scientific reports

OPEN

Identifying links between monsoon variability and rice production in India through machine learning

Christopher Bowden✉, Timothy Foster & Ben Parkes

Climate change poses a major threat to global food security. Agricultural systems that rely on monsoon rainfall are especially vulnerable to changes in climate variability. This paper uses machine learning to deepen understanding of how monsoon variability impacts agricultural productivity. We demonstrate that random forest modelling is effective in representing rice production variability in response to monsoon weather variability. Our random forest modelling found monsoon weather predictors explain similar levels of detrended anomaly variation in both rice yield (33%) and area harvested (35%). The role of weather in explaining harvested rice area highlights that production area changes are an important pathway through which weather extremes impact agricultural productivity, which may exacerbate losses that occur through changes in per-area yields. We find that downwelling shortwave radiation flux is the most important weather variable in explaining variation in yield anomalies, with proportion of area under irrigation being the most important predictor overall. Machine learning modelling is capable of representing crop-climate variability in monsoonal agriculture and reveals additional information compared to traditional parametric models. For example, non-linear yield and area responses of irrigation, monsoon onset and season length all match biophysical expectations. Overall, we find that random forest modelling can reveal complex non-linearities and interactions between climate and rice production variability.

Agriculture is critically important to food security in South Asia, a region that is home to nearly half the malnourished population of the world[1]. Overall, 56% of land area in South Asia is used for agriculture, with over 40% of the population employed in the sector[2,3]. A large proportion of agricultural production in the region occurs during the southwest summer monsoon season[4], with most of the region receiving over 80% of their annual rainfall during this time[5,6]. Rice is one of the most commonly grown crops, with India accounting for over 20% of global rice production, the majority of which occurs during the monsoon season[7]. Agricultural outcomes are closely connected to the monsoon. For example, in India a 1% change in annual monsoon rainfall causes a 0.34% change in India's annual agricultural gross domestic product (GDP)[8], a significant impact in a country where total agricultural production accounts for almost 18% of GDP[7,9,10].

Rice production outcomes are known to be strongly linked to monsoon weather conditions (e.g. total rainfall[11]) and their temporal distribution (e.g. number of rainy days[12]). Changes in monsoon weather dynamics and variability therefore have important implications for rice productivity in India, for which hundreds of millions depend for livelihoods and food security. Past research has shown that the Indian monsoon has shifted towards a pattern of reduced total precipitation and less frequent rainfall events over the second half of the twentieth century, resulting in higher risk of both flood and drought damage[13–15]. Historic changes in monsoon dynamics in South Asia have been attributed to several factors, including increased aerosol concentration, land use changes and urbanisation effects[16,17]. In the future, climate change is likely to further exacerbate monsoon variability and extremes[18]. For example, studies suggest that the monsoon will arrive earlier and retreat later[19], and with increasing annual variability and more frequent extreme precipitation events[20,21].

To support adaptation of rice production to current and future monsoon weather variability, it is critical to understand which aspects of monsoon climate have greatest impacts on agricultural production outcomes. Recent modelling studies have found a range of weather factors to be important in explaining yield variability for rice and other crops in monsoon-dependent agricultural systems. For example, the importance of total precipitation has been well demonstrated[11], with higher seasonal rainfall leading to higher rice yields. Similarly, a more even distribution of rainfall throughout the season has been linked with higher rice yields[12,22], with extreme weather events such as droughts and high temperatures negatively affecting rice yields[23,24]. However, important gaps remain in

Department of Mechanical, Aerospace and Civil Engineering, University of Manchester, Manchester M13 9PL, UK. ✉email: christopher.bowden@postgrad.manchester.ac.uk

our understanding about the effect of monsoon season length and onset timing on crop production[13,25]. For example, delayed monsoon onset and reduced season length have been shown theoretically[25–27] and anecdotally[28,29] to be potentially important drivers of crop yield losses, but are rarely included in statistical empirical crop-climate models in India and South Asia. Furthermore, past crop-climate studies in India and elsewhere exclusively focus on monsoon weather impacts on per-area crop yields, ignoring potential impacts on crop production areas[30]. Yet, emerging evidence suggests that changes to cropped area may also be an important outcome of weather extremes in climate vulnerable farming systems[31]. Failure to consider impacts of weather on crop harvested areas may therefore lead to an underestimation of the impacts of weather variations on agriculture and associated rural economies, with implications for the ability to anticipate future risks posed by climate change.

Relationships between weather and agricultural productivity are complex and non-linear, with production impacts dependent on the magnitude, sequencing, and timing of different weather extremes during the growing season[32–34]. In this context, a further key limitation of previous empirical modelling studies is the reliance on parametric regression approaches for linking production outcomes to observed weather variability. Parametric regression models assume pre-specified relationships between predictors and response variables, often specified as simple linear or quadratic relationships. This limits the ability to detect more complex or non-linear responses of crop yields to different aspects of monsoon weather, which have shown to be important for correctly attributing drivers of yield variability[35–37]. Machine learning (ML) methods, such as Random Forest (RF) models, offer a potential solution to this challenge through their ability to explore relationships between weather predictors and agricultural production outcomes without the need to pre-define the functional form of these relationships. Research has shown that RF models can significantly outperform parametric regression models in terms of ability to explain weather-yield relationships[35,38]. Moreover, RF models are capable of capturing key drivers and disentangling complex non-linear interaction effects[23,39], and thus have significant potential to enhance understanding of weather-related production variability in monsoon-dependent agricultural systems in South Asia and elsewhere.

In this paper, we use RF modelling techniques in conjunction with an extensive long-term dataset of production and weather data from across four states in India (Punjab, Haryana, Uttar Pradesh and Bihar—collectively comprising the Indo-Gangetic Plains or IGP) to assess the dominant monsoon weather-related drivers of variability in rice yields and production area. Our findings provide comprehensive insights about the critical monsoon climate signals, and their non-linear interactions, that explain historic variation in rice yield and production area across the IGP, the latter of which has rarely been considered in past crop-climate impact studies. In doing so, our findings provide key evidence to strengthen understanding of how future changes in monsoon dynamics and variability may impact rice production in India, and help to guide efforts to enhance climate resilience of agriculture and rural livelihoods across South Asia.

Data

Study region. Our study focuses on the IGP, which for the purposes of this study is defined as including 71 total districts within 4 states (according to 1966 boundaries from DLD): Punjab, Haryana, Uttar Pradesh and Bihar (Fig. 1). Agricultural production in the IGP is characterised primarily by a rice-wheat crop rotation, with rice cultivated during the Kharif (monsoon, May to September) season and wheat during the Rabi (dry, November to April) season[40]. The IGP is commonly referred to as the 'breadbasket of India' because it produces 40% and 70% of India's total rice and wheat output respectively[41], and is therefore vital to both national and regional food security. Temporal variability in rice productivity across the IGP is strongly linked with the variability of the Asian summer monsoon, which has large inter-annual variability[42,43]. Thus, the IGP represents an ideal study area to explore links between rice production outcomes and monsoon weather conditions.

Agricultural production data. District-level data on rice yields and harvested production areas for the 71 administrative districts in the four states of the IGP was obtained from ICRISAT's District-Level Database (DLD)[44]. The dataset provides information on yields, harvested area and production for a 52-year period (1966–2017) and has been used in a number of previous studies of yield variability in India, for rice and other crops[24,45,46]. These data have been sourced from quality-controlled state and district surveys and reports collated by the Directorate of Economics and Statistics of the Government of India and made available by ICRISAT. The DLD data was chosen because it provides a robust and trustworthy official record of observed agricultural production, with sufficient spatial resolution (district-level) and temporal extent to enable exploration of climatic drivers of rice yield and production variability. Production data for the four states of the IGP were filtered to remove districts where rice production was minimal (Fig. 1, shaded grey) so as to avoid these data having undue influence on subsequent analyses (see supplementary information).

A large proportion of historic rice yield and production variability in the IGP, and indeed India more broadly, is attributable to non-weather related temporal and spatial trends. For example, rice yields across India have increased over time due to progressive improvements in technologies, new crop varieties, and CO_2 fertilisation[47]. Persistent spatial differences also exist in rice yields and production outcomes between districts and states due to varying levels of agricultural intensification, mechanisation and land suitability. We remove such obfuscating trends in production data prior to running the model using state-specific locally weighted scatterplot smoothing (LOWESS) curves (Supplementary Fig. S2). We remove trends at the state and not district level to avoid overfitting; this is consistent with previous studies[13,25,32].

