

Reviewer 1

This paper proposes a time-block-based cross-validation framework to identify meteorological drivers causing yield failures in the US, demonstrating the portability and interpretability of the derived metrics. The research topic aligns well with GMD while exhibiting innovation. However, several core methodologies present issues. Therefore, revisions to the article are required.

Thank you for taking the time to review our manuscript and providing constructive feedback. We will incorporate the suggestions made, and we respond to the points raised in detail below.

First of all, the portability claim is not evidenced beyond time. All model setup and evaluation rely on time-blocked splits, while training and test sets share the same geographic units. In the absence of a spatio-temporal block hold-out (simultaneously changing time and space), the results substantiate only temporal generalization within locations, not spatial transferability. As such, the central claim of portability across space remains unsubstantiated. Add a primary evaluation with spatio-temporal blocking, e.g., leave-one-region-out for multiple consecutive future years.

We note that the reviewer's comment that 'all model setup and evaluation rely on time-blocked splits' does not fully characterize our methodology, as we have compared the results of the use of different types of data-splitting approaches (random, spatial, temporal and feature-cluster cross-validation) when applying the methodology, and these results are shown in the appendix. However, we agree with the reviewer that the temporal split used for the main results (as shown in Figures 4, 5 and 6) does not provide evidence that the resulting models are able to generalise spatially (i.e., to held-out states). Therefore we have adjusted our methodology in the revised manuscript (the training and test sets used for identifying drivers and evaluating model performance as represented in current figures 4, 5 and 6) accordingly, in a way that we hope will satisfy the reviewer's concerns.

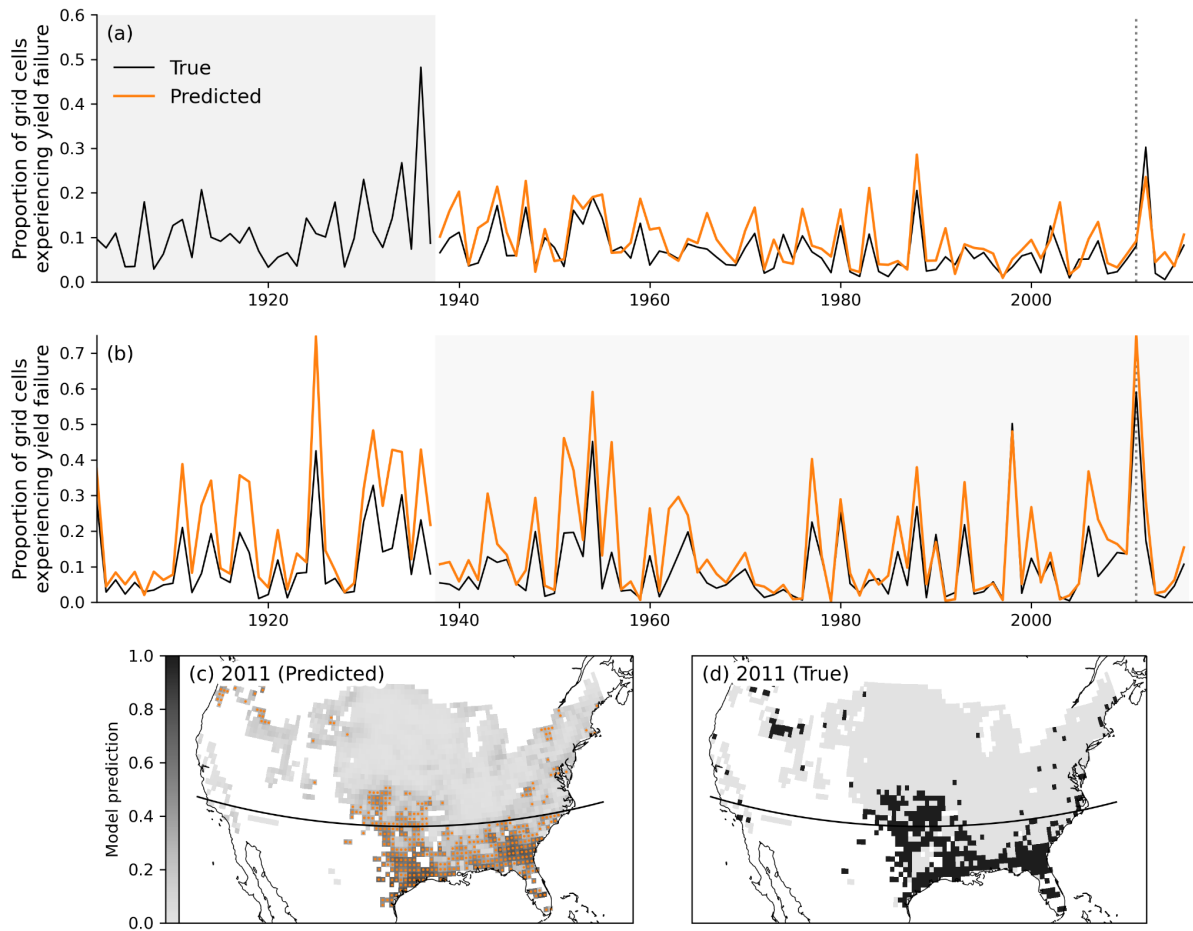
We now split the training set spatially by latitude, obtaining a spatial hold-out test set consisting of datapoints from gridcells with latitude below 35 degrees (which represents chiefly the states California, Texas, Louisiana, Mississippi, Alabama, Georgia and Florida) over the training years. The original test set is now split into a temporal test set (data from held-out years from grid cells with latitude above 35 degrees) and a spatiotemporal test set (data from held-out years from grid cells with latitude below 35 degrees). We then used the same methodology as described in the manuscript to identify climate drivers based on data in the (now reduced) training set (adjusting Figures 5 and 6 accordingly) and have recalculated the resulting model performances on those three separate test sets and report all performance metrics, adjusting the results in Figure 4 accordingly and indicating the spatiotemporal test set in panels 4b-e.

