S1. Yield de-trending methods

We used two different yield de-trending methods: the first-difference [21, 23] and five-year running mean [22]. The de-trending emphasises the change in yield caused by short-term factors, primarily due to seasonal climate, although demand, prices, technology and other factors may also affect yield variability.

S1.1. First-difference of yields

A first-difference time series of grid-cell yields, $Y'_{t,g,c}$ (%), was computed as:

$$Y'_{t,g,c} = \left( \frac{Y_{t,g,c} - Y_{t-1,g,c}}{\bar{Y}_{t,g,c}} \right) \times 100,$$

where the suffixes $t$, $g$ and $c$ indicate the year, grid cell and cropping season of a crop, respectively. $Y_t$ and $Y_{t-1}$ indicate the yields in year $t$ and in the previous year ($t-1$) (t ha$^{-1}$), respectively, and $\bar{Y}$ is the average yield for the interval from year $t-3$ to year $t-1$ (t ha$^{-1}$). Hereafter the average yield is referred to as normal yield. The same normal yield value was used for each of the first four years of the analysis, as in ref. 23. No first-difference of yields was computed for the year 1981 ($t=1$). Hence, the sample size when calculating yield variability for the earlier and letter period was 14 and 15, respectively (i.e., $SDr$, see section S1.3 for more explanation). For maize, rice and wheat, the mean first-difference of yields across the different cropping seasons of a crop was calculated as:

$$Y'_{c,g} = \frac{\sum_{t=1}^{T} P_{g,c} \cdot Y'_{t,g,c}}{\sum_{t=1}^{T} P_{g,c}},$$

$P$ is the production of a crop in each cropping season (tonnes), and $C$ is the number of cropping seasons in a year employed to produce each crop. Two cropping seasons (major and secondary or winter and spring) were used for maize, rice and wheat, whereas a single cropping season was employed for soybean. The production resulting from the different cropping seasons in major crop-producing regions in the 1990s were obtained from the U.S. Department of Agriculture [29].

S1.2. Five-year running mean yields

An anomaly time series of grid-cell yields, $Y'_{t,g,c}$ (%), was calculated as:

$$Y'_{t,g,c} = \left( \frac{Y_{t,g,c} - \bar{Y}_{t,g,c}}{\bar{Y}_{t,g,c}} \right) \times 100,$$

$Y$ indicates the yield (t ha$^{-1}$), and $\bar{Y}$ is the normal yield defined as a five-year running mean yield for the interval from year $t-2$ to year $t+2$ (t ha$^{-1}$). The yield anomaly in the first two years (1982 and 1983) and last two years (2004 and 2005) were not calculated as in ref. 22. Hence, the sample size when calculating yield
variability for each of the earlier and latter periods was 13. Although yield anomaly values are sensitive to the number of averaging years when calculating normal yield, the sample size decreases as the averaging interval elongates. We used a five-year window for this study on the basis of the sensitivity analysis results presented in ref. 22. For maize, rice and wheat, the mean yield anomaly across the cropping seasons was calculated following Eq. S2.

S1.3. Measures of yield variability change

Two different measures were used to characterize yield variability change. The first one is the slope of a linear regression line fitted to a nine-year running window time series of the standard deviation (SD) of yield anomalies (Slp). We first calculated the SD of yield anomalies for every nine years using running window. In this calculation, the SD value for a given window was calculated when at least five samples, out of nine samples, were available. Then a linear regression line was fitted to the SD time series by using the least square method when over 50% of the SD time series were available (note that the sample size in the 30-year period varied by yield de-trending method). Although previous work used a 11-year running window for 92- or 50-year long yield analyses [18, 19], we used the relatively shorter window because of the limited length of yield data.

The second measure is the ratio of the SD of yield anomalies for the later years (1996–2010) relative to that for the earlier years (1981–1995) (SDr). We computed the ratio value only when each of the two periods had over 50% of the yield anomaly time series in the corresponding period. Theoretically, ratio values have a non-Gaussian nature. Thereby, we took natural log of the ratio to increase the visibility when presenting scatter plots (i.e., figures 4, S3, S7–S9). The statistical significance of yield variability change was tested using the bootstrap method [24].