Weather data. To calculate values of each weather variable for each district and year of our agricultural production record, we use the Princeton Global Forcing (PGF) gridded weather dataset[48]. PGF was selected due to its long timespan and relatively high spatial resolution (0.25°x 0.25°), and has been used by other crop-climate
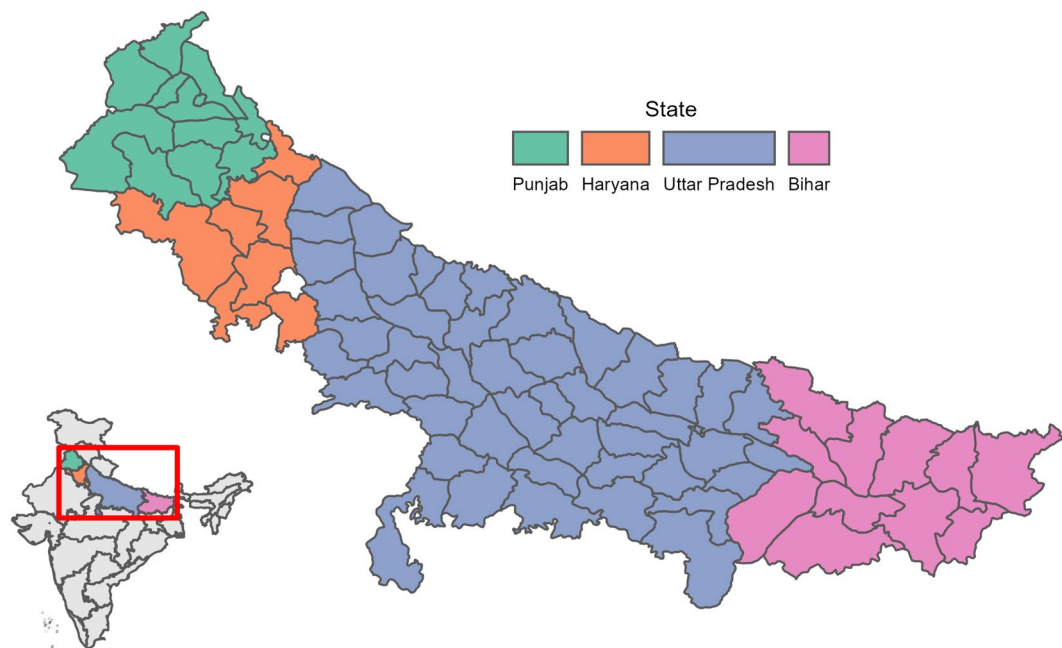
Figure 1. Colours represent the four states of the IGP in this study. Further divisions by black lines represent 1966 district boundaries. Districts in grey have been removed from analysis due to their urban or mountainous geography (see supplementary information).

modelling studies[49–51]. PGF has also been validated for use over India, specifically including the temperature[52] and precipitation[53] variables used in this study. Furthermore, rice yield sensitivity to the PGF weather variables matched that of similar weather datasets in an India-based comparison[54]. Values of each weather variable included in the modelling were aggregated spatially to district-scale using weighted means with DLD boundaries, and temporally to growing-season-scale (1st May to 30th September).

The goal of our study is to understand which aspects of monsoon weather have greatest impacts on rice production variability in the IGP, both individually and in combination. The types of variables included in ML models greatly affect the output and performance of the model. A full list detailing weather variables included in our modelling can be found in Table 1. Basic measures of temperature (aggregated growing degree days, daily minimum and maximum temperatures) and precipitation (total precipitation, proportion of rainy days) were selected due to their importance in many past weather-crop modelling studies[23,39,57,61–63]. In addition, we include several further monsoon weather variables that have largely been neglected in past studies: (1) Downwelling shortave radiation flux (DSWRF), a measure of incoming surface-level solar radiation affected by frequent cloud cover in South Asia[25,64–66]; (2) Monsoon season length, and (3) Monsoon 'onset deviation', both of which can critically affect sowing, transplanting and harvesting efficacy[45,59,67,68]. For additional details on the calculations of monsoon variables, see 'Weather data' section of the supplementary information.

Variable (units)	Description (references)
Total precipitation (mm)	Total growing season precipitation[12,55,56]
Proportion of rainy days	Proportion of growing season days where rainfall is above 2.5 mm, the Indian Meteorological Department's definition of light rain[12]
Growing degree days	A measure of heat accumulation (time temperature is above 10 °C) over the growing season[57]
Mean minimum temperature (°C)	Mean daily minimum temperature over the growing season[55,58]
Mean maximum temperature (°C)	Mean daily maximum temperature over the growing season[55,58]
Downwelling shortwave radiation (W/m²)	Sum of DSWRF over the growing season[25]
Onset deviation (days)	Deviation of monsoon onset from the all-year district mean onset[59]
Monsoon season length (days)	Number of days between monsoon onset and demise[59]
Irrigated rice area proportion	Irrigated proportion of total rice area per district[60]

Table 1. Overview of predictor variables in the random forest models. All variables have been aggregated from gridcell (0.25°x 0.25°) to district scale by weighted means. Yield data has been detrended and normalised prior to use in the model and therefore no spatial or temporal trend terms are included here. Growing season is defined as 1st May to 30th September.

Methods

We used the RF algorithm to quantify the relationships between monsoon weather variables and rice production across the IGP over the period 1966 to 2016. RF models are non-parametric statistical models that work by aggregating the outputs from an ensemble of decision trees. We use a regression RF model where the model output is the mean of each tree's out-of-bag (OOB) output[69]. This relatively simple structure makes RF modelling an ideal method because it enables evaluation of complex non-linear relationships and interactions created during model construction, whilst retaining model transparency that is sometimes lost in other ML approaches. Support vector machine (SVM) modelling was also considered for use in this study but was discarded due to the improved interpretability of relationships formed within RF models. SVM models have often performed similarly to RF models (see review[70]), but lack outputs that enable transparent interpretation of relationships formed within the model[71–73].

In particular, ML is useful for crop-climate modelling where there are many non-linearities and interactions between weather predictors and production outcomes that would be challenging to comprehensively specify *a priori* in a traditional parametric model structure. Similar studies have used RF modelling previously to detect and interpret complex non-linear relationships between predictors and response variables, including in crop-climate modelling[23,55,74].

We develop two separate RF models quantifying relationships between monsoon weather variability and detrended anomalies in margins of production: (1) rice yields (intensive), and (2) harvested rice areas (extensive). Hyperparameter tuning and model runs were performed using 'tidymodels' R package v0.1.3 and 'randomForest' R package v4.6-14 (mtry = 3, min_n = 5, ntrees = 500) in R v4.0.4[69,75,76]. We evaluate the accuracy of each RF model based on the mean OOB root mean square error (RMSE) rate. Relative importance of each predictor variable within each RF model was assessed using '%IncMSE' (aka 'permutation') as the measure of importance (see supplementary information), which was used to create variable importance plots (VIP). Because VIPs are affected by the internal splitting of training data and random selections in a RF model, we ran 100 RF models with different seeds. This was done to provide 95th percentile ranges, representing the variability of model output[77]. The same 100 RF model outputs were also used to calculate mean percentage of explained variance and RMSE for models (1) and (2).

Accumulated local effects (ALE) plots were used to visualise the influence on production outcomes of each predictor, including any non-linearities in responses. For examination of main effects, ALEs were selected over typical partial dependence (PD) plots because they more reliably account for interactions between predictors[60,78–80]. Similar to VIPs, ALE plots are affected by the split of training data. Therefore, results are presented as 95th percentile ranges from 100 RF model runs. The interpretation of second-order ALE plots is difficult and cumbersome, requiring simultaneous assessment of the main-effect ALE plots. Therefore, PD plots were used to assess second-order interactions because they include interaction effects, making for more straightforward interpretation. All pairwise interactions were plotted, but only those pairwise interactions of substantial effect are used for analysis.

Results

Model performance. Weather predictors in the RF models were able to explain similar levels of OOB anomaly variation in both rice yield (33%, RMSE = 0.215) and area harvested (35%, RMSE = 0.792). Furthermore, OOB anomalies estimated by RF models reliably captured instances of both above and below average yield (69% agreement) and area (77% agreement, Fig. 2). When temporal trends and spatial fixed effects terms that were previously removed were added back into the RF models, the amount of OOB variability explained by the models increased to 87% (yield) and 73% (area), comparable to previous fixed and random effects parametric models of crop-climate relationships in India and other smallholder farming systems[11–13,81]. These results highlight the important role that trends in agricultural outcomes over time (e.g. technology or variety improvements) and space (e.g. persistently higher yields in some states) have in explaining the overall variations in outcomes across our sample dataset, while also demonstrating the ability of RF models to explain a meaningful proportion of remaining weather-related variance in agricultural outputs.

