



On the relative importance of climatic and non-climatic factors in crop yield models

Balsher Singh Sidhu¹ · Zia Mehrabi^{2,3} · Milind Kandlikar^{1,4} · Navin Ramankutty^{1,4}

Received: 24 November 2021 / Accepted: 6 July 2022 / Published online: 18 July 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Statistical crop models, using observational data, are widely used to analyze and predict the impact of climate change on crop yields. But choices in model building can drastically influence the outcomes. Using India as a case study, we built multiple crop models (rice, wheat, and pearl millet) with different climate variables: from the simplest ones containing just space and time dummy variables, to those with seasonal mean temperature and total precipitation, to highly complex ones that accounted for within-season climate variability. We observe minimal improvement in overall model performance with increasing model complexity using standard accuracy metrics like the root mean square error and adjusted R^2 , suggesting the simplest models, also the most parsimonious, are often the best. However, we find that simpler models, such as those including only seasonal climate variables, fail to fully capture impacts of climate change and extreme events as they can confound the influence of climate on crop yields with space and time. Automated model and variable selection based on parsimony principles can produce predictions that are not fit for purpose. Statistical models for estimating the impacts of climate change on crop yields should therefore be based on a conjunctive use of domain theory (for example plant physiology) with accuracy and performance metrics.

Keywords Statistical crop models · Model selection · Crop yield · Relative importance · Climate change · India

1 Introduction

Climate is among the major drivers of agricultural output, and predicting crop yields as a function of climate has been a topic of active research for many decades. A common method to predict yields involves building statistical crop models from historical weather and yield data collected over time and/or space. These models can then be used for multiple purposes: estimating the sensitivity of crops to climatic variability (Lobell and Field 2007; Ray et al. 2015; Ortiz-Bobea et al. 2019; Zachariah et al. 2020), predicting crop

✉ Balsher Singh Sidhu
balsher.sidhu@ires.ubc.ca

Extended author information available on the last page of the article

yields under different future climate change scenarios (Birthal et al. 2014; Ray et al. 2015), assessing benefits of crop switching (Rising and Devineni 2020), identifying regions where agricultural interventions like irrigation can help mitigate climate change impacts (Zaveri and Lobell 2019), among others.

A variety of different weather and climate drivers of crop yields have been reported in statistical crop modeling literature. These include the most intuitive and widely used variables like average temperature and total precipitation over the crop growing season, to variables geared towards identifying and incorporating more complex determinants of yield including (but not limited to) number of precipitation days (Fishman 2016), duration of longest dry–wet spell (Tebaldi et al. 2006), growing degree days (Albers et al. 2017), heat or killing degree days (Butler and Huybers 2013), or vapor pressure deficit (Jiang et al. 2021).

In addition to weather and climate, statistical models also need data on non-climatic factors that may impact crop yields, such as soil characteristics, irrigation, use of agrochemicals (fertilizers or pesticides), mechanization and technology uptake, and cultivar varieties. Since not all data are usually available, this task is often accomplished indirectly by including a geographic factor (e.g., lowest appropriate geographical entity like district, county, state, or country) as a dummy variable to account for spatially-variable but time-invariant drivers of crop yield, with temporal variables (often the year of planting or harvest) employed to account for factors with temporal trends (Lobell and Burke 2009). These two categories of independent variables, climatic and non-climatic, together act as input in statistical crop models, which then model the dependent variable (crop yield in this case) as a function of these independent variables.¹

Crop models are built using a wide variety of statistical techniques. These vary from the most popular ordinary least squares (OLS) linear regression (henceforth, linear regression), to advanced machine learning techniques that can extract more resolved climate–yield relationships (Beillouin et al. 2020). Depending on the statistical method used, model accuracy is often measured using statistical parameters like coefficient of variation (R^2), adjusted R^2 , Akaike information criterion (AIC), or root-mean-square error (RMSE) (James et al. 2013). Researchers also use these metrics to build multiple models with different climate variables, compare them, rank them, or identify the most appropriate and accurate one or ensemble from the mix (Kern et al. 2018; Feng et al. 2020); for more examples, refer to SI.1. The model thus selected may then be employed for accomplishing tasks discussed previously. Model and/or variable selection for inference is itself a vast topic; it is in fact inherent in most machine learning applications.

This study focuses on the interplay between the two concepts discussed above: the role of non-climatic variables in explaining crop yield, and the use of standard statistical metrics to estimate and compare crop model accuracy. While including geography and time accomplishes the objective of accounting for non-climatic crop yield determinants, comparing models solely on the basis of overall variance explained (R^2) or overall prediction accuracy (RMSE) has a limitation: these metrics only refer to overall model accuracy, but do not quantify the individual contributions of climatic and non-climatic variables in the model. Consequently, when using these metrics to compare two distinct crop models, there is no specific information about the climatic or non-climatic source of any potential

¹ We use the designation “non-climatic” for district dummies and time, but this is not strictly true because climatic conditions vary across space; hence, district dummies may also carry some climatic signal. This aspect of statistical models is examined in greater detail in this study.

differences in model performance. This complicates studies trying to estimate the utility of including specific climatic variables in their statistical models, because a portion of the anticipated improvement in model performance with the inclusion of a new climate variable to an existing model may be subsumed within the non-climatic (geography and time) component. These geography and time variables may therefore complicate the use of generic model performance metrics like R^2 , adjusted R^2 , or RMSE for model comparison and selection. This is the primary hypothesis examined in this study. While we focus only on OLS linear regression, the hypothesis warrants examination for other statistical techniques as well.

We start by analyzing the cumulative contributions of climatic and non-climatic (geography and time) factors to a crop model's total predictive worth. We then parse and compare the relative importance of these two groups of variables across an array of models. The implications of our findings are then discussed in relation to model utility for predicting the impact of anomalous weather events and long-term climate change on crop yields. Specifically, our study attempts to answer the following two questions:

1. What is the role of geography and time, as proxies of unobserved non-climate variables, in explaining crop yields across multiple models with different sets of climate variables?
2. What are the implications of comparing and selecting models based solely on generic statistical metrics, in light of the role non-climatic factors (geography and time) may play in the models?

2 Data and methods

2.1 Crop production data

This study was designed as a detailed analysis of statistical crop models using India as a case study. We used data from India because of prior familiarity with this region and easy availability of long-term crop production and climate data. Although our study was based in India, we believe that the methods of model selection that we developed are more generally applicable to studies analyzing statistical relationships between crop yields and climate in other parts of the world.

We focused on India's three major cereal crops for this study: rice and pearl millet grown during the summer monsoon (kharif) season, and wheat, which is primarily grown during the winter (rabi) season. Crop production (tonne) and harvested area (ha) data, disaggregated by crop, year, and district, were acquired from the International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) Village Dynamics Studies in South Asia (ICRISAT 2015). This data is reported for 311 districts from 1966 to 2011 using 1966 district boundaries as base. Crop calendar data for crop sowing and harvesting dates at state-level came from Government of India (Ministry of Agriculture and Farmers Welfare 2016). Any aggregation of the climate data from daily to seasonal scale was done after masking it for the growing season for each crop-state combination.

2.2 Weather data

District-level daily minimum temperature, daily maximum temperature, and daily precipitation data were acquired from Indian Meteorological Department (Rajeevan et al.