S2. Agro-climatic index

S2.1. Definition

As shown in figure 1, the agro-climatic index used in this study was the annual time series of the sum of effective radiation intercepted by the crop canopy during the yield formation stage that includes thresholds for extreme temperatures and extreme soil moisture deficit. The index value was crop-specific and calculated by cropping system (rainfed or irrigated) and cropping season (major and secondary or winter and spring). However, for each crop, only a single value of the index averaged over the cropping system and cropping season was used for the analysis.

The sum of effective radiation intercepted by the crop canopy during the yield formation stage in rainfed conditions, Index_{rain} (MJ m^{-2} \text{ season}^{-1}), was calculated:

\[
Index_{\text{rain}, y, g, c, l} = \sum_{d=D_{\text{tr},g,l}}^{D_{\text{e},g,c,l}} R_{y, g, c, l, d}
\]

if \( T_{\text{max}, y, d} < T_{\text{max, tr, g, l}} \), \( T_{\text{min}, y, d} > T_{\text{min, tr, g, l}} \), \( E_{\text{a}, y, d, c, l} / E_{p, y, d, c, l} > E_{\text{a}} / E_{p, tr, g, l} \)

and \( SWE_{y, d} = 0 \), (Eq. S4)
where the suffixes $y$, $g$, $c$, $l$ and $d$ indicates year, grid cell, cropping season of a crop, planting date (see section S2.3) and day of year (DOY), respectively; $R$ is the daily global radiation intercepted by the crop canopy (MJ m$^{-2}$ d$^{-1}$); $D_y$ is the date at which yield formation stage starts ($fr_{PHU} \geq fr_{PHU.tr}$, DOY); $D_h$ is the simulated first date at which $fr_{PHU} \geq 1$ (correspond to the harvesting date, DOY); $T_{\max}$ and $T_{\min}$ is the daily maximum and minimum temperatures ($^\circ$C); $E_a$ and $E_p$ is the actual and potential evapotranspiration rates (mm d$^{-1}$); SWE is the snow water equivalent (mm); and $T_{\max.tr}$ and $T_{\min.tr}$ is the upper and lower bound of the optimal temperature range for yield formation, respectively; and $E_d/E_{p.tr}$ is the lower bound of optimal soil moisture range for yield formation. $T_{\max.tr}$, $T_{\min.tr}$ and $E_d/E_{p.tr}$ describe the crop characteristic to these major abiotic stresses. $D_y$, $T_{\max.tr}$, $T_{\min.tr}$ and $E_d/E_{p.tr}$ were calibrated for each crop, grid cell and planting date (see section S2.4). $R$ is given as:

$$R_{y,g,c,l,d} = 0.5 \cdot SR_{y,g,d} \cdot \left \{ 1 - \exp \left ( -k \cdot LAI_{y,g,c,l,d} \right ) \right \}, \text{ (Eq. S5)}$$

where $SR$ is the global radiation (MJ m$^{-2}$ d$^{-1}$), $0.5 \cdot SR$ is the photosynthetically active radiation (MJ m$^{-2}$ d$^{-1}$) and $k$ is the light extinction coefficient (=0.65, [25]).

For irrigated conditions, Eq. S4 was changed to:

$$Index_{min, y, g, c, l} = D_{y,c,l} \sum_{d=D_{y.c}} R_{y,g,c,l,d}, \text{ (Eq. S5)}$$

if $T_{\max, y, d} < T_{\max.tr, g, l}$, $T_{\min, y, d} > T_{\min.tr, g, l}$ and $SWE_{y, d} = 0$, (Eq. S6)

in order to relax the condition related to soil moisture deficit. Then the mean index value across the cropping seasons of a crop was calculated separately for rainfed and irrigated areas in a similar manner as described in Eq. S2:

$$Index_{min, y, g, l} = \frac{\sum_{c=1}^{C} P_{g,c} \cdot Index_{min, y, g, c,l}}{\sum_{c=1}^{C} P_{g,c}} \text{ and } Index_{min, y, g, l} = \frac{\sum_{c=1}^{C} P_{g,c} \cdot Index_{min, y, g, c,l}}{\sum_{c=1}^{C} P_{g,c}}. \text{ (Eq. S7)}$$

Index values were further averaged across rainfed and irrigated areas as:

$$Index_{y, g, l} = \frac{A_{rain,y} \cdot Index_{rain, y, g, l} + A_{irri,y} \cdot Index_{irri, y, g, l}}{A_{rain,y} + A_{irri,y}}. \text{ (Eq. S8)}$$

where $A_{rain}$ and $A_{irri}$ is the extent of rainfed and irrigation-equipped areas circa 2000 (ha) [28]. Both $A_{rain}$ and $A_{irri}$ are time-constant. Although historical crop-specific information on irrigated area would be preferable, such information is not yet available in a crop-specific manner. A recent study [39] provides historical information on the extent of irrigated area, but it is not crop-specific. The assumption used here is rather unrealistic in a certain location-crop combination and may explain the limited skill of the index in explaining yield variability change.