Yield response model. The proportion of rice cropped area under irrigation within each district was the most important predictor of rice yield anomalies in our RF model (Fig. 3). This was expected due to the dependence of rice yield on water availability, and given the heterogeneous levels of access to irrigation over space and time in the IGP. The most important weather-related predictor of rice yield anomalies was downwelling shortwave radiation flux (DSWRF, Fig. 3). Permutation of the seasonal sum of DSWRF values caused a 55% increase in MSE, notably larger than for other monsoon weather characteristics. Total precipitation and proportion of rainy days were the next most important drivers of rice yield anomalies, reflecting rice's large water requirements and vulnerability to water-related extremes, especially in the absence of accessible supplemental irrigation. Remaining aspects of monsoon variability, which include a range of water and temperature related variables, have similar variable importance scores, with a marginally higher ranking for mean minimum temperature during the growing season. Permutation of any one of these variables results in a meaningful change in %MSE of the model (> 25%), highlighting the complexity and diversity of mechanisms through which rice yields are impacted by monsoon weather variability.

Our analysis identified several monsoon characteristics that were strongly associated with positive rice yield anomalies. Above a DSWRF threshold of approximately 43,500 W/m^2, the positive marginal effect of DSWRF on rice yields increased until around 45,500 W/m^2 (Fig. 4). Positive yield anomalies were also observed for higher total seasonal precipitation above a threshold of 800 mm, beyond which the positive marginal effect increases up to 1000 mm where the effect then plateaus. Season lengths between 80 and 115 days were further found to be an
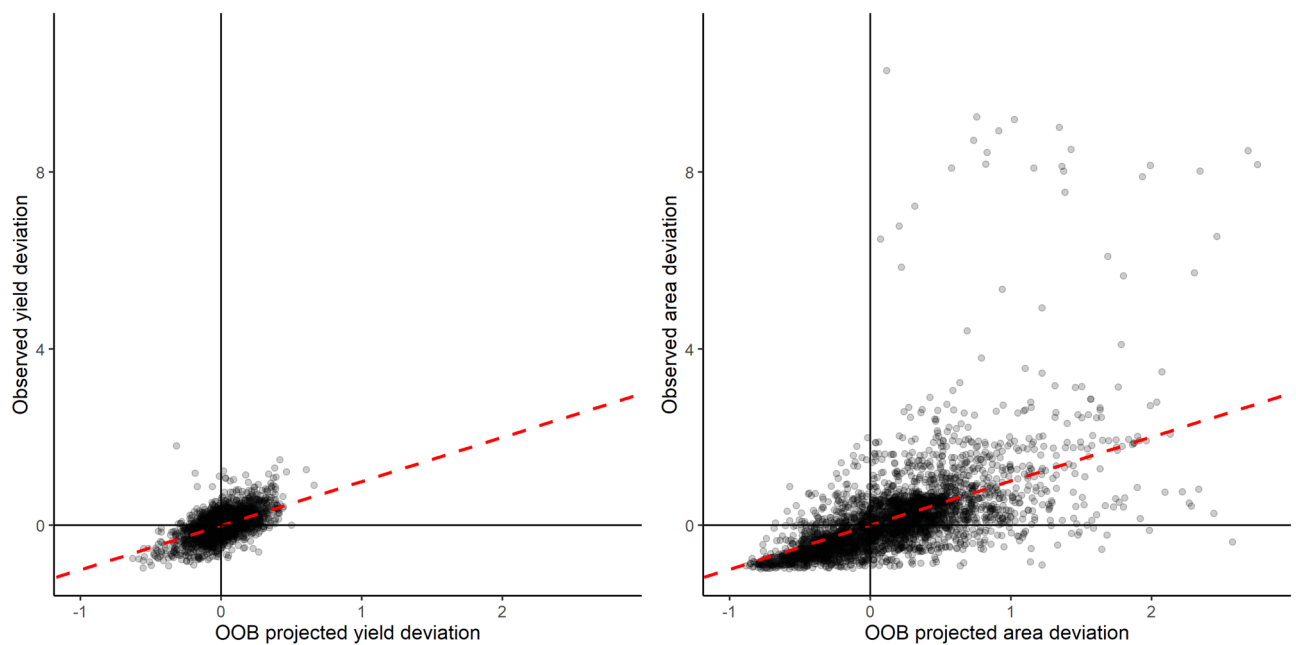
Figure 2. Random forest model projected values with observed yield and area anomalies. Red dashed lines have intercepts of 0 and slopes of 1.
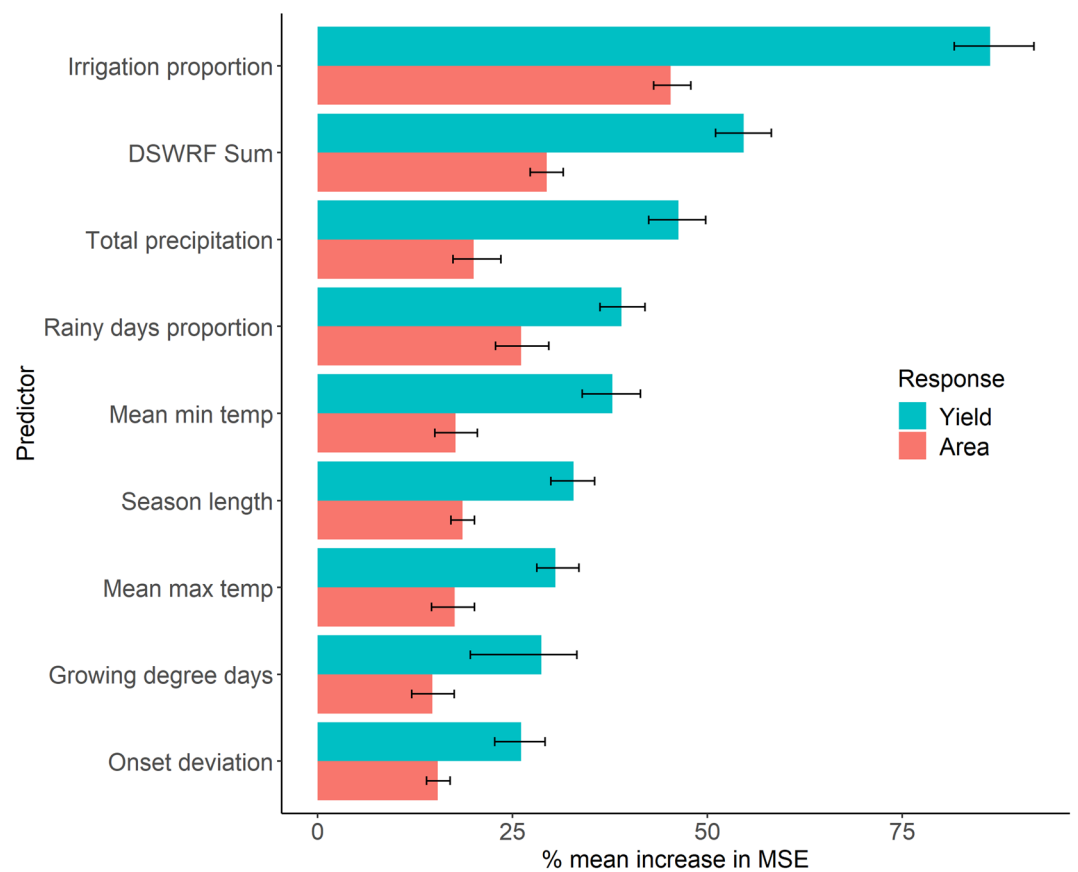