Table 1 Description of variables included in the crop models

Variable name	Description
T_avg_mean	Mean daily average temperature during the growing season
T_min_mean	Mean daily minimum temperature during the growing season
T_max_mean	Mean daily maximum temperature during the growing season
gdd_0, gdd_10, gdd_20, gdd_30	Degree day bins, at 10 °C intervals
P_sum	Total seasonal precipitation
P_days	Total seasonal precipitation days (precipitation > 0.1 mm (May 2004))
P_sum_subseasonal_1, P_sum_subseasonal_2, P_sum_subseasonal_3, P_sum_subseasonal_4	Subseasonal precipitation over the four crop growing stages, as defined by FAO (Allen et al. 1998)
P_days_subseasonal_1, P_days_subseasonal_2, P_days_subseasonal_3, P_days_subseasonal_4	Subseasonal precipitation days over the four crop growing stages, as defined by FAO (Allen et al. 1998)

2006). The temperature data (1961–2015) covered 634 districts using current boundaries, while the precipitation data (1961–2015) had 651 districts. We extracted climate data for 1966–2011 and harmonized it to ICRISAT district boundaries by apportioning data for new districts created after 1966 to their parent districts using area-weighted averaging. With this daily temperature and precipitation data, we derived multiple climate variables for use in our models (Table 1). The concept of degree days is very common in crop modeling research (Roberts et al. 2017). Instead of single thresholds for growing or killing degree days, we adopted a more flexible approach by including multiple degree day bins, which the model could then parameterize independently (detailed methodology in SI.3).

2.3 Statistical analysis

2.3.1 Regression models

All statistical models were built using the stats package in R (R core team 2020). For building crop models with different climate variable sets, we used OLS linear regression model specification:

$$y_{it} = \alpha_i + \beta(t) + \gamma_1(\text{clim_var}_1) + \dots + \gamma_n(\text{clim_var}_n) + \varepsilon_{it}, \quad (1)$$

where y_{it} is crop yield in district i and year t ; α_i is district specific intercept; β is parameter for linear time (harvest year) trend; γ_n is parameter for the n th climate variable (clim_var) included in the model; ε_{it} is the residual error.

We constructed ten models with varying complexity of climate variables. We first built a null model, wherein crop yield was modeled as a function of solely geography and time (only district IDs and year of harvest were used as independent variables). No climate variables were included in the null model. The other nine models ranged from a simple model with only mean seasonal temperature and total seasonal precipitation, to the most complex one with mean daily minimum temperature, mean daily maximum temperature,

degree day bins, subseasonal precipitation amounts, and subseasonal precipitation days. We have divided the climate variables into groups of temperature and precipitation variables, with three sub-groups or levels in each (Table 2). All climate variables used in this study have been shown to affect crop yields in past studies (SI.2), and have physiological basis. Therefore, the increasingly complex models in our study also have increasing physiological basis, and the terms “complex models” and “physiologically-based models” are used interchangeably.

Studies estimating impacts of weather and climate usually include irrigation in their analysis. We began by using percent area irrigated for each crop-year-district combination as a proxy for irrigation (data on actual water used was unavailable), with interactions between irrigation and each precipitation variable. However, in the “relative importance” section of our analysis (discussed later), irrigation would be deemed a non-climatic variable, but its interaction with climate would complicate the disaggregation of total variance explained into climatic and non-climatic portions. So, we left out irrigation in our analysis presented in the main paper from here on. Nonetheless, we present our results with irrigation included in SI.4. The general trends between models’ performance are consistent with our results without irrigation.

2.3.2 Model performance metrics

We used three popular statistical metrics for computing and comparing model accuracy. The first two were R^2 and adjusted R^2 . Both these metrics vary from 0 to 1, and measure how well the model predictions match the actual observed data. A drawback of R^2 is that it always increases (or stays the same at a minimum) with the addition of any variable, regardless of whether that variable has any correlation with the variable of interest (James et al. 2013). Adjusted R^2 fixes this limitation by penalizing the R^2 statistic for the number of variables included in the model. Therefore, adding a variable with little explanatory power can decrease a model’s adjusted R^2 , unlike R^2 which increases monotonically. Nevertheless, we retained R^2 because of its fundamental relationship with the statistical metric called “relative importance” which we introduce and discuss in the next section.

The third statistic we used was RMSE, the square root of the mean of the squared differences between observed and predicted values. To prevent overfitting, we conducted RMSE analysis using out-of-sample tenfold cross-validation with random samples stratified over years (Ortiz-Bobea et al. 2019).

2.3.3 Relative importance of variables

To estimate the individual contribution of different variables in explaining the observed yield variance, we used the concept of “relative importance” (Grömping 2006). This metric refers to the contribution of every independent variable (IV) to a multivariable linear regression model. Specifically, relative importance denotes the portion of total variance explained (or R^2) that can be attributed to a particular variable in the model. For linear regression with uncorrelated data, an individual IV’s contribution is simply the increase in R^2 observed with the addition of that IV to a model with the remaining variables. This, however, is not true in studies where the various IVs usually have some correlation. The various climate variables in our study are not only correlated with each other (SI.11), but also with geography and time (Zaveri and Lobell 2019). Consequently, the increase in R^2 with the addition of a variable is dependent on the variables previously present in the

Table 2 Model names and climate variables included in the ten models

Model name	Temperature		Precipitation	
	Sub-group level	Variables	Sub-group level	Variables
Null	0	-	0	-
Tavg_Psum	1	T_avg_mean	1	P_sum
Tavg_Psumday	1	T_avg_mean	2	P_sum, P_days
Tavg_Psumday_subseasonal	1	T_avg_mean	3	P_sum_subseasonal_1, P_sum_subseasonal_2, P_sum_subseasonal_3, P_sum_subseasonal_4, P_days_subseasonal_1, P_days_subseasonal_2, P_days_subseasonal_3, P_days_subseasonal_4
Tminmax_Psum	2	T_min_mean, T_max_mean	1	P_sum
Tminmax_Psumday	2	T_min_mean, T_max_mean	2	P_sum, P_days
Tminmax_Psumday_subseasonal	2	T_min_mean, T_max_mean	3	P_sum_subseasonal_1, P_sum_subseasonal_2, P_sum_subseasonal_3, P_sum_subseasonal_4, P_days_subseasonal_1, P_days_subseasonal_2, P_days_subseasonal_3, P_days_subseasonal_4
Tminmaxgdd_Psum	3	T_min_mean, T_max_mean, gdd_0, gdd_10, gdd_20, gdd_30	1	P_sum
Tminmaxgdd_Psumday	3	T_min_mean, T_max_mean, gdd_0, gdd_10, gdd_20, gdd_30	2	P_sum, P_days
Tminmaxgdd_Psumday_subseasonal	3	T_min_mean, T_max_mean, gdd_0, gdd_10, gdd_20, gdd_30	3	P_sum_subseasonal_1, P_sum_subseasonal_2, P_sum_subseasonal_3, P_sum_subseasonal_4, P_days_subseasonal_1, P_days_subseasonal_2, P_days_subseasonal_3, P_days_subseasonal_4

model. To disaggregate total variance explained among the regressors, both climatic and non-climatic, we calculated the average increase in R^2 with the addition of a variable to all possible models with distinct permutations of the remaining variables (Grömping 2006).² The resultant relative importances of the variables add up to total R^2 of the model.

We summed up the relative importances of all climate variables to compute the total variance explained by climate in each model formulation. The trends in non-climatic (geography and time) and climatic variables' relative importance with increasing model complexity were then analyzed to ascertain the overlap, if any, between variance explained by these variables.

2.4 Evaluating influence of model choice on yield predictions during extreme weather events

In addition to comparing models based on their spatio-temporally averaged accuracy, we also wanted to analyze model performance in different parts of the country during anomalous periods of extreme weather events. For the time period of our study, 1966–2011, we calculated national mean annual temperature and total annual precipitation from area-weighted average of district-level climate data. The years with least total precipitation and highest mean temperature were designated as “drought year” and “hot year”, respectively. Individual years with conditions closest to the median temperature and median total precipitation respectively were designated as “normal years” for benchmarking purposes. The performance of all models was then compared for the drought, hot and normal years, in terms of RMSE reduction for each year.

