widen Crop Yield Gaps in the Future, Agric. Syst., 220, 104056, 2024.

Islam, S. M., Mount, J., and Demir, I.: Decoding Drought-Yield Relationships in the US Midwest: A Multiscale Analysis Using Climatic Indicators and Random Forests, Jerry and Demir, Ibrahim, Decoding Drought-Yield Relationships in the US Midwest: A Multiscale Analysis Using Climatic Indicators and Random Forests.

Koehler, T., Wankmüller, F. J., Sadok, W., and Carminati, A.: Transpiration Response to Soil Drying Versus Increasing Vapor Pressure Deficit in Crops: Physical and Physiological Mechanisms and Key Plant Traits, J. Exp. Bot., 74, 4789-4807, 2023.

Kumar, R. and Mahapatra, R. P.: Severity of Natural Calamities and Crop Yield Prediction Using Hybrid Deep Learning Model in Uttar Pradesh, World Water Policy, 10, 244-279, 2024.

Li, Q., Liu, Y., Luo, L., Wang, Y., Wang, Q., and Ma, M.: Spatiotemporal Drought Characteristics During Growing Seasons of the Winter Wheat and Summer Maize in the North China Plain, Front. Earth Sci., 11, 1358987, 2024.

Liu, B., Liu, L., Tian, L., Cao, W., Zhu, Y., and Asseng, S.: Post-Heading Heat Stress and Yield Impact in Winter Wheat of China., Glob. Change Biol., 20, 372-381, https://doi.org/10.1111/gcb.12442, 2013.

Liu, L., Xu, H., Liu, S., and Liu, X.: China'S Response to Extreme Weather Events Must be Long Term, Nat. Food, 4, 1022-1023, https://doi.org/10.1038/s43016-023-00892-w, 2023.

Luan, Q., Gu, P., Sun, Q., Lai, B., Zhou, Y., and Weng, B.: Agricultural Drought Evaluation Based on a Soil Moisture Index Coupled Hydrological Model in North China Plain, Ecol. Indic., 166, 112473, 2024.

Nesmith, D. S. and Ritchie, J. T.: Maize (Zea Mays L.) Response to a Severe Soil Water-Deficit During Grain-Filling, Field Crops Res., 29, 23-35, https://doi.org/10.1016/0378-4290(92)90073-i, 1992.

Osman, R., Zhu, Y., Ma, W., Zhang, D., Ding, Z., Liu, L., Tang, L., Liu, B., and Cao, W.: Comparison of Wheat Simulation Models for Impacts of Extreme Temperature Stress on Grain Quality, Agric. For. Meteorol., 288-289, 107995, https://doi.org/10.1016/j.agrformet.2020.107995, 2020.

Oubaha, A., Ongoma, V., Hssaine, B. A., Bouchaou, L., and Chehbouni, A.: Evaluation of the Performance of Meteorological Drought Indices in Morocco: A Case Study of Different Climatic Zones, Int. J. Climatol., 44, 4009-4031, 2024.

Peethani, S., Pathakoti, M., Kheir, A. M., Singh, A., and Govind, A.: Variability of ENSO Teleconnections Indices, and its Impacts on Moroccan Agriculture, J. Atmos. Sol.-Terr. Phys., 265, 106386, 2024a.

Peethani, S., Pathakoti, M., Kheir, A. M., Singh, A., and Govind, A.: Variability of ENSO Teleconnections Indices, and its Impacts on Moroccan Agriculture, J. Atmos. Sol.-Terr. Phys., 265, 106386, 2024b.

Pei, J., Zou, Y., Liu, Y., He, Y., Tan, S., Wang, T., and Huang, J.: Downscaling Administrative-Level Crop Yield Statistics to 1 Km Grids Using Multi-Source Remote Sensing Data and Ensemble Machine Learning, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens., 2024a.

Pei, J., Zou, Y., Liu, Y., He, Y., Tan, S., Wang, T., and Huang, J.: Downscaling Administrative-Level Crop Yield Statistics to 1 Km Grids Using Multi-Source Remote Sensing Data and Ensemble Machine Learning, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens., 2024b.

Porter, J. R. and Gawith, M.: Temperatures and the Growth and Development of Wheat: A Review, Eur. J. Agron., 10, 23-36, 1999.

Sm, S. I., Mount, J., and Demir, I.: Assessment of Drought Impacts on Crop Yields (Corn and Soybeans) Across Iowa During 2000–2022, 2025.