Figure 3. Variable importance plot from the random forest models calculated as percentage increase in mean-square error (MSE). Main bars represent mean value from 100 runs, inset whiskers represent 95th percentile of results.
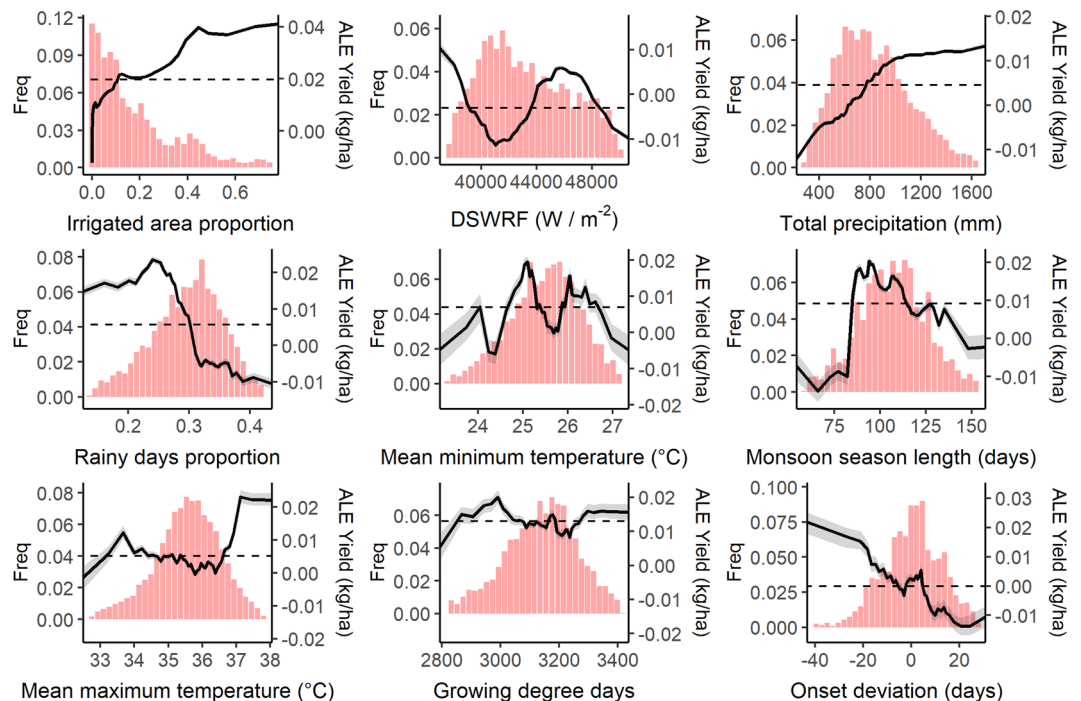
Figure 4. Accumulated Local Effects (ALE) plots of the yield anomaly response model for each predictor variable. 100 random forest models were run with different seeds, solid black lines represent mean ALE and ribbons represent 95th percentile ALE. Histograms show the distribution of the 98th percentile of predictor data in the original full dataset prior to sub-sampling for the random forest modelling. Note that y-axis ALE values are contingent on input data and are not directly comparable. Y axis scale varies between each sub-plot to highlight response curve shapes.

important predictor of positive yield anomalies. A strong interaction effect was detected between the proportion of irrigated area and total precipitation measures (Fig. 5). In particular, high precipitation is linked with positive yield anomalies even when irrigated areas are low (and vice versa), reflecting the fact that irrigation is used to supplement rather than replace rainfall for rice cultivated during the monsoon season.

At other extremes of their observed ranges (e.g. DSWRF below 43,500 W/m^2, total precipitation below 800 mm, season lengths below 80 and above 115 days), these variables were instead found to be associated with negative yield anomalies, reflecting the sensitivity of rice development to appropriate radiation and water-related conditions. Several additional monsoon characteristics were found to be important predictors of negative yield anomalies despite having lower overall variable importance in our RF models. For example, rice yield anomaly
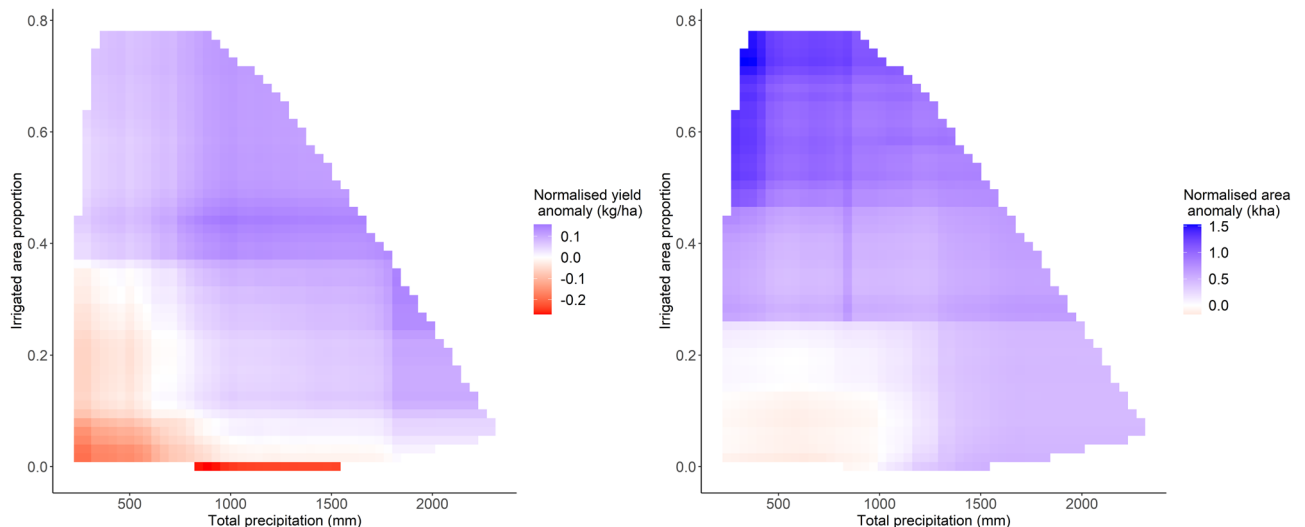

Figure 5. Partial dependence two-way interaction plots for the yield and harvested area response models.

ALE values become negative as monsoon onset is delayed by five days or more relative to the average expected onset date for a given district. Additionally, yield anomalies switched sharply from positive to negative when the proportion of rainy days exceeded 30% (Fig. 4). This unexpected negative effect of well-distributed seasonal precipitation[12] may be due to the correlation between rainy days proportion and total precipitation (Supplementary Fig. S4).

Area response model. The response of harvested area anomalies (Fig. 6) closely matched that of yield anomalies for irrigated area, total precipitation, maximum temperature and monsoon onset deviation predictors. For example, the ALE effect of onset deviation on area harvested switched from positive to negative for onset delayed longer than three days, close to the five-day delay threshold of the yield response model. Similarly, greater total precipitation leads to higher positive marginal effect on area anomaly, but only above a threshold of 1000 mm compared to the 800 mm threshold for yield anomalies. The same interaction effect pattern detected between total precipitation and irrigated area in the yield model was also detected in the area anomaly model (Fig. 5).

Area anomaly responses differed from those found previously for yield anomalies for two of our predictor variables, the proportion of rainy days and monsoon season length. The marginal effect of increasing the proportion of rainy days becomes increasingly positive above a threshold of 30%. Greater monsoon season lengths generally had a positive impact on rice harvested areas, with exception of a small negative effect for season lengths between approximately 75 and 110 days. In addition, the response of area anomalies to minimum temperature exhibited a more distinct negative linear trend in contrast to the mixed signal in the yield anomaly response.