2.5 Simulations of climate change impact

Similar to other studies investigating the impact climate change on crop yields (Lobell et al. 2011), we conducted scenario analysis to estimate the impact long-term climate change over the historical period of 1966–2011 has already had on India's crop yields. Daily minimum temperature, daily maximum temperature, and daily precipitation data was linearly detrended to remove time trend at district-scale. This detrended data was then assumed to denote the weather that would have occurred if climate change had not occurred. Using this detrended daily weather data, we used the same procedure as with the actual weather data to construct all our climate variables of interest. To obtain district-level estimates of climate change impact on crop yields, we conducted residual bootstrapping (Li and Maddala 1996) with 500 repetitions to predict crop yields with and without climate change, and then computed the median value and 95% confidence intervals of the difference between predictions from the two scenarios. For each crop-district-model combination, the average of the ten yield loss values in the last decade in the dataset (2002–2011) is presented as the expected impact of climate change that has occurred since 1966, this study's start year. While our calculation of the climate change impact on crop yields uses 1966 as the baseline year, anthropogenic climate change has been ongoing since long before that, and therefore, our estimated impact of climate change with this simulation is conservative.

² For example, for a model with independent variables IV_1 , IV_2 , and IV_3 , the relative importance of IV_2 would be the average of the increase in R^2 when IV_2 is added to every possible model without IV_2 .

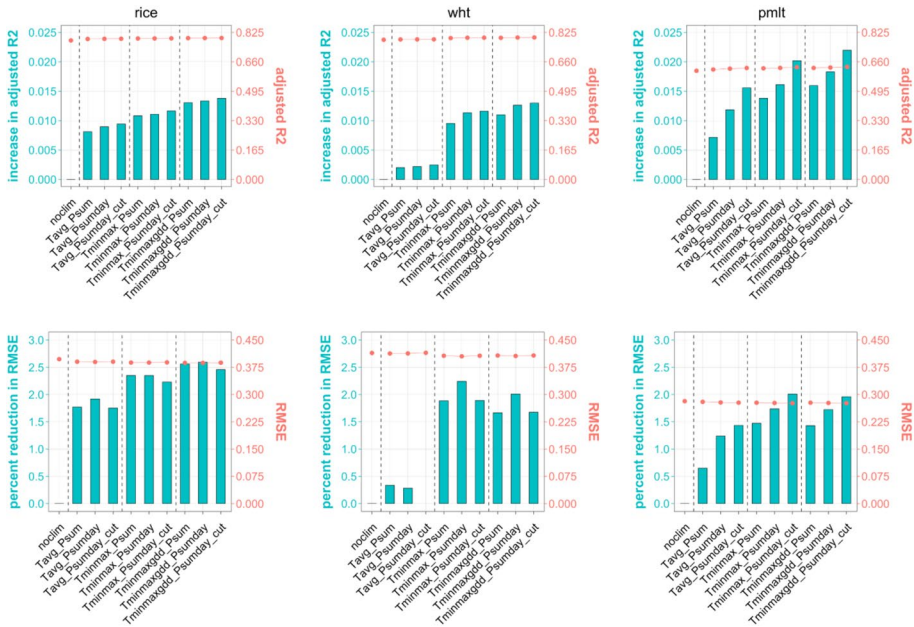


Fig. 1 Model performance measured in terms of adjusted R^2 (top row; absolute values in red, increase compared to null model in blue) and RMSE (bottom row; absolute values in red, percent increase compared to null model in blue). The three crops are rice (left), wheat (center), and pearl millet (right). Within each panel, models include varying levels of climate data, with three levels each of temperature and precipitation (see Table 2 for description of levels). The models are divided into sub-panels with dotted lines and arranged in the following order: null model; temperature level 1 and precipitation levels 1, 2, and 3; temperature level 2 and precipitation levels 1, 2, and 3; and temperature level 3 and precipitation levels 1, 2, and 3

3 Results

3.1 Model performance evaluation using statistical metrics

The performance of the models was first analyzed in terms of adjusted R^2 and RMSE (Fig. 1). Adjusted R^2 depicts similar trends for all three crops: while it increases as more climate variables are added to a model, the increase is only marginal. For rice, the advantage of choosing the best performing model with the most climate variables, compared to the null model without any climate variables, is an increase in adjusted R^2 from 0.780 to 0.794. For wheat, the increase is from 0.784 to 0.797, while pearl millet models outperform the null model by 0.022 units at the most (0.620 to 0.642). This apparently limited utility of climate in our crop models is re-affirmed in the RMSE plots which show that adding more climate variables may even decrease model accuracy, as is visible for both rice and wheat going from level 2 to 3 of precipitation in each sub-panel. In fact, selecting T_avg_Psuumday_subseasonal model for wheat (bar 4 in bottom-centre panel in Fig. 1) provides no benefit over the null model when compared on the basis of RMSE reduction. Pearl millet exhibits a more consistent pattern of improvement in RMSE reduction with more climate variables, even though that trend is broken between levels 2 and 3 of the temperature variables. To summarize Fig. 1, adjusted R^2 and RMSE show that the accuracy and fit of all models for all three crops are not very different from the null model containing only

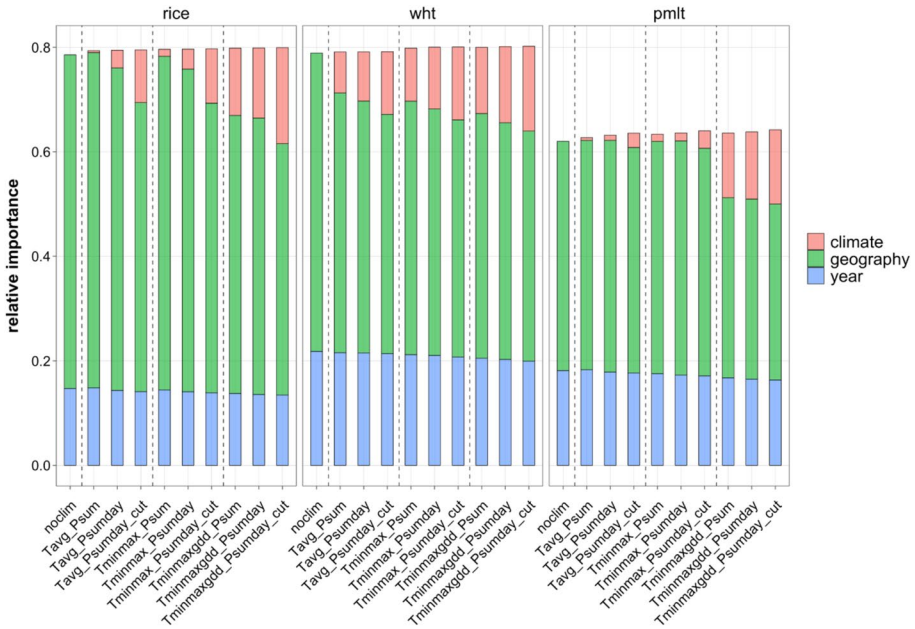


Fig. 2 Relative importance of time (blue), geography (green), and climate (red) variables across the ten models analyzed for rice (left), wheat (center), and pearl millet (right). The plots follow the same arrangement as Fig. 1 for direct comparison. Note that the sum of the relative importances of time, geography, and climate variables equals R^2 , which shows minimal improvement in overall model fit from the simplest null model on the left to the most complex model on the right, for each crop. Similar scheme of arranging models by temperature and precipitation variables’ complexity is followed as Fig. 1

geography and time as the variables of interest, and that a model’s performance does not depend much on what climate variables are included in that model.

