Understanding the weather signal in national crop yield variability

Katja Frieler[1*], Bernhard Schauberger[1], Almut Arneth[2], Juraj Balkovic[3,4], James Chryssanthacopoulos[5], Delphine Deryng[5,6], Joshua Elliott[5,7], Christian Folberth[3], Nikolay Khabarov[3], Christoph Müller[1], Stefan Olin[8], Thomas A. M. Pugh[2,9], Sibyll Schaphoff[1], Jacob Schewe[1], Erwin Schmid[10], Lila Warszawski[1], Anders Levermann[1,11,12]

[1] Potsdam Institute for Climate Impact Research, Potsdam, Germany.

[2] Institute of Meteorology and Climate Research, Atmospheric Environmental Research, Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany.

[3] International Institute for Applied System Analysis, Laxenburg, Austria.

[4] Department of Soil Science, Faculty of Natural Sciences, Comenius University, Bratislava, Slovak Republic.

[5] Center for Climate Systems Research, Columbia University, New York, New York, USA.

[6] Climate Analytics, Berlin, Germany.

[7] ANL Computation Institute, University of Chicago, Chicago, Illinois.

[8] Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden.

[9] School of Geography, Earth and Environmental Sciences and Birmingham Institute of Forest Research, University of Birmingham, Birmingham, UK.

[10] University of Natural Resources and Life Sciences, Vienna, Austria.

[11] Institute of Physics, Potsdam University, Potsdam, Germany.

[12] Lamont-Doherty Earth Observatory, Columbia University, New York, New York.

* Correspondence to Katja Frieler (katja.frieler@pik-potsdam.de)

Key points

- Process-based crop models forced by observational weather data can explain more than 50% of reported yield fluctuations in some major producing countries.
- Water limitation seems to be major driver of the observed variations in most of these countries.

Abstract

Year-to-year variations in crop yields can have major impacts on the livelihoods of subsistence farmers and may trigger significant global price fluctuations, with severe consequences for people in developing

countries. Fluctuations can be induced by weather conditions, management decisions, weeds, diseases, and pests. Although an explicit quantification and deeper understanding of weather-induced crop-yield variability is essential for adaptation strategies, so far it has only been addressed by empirical models. Here we provide conservative estimates of the fraction of reported national yield variabilities that can be attributed to weather by state-of-the-art, process-based crop model simulations. We find that observed weather variations can explain more than 50% of the variability in wheat yields in Australia, Canada, Spain, Hungary, and Romania. For maize, weather sensitivities exceed 50% in seven countries, including the US. The explained variance exceeds 50% for rice in Japan and South Korea and for soy in Argentina. Avoiding water stress by simulating yields assuming full irrigation shows that water limitation is a major driver of the observed variations in most of these countries.

Identifying the mechanisms leading to crop-yield fluctuations is not only fundamental for dampening fluctuations, but is also important in the context of the debate on the attribution of loss and damage to climate change. Since process-based crop models not only account for weather influences on crop yields, but also represent human-management measures, they could become essential tools for differentiating these drivers, and for exploring options to reduce future yield fluctuations.

1 Introduction

Year-to-year variations in crop yields pose a significant risk to subsistence farmers or people depending on local supply. In addition, yield variations on a national level can trigger chain reactions on the global market, leading to large increases in crop prices, such as occurred in 2008. These global fluctuations can, in turn, affect local prices and food supply in developing countries, which are not directly affected by the initial crop failure [*Ivanic and Martin*, 2008; *Headey*, 2011; *Schewe et al.*, 2017].

Annual crop yields depend on several factors. In addition to weather conditions, the occurrence of weeds, diseases and pests can result in yield fluctuations [*Gregory et al.*, 2009]. Management decisions regarding fertilizer use, irrigation, changes in land-use patterns and crop rotations [*Brisson et al.*, 2010] are expected to cause changes not only in the long-term-yield means, but also in year-to-year crop yields. For example, irrigation will help to mitigate yield reductions in particularly dry and hot years, or crop yields may be increased by investments in intensification in response to high crop prices [*Haile et al.*, 2016].