As with yield, we calculated Slp and SDr values for the index to characterize changes in the variability of agro-climatic condition. First-difference of the index was used to compute these measures to be compared with yield variability change derived using FD de-trending method. Anomaly of the index, deviated from the mean index in 1981–2010, was used to calculate the measures to be compared with yield variability change derived based on RM de-trending method.

S2.2. Inputs

As shown in figure 1, the index calculation procedure requires planting date as well
as daily weather data (daily maximum and minimum 2-m air temperatures, precipitation, global radiation, relative humidity and 10-m wind speed) as the inputs. The maps of harvested area [43] as well as the map of irrigated and rainfed areas [28], both are time-constant and crop-specific information, are also used as the inputs.

The procedure uses dynamic and interacting models on crop phenology, potential leaf area and canopy height, evapotranspiration, soil water and snow cover. The daily crop development was modelled using the fraction of heat units accumulated for the crop in a given day ($fr_{PHU}$) to the total heat units required for crop maturity ($PHU$) and crop-specific base temperature (maize and rice, 8°C; soybean, 10°C; and wheat, 0°C) [25]. The daily increment of potential leaf area and canopy height under optimal condition was simulated based on $fr_{PHU}$. The approach used is based on the Soil and Water Assessment Tool (SWAT) [25].

The potential evapotranspiration rate under potential leaf area and canopy height is estimated using a variant of Penman-Monteith method [25]. The potential evapotranspiration rate is adjusted for high vapour pressure deficit and varying atmospheric concentration of carbon dioxide ($CO_2$) [25]. The actual evapotranspiration rate is derived from the potential evapotranspiration rate and available root-zone soil moisture. The root-zone soil moisture is computed using the soil water balance model [26] coupled with the snow cover model [27] that accounted for the plant-extractable soil water capacity [44], precipitation, evapotranspiration, snow accumulation and melting, surface and subsurface runoffs, and ground water loss through deep percolation. The plant-extractable soil water capacity is an estimate based on soil texture, soil organic content, plant-root (or soil-profile) depth derived from the FAO/UNESCO (FAO/United Nations Educational, Scientific and Cultural Organization) soil database [44]. The root-zone soil moisture and snow water equivalent simulation was started at January 1st 1970 to complete the 11-year long (1970–1980) spin-up before providing data for the index calculation.

For this study, the meteorological forcing data set for global crop modelling [16] was used as the daily weather inputs because recent work [16, 45] suggests that the data set provides more reliable estimates of some climatic variables (vapour pressure and wind speed in particular which affect simulated potential and actual evapotranspiration rates) than other data sets. The observed annual atmospheric $CO_2$ concentration data at Mauna Loa, Hawaii [46] was uniformly used for all grid cells over the global cropland. Because $CO_2$ data for 2009–2010 were not yet available from this source, we estimated data values in these years by using the liner regression line calculated based on the data in the latest five years.

S2.3. Calibration of total heat units

The global crop calendar [47] provides grid-cell dates of planting and harvesting circa 2000 by crop and cropping season of a crop. The crop duration (planting till harvesting) derived from this data set was used to calibrate the crop phenology model. The planting date was also used as the input to the index calculation procedure. We used time-constant planting date throughout this study. This assumption is rather unrealistic for some location-crop combinations [30, 32, 33, 48], but historical information is not available.

For a given crop and cropping season of a crop, we calculated grid-cell $PHU$ value
year by year for the period 1996–2005 using the reported typical dates of planting and harvesting and daily mean temperature \((=\frac{T_{\text{max}} + T_{\text{min}}}{2})\). Then the mean \(PHU\) value in that period was computed and used as the calibrated \(PHU\) value because the crop calendar [47] reports dates of planting and harvesting in the 1990s or 2000s and users cannot identify which year is used for a given location and crop.