A logistic regression model with ten identified climate drivers as predictive variables, based on non-detrended pDSSAT data, using data from the (now reduced) training set only, achieves the following scores on the different test sets (see table below). The best

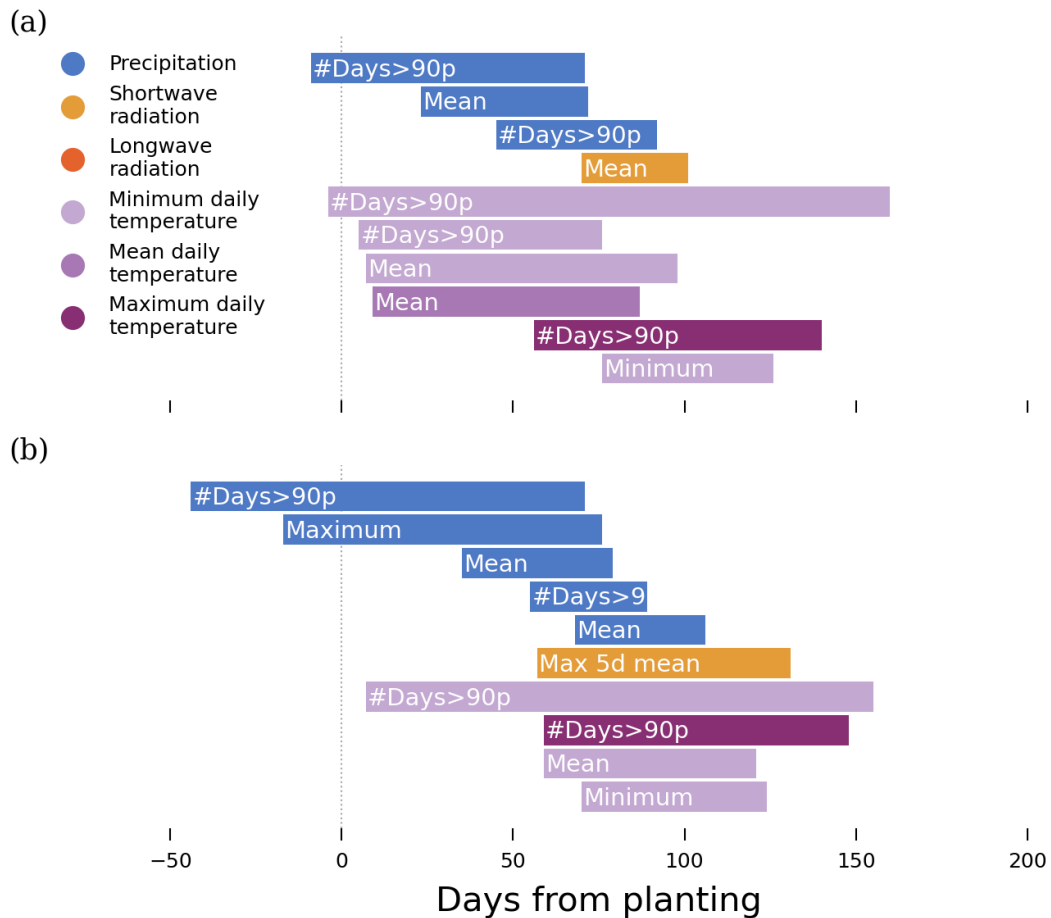
performance for each metric achieved by any of the baseline feature sets (annual, three-month or monthly averages of climate variables, with and without extreme indicators) using a lasso logistic regression model, using 5-fold temporal CV to select the value of alpha, is reported in italics for comparison. Notably, the model using ten climate drivers obtained using our method is able to outperform models using all baseline feature sets tested for every test set and for all metrics calculated. In terms of year-to-year variability in the proportion of grid cells experiencing yield failure, the model achieves a Pearson correlation between the true and predicted annual proportions of grid cells experiencing yield failure of 0.84 in the temporal test set, 0.95 in the spatial test set and 0.91 in the spatiotemporal test set (in comparison to 0.84 on the original test set as reported previously in the manuscript). Similar results are obtained for LPJmL.