3.2 Relative importance of variables

We used the previously discussed metric of relative importance to apportion variance explained by a crop model to different explanatory variables included in the model. Unlike with the standard metrics, for all three crops analyzed, the relative importance of geography, and time to a smaller extent, reduces as more climatic variables are added to the models to account for subseasonal climate variability (Fig. 2). Hence, even though the total variance explained, or R^2 , may not increase by the same amount, the portion of the variance explained that can be attributed to climate is increasing disproportionately more compared to the change in model R^2 . For all crops, and for all model progressions within each sub-panel, as more climate variables are added to account for precipitation availability (going from total seasonal precipitation to subseasonal precipitation and precipitation days), the relative importance of climate goes up, while that of time and geography goes down. From a low value of 0.004, 0.078, and 0.005 in the simplest models (seasonal temperature and precipitation) for rice, wheat, and pearl millet, the relative importance of climate goes up to 0.184, 0.162, and 0.142 in the most complex models on the right. While the maximum increase in adjusted R^2 or RMSE over the null model is less than 0.02 units

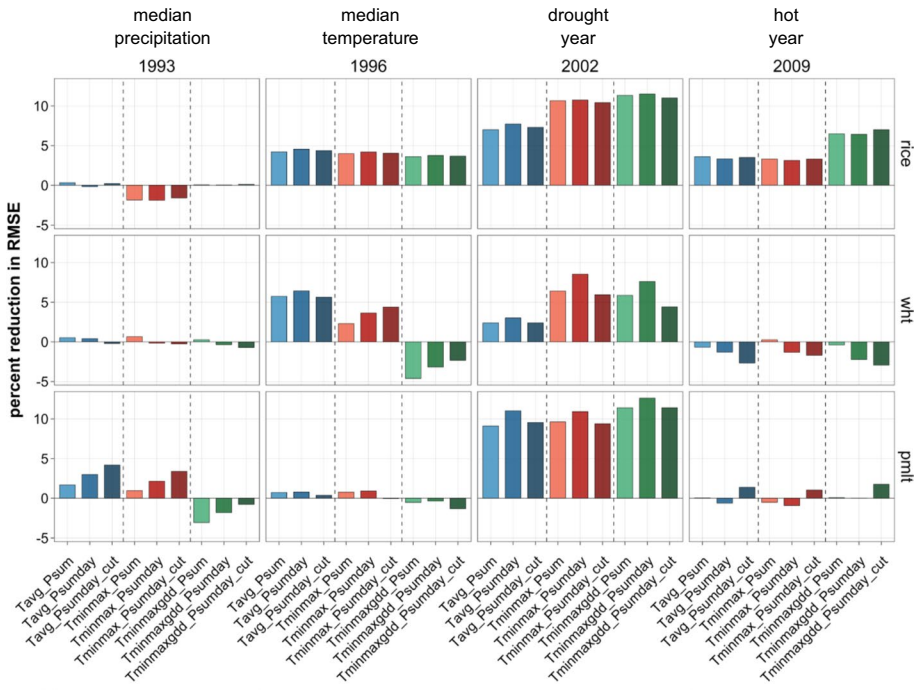


Fig. 3 Improvement in model performance (in terms of RMSE reduction compared to the null model with no climate variables) for median precipitation (1993), median temperature (1996), drought (2002), and hot (2009) years

and 3%, respectively (Fig. 1), relative importance analysis shows that the contribution of climate can be more than 20% of the total variance explained by a model. Supplemental results of relative importance analysis at the level of individual climate variables are available in SI.5.

3.3 Model sensitivity to extreme weather events

In the timeframe of our study, the least amount of rainfall fell during 2002, which we designated as a “drought year”; 2009 because of its highest mean annual temperature was designated as a “hot year”. Our method matches the results of Aadhar and Mishra (2021) who analyzed South Asian climate data from 1951 to 2016 and found that the worst drought during this period occurred in 2002, affecting more than 65% of the region. The years with median precipitation (1993) and median temperature (1996) constituted “normal years”. Models’ performance in these years was compared by calculating national RMSE of model predictions for each of these years from the tenfold out-of-sample cross-validation results described previously (Fig. 3). We also conducted this analysis at a more local-scale by calculating state-level RMSE for each model in a similar manner. Nationally aggregated RMSE reduction for all models and crops (compared to respective null models) is shown in Fig. 3. Similar plots, but with RMSE aggregated at state-level, are available in SI.8.

Compared to the null model with only geography and time, the improvement in performance from the simplest model to the most complex models depends a lot on the year in

question. In the drought year of 2002, all models exhibit an enhanced performance compared to the other years. There is a general trend of all models performing better in 2002 than the other years, irrespective of the levels of climate variables included in them. While the overall improvement in model performance when measured for the full time period hovers around 2–3% reduction in RMSE compared to null model (Fig. 1), the more complex models exhibit performance improvement in excess of 10% during the drought year of 2002 (Fig. 3).

For rice, the more complex models have a markedly better performance than the simpler models for both the anomalous years (2002 and 2009), as opposed to the normal years where additional climate variables have little impact on model performance. This trend is also exhibited by wheat but only for 2002. Pearl millet shows a more subdued difference between simple and complex models in 2002. In contrast, for all three crops, the performance of the complex models is either similar to, or worse than the simpler model in 1993 and 1996, leading us to infer that the complex models can be better suited than the simpler models at accounting for anomalous weather patterns. This is especially evident in the drought year (Fig. 3).

The difference in the performance of the models in anomalous years is more pronounced when model predictions are analyzed at state-level (Figs. SI.8.1 to SI.8.3). There are some important crop-state combinations, like rice in Madhya Pradesh and Punjab; wheat in Gujarat, Haryana, Maharashtra, and Punjab; and pearl millet in Karnataka and Rajasthan, where the RMSE reduction is highest for the more complex models (over 25% in some cases) during the drought year of 2002; simpler models are unable to match this accuracy. Figure 4 shows the difference between predicted and observed pearl millet yield in the state of Rajasthan, the biggest producer of this crop in India.³ In the anomalous years of 2002 and 2009, the complex model performs better than the simple model: blue points are located closer to zero than the red points. This is distinctly visible in the drought year of 2002; the signal in 2009 is weaker. However, during the median precipitation and temperature years, the complex model performs either similar to the simple model (1993), or exhibits slightly worse predictions (1996).

While there are instances where the simpler models outperform the complex models in the drought or hot years, or when the complex models outperform the simpler models during normal years too, the trend is biased towards complex models having higher utility than simpler models in 2002 and 2009. For quantitative evidence of this trend, we scored and ranked our models according to the level of climate complexity (Table 2), with scores of 1, 2, and 3 for each level of complexity of temperature and precipitation variables. The scores for the best performing model for each crop, year, and state were averaged to get a national score for each crop-year combination. For rice, the average scores of both the normal years are 2.9, while the drought and hot years' scores average 4.2 and 3.2. More complex models performed better than the simpler models in the drought year of 2002; this was also observed for the hot year (score of 2.9 vs 3.2), although the difference was a lot smaller. This trend was visible across the other two crops too, and these scores for wheat and pearl millet were 2.3, 3.1, 4.0, 2.4 and 3.2, 2.6, 4.1, 3.2, respectively (Fig. 5). Overall, the drought year exhibited a bigger jump in model complexity score, while the hot year was more muted. We redid the analysis with a different scoring scheme based on number

³ We use pearl millet in Rajasthan as an example. But this trend was visible in other state-crop combinations as well. For example, rice in Odisha (SI.6).