Identifying the main causes of yield variability and designing strategies to minimise them, is essential for reducing the associated risks. Our analysis is mainly dedicated to a quantification of weather-induced yield variability - a component that is expected to increase under climate change [*Challinor et al.*, 2014].

There have been several approaches to attribute long-term trends in historical yield data to climate change based on process-based or statistical models [*Maltais-landry and Lobell*, 1998; *Lobell and Field*, 2007; *Brisson et al.*, 2010; *Lobell et al.*, 2011a]. However, an explicit quantification of the weather-induced variability around the long-term mean change has so far been limited to empirical studies [*Lobell and Field*, 2007; *Ray et al.*, 2015; *Lesk et al.*, 2016]. Here, we introduce historical simulations of eight process-based Global Gridded Crop Models (GGCMs) as an alternative tool. In contrast to simplified empirical models, the process-based models represent implementations of our current

knowledge about crop phenology, in particular accounting for: 1) potentially complex non-linearities in crop responses to weather variables on a daily time scale; 2) variations in crop responses in terms of the phenological state of the crop; 3) interaction between temperature, precipitation and other weather variables, such as radiation, as well as delayed effects due to e.g. water storage in the soil. On the other hand, empirical models allow for the detection of simplified relationships between weather indicators and crop yields that may go beyond our current understanding or implementation of the underlying physiological processes. For example, empirical methods may detect the effects of pests and diseases on crop-yield variability, insofar as they can be linked to weather fluctuations. To this end and for comparison we also consider four different empirical approaches to estimate the faction of variability attributable to weather.

The process understanding embedded in the GGCMs offers additional opportunities to identify the main processes inducing the observed weather sensitivity of crop yields. For example, the sharp drop of US maize yields above a certain temperature threshold found in an empirical analysis [*Schlenker and Roberts*, 2009] was traced back to associated variations in water supply and demand within a process-based model [*Lobell et al.*, 2013]. Here, we use the GGCMs to quantify the reduction in the fraction of the variability that is attributable to weather when assuming full irrigation. The reduction can be taken as a measure of the degree to which observed yield fluctuations are driven by water availability. In this way, we can identify the countries where irrigation has a high potential to reduce crop yield fluctuations. We apply our approach to average national yields. At this scale, adaptation measures can be implemented that are capable of reducing annual yield fluctuations in the form of regulations, recommendations and subsidies (see for example the National Adaptation Strategies of EU member states [*Biesbroek et al.*, 2010], or National Adaptation Plans by developing countries as part of the Cancun Adaptation Framework [*Expert Group*, 2012]).

2 Methods

2.1 Process-based crop model simulations

We analyse an ensemble of crop-model simulations generated within Agricultural Model Intercomparison and Improvement Project AgMIP and the Inter-Sectoral Impact Model Intercompsiron Project, Phase 2a (ISIMIP2a). Six of the eight GGCMs also participated in the ISIMIP Fast Track (EPIC-BOKU [*Williams*, 1995], GEPIC [*Williams et al.*, 1989], LPJ-GUESS [*Lindeskog et al.*, 2013], LPJmL [*Bondeau et al.*, 2007], pDDSAT [*Jones et al.*, 2003], PEGASUS [*Deryng et al.*, 2011, 2014]). EPIC-IIASA and pAPSIM [*Keating et al.*, 2003] later joined the evaluation exercise of the crop-model intercomparison. Models are classified as either 1) site-based crop models (EPIC-BOKU, EPIC-IIASA, GEPIC, pAPSIM, and pDSSAT) originally designed to simulate processes at the field scale, including dynamic interactions between crop, soil, atmosphere, and management components; and 2) agro-ecosystem models (LPJ-GUESS, LPJmL, and PEGASUS), which were originally developed to simulate terrestrial carbon dynamics, surface energy balance, and soil water balance [*Rosenzweig et al.*, 2014].