We generated four additional planting dates, -30 days, -15 days, +15 days and +30 days of the typical date. As with planting date, four different harvesting dates were provided. This led to five different crop durations indicated by \{planting date, harvesting date\} ranging from \{-30 days, -30 days\} to \{+30 days, +30 days\}. We did not consider other combinations, such as \{-30 days, +30 days\}, because of the lack of information to justify such irregular combinations. These were used to account for the uncertainty of index value associated with different planting management.

S2.4. Calibration of other empirical parameters

Values of the four empirical parameters, \(D_y\), \(T_{\text{max, tr}}\), \(T_{\text{min, tr}}\) and \(E/E_{p, tr}\), as well as the variance of error (a normal distribution with zero mean was assumed as in previous work [35, 49, 50]) were determined in a probabilistic manner separately for each grid cell using the Markov Chain Monte Carlo (MCMC) method [35, 49, 50] and data in the calibration period (1989–2001) to best fit the time variation pattern between the yield anomaly and the index anomaly. Hence the posterior distributions of the parameters were specific to crop, location and planting date. This grid-cell specific calibration procedure was done to account for the major characteristics of local agronomic technology used by producers, as done in ref. 14.

The prior distributions used here were uniform distributions, of which upper and lower bounds were given by literature reviewing the critical temperature thresholds for the crops at various growth stages [51, 52]. The lower and upper bounds were 0.4–0.9 for \(f_rPHU, tr\), 25.0–50.0 for \(T_{\text{max, tr}}\), 0.0–20.0 for \(T_{\text{min, tr}}\) and 0.0–1.0 for \(E/E_{p, tr}\). By using these prior distributions, the MCMC method could incorporate the information derived from the field and chamber experimental results into the estimation of the parameters values. Monte Carlo steps of 10,000 with six chains were used. The convergence to posterior distributions was assessed using the Gelman-Rubin statistics (<1.05) [53].

Twenty different sets of these parameters were randomly sampled from the posterior distributions for each of five different planting dates and formed a perturbed-parameter ensemble simulation. The mean value and coefficient of variation (CV) for each of the parameters across the 100 ensembles of five planting dates and the 20 parameter sets are shown in figures S11–S14. It is consistently evident for all crops examined here that: 1) the parameter values varied by location; 2) the range of posterior parameter values was far narrower than that of the corresponding prior distribution (note that the range of the colour bars in the left panels of figures S11–S14 was arranged to be the same as that of the corresponding prior distributions mentioned earlier); and 3) the parametric uncertainty of \(D_y\) and \(T_{\text{max, tr}}\) (represented by the CV) was relatively small compared to that of \(T_{\text{min, tr}}\) and \(E/E_{p, tr}\). Then, as the last part of the calibration procedure, seven members from the 100 ensembles (planting date could vary by member) were selected according to the correlation coefficient values between the yield anomaly and the calculated index in the calibration period (1989–2001) and used in further analysis. The calibration using
the limited portion of the data ensured that variability changes in the yield and agro-climatic condition for the whole analysis period were not a result of the calibration. More than seven ensemble members may be preferable, but the percentage of the global harvested area with skillful index gradually decreased with increase in the number of ensemble members. We set the number of ensemble members to be seven to cover at least 60% of the global harvested area for all crops examined in this study.

S2.5. Limitations

There are some limitations in the models used in this study to develop the index. First, at the moment of writing, the LAI sub-model simulates potential leaf area instead of the actual one that is often regulated by abiotic and biotic stresses. Generally, more transpiration demand is expected in higher LAI conditions, suggesting the risk that simulated water deficit could be overestimated. However, as shown in figures 5, S9 and S10, our result showed that the variability change in soil water deficit was less capable in explaining the yield variability change than other abiotic stresses. Therefore, we infer that the assumption of potential LAI is conservative and that this limitation does not affect the findings of our study.

Second, we did not evaluate the reliability of the simulated potential LAI. The modeling of potential LAI dynamics and crop-specific maximum LAI value used here were based on SWAT [25]; the crop growth simulations of SWAT have been tested in many locations of the world. Based on our review of the literature evaluating SWAT-simulated crop LAI (for instance, refs. 54 and 55), it is possible to assume that the simulated potential LAI is reasonable. The potential LAI in some location and crop may be unreasonable, but no LAI observation networks are available to test this globally and historically.