Discussion

Our study leverages data-driven machine learning approaches to deepen understanding of the relationship between rice production outcomes and climate variability in monsoon-dependent agricultural systems in South Asia. Our results show that RF machine learning models are capable of identifying and attributing historical relationships between weather and rice production variability in India, consistent with studies for other crops and in different regions[35,82]. Our RF model simulations capture the complex non-linear responses of rice production to interacting characteristics of the South Asian monsoon, providing novel insights about non-linear crop-climate responses and variable interactions that would be challenging to identify in traditional parametric regression analyses[57,83,84]. Critically, our findings demonstrate the important role that adjustments to harvested area play in determining the overall impact of monsoon weather shocks on rice production outputs. This is an important and novel result because harvested area responses have largely been neglected in previous crop-climate
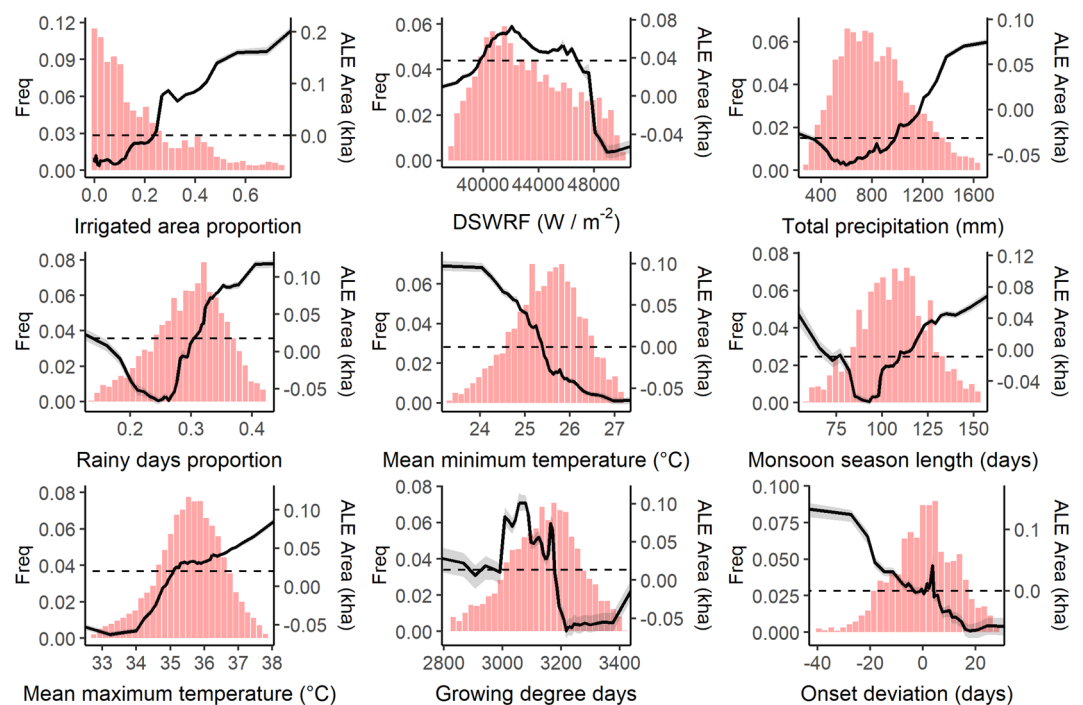

Figure 6. Accumulated Local Effects (ALE) plots of the harvested area anomaly response model for each predictor variable. 100 random forest models were run with different seeds, solid black lines represent mean ALE and ribbons represent 95th percentile ALE. Histograms show the distribution of the 98th percentile of predictor data in the original full dataset prior to sub-sampling for the random forest modelling. Note that y-axis ALE values are contingent on input data and are not directly comparable. Y axis scale varies between each sub-plot to highlight response curve shapes.

studies in the region—both using statistical and process-based models—that focus solely on impacts of weather variability on per-area yields[85–87].

In line with previous studies, our findings support the conclusion that total seasonal precipitation and irrigated area explain a significant proportion of weather-related rice yield variability in India and South Asia more widely[11,13,88,89]. The high relative importance of irrigation reflects the heterogeneity in access to and use of irrigation in space and time across the IGP[90]. Lower levels of irrigation in some district-years limit farmers' abilities to buffer rice against water and heat-related weather extremes (Fig. 5). Beyond a seasonal precipitation threshold of around 1,000 mm, the positive marginal effect of increased precipitation is reduced significantly. This type of precipitation insensitivity threshold has been found previously in other regions, for example for maize and sorghum in sub-Saharan Africa and North America[39,55] as well as for rice in China[91]. The importance of solar radiation as a driver of yield outcomes in this study was also consistent with past studies in other regions[58,92], although it is notable that solar radiation has often been neglected as a driver of crop yield variability in previous studies in India and South Asia[11,93,94].

A key novel finding from our analysis is that negative yield and harvested area anomalies occur when the monsoon onset is late by more than a few days, relative to expected local onset timings. This finding is likely to reflect the negative impact of delayed transplantation on rice yields and production. Delayed transplantation occurs when the monsoon is delayed and irrigation is absent or underutilised[45], meaning that there is insufficient water to move plants from the nursery into the main fields. This delay in transplanting leads to rice plants being transplanted at an older age, a factor that is associated with reductions in productivity due to combined impacts of a larger transplantation shock and greater vulnerability to cold stress later in the season[45]. Given the increasing variation in monsoon onset, demise, and length[18,27,95,96], failure to account explicitly for these factors in crop-climate models may lead to errors in the estimation of yield extremes (including losses), in particular when models are used to project impacts of increasingly variable future climate scenarios. Through exploration of interaction effects, we show that negative effects from a delayed onset can be mitigated through increasing irrigation, suggesting importance of expanding irrigation to ensure timely crop establishment and buffer against dry spells. This matches expectations based on previous studies, that yield losses from a delayed onset are often associated with intraseasonal dry spells[67,97,98].

Our findings also highlight the importance of a number of additional monsoon characteristics in explaining rice production variability, which have not been included in prior crop-climate studies for rice and other crops in South Asia. Monsoon season length, for example, is not commonly included as a predictor variable in modelling studies, with weather variables instead computed over arbitrary fixed calendar windows[12,22,81,88]. We find that monsoon season length is an important determinant of both area and yield anomalies, and that its effects are bounded by two distinct thresholds of season length (75 and 115 days). However, the response of monsoon season length may be complicated by non-weather factors and interactions. For example, a longer growing season may be beneficial for longer duration rice varieties, but may also increase risks of the crop being impacted by erratic or extreme rainfall at the time of harvest.

We note that the negative yield effect of a greater proportion of rainy days found in this study is contrary to expectations. Generally, a greater proportion of rainy days would be expected to reduce the risk of crop yield losses due to dry spells and associated plant water stress[12]. At the same time, the consistently positive relationship between total precipitation and rice yield and area anomalies found in this study implies that even at high levels of precipitation, yield and harvested area are not negatively affected, as might be expected from flooding or lodging damage[99]. However, these two unexpected relationships (total precipitation and rainy days with yield anomalies) may be explained by a mixed precipitation signal within the model. The proportion of rainy days data is positively correlated with total precipitation data (0.57, Supplementary Fig. S4). High values of rainy days may therefore represent an excess rainfall signal, where submergence and physical flood damage reduce yield, in particular if occurring during critical crop development periods.

While offering a powerful tool for crop-climate analysis, it is important to acknowledge that the non-linear relationships formed within RF models may also pose challenges for interpretation and identification as demonstrated in parts of our results. For both temperature- and precipitation-derived predictors, the weather signal may be split unevenly and unpredictably between multiple variables. For example, the irrigated area, total precipitation, rainy days proportion, onset deviation and season length variables were all selected to capture different aspects of the precipitation signal. The many complex interactions between these variables within the model may obfuscate the true precipitation signal for each individual predictor. This may also explain the unexpected impacts of rainy days found in our analysis, for which a positive effect on yields was expected[11,12]. Similarly, using temporal aggregations of weather variables over the monsoon season also limits the precision of response curves that can be derived using RF modelling. Future work should seek to explore the benefits of considering the intraseasonal distribution and sequencing of weather variables in order to improve the ability of RF models to disentangle monsoon crop-climate relationships and to support in-season forecasting of crop production.

District scale data used in our study reveals broad patterns in the response of rice production outcomes to monsoon weather predictors. However, our conclusions are limited by relatively coarse spatial resolution of the rice production data and accompanying weather data used as inputs to our RF models. Aggregation of data to the district scale is likely to obscure important local heterogeneity in agricultural productivity, potentially resulting in an underestimation of the true impacts of spatially heterogeneous weather-related extremes on crop yields and harvested areas[100]. For example, our RF model reveals a negative impact of delayed monsoon onset on rice yield, but this effect may be higher if it were possible to account for sub-district level heterogeneity in rainfall and associated rice yield outcomes.

Conclusions

Our study contributes to the growing body of literature that seeks to identify and assess impacts of climate variability and change on agricultural production outcomes in smallholder farming systems[101–103]. Our findings demonstrate the capability of RF machine learning models to capture and explain variability in rice production outcomes in response to monsoon climate signals. Critically, we demonstrate the additional impacts of monsoon weather variability on crop production area, highlighting the importance of considering changes in both per-area crop yields and harvested areas when assessing impacts of weather on agricultural production. Our modelling identifies complex non-linear yield and harvested area response signals from weather variables which can be used to better guide future efforts to support adaptation of rice production to extreme weather and climate change. In particular, we demonstrate the importance of considering additional characteristics of monsoon variability beyond standard measures of precipitation and temperature, including variations in the onset and duration of the monsoon which are expected to be impacted significantly by climate change in the region[21]. More broadly, our findings reaffirm the importance of irrigation access as a critical buffer against monsoon variability, both historically and under future climate change. In this context, there are urgent needs to intensify and expand access to water in areas of the IGP with currently low levels of irrigation use (e.g. Bihar), while also simultaneously implementing measures to improve long-term sustainability of water management in other parts of the region (e.g. NW India) to mitigate future negative impacts on agricultural productivity and climate risk[104].