For the historical simulations considered here, all crop models were forced by the same daily climate-input data (AgMERRA, Ruane *et al* 2015). The data set is a bias-corrected version of re-analysis data (NASA's Modern-Era Retrospective analysis for Research and Applications, MERRA [*Rienecker et al.*, 2011] and NCEP's Climate Forecast System Reanalysis, CFSR [*Saha et al.*, 2010]) on a 0.5 x 0.5 degree grid, where models use information about daily mean, maximum and minimum temperatures, precipitation, solar radiation, relative humidity, and wind speed (see Section 1 of the SI for the model specific input). The bias correction is based on an ensemble of gridded observational data from weather stations and satellites. To test the sensitivity of the maximum variance explained by the GGCMs to the climate input data, we also used GGCM simulations forced by an alternative set of observational climate input data (WFDEI, *Weedon et al.*, 2014) based on the ERA-Interim re-analysis product [*Dee et al.*, 2011]. While most of the model simulations are based on a daily time step, temperature and radiation are downscaled to hourly for all calculations within pDSSAT, and temperature is downscaled to 3-hourly for the determination of phenology within pAPSIM (see Section 1 of the SI). Models are designed to represent the cumulative effect of weather variations across the growing season on annual crop yields. However, they also account for effects of weather conditions before planting via soil water conditions. In many cases, this is achieved by conducting transient simulations, which propagate water deficit or excess through the entire simulation period, or at least from the 3 months prior to planting (see section 1.2 of the SI for more details). Additional details about model-specific representation of CO_2 fertilization, management, and calibration are provided in Section 1 of the SI and in the online documentation (www.isimip.org).

Model performance on the global scale is further evaluated by Müller *et al* 2016. In addition, many of the models have also been evaluated at field-scale trials and have contributed to site-specific model-intercomparison exercises [*Asseng et al.*, 2013, 2014; *Bassu et al.*, 2014]. The ensemble of models considered here has also been shown to reproduce the strong drop observed in reported US crop yields at high temperatures [*Schlenker and Roberts*, 2009; *Schauberger et al.*, 2017].

Given the lack of temporally and spatially-explicit observational data sets, modelling groups used their own default settings for uncertain management choices such as sowing dates, harvesting dates, fertilizer applications rates, and crop varieties and soil conditions, to imitate present day conditions. The model settings were deliberately not harmonized in order to capture the associated uncertainties as present in the multi-model ensemble. Fertiliser application rates were held constant (PEGASUS, pDSSAT, pAPSIM) or were adjusted flexibly according to N stress (EPIC-IIASA, EPIC-BOKU, GEPIC) through time, whilst planting dates decisions were based on daily weather conditions within a fixed (pDSSAT and LPJmL) or dynamic planting window (EPIC-BOKU, LPJ-GUESS, GEPIC, and PEGASUS). Harvesting dates were specified by the choice of cultivars (in terms of heat unit requirement to reach maturity). For PEGASUS, LPJ-GUESS, and to some extent in GEPIC (regarding the selection of winter and spring wheat), the cultivar choice was allowed to vary in time according to changes in temperatures. Cultivars in LPJ-GUESS and GEPIC are only adjusted according to decadal temperature change, and the adjustment in PEGASUS is made in response to inter-annual temperature variability (see SI). All models describe yield responses (for wheat, maize, rice, and soy) to weather conditions (see SI for the list of variables considered by the individual models) within the limitations imposed by global coverage. They do not account for yield reductions due to floods, storms and hail. None of the models is systematically adjusted to reproduce the reported variability of national crop yields [*FAO*, 2013]. In contrast,

management settings were only selected to reproduce average levels of observed yields (but not yield variability) within a reference period. LPJ-GUESS and EPIC-BOKU provide uncalibrated yields assuming minimal nutrient constraints (Table S1 in the SI). All models provide "pure crop runs" where the considered crop is grown everywhere on the global land area where permitted by soil and weather conditions. For each crop, two different simulations were provided: one assuming rain-fed conditions and another assuming full irrigation without accounting for water constraints. To calculate country-specific average yields (t/ha), the associated annual yield patterns were multiplied by fixed land-use and irrigation patterns for the year 2000 (MIRCA2000, Portmann *et al* 2010). To test to what degree the simulated yields are constrained by water supply, we also calculate the yield fluctuations under full irrigation. This was done by multiplying the simulated patterns of crop yields (t/ha) under full irrigation by the total harvested areas (rain-fed + irrigated) for the considered crop, before calculating the national averages.