Test set	ROC AUC	Average precision	Brier score	Log loss
Temporal	0.83 <i>0.79</i>	0.32 <i>0.27</i>	0.056 <i>0.059</i>	0.20 <i>0.22</i>
Spatial	0.86 <i>0.85</i>	0.51 <i>0.42</i>	0.075 <i>0.081</i>	0.26 <i>0.27</i>
Spatio-temporal	0.85 <i>0.81</i>	0.45 <i>0.32</i>	0.068 <i>0.077</i>	0.25 <i>0.27</i>
Previously reported scores (training on all spatial regions with a temporal test set)	0.84	0.36	0.059	0.21

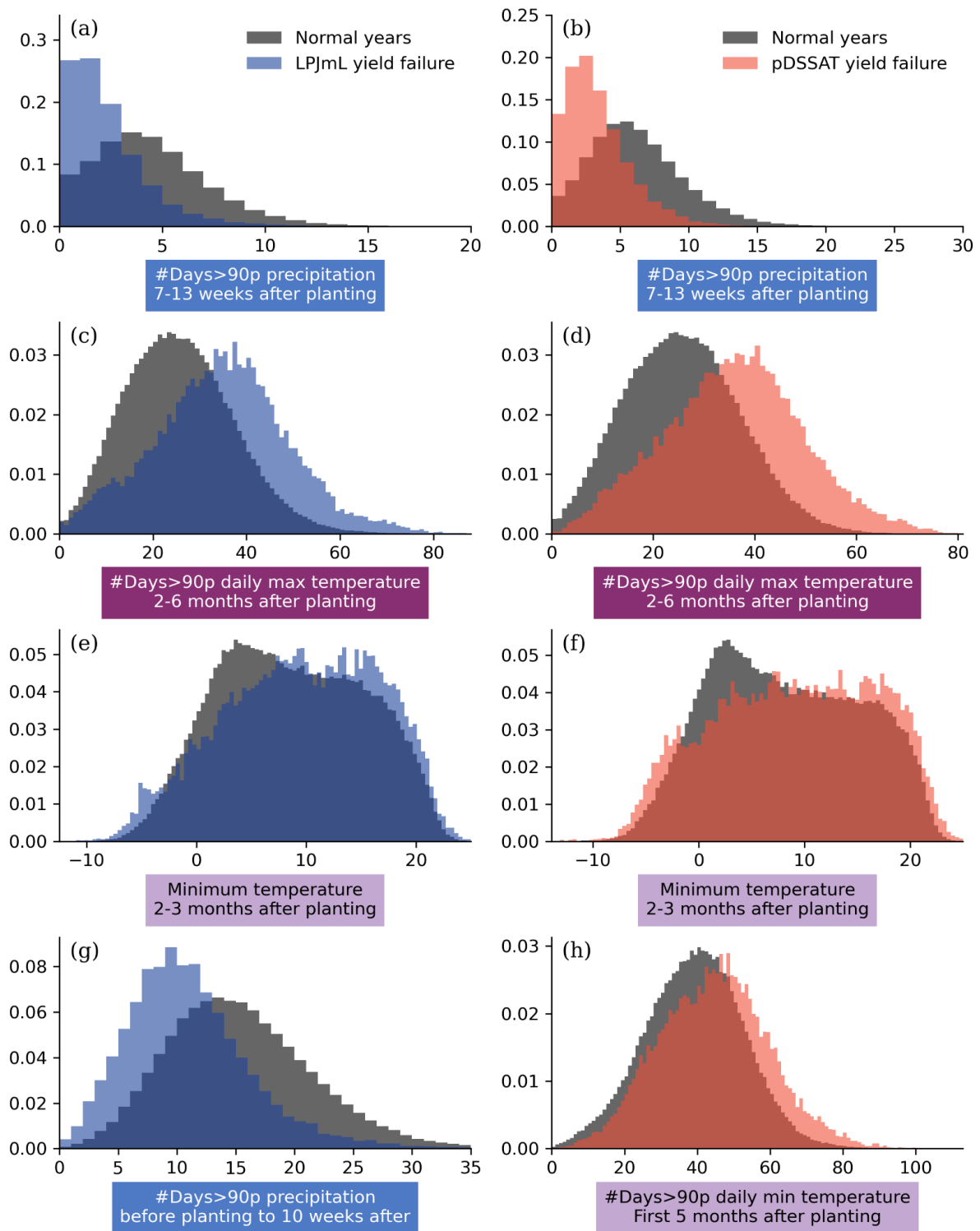
These new results show the spatial (and spatio-temporal) transferability of the drivers obtained using our method. We will update the methodology section of the manuscript to describe the updated dataset splitting procedure, and report the new performance scores in the results section (Section 4.2). We will update Figures 4, 5, 6, A2, B3, B4, B5 and Table 1 to illustrate the climate drivers and corresponding odds ratios obtained using the updated training set. We find that the obtained drivers from the reduced training set, for both crop models, are very similar to those previously obtained (based on data covering all regions).



