Author’s response

We sincerely thank the reviewer for their careful evaluation and insightful suggestions, which have greatly contributed to enhancing the clarity and quality of our manuscript. Detailed responses to each of the reviewer’s comments are provided below. All line numbers cited refer to the clean version of the revised manuscript rather than the tracked changes version.

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

Dear Reviewer,

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 line number is based on the clean version of the revised manuscript, not the track change version. The original *reviewer comments are presented in italic*, while the authors' responses are provided in blue.

Thank you for the detailed responses and the clarifications you added to the manuscript; they were very helpful. I also appreciate that you expanded the discussion of limitations, which adds important context. I have two remaining comments:

1) Thank you for providing Figure S1 and the additional details in the supplement; I had overlooked this figure in the first revision. If I understand correctly, during the training phase, SCE-UA calibration is used both to adjust WOFOST parameters (to fit phenology and yield) and to derive the "extreme weather factor" F(EW), which then serves as the target for the LSTM. In other words, F(EW) is treated as an optimizable factor during calibration, and the calibrated values are subsequently mapped back to extreme weather indices using the LSTM. Could you confirm whether this is indeed the procedure? Figure S1 clarifies the workflow well, but it is not immediately apparent from the main text. I suggest briefly summarizing this in the manuscript, as understanding how F(EW) is derived and how calibration is involved seems central to the approach.

Response:

We sincerely thank you for your careful review and insightful comments. We greatly appreciate your recognition of the supplementary materials (including Fig. S1) and your valuable suggestion to clarify the derivation of the extreme weather factor F(EW), which indeed strengthens both the clarity and rigor of the manuscript. We fully agree with your understanding of the technical workflow, and we have added key details in the main text (lines 215–220) to address the information gap you noted. Below, we provide a detailed response:

Confirmation of the derivation and calibration process of F(EW)

First, we would like to confirm that your understanding of our research workflow is fully consistent with our study design. The Shuffled Complex Evolution-University of Arizona (SCE-UA) algorithm plays a dual role in model development, and the derivation of the F(EW) factor follows a three-step route: calibration, target value generation, and Long Short-Term Memory (LSTM) mapping. The key steps are as follows:

Step 1. Dual-objective optimization via SCE-UA calibration (1980–2000 period)

During the calibration phase, the SCE-UA algorithm was used to simultaneously optimize two components:

- **Optimization of baseline WOFOST parameters:** The objective function was defined as minimizing errors in both phenology (heading and maturity dates) and yield, with priority given to ensuring phenology accuracy for biological consistency. In this process, core WOFOST parameters were calibrated. These parameters established the baseline framework for crop growth simulation, providing a foundation for subsequent integration of F(EW).

- **Derivation of calibrated F(EW) values:** In the same calibration process, F(EW) was treated as an optimizable adjustment factor. By iteratively tuning F(EW), we obtained site- and year-specific values for each agro-meteorological station. These calibrated values essentially quantify the correction strength required to offset the impacts of extreme weather. These values were subsequently used as the target output variables for LSTM training.

Step 2. LSTM training

Seven extreme weather indices (HDD: high temperature days, LDD: low temperature days, R95P: precipitation on very wet days, R10mm: number of heavy rainfall days, Rx1day: maximum 1-day precipitation, PDSI: Palmer Drought Severity Index, VPD: vapor pressure deficit) were used as input features, with the SCE-UA–derived F(EW) values serving as targets for training the LSTM network. The LSTM was designed to capture the nonlinear and spatiotemporally dynamic relationship between extreme weather conditions and F(EW). This process effectively transformed the “empirically calibrated F(EW)” into a “data-driven generator of F(EW),” enabling application to unseen extreme weather scenarios.

Step 3. Integration into the WOFOST-EW model

After training, the LSTM network was embedded into WOFOST. During simulation, the model first inputs real-time extreme weather indices into the LSTM, which dynamically generates F(EW). This factor is then multiplied with the original development rate (DVR) to produce the corrected rate:

$$DVR_{EW} = \frac{F(T)}{\sum T_i} \times F(V) \times F(P) \times F(EW)$$

(Equation 9). This integration ensures that the model can dynamically respond to extreme weather impacts without requiring repeated manual recalibration.

Once again, we are sincerely grateful for your constructive suggestions, which have greatly helped us to improve the clarity and rigor of the manuscript.

2) I appreciate the clarification that stage-specific extreme weather indices are included as inputs to the LSTM. However, since the model outputs a single season-level F(EW), I am concerned that extreme events occurring during the reproductive phase could still influence simulated development *in earlier vegetative stages, because the correction is applied uniformly. Could the authors confirm whether this is the case?*

Response:

We thank the reviewer for raising this critical point. Your understanding of the model design is entirely correct: in the current study, the Long Short-Term Memory (LSTM) network is indeed trained with the seasonal F(EW) values as the target output. In other words, the model ultimately generates F(EW) as a season-scale scalar correction factor, which is applied to the calculation of crop development rates throughout the entire growing period. This design inevitably implies the concern you raised: extreme events occurring during the reproductive stage may influence the season-long F(EW), thereby affecting the developmental process across the entire growth period. This is a limitation of the model, which we have acknowledged in the revised manuscript (Section 4.4, lines 444–450).

However, in the practical construction of the model, we do not entirely ignore stage-specific differences. Different threshold conditions are set to reflect the varying sensitivity of distinct growth stages to extreme weather. Although F(EW) is formally uniform, its derivation process implicitly incorporates these stage-specific sensitivity thresholds. Therefore, the mapping learned by the LSTM has already partially captured the differences among growth stages. In other words, while F(EW) is a season-level metric, its formation process integrates the response characteristics of extreme weather across different stages. This approach can, to some extent, mitigate the concern you raised.

The primary reason for using a season-level F(EW) is to ensure model stability under large-scale, multi-year, and multi-site conditions during training and validation with historical data (1980–2020). Dividing F(EW) into multiple stage-specific parameters would not only substantially increase the number of parameters to estimate but also introduce a “sample insufficiency—overfitting” risk. Considering the sparse temporal and spatial occurrence of extreme weather events, directly modeling stage-level F(EW) would be difficult to achieve robust training under the current data conditions.

During the independent validation period (2001–2020), we found that even when F(EW) is used as a single season-level factor, the model still significantly improves the simulation accuracy of phenology and yield. This indicates that, although F(EW) is mathematically “uniform across the season,” its practical effect can still capture, to some extent, the cumulative and integrated impact of extreme events across different stages. Nevertheless, we acknowledge that this remains a limitation of the study, and stage-specific modeling could potentially improve model accuracy.

We fully agree with the reviewer that a season-level F(EW) cannot finely reflect the differentiated impacts of extreme weather across phenological stages. This is an important limitation of our research. In the revised Discussion section (Section 4.4, lines 444–450), we have explicitly addressed this shortcoming. While the current F(EW) is indeed a season-scale scalar correction, this design still captures some degree of extreme weather effects. In future work, we aim to develop a more refined stage-specific correction mechanism to overcome this limitation.