2.2 Weather indicators used in the empirical models

In addition to the process-based crop models, we also calculated three sets of weather indicators to quantify the fraction of the reported yield variability that can be explained by these indicators. All indicators were calculated based on daily air surface temperature (tas) and precipitation (pr, GPCC), as provided in the AgMERRA data set. We used the fixed sowing and harvesting dates described in the AgMIP protocol [*Elliott et al.*, 2014] to calculate 1) growing season mean values of temperature and precipitation (T, P); 2) mean values of temperature and precipitation over the reproductive growth period, defined as the second half of the growing season (T^{rep}, P^{rep}); 3) the 10th and 90th percentile of the growing season temperatures (T^{min}, T^{max}); and 4) the integral of temperatures below and above a crop specific threshold (Growing and Extreme Degree Days, GDD and EDD see section 2 of the SI for more details). The temporal aggregation was applied at each grid point. Spatial aggregation on the national level was done after application of the MIRCA2000 land-use patterns [*Portmann et al.*, 2010], i.e. national averages include a spatial weighting according to the area where the crop is grown.

2.3 Statistical analysis

To calculate the fraction of the reported variability [*FAO*, 2013] that can be explained by the purely weather-dependent GGCM simulations or climate indicators, all time series, observed (Y_{obs}) and simulated yields (Y_{sim}), as well as the climate indicators, were de-trended at country level (quadratic fit). We use the Δ symbol to refer to remaining fluctuations after de-trending. The reported explained variances are derived from the following statistical models: $\Delta Y_{obs,i} = \alpha_0 + \alpha \, \Delta Y_{sim,i} + \varepsilon_i$, $\Delta Y_{obs,i} = \alpha_0 + \alpha \, \Delta T_i + \beta \, \Delta P_i + \varepsilon_i$ (following Osborne and Wheeler 2013), $\Delta Y_{obs,i} = \alpha_0 + \alpha \, \Delta T_i^{rep} + \beta \, \Delta P_i^{rep} + \varepsilon_i$ (following Iizumi *et al* 2013), $\Delta Y_{obs,i} = \alpha_0 + \alpha \, \Delta T_i^{min} + \beta \, \Delta T_i^{max} + \gamma \, \Delta P_i + \varepsilon_i$ (similar to Lobell and Field 2007), and $\Delta Y_{obs,i} = \alpha_0 + \alpha \, \Delta GDD_i + \beta \, \Delta EDD_i + \gamma \, \Delta P_i + \varepsilon_i$ (following Schlenker and Roberts 2009 and Lobell *et al* 2011a), where i indicates the year and ε describes the residual variations not explained by the models.

Some of the divergence between reported and simulated yields could be due to differences between actual growing season and simulated growing season. Since there is insufficient information about the

temporally-varying growing seasons, they cannot be fully constrained by observations. Therefore, if harvest is close to the end of the year, the correlation between reported and simulated yields might be low if individual yields were reported for different calendar years due to differences in the timing of simulated and actual harvests. On the other hand, some of the divergences may be due to systematic differences in the reporting of the yields. For example, yields might routinely be counted as yields for the previous year if harvest took place at the beginning of a year or vice versa. To test for the sensitivity of the results to these potentially systematic shifts, we also shifted the simulated time series by one year back and forth (see section 3 of the SI).

2.4 Selection of countries

Analysis was done for main producers, which accounted for at least 90% of the global production averaged over 2000-2011. Some countries have been excluded because their time series ended before 2000 or had only sporadic data. For wheat, we compared the explained variances for the following 22 countries: Argentina, Australia, Brazil, Canada, China, Denmark, Egypt, France, Germany, Hungary, India, Iran, Italy, Morocco, Pakistan, Poland, Romania, Spain, Syria, Turkey, UK, and the USA. For maize we included 21 countries: Argentina, Brazil, Canada, China, Egypt, France, Germany, Hungary, India, Indonesia, Italy, Mexico, Nigeria, Philippines, Romania, South Africa, Spain, Thailand, Tanzania, the USA, and Vietnam. For rice we considered 14 countries: Bangladesh, Brazil, China, Egypt, India, Indonesia, Japan, Myanmar, Pakistan, Philippines, South Korea, Thailand, the USA, and Vietnam, and five for soy: Argentina, Brazil, China, India and the USA.