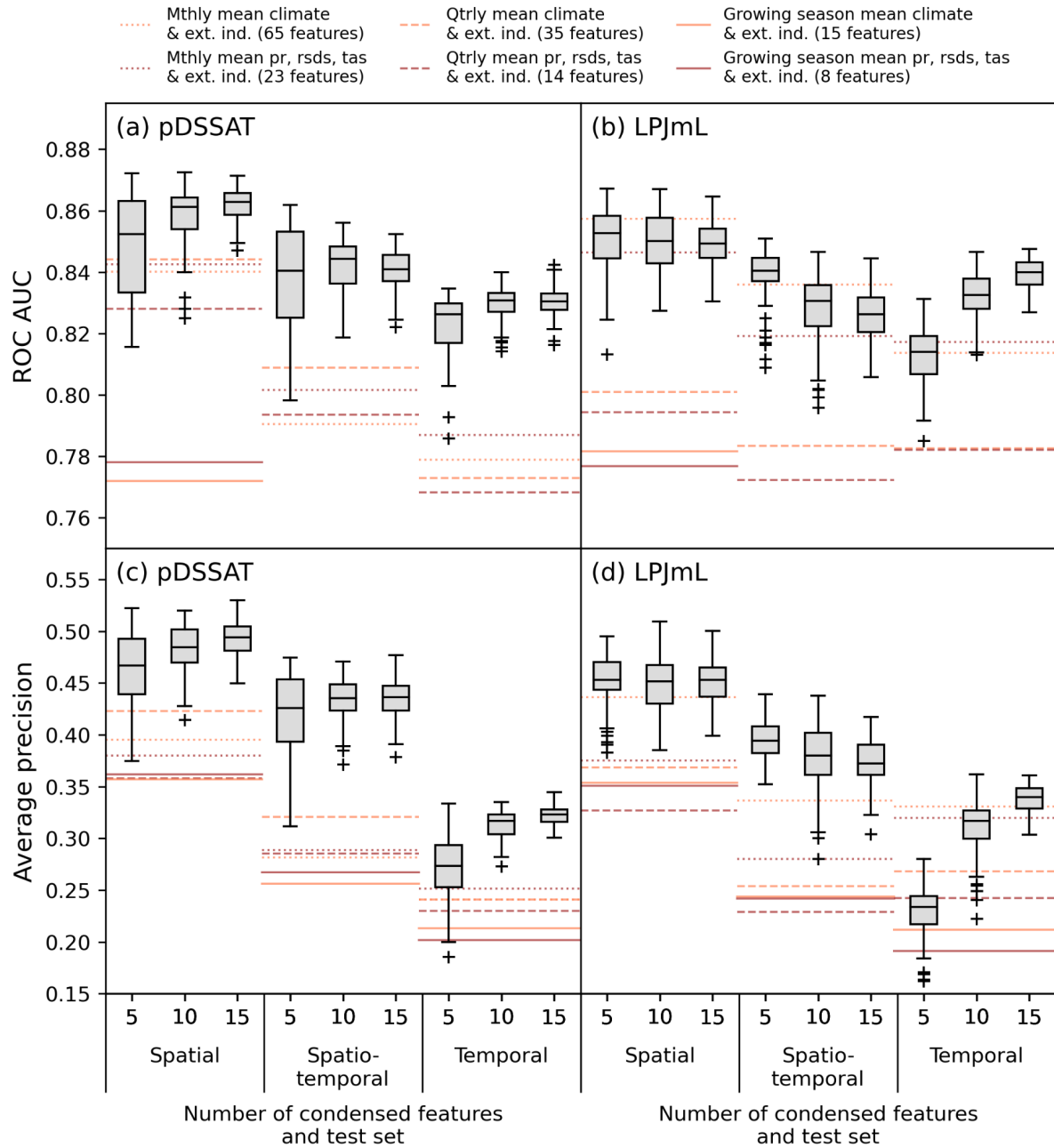
Updated Figure 4. Temporal and spatial predictions of a logistic regression model using ten identified pDSSAT drivers. (a) True (black) and predicted (orange) proportion of grid cells experiencing yield failure each year over the grid cells used for training (latitudes above 35 degrees), using a threshold of 0.235 to define a yield failure prediction (selected based on achieving the maximal f-score of 0.405 over the training years, which are marked by the grey shaded area). (b) True (black) and predicted (orange) proportion of grid cells experiencing yield failure each year in the held-out grid cells (latitude below 35 degrees). Model predictions for 2011 are illustrated in panel (c), with orange dots marking grid cells where the prediction exceeds the threshold and therefore yield failure is predicted. The corresponding ground truth is shown in panel (d), with black areas indicating locations experiencing yield failure.