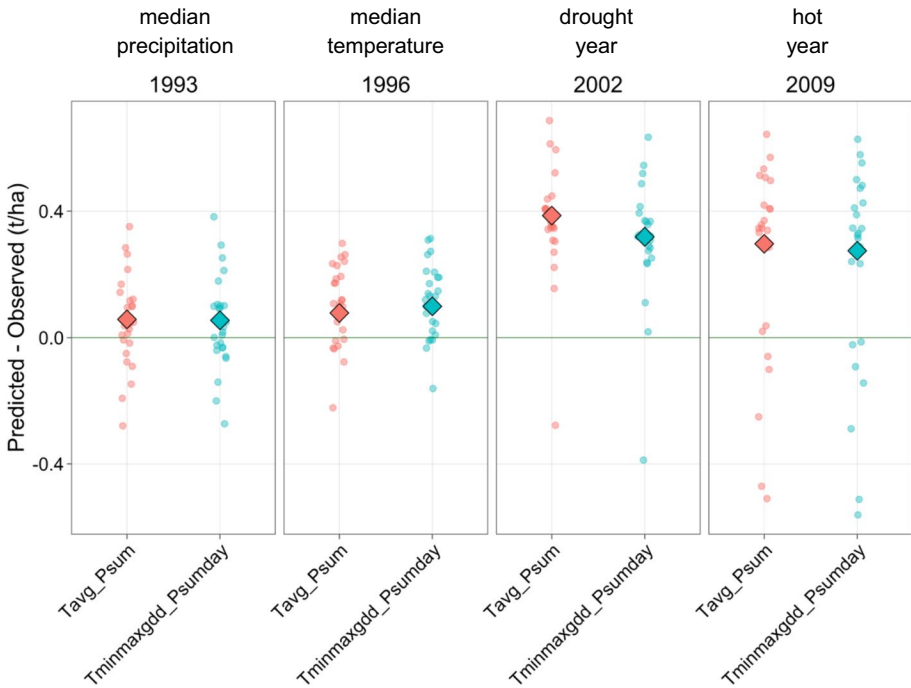


Fig. 4 Difference between predicted and observed pearl millet yield for all districts of the state of Rajasthan for 1993, 1996, 2002, and 2009. The advantage of the most complex model (blue) over the simplest model (red) is most pronounced in 2002 and to a lesser extent in 2009. Plot shows district-level (round) and average state-level values (district values weighted by crop harvested area; diamond)

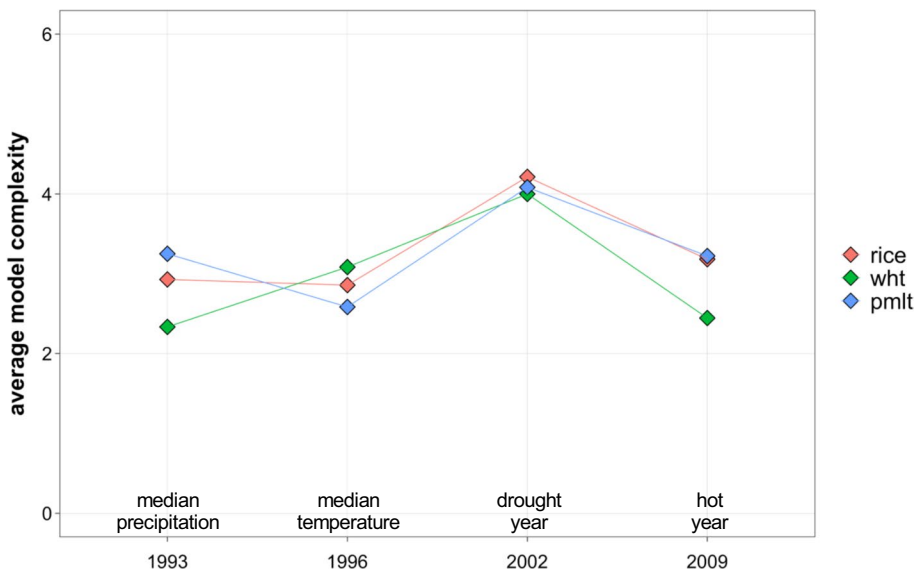


Fig. 5 Nationally averaged score of the best performing model for each crop-state combination for 1993, 1996, 2002, and 2009. The score (metric of complexity level of climate variables) is markedly higher in the drought year of 2002

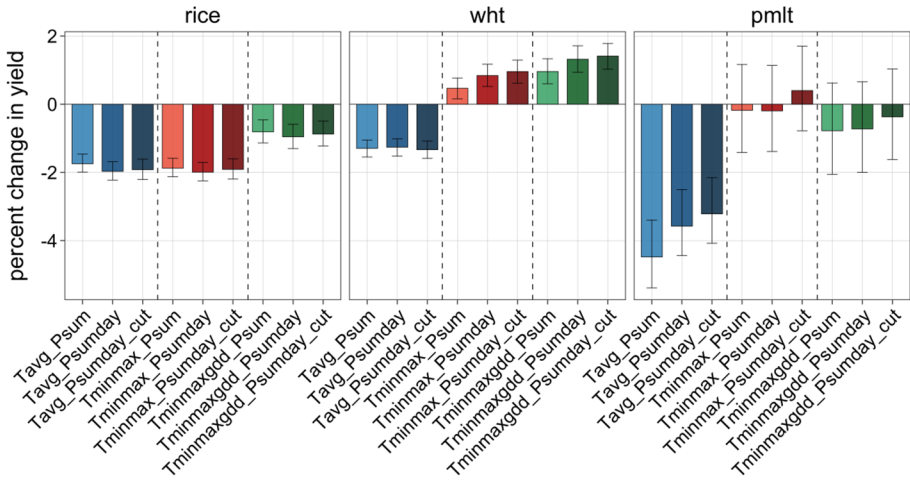


Fig. 6 Nationally averaged yield change due to historical climate change since 1966. The plots show the change in crop yields due to historical climate change (since 1966) in the last decade of this study (2002–2011). The error bars depict 95% confidence intervals from 500 bootstrap simulations

of climate variables, and the difference between model scores for normal and anomalous years is even starker (SI.7).

3.4 Simulations of climate change impact

In terms of our simulated impact of climate change, nationally averaged yield change results estimate rice yield losses from all nine models (Fig. 6). In contrast, more variation is observed for wheat, where the most complex models predict a net gain in nationally averaged yield for the crop. Our estimates with mean seasonal temperatures show that national pearl millet yield has witnessed a reduction due to climate change, although there is no significant change observed from the predictions of models with more granular temperature variables. One common pattern among all three crops, especially rice and pearl millet, is that the estimated impact of mean climate change on crop yields is more dependent on the complexity of temperature variables included as opposed to the level of precipitation variables added.

Figure 7 shows the simulated impact of climate change on rice yield during the last decade of this study’s time period (2002–2011); similar plots for wheat and pearl millet are available in SI.9. The panels from top to bottom depict an increasing number of variables to account for temperature variability, and panels from left to right denote models with increasing levels of precipitation variables. For all three crops, there are significant differences between the predictions by the nine different models we analyzed.

Rice plots show a negative impact for most of the country for the simpler models containing only mean seasonal temperature (top row), except a small patch in eastern India (region A) where rice yields are predicted to have benefited from climate change. As more precipitation variables are added, there is a region in south India covering the states of Kerala and Tamil Nadu (region B) where the predicted impacts of climate change become more drastic. In fact, the bottom row depicts a reversal from a small net positive to a net negative impact of climate change in this region when subseasonal precipitation variables