Finally, yield losses due to extreme excess soil water or floods were not considered in this study. This is because intense precipitation events do not necessarily indicate the incidence of flood, and further, global relationships between intense precipitation events and crop yields are less known (an exception is ref. 56. A regional study is also available, for instance, ref. 57). However, this does not mean that extreme precipitation have no impact on crop yields. A study presented the modeling of negative impacts of excess soil water on crop growth [58]. Such modeling may be worth including in future work.

References
43. Monfreda C, Ramankutty N and Foley J A 2008 Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000 Glob. Biogeochem. Cycles 22 GB1022
44. Dunne K A and Willmott C J 1996 Global distribution of plant-extractable water capacity of soil Int. J. Climatol. 16 841–859
45. Ruane A C, Goldberg R and Chryssanthacopoulos J 2015 Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation Agric. For. Meteorol. 200 233–248
46. Keeling R F, Piper S C, Bollenbacher A F and Walker J S 2009 Atmospheric CO2 records from sites in the SIO air sampling network Available: http://cdiac.ornl.gov/trends/co2/sio/mlo.html (accessed October 2015)
47. Sacks W J Deryng D, Foley J A and Ramankutty N 2010 Crop planting dates: An analysis of global patterns *Glob. Ecol. Biogeogr.* **19** 607–620
48. Kucharik C J 2006 A multidecadal trend of earlier corn planting in the central USA *Agron. J.* **98** 1544–1550
49. Iizumi T, Yokozawa M and Nishimori M 2009 Parameter estimation and uncertainty analysis of a large-scale crop model for paddy rice: Application of a Bayesian approach *Agric. For. Meteorol.* **149** 333–348
50. Sakurai G, Iizumi T, Nishimori M and Yokozawa M 2014 How much has the increase in atmospheric CO$_2$ directly affected past soybean production? *Sci. Rep.* **4** 4978
51. Sánchez B, Rasmussen A and Porter J R 2014 Temperatures and the growth and development of maize and rice: a review *Glob. Change Biol.* **20** 408–417
52. Luo Q 2011 Temperature thresholds and crop production: a review *Clim. Change* **109** 583–598
53. Gelman A and Rubin D B 1992 Inference from iterative simulation using multiple sequences *Stat. Sci.* **7** 457–511
54. Cavero J, Farre I, Debaeke P, Faci J M 2000 Simulation of maize yield under water stress with the EPICphase and CROPWAT models *Agron. J.* **92** 679-690
55. Luo Y et al 2008 Assessment of crop growth and soil water modules in SWAT2000 using extensive field experiment data in an irrigation district of the Yellow River Basin *J. Hydrol.* **352** 139–156
56. Lesk C, Rowhani P, Ramankutty N 2016 Influence of extreme weather disasters on global crop production *Nature* **529** 84–87
57. Iizumi T, Yokozawa M, Hayashi Y, Kimura F 2008 Climate change impact on rice insurance payouts in Japan *J. Appl. Meteor. Climatol.* **47** 2265–2278
58. Rosenzweig C, Tubiello F N, Goldberg R, Mills E and Bloomfield J 2002 Increased crop damage in the US from excess precipitation under climate change *Glob. Environ. Change* **12** 197–202
Figure S1. Yield variability change in 1981–2010 for four crops. Results based on FD-SDr and RM-SDr methods and hybrid yield estimates are presented. The significance level of variability change was set to be 5% (two-tailed test; and the bootstrap replication of 1,000 times). The sample size per grid cell was ~29 yield anomalies for FD-SDr method and ~26 for RM-SDr method. For FD-SDr method, the analysis was performed only when eight or more samples were available for each of the earlier and latter periods, whereas for RM-SDr method it was done so only when seven or more samples were available for each period. The pie diagrams indicate the percentages of global harvested area in the colored areas, normalized to the global harvested area in 2000.
Figure S2. Yield variability change in 1981–2010 for four crops. Results based on FD-Slp and RM-Slp methods and hybrid yield estimates are presented. The significance level of variability change was set to be 5% (two-tailed test; and the bootstrap replication of 1,000 times). The sample size per grid cell was ~21 of nine-year-running-window-derived standard deviation of yield anomalies for FD-Slp method and ~18 for RM-Slp method. For FD-Slp method, the analysis was performed only when 11 or more samples were available, whereas for RM-Slp method it was did so only when 10 or more samples were available. The pie diagrams indicate the percentages of global harvested area in the colored areas, normalized to the global harvested area in 2000.