Updated Figure 5. Identified climate drivers of maize yield failure for (a) pDSSAT and (b) LPJmL, based on using 30 sampled time periods to generate 100 pools of candidate features, using ten-fold temporal cross-validation to select ten features from each pool, and condensing the resulting 1000 features to ten drivers. Each climate driver is denoted by a coloured horizontal bar, and consists of an aggregated daily climate variable over a time interval which is defined relative to the planting date (illustrated by the bar's length and position). The caption overlaid on each bar describes the aggregation method used (i.e., 'mean' indicates that the corresponding daily climate values are averaged over the time period selected; '#Days>90p' means the number of days in which the respective variable exceeds the 90th percentile for that location; 'Max 5d mean' takes the maximum value of the five-day rolling mean of the respective climate variable over the selected time period). Drivers are ordered by variable and start date.



Updated Figure 6. Distributions for normal and yield-failure years of some of the identified climate drivers of maize yield failure shown in Fig. 5, for LPJmL (panels (a), (c), (e) and (g)) and pDSSAT (panels (b), (d), (f) and (h)).



Updated Figure 7. Comparison of the test-set performance of Lasso models for pDSSAT ((a) and (c)) and LPJmL ((b) and (d)). Box plots show the scores for models using sets of five, ten or fifteen drivers using our proposed method, and horizontal lines denote the scores for the same models using baseline feature sets as predictors (made up of mean aggregate climate variables over the growing season or in monthly/quarterly intervals as well as five extreme indicators).

Additionally, fixed phenology undermines window-based interpretability. Planting dates in the crop-model experiments vary across space but not across years,

whereas real-world phenology adjusts with weather, cultivars, and management. The method's reliance on calendar time windows makes the selected "critical periods" vulnerable to misalignment with true growth stages, which compromises the mechanistic meaning attributed to the discovered drivers. This concern is amplified by extensive evidence of phenological shortening under warming and by indications of misfit in the county-level results.

This is a good point - in the simulations used, planting dates are held constant, while in reality farmers may adjust their planting date in response to weather conditions and other factors. However, keeping constant the planting dates used for shifting the input data allows one to theoretically capture the impact of climate variability on yields via delayed or early planting, and means that we do not need to obtain yearly planting date data (which is not easily available). On the other hand, trends in planting dates over longer time horizons in response to climate change would make interpretation of results difficult due to confounding effects.

In general, maize cropping areas in the US have experienced relatively little warming over the last fifty years (Lobell and Di Tommaso, 2025). Maize planting dates in the US have been observed to advance between 1985 and 2005 (Sacks and Kucharik, 2011), but this trend may have subsequently stalled or reversed (Deines et al. 2023 found no meaningful trend in US maize sowing dates over 2000-2020). Therefore, we assume that this effect would not meaningfully impact the results from observational data shown in this manuscript.

However, we do agree that this is an important point, and we will revise the text of the discussion section to include our justification for this assumption. Future work to explore the robustness of this method to these effects by using crop model simulations with changing planting dates in response to climate conditions would be very beneficial. Furthermore, as data on planting dates is harder to obtain, and may be noisy, it would be useful to explore the impact of injected noise to these simulated dates on the results of the method. However, we feel that it is outside of the scope of the current study.

Revised text in the Discussion (lines 441 - 447):

While model performance using the identified drivers over the studied period is good, and the identified drivers appear plausible, it is difficult to assess the robustness of the methodology on observational data. Crop yields are affected by many types of weather conditions during different times of the growing season, making it challenging to argue that any obtained drivers are not plausible in some way. **In particular, our method identified climate drivers relative to the planting date, which were held constant in the simulations used for validating our method and assumed to be constant in the application to observational data. In reality, farmers may adjust their planting date in response to climatic and/or other conditions. While keeping planting dates constant could allow for capturing the impact of climate variability on yields via delayed or early planting, long-term trends in planting dates due to climate change would make interpretation of the results difficult due to confounding effects. Maize cropping areas in the US have experienced relatively little warming over the last half-century**

(Lobell and Di Tommaso, 2025), and although planting dates advanced between 1985 and 2005 (Sacks and Kucharik, 2011), no meaningful trend was observed over 2000-2020 (Deines et al. 2023). We therefore assume that shifts in planting dates would not have a significant impact on our results. However, this assumption may not hold in regions more affected by warming. Future work exploring the robustness of our proposed method to changing planting dates in response to climate change trends, or to uncertainties in planting date data, would be beneficial. Furthermore, observational data is only available over a limited period of time, during which agricultural practices have evolved, making robust model evaluation challenging. Therefore, while the strength of our results suggest that this approach could be applied to observation data, we advise conducting careful sanity checks and validating that the results are in agreement with scientific understanding.

D.B. Lobell, & S. Di Tommaso, A half-century of climate change in major agricultural regions: Trends, impacts, and surprises, *Proc. Natl. Acad. Sci. U.S.A.* 122 (20) e2502789122, <https://doi.org/10.1073/pnas.2502789122> (2025).

William J. Sacks, Christopher J. Kucharik, Crop management and phenology trends in the U.S. Corn Belt: Impacts on yields, evapotranspiration and energy balance, *Agricultural and Forest Meteorology* 151 (7), <https://doi.org/10.1016/j.agrformet.2011.02.010> (2011).

J.M. Deines et al., Field-scale dynamics of planting dates in the US Corn Belt from 2000 to 2020, *Remote Sensing of Environment* 291 (113551), <https://doi.org/10.1016/j.rse.2023.113551> (2023).