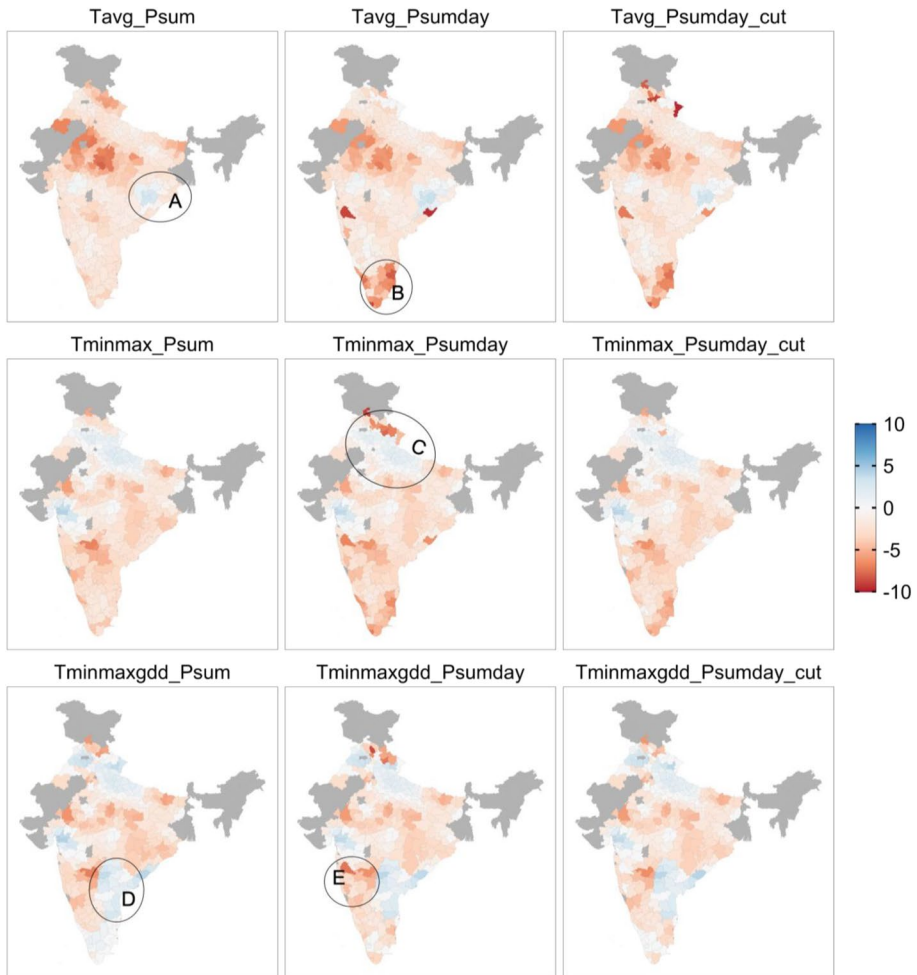


Fig. 7 Simulated impact of long-term climate change (since 1966) on rice yield, in terms of percent change compared to no climate change hypothetical scenario, in the last decade (2002–2011) of the study time period. The climate data was linearly detrended to remove time trend at district-scale. District-level estimates of median value and 95% confidence intervals of climate change impact on yield were obtained through residual bootstrapping ($n=500$). The average district-level yield loss during the last decade in the dataset (2002–2011) is presented here as the expected impact of climate change that has occurred since 1966. Only results with 95% significance of the confidence intervals are shown; insignificant results are shown in gray. The null model is not shown because it is climate invariant and predicts zero impact of climate change

are included. From top to bottom, as more temperature variables are added to the models in the middle and bottom rows, a bigger range of predicted impacts is visible: compared to the first row, there are larger regions where climate change is predicted to have benefited rice yields. These include the highly mechanized Indo-Gangetic belt comprising the states of Punjab, Haryana, and Uttar Pradesh (region C). In the third row with the most complex models, even the state of Andhra Pradesh (region D), a big rice producer, turns blue from red in the previous panels. Most parts of the country seem to show more drastic impacts

of climate change with the simpler models, with the exception of some districts in south-western India (region E) where the more complex models predict a more drastic impact of climate change on rice yields, compared to simpler models in the top row.

For wheat (Fig. SI.9.1), while simpler models predict a more consistent impact throughout the country, the more complex models in the middle and bottom rows show more variation; there are regions where climate change has positively impacted wheat yields, and these include major wheat producing states of Punjab, Haryana, Uttar Pradesh (region A). However, there are also districts in eastern and southern India where climate change seems to have had a more detrimental impact than that predicted by the simpler models in the top row. In fact, certain districts in southern India show a reduction of up to 14% in wheat yields because of climate change. The patterns observed in pearl millet panels (Fig. SI.9.2) are similar to wheat. The simpler models in the top row predict a more consistent negative impact throughout the country with some blue patches in eastern parts of India (region A); the middle and bottom rows depict a higher contrast in the expected impacts of climate change on pearl millet yields. Huge parts in north and central India that were depicted in red in the top row now seem to show a net positive impact of climate change, while southern India has turned an even darker shade of red, denoting a more serious negative impact of climate change than one would observe if the analysis is limited to simpler models with only mean seasonal temperature.

4 Discussion

4.1 Role of a priori climate-crop relationship knowledge in building statistical models

Using statistical metrics of adjusted R^2 and RMSE, we observed that the model performance does not vary noticeably between models with various levels of climate variables. This is consistent with results from the USA where Schlenker and Roberts (2009) reported minimal increase in R^2 when the growing season was divided into sub-periods for analysis, and Ortiz-Bobea et al. (2019) achieved a maximum reduction of around 6% in RMSE from amongst six different winter wheat yield models. This marginal role of climate in improving model performance has also been reported by previous studies conducted on Indian agriculture. Fishman (2016) analyzed rice yields and reported an increase in adjusted R^2 from 0.735 for a null model (with no climate data) to 0.758–0.772 with different combinations of climate variables including precipitation, degree days, and rainy days as measures of climate variability. Davis et al. (2019) similarly reported a decline in their crop model explanatory power with the addition of potential climate variables (number of monsoon dry days, ratio of precipitation to number of monsoon rain days, and squared terms of temperature and precipitation), and did not include those variables in their final model.

A parsimonious⁴ model selection process based purely on examining R^2 , RMSE, or related measures as above can advocate for the simplest models as being most appropriate, as the difference in model performance can be small. However, relative importance analysis showed that as more climate variables were added, climate occupied an increasingly

⁴ Principle of statistical parsimony for model selection states that the model with fewest variables generally has the highest predictive ability for new data (Seasholtz and Kowalski 1993).

important role in predicting crop yields, even if that was not reflected fully in the increase in total variance explained by a model. Crop yield signal that would otherwise be explained by subseasonal climate is subsumed by geography and time in the absence of those climate variables, a trend that gets amplified in periods of anomalous weather. This is a classic case of confounding which is a common problem in observational studies (Bakker et al. 2005; Ogundari and Onyeaghala 2021).

This result has important implications for the study of climate-crop relationships using statistical models. There are usually multiple climate variables that can be included in a model, and choosing the best model based on generic model performance metrics like R^2 or RMSE may lead to selection of models which downplay the role of climate.⁵ This is especially true if model selection were to happen without adequate domain knowledge about important variables that need to be included in a model irrespective of their role in increasing overall model performance. While our analysis was limited to OLS linear regression, this error of omission could easily occur in advanced machine learning based methods as well, where the variables are automatically selected by the algorithm.

Adequate importance needs to be given to fundamental plant physiological understanding of how weather and climate affect crop yields while building crop models even if those climate variables may initially seem insignificant during model selection. For example, in our study, if models are selected based on statistical metrics like R^2 , adjusted R^2 , or RMSE, Occam's razor or the principle of parsimonious model selection might dictate that we choose the simplest model with only seasonal average temperature and total precipitation. However, this would ignore the increasingly significant role climate plays in the more complex models. An argument can thus be made for including key a priori variables which are theoretically expected to impact crop yields. For example, field experiments have shown that rice yield can decline by up to 10% for every degree Celsius rise in night temperature, but no significant impacts were observed for rising day temperatures (Peng et al. 2004). This is backed up physiologically by evidence of high night temperature adversely impacting movement of carbohydrates and nitrogen within the rice plant (Singh et al. 2020). In this case, separately including mean daily minimum temperature and mean daily maximum temperature in statistical crop models makes more sense physiologically, than including just mean daily average temperature. In our study too, while the models accounting for both mean daily minimum and mean daily maximum did not have drastically different adjusted R^2 or RMSE compared to the model containing only mean daily average temperature, climate change simulations showed opposite results for some regions as discussed earlier.