Figure S3. Comparison of (x-axis) the grid-cell yield variability change derived from the hybrid yield estimates with (y-axis) those derived from the subnational yield statistics for four crops. Only data in areas where the subnational yield statistics were available were compared. Data derived from four different methods (RM-Slp, RM-SDr, FD-Slp and FD-SDr) are presented. The measure of yield variability change was calculated only when the following criteria were met: FD-SDr method, eight or more samples of yield anomalies were available for each of the earlier and latter periods; RM-SDr method, seven or more samples of yield anomalies were available for each of the two periods; FD-Slp method, 11 or more samples of nine-year-running-window-derived SD of yield anomalies were available; and RM-Slp method, 10 or more samples of nine-year-running-window-derived SD of yield anomalies were available. The coefficient of determination ($R^2$), p-value (p) and sample size (n) are presented. Dashed diagonal line indicates one-to-one line.
Figure S4. Locations where the major characteristics of yield variability change could be reliably explained by the agro-climatic index. Data derived from RM and FD de-trending methods and hybrid yield estimates are presented. The nine-year running window time series of the standard deviation of yield anomalies and those of the index were compared. If the correlation coefficient was significant at the 5% level (two-tailed test; and the bootstrap replication of 1,000 times) and the significant results were consistent across over two-third (>5) of the seven members, then the index was assumed reliable. Otherwise, the index was assumed less reliable. For FD method, the sample size per grid cell and member was ~21 because the nine-year running window was applied to the first-difference-derived 29 yield anomalies in the 30-year period. The analysis was performed only when 11 or more samples were available. In contrast, for RM method, the sample size per grid cell and member was ~18 because the nine-year running window was applied to the five-year-running-mean-derived 26 yield anomalies in the 30-year period. The analysis was performed only when 10 or more samples were available. The pie diagrams indicate the percentages of harvested area in the colored areas, normalized against the world harvested area in 2000.
Figure S5. Change in the variability of agro-climatic index in 1981–2010 for four crops. Results based on FD-SDr and RM-SDr methods are presented. The significance level of variability change was set to be 5% (two-tailed test; the bootstrap replication of 1,000 times). The sample size per grid cell and member was ~29 of yield anomalies for FD-SDr method and ~26 for RM-SDr method. For FD-SDr method, the analysis was performed only when eight or more samples were available for each of the earlier and latter periods, whereas for RM-SDr method it was did so only when seven or more samples were available for each period. The result was assumed consistent only when over two-third (>5) of the seven members showed the significant results with the same direction of change. The pie diagrams indicate the percentages of global harvested area in the colored areas, normalized to the global harvested area in 2000.
Figure S6. Change in the variability of agro-climatic index in 1981–2010 for four crops. Results based on FD-Slp and RM-Slp methods are presented. The significance level of variability change was set to be 5% (two-tailed test; and the bootstrap replication of 1,000 times). The sample size per grid cell and member was ~21 of nine-year-running-window-derived standard deviation of yield anomalies for FD-Slp method and ~18 for RM-Slp method. For FD-Slp method, the analysis was performed only when 11 or more samples were available, whereas for RM-Slp method it was done only when 10 or more samples were available. The result was assumed consistent only when over two-third (>5) of the seven members showed the significant results with the same direction of change. The pie diagrams indicate the percentages of global harvested area in the colored areas, normalized to the global harvested area in 2000.
Figure S7. Smoothed density scatter plots of (x-axis) the grid-cell change in the variability of agro-climatic index indicated by the ensemble median of the index against (y-axis) grid-cell yield variability change for four crops. Data derived from four different methods (RM-Slp, RM-SDr, FD-Slp and FD-SDr) and hybrid yield estimates are presented. Only the data at the locations where the index could reliably explain the major characteristics of yield variability change were used. The coefficient of determination ($R^2$), p-value (p) and sample size (n) are presented. Dashed diagonal line indicates one-to-one line. Solid black line indicates the best-fitted liner regression line.