Labeling and detrending choices compromise internal. The definition of “failure” via a 10% threshold is pivotal yet theoretically unmotivated in the text, and no sensitivity to alternative quantiles is demonstrated. In the non-detrended analyses, the threshold is anchored to the earliest 30% of years and then applied to later decades, conflating long-term shifts with interannual shocks and biasing the failure rate. In addition, detrending windows differ between datasets (7-year for simulated yields versus 5-year for county yields), rendering labels non-comparable and further obscuring inference about drivers.

We agree that ‘crop failure’ is not a well-defined target variable, and that our decision to use a threshold of the tenth percentile of yields over the training data period is not clearly motivated within the text. We did perform some sensitivity analysis to this threshold (adjusting to 5% and 20%) and did not observe large differences in the results.

The decision of anchoring the threshold to the first 30% of years and applying it to later decades is intended to capture the real-world challenge of modelling yield failure risk in future decades when only observational data from recent years is available; we will rewrite the text to explain this more clearly. We agree with the reviewer that not detrending yield data when defining yield failure years can conflate long-term shifts with interannual shocks, and interannual variability may be more relevant when assessing climate change impacts to yields than the absolute value of yields. However, we also think that for some climate impacts and studies, it may be more relevant to identify when a variable of interest falls

below an absolute threshold, rather than a relative drop in comparison to recent years. Therefore, we chose to repeat the analysis and validation of the method for both detrended and non-detrended yields in order to assess the utility of the method in both cases.

Revised text (Lines 119-122):

Each datapoint in our compiled dataset consists of a binary target variable and the corresponding daily multivariate climate input for one growing season at one grid cell. The target variable is maize yield failure, which we define as any yield below the 10th percentile observed during the training period at that grid cell. **Due to the trends in yields due to improved management practices and breeding that have occurred over the last decades, researchers often consider relative yields (after detrending) rather than their absolute values. We test our method on both non-detrended and detrended yields (based on subtracting a seven-year rolling mean, meaning that we discard datapoints from the first six years), in order to assess the utility of the method for analysing both relative yield failure and the occurrence of yields below an absolute threshold.**

The reviewer correctly observes that different detrending windows were used for simulated and observational datasets; however, we make no attempt in this study to compare the identified drivers between simulations and observational data, so this difference should not impact the interpretation of the results.

Reviewer 2

Prediction of crop yield failure is a fundamental to food security but challenging. Both statistical and process based crop models are widely used for crop yield projection and crop failure detection. This article presents a novel approach to identifying the climate drivers of crop yield failure. The authors first developed a machine learning procedure to identify the climate drivers of yield failure and tested it on maize yield data from simulations of two process based crop models (LPJmL and pDSSAT). Since the method well captured the crop failure for the two crop model outputs, the authors further applied it to U.S. county level maize yield data and found four climate drivers for US maize yield failure. I think this approach is very interesting , as it is useful not only for crop yield failure prediction, but also for diagnosing process based models diagnose. It is also has the potential to be used in many other climate related studies. The article is well written, however, there are some non-results descriptions in the results section that could distract from the main point and confuse readers. I suggest moving some of the texts (specified in the following comments) in the results could be moved to the method or discussion sections.

Thank you for taking the time to review our manuscript and providing detailed and constructive feedback. We respond to individual comments line-by-line below.

The authors mentioned several times that it matters a lot whether the climate drivers are the input of the crop models, and they want to avoid using the selected climate drivers if they were not used as inputs of a crop model. I understand that the authors want to guide the machine learning procedure to capture the actual climate factors in each crop model. However, even if a climate driver was not an input of a crop model, some crop models may have internal processes that calculate certain climate variables. For example, VPD is normally not an input climate variable, but it can be calculated using humidity and air temperature. Therefore, I think the authors should also check if a climate driver is calculated or used indirectly in the crop model. If a non-input climate driver is used indirectly in a crop model, then it could be included in the analysis. By the way, I'm a land surface modeler. I'm wondering, if there is no wind speed in PDSSAT, how does this model calculate evapotranspiration? If there is no specific relative humidity in LPJmL, how does it determine near-surface dryness and wetness, the conditions required by radiation and surface energy processes? Do these two crop models not need to calculate plant transpiration for their crop growth simulation?

Thank you for this feedback. We consider matches between climate variables identified by our method as important and actual driver variables considered in the modeling that generated the data we worked with as a central evaluation aspect. Ideally, no driver variables would be selected that were not used to generate the yield data on which we base the analysis. However, as the reviewer rightly pointed out, some variables may be computed or approximated from others or they may also not affect the results much as they are considered only in processes not central to the yield formation in the crop model. LPJmL uses the Priestley-Taylor approach to compute potential evapotranspiration (PET), which does not consider air humidity, but implicitly represents atmospheric demand through available energy and an empirical coefficient, i.e. evaporation is controlled mainly by net radiation. Wind speed is also not considered in the Priestley-Taylor equation, yet it is used in LPJmL for the volatilization of ammonium.