4.2 Model performance for extreme weather events and long-term climate change

Even when extra climate variables may not noticeably improve the model performance measured in terms of adjusted R^2 or RMSE, we showed that in more complex models, climate plays an increasingly crucial role in explaining crop yield variance. Hence, while model performance averaged over time was not significantly impacted by the levels of climate variables included, more complex models were better able to account for anomalous

⁵ For example, in pearl millet models with and without degree day bins, while adjusted R^2 improves marginally, RMSE shows a reduction in model accuracy (Fig. 1). However, Fig. 2 depicts a sharp increase in climate's relative importance with addition of degree day bins. Model selection based on adjusted R^2 or RMSE would ignore degree day bins, while relative importance shows the importance of including them in the models.

weather patterns. Complex models performed particularly well (in terms of RMSE reduction) for some important crop-state combinations like rice in Punjab, wheat in Haryana and Punjab, and pearl millet in Rajasthan. The importance of this improved performance is underscored by the fact that Punjab and Haryana are among the biggest producers of wheat and rice in India.⁶ Meanwhile, Rajasthan is the largest pearl millet producer in India.

The accuracy of model predictions is especially critical when inaccurate predictions are biased upwards (predictions are higher than observed values), something RMSE does not factor in since it is insensitive to the direction of change. Models prone to over-predicting crop yields may provide a false sense of security to policymakers when they use these models to predict season-end yields and formulate food policies during extreme weather events. For example, in 2002 in the state of Madhya Pradesh, the simplest model (levels 1 of both temperature and precipitation) over-predicted wheat yields by 15.5%, while the model with levels 3 and 2 of temperature and precipitation variables over-predicted by 11.9%. Similarly, for national pearl millet production, the difference between predictions from the simplest model and a 3/2 level model was 769 thousand tonnes, or 15.5% of the national production in 2002. This susceptibility to over-predict production during anomalous weather events strengthens the case for examining model performance more closely under different conditions instead of making a selection based on standard statistical criteria.

A theoretically grounded (rather than statistically selected) model can also allow researchers to detect patterns of long-term climate change impacts on crop yields that may otherwise not be visible in simpler models. In our study, for all three crops, the simulated impact of climate change on crop yields in the last decade of our dataset's timeline, from 2002 to 2011, depicts stark differences in the predictions of various models. Simpler models predict more uniform yield losses across the country from climate change, whereas the complex models predict more variegated patterns of both losses and gains depending on the geographic region being analyzed. In addition to these distinct patterns, the more complex wheat models predict yield losses of up to 14% in some parts of the country, as opposed to the simpler models in the top row of Fig. SI.9.1 where the predicted losses peak at 5%. This observation further underscores the importance of considering an ensemble of models for making future yield predictions instead of selecting one solely on the basis of statistical parameters. For example, when a certain constituent of a group of models predicts negative impacts of climate change but is not the most accurate based on standard statistical metrics, the predictions from that model should not be dismissed without adequate examination.

4.3 Implications

By breaking up yield variance explained by crop models into climatic and non-climatic components, our study shows the potential pitfalls of building and selecting crop models based only on generic statistical tests without paying adequate attention to physiological processes that may mandate the inclusion of specific climate variables. An analyst may ignore marginal improvements in adjusted R^2 and RMSE and choose a simpler model to save data acquisition and computational costs. But as we show in this study, such a model selection at this stage would ignore the significantly different results obtained from complex models with respect to impacts of climate change and extreme weather events. This is especially significant given

⁶ In 2019–2020, they together accounted for 50% wheat and 30% rice to the central food reserves, a crucial source of subsidized food for the economically weaker sections of society.

the sufficient evidence of anthropogenic climate change making weather more unpredictable and increasing the frequency and intensity of extreme events in this region (Murari et al. 2015; Rohini et al. 2019; Das and Umamahesh 2021). Simultaneously, there has already been a significant increase in frequency of dry spells and intensity of wet spells during the monsoon season (Singh et al. 2014), and future predictions estimate a further increase in the frequency and magnitude of hot and dry extreme events (Mishra et al. 2020).

This study is also important because geography and time are not the only variables that can subsume climate signal; it is possible that some other non-climatic variables which vary across time or space (for example chemical inputs, mechanization, development of roads, or atmospheric carbon dioxide concentration) may have correlation with climate leading to subsequent conflation of the non-climatic and climatic signals. This nuance needs to be paid attention to when including such variables in crop models. Our study presents an analytical framework that can be used in such scenarios.

4.4 Limitations and future work

There are some caveats and limitations in our study that warrant discussion. One, we only report results for three crops. We did this to focus the discussion on the mechanics of statistical models with three representative crops from the two major growing seasons. With their contrasting results, these three crops serve as examples of different crop-dependent outcomes of our analysis. Nonetheless, the analysis can be easily extended to other crops. Two, as discussed and rationalized earlier, we excluded irrigation in our analysis, even though it is a big determinant of crop yields. Three, India is a large country, and national level studies like ours may ignore important trends and patterns that have been reported in more granular studies (Zachariah et al. 2020). This limitation applies to all studies conducted over a large but heterogeneous nation state. A case can therefore be made for building more local models, and assessing variable relative importances in those models.

Some salient questions arose from our study that warrant further research. We observed that in anomalous weather years, our complex models (with most climate variables) had significantly lower RMSE than the simpler models (Fig. 3). Simultaneously, the overall RMSE analysis shows little difference in model performance over the full time period (Fig. 1). It is worth investigating if the improvement in model performance is minimal (or zero) in normal years and just amplified during periods of anomalous weather, or if the simple model performs better than the more complex models in normal years and this trend flips in anomalous years. We saw evidence supporting both these possibilities: in 1996, simple rice models outperformed complex ones but all wheat models exhibited similar performance (Fig. 3), while 2002 saw the complex models perform better than simple ones for both these crops. It may be worthwhile to look into hybrid models that are trained on two distinct datasets: the normal weather years, and anomalous weather years. The predictions from both these models may then be combined with pre-determined probabilities to arrive at more accurate predictions.

5 Conclusion

Researchers using crop yield models have a vast array of climate variables to choose from for inclusion in their models. Without adequate domain knowledge about plant physiology or critical climate factors, climate variables are sometimes chosen based solely on overall model performance using common statistical techniques like R^2 , adjusted R^2 , or RMSE.

However, this study demonstrates that obfuscation of the signal between non-climatic and climatic variables may cause the performance thus measured to improve only marginally with the inclusion of new climate variables, even though those omitted climate variables may be explaining important climate-yield relationships. This was seen for the state of Rajasthan in our study, where the seasonal model failed to capture the impact of exceptionally dry or hot weather on pearl millet yield. In contrast the subseasonal model, even though its overall accuracy was similar to the subseasonal model, performed significantly better at capturing yield losses in those anomalous years.

Model selection based on parsimony criteria can seriously fail to parameterize important climate effects and lead to poor predictions of the impact of extreme weather events and long-term climate change. For example, our results showed that the assessment of historical impact of climate change, as measured by the model containing only seasonal variables, may not capture the more drastic impacts predicted at a subnational level by the more complex subseasonal models, as was seen in the case of wheat or pearl millet. Researchers are advised to use statistical metrics in combination with theoretical or process-based knowledge for choosing variables to include in their crop models.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10584-022-03404-0>.

Author contribution All authors conceived and developed the project idea. BSS and NR worked on data acquisition. BSS led data analysis, result interpretation, and writing. All authors contributed equally towards editing the manuscript.

Data availability All data used in this study is available open-source.

Code availability Available on request.

Declarations

Competing interests The authors declare no competing interests.