2.5 Testing the sensitivity of annual crop yield variability to underlying long term changes in management

Annual weather-induced crop-yield fluctuations are expected to depend on the underlying management assumptions. To test the sensitivity of the weather-induced yield fluctuations on long-term changes in management, we conducted additional simulations with the GGCM LPJmL, for which model parameters were adjusted to reproduce long-term trends in reported national yield data. LPJmL does not have an explicit representation of the nitrogen cycle. Instead, the Leaf Area Index (LAI), the Harvest Index (HI), and a scaling factor representing the homogeneity and density of the crops in the field, were varied to simulate changing management. In the default experiments, these parameters were only adjusted once to reproduce mean national yield levels in the reference period 1996-2000. For the sensitivity experiment, the parameters were repeatedly adjusted to reproduce decadal average yields at country level. To generate a continuous time series of historical yields under changing management, the parameters for each decade were interpolated to annual time steps. The simulated time series, based on these temporally-varying parameters, were de-trended and correlated to the reported yields analogously to the default "fixed management" case (see section 9 of the SI).

In a second sensitivity experiment, we varied the irrigation fraction according to the temporally resolved "fraction of the cultivated area equipped for irrigation" reported by the FAO, while the irrigated land was kept constant in the default analysis (see section 10 of the SI). In a third experiment we also accounted for historical expansion or shrinkage of agricultural land based on HYDE3 [*Klein Goldewijk and van Drecht*, 2006] (see section 11 of the SI). We also tested the sensitivity to a flexibilisation of the planting dates by comparing the explained variances of the default LPJmL configuration (based on fixed

sowing dates) to the explained variances (based on simulations allowing for grid-cell-specific adjustments) (see section 12 of the SI).

2.6 Testing the sensitivity of simulated crop yield variability to spatial resolution of climate input data

The climate data used as input for the crop-model simulations (AgMERRA, Ruane *et al* 2015) do not capture weather fluctuations within the 0.5°x0.5° grid cells. To our knowledge, there is no higher-resolution observational data set with global coverage that provides all the variables needed as input for the GGCMs. However, for Europe there is a high-resolution climate data set generated by dynamically downscaling observation-based re-analysis data [*Lucas-Picher et al.*, 2013]. To test the sensitivity of the results to the spatial resolution of the climate input data, we compared national yields simulated by LPJmL when forced by the high-resolution (12 km) climate data, to time series of national yields based on model simulations forced by a low-resolution version of the same data. The climate data were remapped to a 0.5°x0.5° grid by: 1) linear interpolation, and 2) a conservative remapping that computes the weighted sum of all contributing source cells (by contributing area) for each target cell (see section 6 of the SI).

3 Results

3.1 Weather variations can explain more than half of the observed variability in crop yields in individual major-producing countries

Individual process-based models driven by weather variations produce yield fluctuations ΔY_{sim} that closely reproduce the observed yield variability ΔY_{obs} [*FAO*, 2013] (Fig. 1). Overall, these models explain more than half of the year-to-year variability in 15 countries (Fig. 2): Australia, Canada, Spain, Hungary, and Romania for wheat; South Africa, Romania, France, the US, Hungary, Germany, and Italy for maize; Japan and South Korea for rice; Argentina for soy (see world maps of the maximum explained variances per country in section 13 of the SI). For most countries, the maximum explained variances derived from the WFDEI-forced model simulations are similar to those derived from the AgMERRA forcing (see Figure S7 of the SI).

The explained variances are a conservative estimate of the weather-induced variability of actual yields, since several factors are expected to reduce the correlation between observed and simulated yield variations: i) the potential within-grid-cell structure of relevant weather events not resolved in the climate input data set (0.5 x 0.5 degree resolution); ii) deficiencies of the crop models in reproducing yield responses such as e.g. underestimated sensitivities to extreme heat [*Lobell et al.*, 2012]; iii) potential influence of floods, storm, hail or weather induced pests etc.; iv) short-term weather induced management adjustments not included in the models (e.g. variations in planting and harvesting dates beyond the adjustments resolved in the models (see Table S1 of the SI)); and v) reporting errors in the

FAO time series. Therefore, low explained variances do not necessarily imply a low sensitivity, but high explained variances mean that there a process-based crop model can plausibly explain the reported crop-yield fluctuations using weather fluctuations.