Similarly, pDSSAT uses the Priestley-Taylor radiation-based approximation that implicitly assumes typical atmospheric demand for calculating PET. In case wind speed and humidity can be provided, DSSAT will use the FAO-56 Penman-Monteith method for a full physically based PET calculation. For global gridded simulations, and for GCM scenario-based model setups with these process-based crop models, climate inputs variables are usually limited to T_{min}, T_{max}, precipitation, and shortwave downwelling radiation.

We now discuss this point more explicitly in the discussion section, stating that the match between selected variables and input variables for the process-based crop models is only one aspect in the evaluation of the method's fidelity and that it can be blurred by a) functional relationships or correlations between individual variables (e.g. temperature and humidity), b) internal computation of secondary variables that are not explicitly used as inputs (e.g. PET), c) the sensitivity of model processes used to compute yield to individual driver variables (e.g. windspeed in LPJmL).

Line 179-206. These paragraphs in section 3.3 are more like method descriptions rather than results. I suggest reorganizing these texts to only show

the model evaluation and sensitivity analysis results, not how to perform these analysis.

Section 3.3 is part of the Methods section, and the results of the described evaluation and sensitivity analysis are described in section 4.3.

Line 326-338. The text of crop-validation could be moved to the discussion section.

The text indicated reports the results of the sensitivity analyses as described in Methods section 3.3. We feel that, as we are publishing this method so that it can potentially be used by other researchers, it is appropriate to assess its robustness to parameter choices, and that how we conducted these sensitivity analyses and their results are relevant to include in the Methods and Results sections.

Figure 3: In the caption of Figure 3, the authors mention that "solid lines denote the median daily climate conditions over all years and locations for normal or yield-failure years," I'm still confused about the normal and yield-failure years, as well as the data shown in Figures 3c–h. Based on Figures 3a and 3b, some grid cells experience yield failure each year. Therefore, crop failure occurs in some grid cells every year. Did you calculate the average of a climate variable for grid cells showing yield failure in the U.S. to represent crop failure years? Please clarify this in the figure caption.

Thank you for drawing attention to the lack of clarity in the manuscript here. We do indeed calculate the median of a climate variable for grid cells experiencing yield failure in the US to represent crop failure years in these figures. We will adjust this text to be more explicit.

Revised text (Figure 3 caption):

(c)-(h) Composite plots of mean daily temperature, five day precipitation sums and near-surface humidity ((c), (e), and (g) for LPJmL and (d), (f) and (h) for pDSSAT). **Solid lines denote the median daily climate conditions over all datapoints (year and gridcell combinations) where yield failure did not occur (grey) or all datapoints where yield failure did occur (blue/red), and shaded regions indicate the interquartile range.** Mean daily temperature and near-surface relative humidity are smoothed by taking the seven-day rolling mean.

Line 96. Why selecting these ten climate variables? And why using different climate variables for the US county level yield failure analysis? Please explain these different selections in the method section.

The ten climate variables were selected as these are the variables provided to all impact models in the ISIMIP simulation protocol. For the US county-level yield analysis, we made use of PRISM because it is a high-quality dataset which has been frequently used in studies analysing climate drivers of agricultural yields in the US (e.g., Roberts et al. 2012, Bobea et al. 2019, Hogan & Schlenker 2024), and we made use of all variables available in the dataset for the same reason as with the simulated dataset: our method is intended to be able

to identify the most relevant variables and time periods ‘automatically’, and selecting variables based on expert knowledge would run counter to this intention.

Revised text (Lines 96-102):

We use all ten climate variables provided to impact models in the ISIMIP simulation protocol at daily resolution: near-surface relative humidity (*hurs*, %), near-surface specific humidity (*huss*, kgkg⁻¹), precipitation (*pr*, mm), surface air pressure (*ps*, Pa), surface downwelling longwave radiation (*rlds*, Wm⁻²) and shortwave radiation (*rsds*, Wm⁻²), near-surface windspeed (*sfcwind*, ms⁻¹), near-surface air temperature (*tas*, °C), daily minimum (*tasmin*, °C) and maximum near-surface air temperature (*tasmax*, °C). However, each crop model only considers a subset of those variables; both models take *pr*, *tasmin*, *tasmax*, *rsds*, and *rlds* as input, and LPJmL additionally uses *sfcwind* and *tas*. pDSSAT does not use *tas* directly, but internally estimates hourly temperatures based on daily *tasmin* and *tasmax* values.

Revised text (Lines 112-116):

Observed daily meteorological data over the same time period, at 4 km spatial resolution, is obtained from the PRISM climate group (PRISM Group, 2018). **This dataset has been used in a number of previous studies analysing the climate drivers of agricultural yields in the US (Roberts et al. 2012, Ortiz-Bobea et al. 2019, Hogan et al. 2024).** The variables are aggregated to county-level using a weighted average of overlapping grid cells based on enclosed cropping area. Variables used consist of daily minimum and maximum temperature (°C), minimum and maximum vapour pressure deficit (hPa), precipitation (mm) and dewpoint temperature (°C). Daily mean temperature and vapour pressure deficit is estimated by averaging the minimum and maximum values, resulting in a total of eight meteorological variables.