References


- Aadhar S, Mishra V (2021) On the occurrence of the worst drought in South Asia in the observed and future climate. *Environ Res Lett* 16. <https://doi.org/10.1088/1748-9326/abd6a6>
- Albers H, Gornott C, Hüttel S (2017) How do inputs and weather drive wheat yield volatility? The example of Germany. *Food Policy* 70:50–61. <https://doi.org/10.1016/j.foodpol.2017.05.001>
- Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration - guidelines for computing crop water requirements. FAO Irrigation and drainage paper 56. FAO, Rome 300(9):D05109
- Bakker MM, Govers G, Ewert F et al (2005) Variability in regional wheat yields as a function of climate, soil and economic variables: assessing the risk of confounding. *Agric Ecosyst Environ* 110:195–209. <https://doi.org/10.1016/j.agee.2005.04.016>
- Beillouin D, Schauburger B, Bastos A et al (2020) Impact of extreme weather conditions on European crop production in 2018. *Philos Trans R Soc B Biol Sci* 375. <https://doi.org/10.1098/rstb.2019.0510>
- Birthal PS, Khan T, Negi DS, Agarwal S (2014) Impact of climate change on yields of major food crops in India: implications for food security. *Agric Econ Res Rev* 27:145. <https://doi.org/10.5958/0974-0279.2014.00019.6>
- Butler EE, Huybers P (2013) Adaptation of US maize to temperature variations. *Nat Clim Chang* 3:68–72. <https://doi.org/10.1038/nclimate1585>
- Das J, Umamahesh NV (2021) Heat wave magnitude over India under changing climate: projections from CMIP5 and CMIP6 experiments. *Int J Climatol* 1–21. <https://doi.org/10.1002/joc.7246>

- Davis KF, Chhatre A, Rao ND et al (2019) Sensitivity of grain yields to historical climate variability in India. *Environ Res Lett* 14. <https://doi.org/10.1088/1748-9326/ab22db>
- Feng P, Wang B, Liu DL et al (2020) Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique. *Agric for Meteorol* 285–286:107922. <https://doi.org/10.1016/j.agrformet.2020.107922>
- Fishman R (2016) More uneven distributions overturn benefits of higher precipitation for crop yields. *Environ Res Lett* 11:24004. <https://doi.org/10.1088/1748-9326/11/2/024004>
- Grömping U (2006) Relative importance for linear regression in R: the package relaimpo. *J Stat Softw* 17:1–27. <https://doi.org/10.18637/jss.v017.i01>
- ICRISAT (2015) Village dynamics in South Asia: meso level data for India: 1966–2011. Hyderabad
- James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning. Springer
- Jiang H, Hu H, Li B et al (2021) Understanding the non-stationary relationships between corn yields and meteorology via a spatiotemporally varying coefficient model. *Agric For Meteorol* 301–302. <https://doi.org/10.1016/j.agrformet.2021.108340>
- Kern A, Barcza Z, Marjanović H et al (2018) Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices. *Agric for Meteorol* 260–261:300–320. <https://doi.org/10.1016/j.agrformet.2018.06.009>
- Li H, Maddala GS (1996) Bootstrapping time series models. *Econom Rev* 15:115–158
- Lobell D, Burke M (2009) Climate change and food security: adopting agriculture to a warmer world. Springer Science & Business Media
- Lobell D, Field C (2007) Global scale climate-crop yield relationships and the impacts of recent warming. *Environ Res Lett* 2. <https://doi.org/10.1088/1748-9326/2/1/014002>
- Lobell D, Schlenker W, Costa-Roberts J (2011) Climate trends and global crop production since 1980. *Science* 333(6042):616–620
- Ministry of Agriculture and Farmers Welfare (2016) Agricultural statistics at a glance. Ministry of Agriculture and Farmers Welfare, New Delhi
- Mishra V, Thirumalai K, Singh D, Aadhar S (2020) Future exacerbation of hot and dry summer monsoon extremes in India. *npj Clim Atmos Sci* 3. <https://doi.org/10.1038/s41612-020-0113-5>
- Murari KK, Ghosh S, Patwardhan A et al (2015) Intensification of future severe heat waves in India and their effect on heat stress and mortality. *Reg Environ Chang* 15:569–579. <https://doi.org/10.1007/s10113-014-0660-6>
- Ogundari K, Onyeaghala R (2021) The effects of climate change on African agricultural productivity growth revisited. *Environ Sci Pollut Res* 28:30035–30045. <https://doi.org/10.1007/s11356-021-12684-5>
- Ortiz-Bobea A, Wang H, Carrillo CM, Ault TR (2019) Unpacking the climatic drivers of US agricultural yields. *Environ Res Lett* 14:64003. <https://doi.org/10.1088/1748-9326/ab1e75>
- Peng S, Huang J, Sheehy JE et al (2004) Rice yields decline with higher night temperature from global warming. *Proc Natl Acad Sci U S A* 101:9971–9975. <https://doi.org/10.1073/pnas.0403720101>
- R core team (2020) R: A language and environment for statistical computing
- Rajeevan M, Bhatte J, Kale JD, Lal B (2006) High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells. *Curr Sci* 91:296–306
- Ray DK, Gerber JS, Macdonald GK, West PC (2015) Climate variation explains a third of global crop yield variability. *Nat Commun* 6:1–9. <https://doi.org/10.1038/ncomms6989>
- Rising J, Devineni N (2020) Crop switching reduces agricultural losses from climate change in the United States by half under RCP 8.5. *Nat Commun* 11:1–7. <https://doi.org/10.1038/s41467-020-18725-w>
- Roberts MJ, Braun NO, Sinclair TR et al (2017) Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environ Res Lett* 12. <https://doi.org/10.1088/1748-9326/aa7f33>
- Rohini P, Rajeevan M, Mukhopadhyay P (2019) Future projections of heat waves over India from CMIP5 models. *Clim Dyn* 53:975–988. <https://doi.org/10.1007/s00382-019-04700-9>
- Schlenker W, Roberts MJ (2009) Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proc Natl Acad Sci* 106:15594–15598. <https://doi.org/10.1073/pnas.0906865106>
- Seasholtz MB, Kowalski B (1993) The parsimony principle applied to multivariate calibration. *Anal Chim Acta* 277:165–177. [https://doi.org/10.1016/0003-2670\(93\)80430-S](https://doi.org/10.1016/0003-2670(93)80430-S)
- Singh D, Tsiang M, Rajaratnam B, Diffenbaugh NS (2014) Observed changes in extreme wet and dry spells during the south Asian summer monsoon season. *Nat Clim Chang* 4:456–461. <https://doi.org/10.1038/nclimate2208>

- Singh A, Chaudhuri B, Roychoudhury A (2020) Influence of night temperature on rice yield and quality. In: Roychoudhury A (ed) Rice research for quality improvement: genomics and genetic engineering, vol 1. breeding techniques and abiotic stress tolerance. Springer Singapore, Singapore, pp 579–590
- Tebaldi C, Hayhoe K, Arblaster JM, Meehl GA (2006) Going to the extremes: an intercomparison of model-simulated historical and future changes in extreme events. *Clim Change* 79:185–211. <https://doi.org/10.1007/s10584-006-9051-4>
- Zachariah M, Mondal A, Das M et al (2020) On the role of rainfall deficits and cropping choices in loss of agricultural yield in Marathwada, India. *Environ Res Lett* 15. <https://doi.org/10.1088/1748-9326/ab93fc>
- Zaveri E, Lobell D (2019) The role of irrigation in changing wheat yields and heat sensitivity in India. *Nat Commun* 10. <https://doi.org/10.1038/s41467-019-12183-9>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Balsher Singh Sidhu¹  · Zia Mehrabi^{2,3}  · Milind Kandlikar^{1,4}  · Navin Ramankutty^{1,4} 

Zia Mehrabi
zia.mehrabi@colorado.edu

Milind Kandlikar
mkandlikar@ires.ubc.ca

Navin Ramankutty
navin.ramankutty@ubc.ca

- ¹ Institute for Resources, Environment and Sustainability, The University of British Columbia, Vancouver, BC, Canada
- ² Department of Environmental Studies, The University of Colorado Boulder, Boulder, CO, USA
- ³ Mortenson Center in Global Engineering, The University of Colorado Boulder, Boulder, CO, USA
- ⁴ School of Public Policy and Global Affairs, The University of British Columbia, Vancouver, BC, Canada