Figure S8. Smoothed density scatter plots of (x-axis) the grid-cell change in the variability of agro-climatic index indicated by the ensemble median of the index against (y-axis) grid-cell yield variability change for four crops. Data derived from four different methods (RM-Slp, RM-SDr, FD-Slp and FD-SDr) and subnational yield statistics are presented. Only the data at the locations where the index could reliably explain the major characteristics of yield variability change were used. The coefficient of determination ($R^2$), p-value ($p$) and sample size ($n$) are presented. Dashed diagonal line indicates one-to-one line. Solid black line indicates the best-fitted liner regression line.
Figure S9. Smoothed density scatter plots of (x-axis) the grid-cell variability change of relative frequencies of sub-optimal conditions indicated by the ensemble median against (y-axis) grid-cell yield variability change for four crops. Sub-optimal conditions include temperature above and below of the optimal range for yield formation (Above-$T_{\text{opt}}$ and below-$T_{\text{opt}}$, respectively), soil water deficit during the yield formation stage and either (or all) of the three suboptimal conditions (All sub-opt). Data derived from RM-Slp method and hybrid yield estimates are presented. Only the data at the locations where the index could reliably explain the major characteristics of yield variability change were used. The coefficient of determination ($R^2$), p-value (p) and sample size (n) are presented. Dashed diagonal line indicates one-to-one line. Solid black line indicates the best-fitted liner regression line.
Comparison of the skill in explaining yield variability change across different sub-optimal conditions for four crops. The skill was indicated by the coefficient of determination ($R^2$) calculated between the grid-cell yield variability change (derived using the subnational yield statistics) and the grid-cell change in variability of each sub-optimal condition. The sub-optimal conditions include the temperature above and below the optimal range for yield formation (Above-$T_{opt}$ and below-$T_{opt}$, respectively), soil water deficit during the yield formation stage and all three sub-optimal conditions (All sub-opt). Data for the change in variability of the agro-climatic index are presented as the reference. Red circle and blue cross indicates significant and insignificant $R^2$ value at the 5% level. A box indicates 50% probability interval. Black and green horizontal line within a box indicates median and mean value, respectively. Data for each box plot were derived from four different methods (RM-Slp, RM-SDr, FD-Slp and FD-SDr) and seven different members (thus, the sample size for a box was 28).
Figure S11. Mean and coefficient of variation (CV, in percent) values of the location-specific parameters for maize across the five different planting dates and 20 different parameter sets. Data at the location where the agro-climatic index could reliably explain the major characteristics of yield variability change are presented. The color bars in the left panels were arranged to match the range of the corresponding prior distributions (see section S2.4). The unit is as follows: \(D_y\), fraction; \(T_{\text{max, tr}}\), °C, \(T_{\text{min, tr}}\), °C; and \(E_a/E_p, \text{tr}\), fraction.
Figure S12. Mean and coefficient of variation (CV, in percent) values of the location-specific parameters for soybean across the five different planting dates and 20 different parameter sets. Data at the location where the agro-climatic index could reliably explain the major characteristics of yield variability change are presented. The color bars in the left panels were arranged to match the range of the corresponding prior distributions (see section S2.4). The unit is as follows: $D_y$, fraction; $T_{\text{max, tr}}$, °C, $T_{\text{min, tr}}$, °C; and $E_a/E_p$, tr, fraction.
Figure S13. Mean and coefficient of variation (CV, in percent) values of the location-specific parameters for rice across the five different planting dates and 20 different parameter sets. Data at the location where the agro-climatic index could reliably explain the major characteristics of yield variability change are presented. The color bars in the left panels were arranged to match the range of the corresponding prior distributions (see section S2.4). The unit is as follows: $D_Y$, fraction; $T_{\text{max, tr}}$, °C, $T_{\text{min, tr}}$, °C; and $E_a/E_p$, fraction.
Figure S14. Mean and coefficient of variation (CV, in percent) values of the location-specific parameters for wheat across the five different planting dates and 20 different parameter sets. Data at the location where the agro-climatic index could reliably explain the major characteristics of yield variability change are presented. The color bars in the left panels were arranged to match the range of the corresponding prior distributions (see section S2.4). The unit is as follows: $D_y$, fraction; $T_{\text{max, tr}}, ^\circ C$, $T_{\text{min, tr}}, ^\circ C$; and $E_{\text{a}}/E_{\text{p, tr}}$, fraction.