Data availability

Crop production data supporting this publication are freely available from ICRISAT's district-level database, http://data.icrisat.org/dld/src/crops.html[44]. Raw weather data is available from the Terrestrial Hydrology research group at Princeton University, http://hydrology.princeton.edu/data.pgf.php[48]. Sample data and code are available on fig share (https://figshare.com/projects/Identifying_links_between_monsoon_variability_and_rice_production_in_India_through_machine_learning/151362) with further information available upon reasonable request from the authors.

References

1. Pandey, V. L., Mahendra Dev, S. & Jayachandran, U. Impact of agricultural interventions on the nutritional status in South Asia: A review. *Food Policy* **62**, 28–40. https://doi.org/10.1016/j.foodpol.2016.05.002 (2016).
2. World Bank. in *World Development Indicators 2017* (Tech. Rep., World Bank, Washington, DC, 2017). https://doi.org/10.1596/26447.
3. World Bank. Agricultural land (% of land area)—South Asia | Data. https://data.worldbank.org/indicator/AG.LND.AGRI.ZS?locations=8S (2018).
4. Amrith, S. S. Risk and the South Asian monsoon. *Clim. Chang.* **151**, 17–28. https://doi.org/10.1007/s10584-016-1629-x (2018).
5. Bookhagen, B. & Burbank, D. W. Toward a complete Himalayan hydrological budget: Spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge. *J. Geophys. Res. Earth Surface* **115**, F03019. https://doi.org/10.1029/2009JF001426 (2010).
6. Malik, N., Bookhagen, B. & Mucha, P. J. Spatiotemporal patterns and trends of Indian monsoonal rainfall extremes. *Geophys. Res. Lett.* **43**, 1710–1717. https://doi.org/10.1002/2016GL067841 (2016).
7. FAOSTAT. FAOSTAT. http://www.fao.org/faostat/en/data/QCL/metadata (2018).
8. Gulati, A., Saini, S. & Jain, S. Monsoon 2013: Estimating the Impact on Agriculture. Tech. Rep. id:5608, eSocialSciences (2013).
9. Jat, M. L. *et al.* Chapter Three–climate change and agriculture: Adaptation strategies and mitigation opportunities for food security in South Asia and Latin America. *Adv. Agron.* **137**, 127–235. https://doi.org/10.1016/bs.agron.2015.12.005 (2016).
10. World Bank. Employment in agriculture (% of total employment) (modeled ILO estimate)—India | Data. https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS?locations=IN (2019).
11. Davis, K. F., Chhatre, A., Rao, N. D., Singh, D. & DeFries, R. Sensitivity of grain yields to historical climate variability in India. *Environ. Res. Lett.* **14**, 064013. https://doi.org/10.1088/1748-9326/ab22db (2019).
12. Fishman, R. More uneven distributions overturn benefits of higher precipitation for crop yields. *Environ. Res. Lett.* **11**, 024004. https://doi.org/10.1088/1748-9326/11/2/024004 (2016).
13. Auffhammer, M., Ramanathan, V. & Vincent, J. R. Climate change, the monsoon, and rice yield in India. *Clim. Chang.* **111**, 411–424. https://doi.org/10.1007/s10584-011-0208-4 (2012).
14. Jin, Q. & Wang, C. A revival of Indian summer monsoon rainfall since 2002. *Nat. Clim. Chang.* **7**, 587–594. https://doi.org/10.1038/nclimate3348 (2017).
15. Singh, D., Ghosh, S., Roxy, M. K. & McDermid, S. Indian summer monsoon: Extreme events, historical changes, and role of anthropogenic forcings. *WIREs Climate Change* **10**, e571. https://doi.org/10.1002/wcc.571 (2019).
16. Bollasina, M. A., Ming, Y. & Ramaswamy, V. Anthropogenic aerosols and the weakening of the South Asian summer monsoon. *Science* **334**, 502–505. https://doi.org/10.1126/science.1204994 (2011).
17. Vinnarasi, R. & Dhanya, C. T. Changing characteristics of extreme wet and dry spells of Indian monsoon rainfall. *J. Geophys. Res. Atmos.* **121**, 2146–2160. https://doi.org/10.1002/2015JD024310 (2016).
18. Kitoh, A. *et al.* Monsoons in a changing world: A regional perspective in a global context. *J. Geophys. Res. Atmos.* **118**, 3053–3065. https://doi.org/10.1002/jgrd.50258 (2013).
19. Kajikawa, Y., Yasunari, T., Yoshida, S. & Fujinami, H. Advanced Asian summer monsoon onset in recent decades. *Geophys. Res. Lett.* **39**, L03803. https://doi.org/10.1029/2011GL050540 (2012).
20. Mukherjee, S., Aadhar, S., Stone, D. & Mishra, V. Increase in extreme precipitation events under anthropogenic warming in India. *Weather Clim. Extrem.* **20**, 45–53. https://doi.org/10.1016/j.wace.2018.03.005 (2018).
21. Ha, K.-J., Moon, S., Timmermann, A. & Kim, D. Future changes of summer monsoon characteristics and evaporative demand Over Asia in CMIP6 simulations. *Geophys. Res. Letters* **47**, e2020GL087492. https://doi.org/10.1029/2020GL087492 (2020).
22. Revadekar, J. V. & Preethi, B. Statistical analysis of the relationship between summer monsoon precipitation extremes and foodgrain yield over India. *Int. J. Climatol.* **32**, 419–429. https://doi.org/10.1002/joc.2282 (2012).
23. Vogel, E. *et al.* The effects of climate extremes on global agricultural yields. *Environ. Res. Lett.* **14**, 054010. https://doi.org/10.1088/1748-9326/ab154b (2019).