Wu, X., Wang, X., Liang, X., Yang, J., Liu, Z., Wang, H., Wu, S., and Wu, J.: Differences in Winter Wheat Net Primary Productivity in Response to Agricultural Drought, Int. J. Plant Prod., 18, 639-653, 2024.

Yan, H., Wang, S. Q., Wang, J. B., Lu, H. Q., Guo, A. H., Zhu, Z. C., Myneni, R. B., and Shugart, H. H.: Assessing Spatiotemporal Variation of Drought in China and its Impact on Agriculture During 1982–2011 by Using PDSI Indices and Agriculture Drought Survey Data, J. Geophys. Res.-Atmos., 121, 2283-2298, 2016.

Yang, H., Gao, X., Sun, M., Wang, A., Sang, Y., Wang, J., Zhao, X., Zhang, S., and Ariyasena, H.: Spatial and Temporal Patterns of Drought Based on RW-PDSI Index on Loess Plateau in the Past Three Decades, Ecol. Indic., 166, 112409, 2024.

Yu, L., Du, Z., Li, X., Zheng, J., Zhao, Q., Wu, H., Weise, D., Yang, Y., Zhang, Q., Li, X., Ma, X., and Huang, X.: Enhancing Global Agricultural Monitoring System for Climate-Smart Agriculture, Climate Smart Agriculture, 2, 100037, https://doi.org/https://doi.org/10.1016/j.csag.2024.100037, 2025.

Yu, X., Niu, L., Zhang, Y., Xu, Z., Zhang, J., Zhang, S., and Li, J.: Vapour Pressure Deficit Affects Crop Water Productivity, Yield, and Quality in Tomatoes, Agric. Water Manag., 299, 108879, 2024.

Zhang, L., Zhang, Z., Chen, Y., Wei, X., and Song, X.: Exposure, Vulnerability, and Adaptation of Major Maize-Growing Areas to Extreme Temperature, Nat. Hazards, 91, 1257-1272, https://doi.org/10.1007/s11069-018-3181-7, 2018.

Zhang, Q. and Miao, C.: A New High-Resolution Multi-Drought Indices Dataset for Mainland China; CHM_Drought, in, edited, figshare, 10.6084/m9.figshare.25656951.v2, 2024.

Zhang, Q., Miao, C., Su, J., Gou, J., Hu, J., Zhao, X., and Xu, Y.: A New High-Resolution Multi-Drought-Index Dataset for Mainland China, Earth Syst. Sci. Data, 17, 837-853, 2025.

Zhang, S. and Tao, F.: Improving Rice Development and Phenology Prediction Across Contrasting Climate Zones of China, Agric. For. Meteorol., 268, 224-233, https://doi.org/10.1016/j.agrformet.2019.01.019, 2019.

Zhang, S., Tao, F., and Zhang, Z.: Changes in Extreme Temperatures and their Impacts on Rice Yields in Southern China from 1981 to 2009, Field Crops Res., 189, 43, https://doi.org/10.1016/j.fcr.2016.02.008, 2016.

Zhao, Y., Xu, X., Li, J., Zhang, R., Kang, Y., Huang, W., Xia, Y., Liu, D., and Sun, X.: The Large-Scale Circulation Patterns Responsible for Extreme Precipitation Over the North China Plain in Midsummer, J. Geophys. Res.-Atmos., 124, 12794-12809, 2019.

Zheng, J. and Zhang, S.: Improving Rice Phenology Simulations Based on the Bayesian Model Averaging Method, Eur. J. Agron., 142, 126646, https://doi.org/10.1016/j.eja.2022.126646, 2023.

Zheng, J. and Zhang, S.: Decomposing the Total Uncertainty in Wheat Modeling: An Analysis of Model Structure, Parameters, Weather Data Inputs, and Squared Bias Contributions, Agric. Syst., 224, 104215, https://doi.org/10.1016/j.agsy.2024.104215, 2025.

Zheng, J., Yu, Y., Zhang, X., and Hao, Z.: Variation of Extreme Drought and Flood in North China

Revealed by Document-Based Seasonal Precipitation Reconstruction for the Past 300 Years, Clim. Past., 14, 1135-1145, 2018.

Zhong, Z., He, B., Guo, L., and Zhang, Y.: Performance of Various Forms of the Palmer Drought Severity Index in China from 1961 to 2013, J. Hydrometeorol., 20, 1867-1885, 2019.