Roberts et al., Agronomic Weather Measures in Econometric Models of Crop Yield with Implications for Climate Change, *American Journal of Agricultural Economics* 95 (2), <https://doi.org/10.1093/ajae/aas047> (2012).

Ariel Ortiz-Bobea et al., Unpacking the climatic drivers of US agricultural yields, *Environ. Res. Lett.* 14 (6), <https://doi.org/10.1088/1748-9326/ab1e75> (2019).

Hogan, D., Schlenker, W., Non-linear relationships between daily temperature extremes and US agricultural yields uncovered by global gridded meteorological datasets, *Nat Commun* 15 (4638), <https://doi.org/10.1038/s41467-024-48388-w> (2024).

Line 137. Please specify the multiple pools and candidate features in the text. It is clear in figure 1 but not specified in the text.

Thank you; we will specify the aggregations and settings more clearly in the text.

Revised text (Lines 137-142):

The method for identifying climate drivers consists of three steps (Fig. 1): first, we generate **100** pools of candidate features by using a set of chosen aggregations (**mean, minimum, maximum, number of days below 0, number of days above the 90th percentile, minimum and maximum 5-day mean**) **over each daily climate variable from 30 randomly-sampled time intervals (with minimum duration of two weeks) during the growing season**; second, we select ten features from each pool using sequential forward feature selection, based on the predictive performance of ML models on held-out time periods; finally, we collect the selected features from the pools, and extract a set of condensed climate drivers from that collection using agglomerative clustering, based on the aggregation methods, variables and periods of the growing season which are selected most frequently.

Line 41, it is unclear to which reference Sweet et al., 2025 are referring because there are two references by Sweet in 2025.

Our revised version of the manuscript will cite a revised code and data publication (Sweet 2026) in replacement to one of those references, so this will be clearer.

Line 272. Why are the two most influential climate drivers are based on mean precipitation? Based on which data? The tasmin and tasmax show the largest odds ratio. Doesn't higher odds ratios imply higher impacts?

The distance of the odds ratios from 1 implies the level of positive or negative influence. For both crop models, precipitation-based drivers have odds ratios between approximately 0.5 and 0.7 which imply a strong association with yield failure. We will add an explanation of this to the manuscript to make this clearer for readers unfamiliar with odds ratios.

Line 276. "both positive and negative associations" How do you tell the positive and negative association? Does it base on the odd ratios?

Yes, this assessment is based on the odds ratios being higher or lower than 1. We will add an explanation of this to the manuscript to make this clearer.

Revised text (Lines 270-278):

Analysis of the relationships between identified climate drivers and yield failure is made possible by the use of simple, interpretable models. Odds ratios associated with the climate drivers for fitted logistic regression models are reported in Table 1. **Odds ratios greater (smaller) than one indicate higher (lower) odds of yield failure given a one-unit increase in the predictor, conditional on the other predictors of the model, while an odds ratio equal to one would suggest no association of that predictor with yield failure, in this model.** For both crop models, the two most influential climate drivers are based on mean precipitation, with increased rainfall strongly associated with lower yield failure probability. Additionally, the climate driver with the strongest positive association with yield failure is temperature-related (for pDSSAT, the mean minimum daily temperature in the first three months after planting, and for LPJmL, the number of days where the maximum temperature exceeds the 90th percentile between two and six months after planting).

For both crop models, however, both positive and negative associations with yield failure are identified for temperature-related drivers, pointing towards nonlinear relationships between growing-season temperature and yield failure probability.

Line 281-283. “However, when yields are detrended, windspeed and long-wave radiation is not used by any of the climate drivers identified for LPJmL, nor is shortwave radiation for pDSSAT.” How to explain this?

In response to another reviewer’s suggestion, we have now changed the data-splitting procedure so that we hold out some spatial regions, reducing the training set used to identify drivers. This has resulted in largely similar drivers being identified for each crop model, but windspeed and long-wave radiation are no longer in the ten selected climate drivers for LPJmL (although do appear if a greater number of drivers is identified). While this has conveniently removed the discrepancy between non-detrended and detrended drivers for LPJmL, the discrepancy in shortwave radiation not appearing for pDSSAT after detrending is still there. However, the driver making use of shortwave radiation is the last driver of the ten to be selected (it does not appear if only nine drivers are identified), and shortwave radiation does appear as a driver for detrended pDSSAT if more than 10 drivers are requested. So, while there is a difference, it is not as strong as it appears.

However, the point made by the reviewer is an important one that we should, and will now, discuss further in the manuscript. It is very possible that there would be differences in the drivers of detrended and non-detrended yield failures, as the underlying definition of yield failure is not the same. This needs to be considered when interpreting the results of such analyses.

Line 295. Please explain the purpose of the 100 bootstrap repeat. Is 100 times enough?

We selected $n=100$ for practicality reasons but initial testing indicated that 100 repeats sufficiently approximate larger sample sizes.