24. Fontes, F., Gorst, A. & Palmer, C. Threshold effects of extreme weather events on cereal yields in India. *Clim. Chang.* **165**, 26. https://doi.org/10.1007/s10584-021-03051-x (2021).
25. Boehm, R. *et al.* Association between empirically estimated monsoon dynamics and other weather factors and historical tea yields in China: Results from a yield response model. *Climate* **4**, 20. https://doi.org/10.3390/cli4020020 (2016).
26. Guan, K., Sultan, B., Biasutti, M., Baron, C. & Lobell, D. B. What aspects of future rainfall changes matter for crop yields in West Africa?. *Geophys. Res. Lett.* **42**, 8001–8010. https://doi.org/10.1002/2015GL063877 (2015).
27. Balwinder-Singh, *et al.* Taking the climate risk out of transplanted and direct seeded rice: Insights from dynamic simulation in Eastern India. *Field Crops Res.* **239**, 92–103. https://doi.org/10.1016/j.fcr.2019.05.014 (2019).
28. BBC. Food prices rise on late monsoon. *BBC News* (2012).
29. Express, I. UP: After monsoon delay, excess rain, strong winds hit crops (2022).
30. Iizumi, T. & Ramankutty, N. How do weather and climate influence cropping area and intensity?. *Glob. Food Secur.* **4**, 46–50. https://doi.org/10.1016/j.gfs.2014.11.003 (2015).
31. Rezaei, E. E., Ghazaryan, G., Moradi, R., Dubovyk, O. & Siebert, S. Crop harvested area, not yield, drives variability in crop production in Iran. *Environ. Res. Lett.* **16**, 064058. https://doi.org/10.1088/1748-9326/abfe29 (2021).
32. Schlenker, W. & Roberts, M. J. 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 (2009).
33. Holmgren, M., Gómez-Aparicio, L., Quero, J. L. & Valladares, F. Non-linear effects of drought under shade: Reconciling physiological and ecological models in plant communities. *Oecologia* **169**, 293–305. https://doi.org/10.1007/s00442-011-2196-5 (2012).
34. Lobell, D. B., Sibley, A. & Ortiz-Monasterio, J. I. Extreme heat effects on wheat senescence in India. *Nat. Clim. Chang.* **2**, 186–189. https://doi.org/10.1038/nclimate1356 (2012).
35. Feng, *et al.* Impacts of rainfall extremes on wheat yield in semi-arid cropping systems in eastern Australia. *Clim. Chang.* **147**, 555–569. https://doi.org/10.1007/s10584-018-2170-x (2018).
36. Elavarasan, D. & Vincent, P. M. D. R. A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters. *J. Ambient Intell. Humaniz. Comput.* https://doi.org/10.1007/s12652-020-02752-y *(2021)*.
37. Shahhosseini, M., Hu, G., Huber, I. & Archontoulis, S. V. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Sci. Rep.* **11**, 1606. https://doi.org/10.1038/s41598-020-80820-1 (2021).
38. Leng, G. & Hall, J. W. Predicting spatial and temporal variability in crop yields: An inter-comparison of machine learning, regression and process-based models. *Environ. Res. Lett.* **15**, 044027. https://doi.org/10.1088/1748-9326/ab7b24 (2020).
39. Hoffman, A. L., Kemanian, A. R. & Forest, C. E. The response of maize, sorghum, and soybean yield to growing-phase climate revealed with machine learning. *Environ. Res. Lett.* **15**, 094013. https://doi.org/10.1088/1748-9326/ab7b22 (2020).
40. Gupta, D. K. *et al.* Global warming potential of rice-wheat cropping system of the indo-gangetic plains. *Indian J. Agricult. Sci.* **85**(6), 807–816 (2015).
41. Sekar, I. & Pal, S. Rice and wheat crop productivity in the indo-gangetic plains of India: Changing pattern of growth and future strategies. *Indian J. Agric. Econ.* **67**, 238–252 (2012).
42. Singh, N. & Sontakke, N. A. On climatic fluctuations and environmental changes of the indo-gangetic plains India. *Clim. Chang.* **52**, 287–313. https://doi.org/10.1023/A:1013772505484 (2002).
43. Preethi, B., Mujumdar, M., Kripalani, R. H., Prabhu, A. & Krishnan, R. Recent trends and tele-connections among South and East Asian summer monsoons in a warming environment. *Clim. Dyn.* **48**, 2489–2505. https://doi.org/10.1007/s00382-016-3218-0 (2017).
44. Swamikannu, N. *et al.* District level data (DLD) and dashboard for agriculture and allied-sectors in India, https://doi.org/10.21421/D2/XFB1BZ (2021).
45. Davis, K. F. *et al.* Assessing the sustainability of post-Green Revolution cereals in India. *Proceedings of the National Academy of Sciences* **116**, 25034–25041. https://doi.org/10.1073/pnas.1910935116 (2019).
46. Amale, H. S., Birthal, P. S. & Negi, D. S. Delayed monsoon, irrigation and crop yields. *Agric. Econ. n/a* https://doi.org/10.1111/agec.12746 *(2022)*.
47. Grassini, P., Eskridge, K. M. & Cassman, K. G. Distinguishing between yield advances and yield plateaus in historical crop production trends. *Nat. Commun.* **4**, 2918. https://doi.org/10.1038/ncomms3918 (2013).
48. Sheffield, J., Goteti, G. & Wood, E. F. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Clim.* **19**, 3088–3111. https://doi.org/10.1175/JCLI3790.1 (2006).
49. Shrivastava, S., Kar, S. C. & Sharma, A. R. Intraseasonal variability of summer monsoon rainfall and droughts over central India. *Pure Appl. Geophys.* **174**, 1827–1844. https://doi.org/10.1007/s00024-017-1498-x (2017).
50. Heino, M., Guillaume, J. H. A., Müller, C., Iizumi, T. & Kummu, M. A multi-model analysis of teleconnected crop yield variability in a range of cropping systems. *Earth Syst. Dyn.* **11**, 113–128. https://doi.org/10.5194/esd-11-113-2020 (2020).
51. Jackson, N. D., Konar, M., Debaere, P. & Sheffield, J. Crop-specific exposure to extreme temperature and moisture for the globe for the last half century. *Environ. Res. Lett.* **16**, 064006. https://doi.org/10.1088/1748-9326/abf8e0 (2021).
52. Manzoor, S. & Ahmad Ahanger, M. Spatio-temporal trends in precipitation and temperature means/extremes in the Himalayan states of India. *J. Water Clim. Chang.* **13**, 2531–2558. https://doi.org/10.2166/wcc.2022.395 (2022).
53. Bhattacharyya, S. & Sreekesh, S. Assessments of multiple gridded-rainfall datasets for characterizing the precipitation concentration index and its trends in India. *Int. J. Climatol.* **42**, 3147–3172. https://doi.org/10.1002/joc.7412 (2022).
54. Parkes, B. *et al.* Weather dataset choice introduces uncertainty to estimates of crop yield responses to climate variability and change. *Environ. Res. Lett.* **14**, 124089. https://doi.org/10.1088/1748-9326/ab5ebb (2019).
55. Hoffman, A. L., Kemanian, A. R. & Forest C. E. (2018) Analysis of climate signals in the crop yield record of sub-Saharan Africa. *Glob. Chang. Biol.* **24**, 143–157, https://doi.org/10.1111/gcb.13901.
56. Roell, Y. E., Beucher, A., Møller, P. G., Greve, M. B. & Greve, M. H. Comparing a random forest based prediction of winter wheat yield to historical yield potential. *Agronomy* **10**, 395. https://doi.org/10.3390/agronomy10030395 (2020).
57. Wang, P. *et al.* Temperature variations and rice yields in China: Historical contributions and future trends. *Clim. Chang.* **124**, 777–789. https://doi.org/10.1007/s10584-014-1136-x (2014).
58. Jeong, J. H. *et al.* Random forests for global and regional crop yield predictions. *PLOS ONE* **11**, e0156571. https://doi.org/10.1371/journal.pone.0156571 (2016).
59. Misra, V., Bhardwaj, A. & Mishra, A. Local onset and demise of the Indian summer monsoon. *Clim. Dyn.* **51**, 1609–1622. https://doi.org/10.1007/s00382-017-3924-2 (2018).
60. Partridge, T. F. *et al.* Mid-20th century warming hole boosts US maize yields. *Environ. Res. Lett.* **14**, 114008. https://doi.org/10.1088/1748-9326/ab422b (2019).
61. Ben-Ari, T. *et al.* Identifying indicators for extreme wheat and maize yield losses. *Agric. For. Meteorol.* **220**, 130–140. https://doi.org/10.1016/j.agrformet.2016.01.009 (2016).
62. Zhang, S., Tao, F. & Zhang, Z. Changes in extreme temperatures and their impacts on rice yields in southern China from 1981 to 2009. *Field Crops Res.* **189**, 43–50. https://doi.org/10.1016/j.fcr.2016.02.008 (2016).
63. Bachmair, S., Tanguy, M., Hannaford, J. & Stahl, K. How well do meteorological indicators represent agricultural and forest drought across Europe?. *Environ. Res. Lett.* **13**, 034042. https://doi.org/10.1088/1748-9326/aaafda (2018).
64. Peng, S. *et al.* Rice yields decline with higher night temperature from global warming. *Proc. Natl. Acad. Sci. USA* **101**, 9971–9975. https://doi.org/10.1073/pnas.0403720101 (2004).

65. Tao, F. *et al.* Single rice growth period was prolonged by cultivars shifts, but yield was damaged by climate change during 1981–2009 in China, and late rice was just opposite. *Glob. Chang. Biol.* **19**, 3200–3209. https://doi.org/10.1111/gcb.12250 (2013).

66. Deng, N. *et al.* Influence of temperature and solar radiation on grain yield and quality in irrigated rice system. *Eur. J. Agron.* **64**, 37–46. https://doi.org/10.1016/j.eja.2014.12.008 (2015).

67. Jain, M., Naeem, S., Orlove, B., Modi, V. & DeFries, R. S. Understanding the causes and consequences of differential decision-making in adaptation research: Adapting to a delayed monsoon onset in Gujarat, India. *Glob. Environ. Chang.* **31**, 98–109. https://doi.org/10.1016/j.gloenvcha.2014.12.008 (2015).

68. Noska, R. & Misra, V. Characterizing the onset and demise of the Indian summer monsoon. *Geophys. Res. Lett.* **43**, 4547–4554. https://doi.org/10.1002/2016GL068409 (2016).

69. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32. https://doi.org/10.1023/A:1010933404324 (2001).

70. Kok, Z. H., Mohamed Shariff, A. R., Alfatni, M. S. M. & Khairunniza-Bejo, S. Support vector machine in precision agriculture: A review. *Comput. Electron. Agric.* **191**, 106546. https://doi.org/10.1016/j.compag.2021.106546 (2021).

71. Chen, H., Wu, W. & Liu, H.-B. Assessing the relative importance of climate variables to rice yield variation using support vector machines. *Theor. Appl. Climatol.* **126**, 105–111. https://doi.org/10.1007/s00704-015-1559-y (2016).

72. Gandhi, N., Armstrong, L. J., Petkar, O. & Tripathy, A. K. Rice crop yield prediction in India using support vector machines. In *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 1–5, https://doi.org/10.1109/JCSSE.2016.7748856 (2016).

73. Guo, Y. *et al.* Integrated phenology and climate in rice yields prediction using machine learning methods. *Ecol. Indic.* **120**, 106935. https://doi.org/10.1016/j.ecolind.2020.106935 (2021).

74. Feng, P., Wang, B., Liu, D. L., Waters, C. & Yu, Q. Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia. *Agric. For. Meteorol.* **275**, 100–113. https://doi.org/10.1016/j.agrformet.2019.05.018 (2019).

75. Liaw, A. & Wiener, M. Classification and regression by random forest. *R News* **2**, 18–22 (2002).

76. R Core Team. R: A language and environment for statistical computing. *R Found. Stat. Comput.* (2021).

77. Feng, S., Li, J., Qian, G. & Feng, B. Association between the yield and the main agronomic traits of tartary buckwheat evaluated using the random forest model. *Crop Sci.* **60**, 2394–2407. https://doi.org/10.1002/csc2.20243 (2020).

78. Apley, D. W. & Zhu, J. Visualizing the effects of predictor variables in black box supervised learning models. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **82**, 1059–1086. https://doi.org/10.1111/rssb.12377 (2016).

79. Spera, S. A., Winter, J. M. & Partridge, T. F. Brazilian maize yields negatively affected by climate after land clearing. *Nat. Sustain.* **3**, 845–852. https://doi.org/10.1038/s41893-020-0560-3 (2020).

80. Antoniadis, A., Lambert-Lacroix, S. & Poggi, J.-M. Random forests for global sensitivity analysis: A selective review. *Reliab. Eng. Syst. Saf.* **206**, 107312. https://doi.org/10.1016/j.ress.2020.107312 (2021).

81. Zaveri, E. & Lobell, D. B. The role of irrigation in changing wheat yields and heat sensitivity in India. *Nat. Commun.* **10**, 1–7. https://doi.org/10.1038/s41467-019-12183-9 (2019).

82. Crane-Droesch, A. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environ. Res. Lett.* **13**, 114003. https://doi.org/10.1088/1748-9326/aae159 (2018).

83. Asseng, S. *et al.* Hot spots of wheat yield decline with rising temperatures. *Glob. Chang. Biol.* **23**, 2464–2472. https://doi.org/10.1111/gcb.13530 (2017).

84. Tack, J., Barkley, A. & Hendricks, N. Irrigation offsets wheat yield reductions from warming temperatures. *Environ. Res. Lett.* **12**, 114027. https://doi.org/10.1088/1748-9326/aa8d27 (2017).

85. Reidy, S. Timely monsoon helps India's rice, corn crop | World-grain.com | August 09, 2017 13:44 | World Grain. https://www.world-grain.com/articles/8518-timely-monsoon-helps-india-s-rice-corn-crop (2017).

86. Krar, P. & Chaudhury, D. R. Rice production in India to shrink due to late monsoon. *Econ. Times* (2019).

87. Saini, S. & Das, S. 2021 monsoon shows impact of climate change. Here's what it's doing to kharif crops (2021).

88. Mishra, A., Singh, R., Raghuwanshi, N. S., Chatterjee, C. & Froebrich, J. Spatial variability of climate change impacts on yield of rice and wheat in the Indian Ganga basin. *Sci. Total Environ.* **468–469**, S132–S138. https://doi.org/10.1016/j.scitotenv.2013.05.080 (2013).

89. Soora, N. K. *et al.* An assessment of regional vulnerability of rice to climate change in India. *Clim. Chang.* **118**, 683–699. https://doi.org/10.1007/s10584-013-0698-3 (2013).

90. Zampieri, M. *et al.* Surface freshwater limitation explains worst rice production anomaly in India in 2002. *Remote Sens.* **10**, 244. https://doi.org/10.3390/rs10020244 (2018).

91. Hsiang, S., Lobell, D., Roberts, M. & Schlenker, W. Climate and crop yields in Australia, Brazil, China, Europe and the United States. SSRN Scholarly Paper ID 2977571 (Social Science Research Network, Rochester, NY, 2013). https://doi.org/10.2139/ssrn.2977571.

92. Carter, E. K., Melkonian, J., Riha, S. J. & Shaw, S. B. Separating heat stress from moisture stress: Analyzing yield response to high temperature in irrigated maize. *Environ. Res. Lett.* **11**, 094012. https://doi.org/10.1088/1748-9326/11/9/094012 (2016).

93. Mondal, P., Jain, M., DeFries, R. S., Galford, G. L. & Small, C. Sensitivity of crop cover to climate variability: Insights from two Indian agro-ecoregions. *J. Environ. Manag.* **148**, 21–30. https://doi.org/10.1016/j.jenvman.2014.02.026 (2015).

94. Prasad, N. R., Patel, N. R. & Danodia, A. Crop yield prediction in cotton for regional level using random forest approach. *Spat. Inf. Res.* **29**, 195–206. https://doi.org/10.1007/s41324-020-00346-6 (2021).

95. Cornish, P. S., Karmakar, D., Kumar, A., Das, S. & Croke, B. Improving crop production for food security and improved livelihoods on the East India Plateau. I. Rainfall-related risks with rice and opportunities for improved cropping systems. *Agric. Syst.* **137**, 166–179. https://doi.org/10.1016/j.agsy.2015.01.008 (2015).

96. Lee, J.-Y. *et al.* Future global climate: Scenario-based projections and near-term information. In Masson-Delmotte, V. *et al.* (eds.) *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2021).

97. Sultan, B., Baron, C., Dingkuhn, M., Sarr, B. & Janicot, S. Agricultural impacts of large-scale variability of the West African monsoon. *Agric. For. Meteorol.* **128**, 93–110. https://doi.org/10.1016/j.agrformet.2004.08.005 (2005).

98. Chhogyel, N., Kumar, L. & Bajgai, Y. Rainfall anomalies and their impacts on Bhutan's agro-ecological landscape. *Reg. Environ. Chang.* **21**, 125. https://doi.org/10.1007/s10113-021-01851-6 (2021).

99. Lausier, A. M. & Jain, S. Overlooked trends in observed global annual precipitation reveal underestimated risks. *Sci. Rep.* **8**, 16746. https://doi.org/10.1038/s41598-018-34993-5 (2018).

100. Shi, W., Tao, F. & Zhang, Z. A review on statistical models for identifying climate contributions to crop yields. *J. Geogr. Sci.* **23**, 567–576. https://doi.org/10.1007/s11442-013-1029-3 (2013).

101. Hay, J. Extreme weather and climate events, and farming risks. In Sivakumar, M. V. K. & Motha, R. P. (eds.) *Managing Weather and Climate Risks in Agriculture*, 1–19, (Springer, Berlin, Heidelberg, 2007) https://doi.org/10.1007/978-3-540-72746-0_1

102. Bhatta, G. D. & Aggarwal, P. K. Coping with weather adversity and adaptation to climatic variability: A cross-country study of smallholder farmers in South Asia. *Clim. Dev.* **8**, 145–157. https://doi.org/10.1080/17565529.2015.1016883 (2016).

103. Sikka, A. K., Rao, B. B. & Rao, V. U. M. (2016) Agricultural disaster management and contingency planning to meet the challenges of extreme weather events. *MAUSAM*, **67**, 155–168. https://doi.org/10.54302/mausam.v67i1.1173

104. Panda, D. K., Ambast, S. K. & Shamsudduha, M. Groundwater depletion in northern India: Impacts of the sub-regional anthropogenic land-use, socio-politics and changing climate. *Hydrol. Process.* **35**, e14003. https://doi.org/10.1002/hyp.14003 (2021).

Author contributions
All authors formulated the research. C.B. conducted the data preparation, modelling and wrote the manuscript. T.F. and B.P. provided guidance for the research and reviewed and edited the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-27752-8.

Correspondence and requests for materials should be addressed to C.B.

Reprints and permissions information is available at www.nature.com/reprints.

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