No single model provides the highest explained variance in all regions. Simulation teams were asked to run their models assuming their model's "best guess" representation of "present-day conditions" without any harmonization, except for the climate input data (see section 1 of the SI). Therefore, simulations cover a broad range of model choices with regard to uncertain inputs and processes such as fertilizer application rates, crop varieties, representation of soil characteristics, biological processes, and plant parameters, which may lead to model-specific strengths in some regions and poorer performance in others. For each country, the different model settings represent a range of potential real-world management conditions. In this setting, the time series of the multi-model means calculated for each year provides a relatively high explanatory power (Fig. 2, black line) because model-specific biases can cancel out. While countries in Fig. 2 are ordered according to the highest explained variance provided by individual models, deriving the explained variances from the multi-model mean time series does not change this order significantly. The general conclusion is not affected by the temporal shifting of the time series to account for potentially systematic differences in the timing of the simulated and actual harvests or the associated reporting. The temporal shifting of the simulated time series by one year back and forth implies correlations increases by more than 0.2 for only a very limited number of crop models and the correlation between the time series of the multi-model mean yields and the reported data is rarely affected (see Fig. S3 of the SI). Sub-grid-scale weather events do not seem to have a major influence on the national yield statistics (see section 6 of the SI for a sensitivity experiment based on higher resolution data for Europe). Results are not strongly affected by the considered modifications in long-term management (see Fig. S10 - S14 of the SI). Furthermore, maximum explained variances do not seem to be related to the crop-specific harvested areas within a country and are only to a limited extent linked to the standard deviations of the reported crop yields (Fig. S5 and S6 of the SI, respectively).

3.2 Individual process-based models explain more of the observed variability than the considered empirical approaches

As a comparison to process-based crop yield simulations, we also provide results from four empirical approaches. One of these empirical approaches addresses the issue of non-linear crop responses to temperature fluctuations by integrating temperatures below and above a certain threshold over time, and allowing for differential responses [*Schlenker and Roberts*, 2009; *Lobell et al.*, 2011b] (see Methods and SI). The empirical approaches are restricted to the derivation of highly-simplified relationships between yields and spatially and temporally-aggregated weather indicators, but could indirectly account for the effects of flood, storms, hail or weather-related occurrence of weeds, diseases or pests, in so far as they are correlated with weather indicators. The highest fraction of weather-induced variability derived from the four empirical models (Fig. 2, red diamonds) is generally lower than the highest explained variance provided by the process-based models, except for soy yields. This is remarkable and underlines the good performance of the process-based models for the following reason: the regression models include two to three predictors and associated fit parameters (see section 2.2.and 2.3), compared to one predictor ΔY_{sim}

and the associated fit parameter used in the regression based on the process-based crop models' simulations. This higher number of fitting parameters affords a purely statistical "advantage" when estimating the explained variances. This artifact most likely explains the generally superior performance of the statistical models based on three indicators compared to the models using only two climate indicators (Fig. S4 in the SI), but does not seem to be sufficient to outperform the process-based models except for soy.

3.3 Water stress appears to be a major driver of crop yield fluctuations in most considered countries

With regard to adaptation options, it is important to identify the critical drivers of yield variations. GGCMs offer alternative ways of looking into this question compared to simplified statistical models [*Lobell et al.*, 2013]. In particular, they allow for an artificial reduction of water stress by assuming optimal irrigation at each grid point. Here we calculate the explained variances assuming full irrigation for the model providing the highest variance in Fig. 2. The reduction in water stress reduces the explained variances to 20% or below for all countries where the assumed present-day irrigation fraction is below 40%, except for maize in the US and rice in Brazil (see Fig. 3).

For wheat in Australia, Spain, Iran and Morocco and maize in South Africa, Romania, France, Hungary, Argentina, and China there is no positive correlation. This supports the hypothesis that water stress is a key driver of the observed yield variations, or at least mediates the effects of temperature fluctuations in a way that can be strongly reduced by irrigation [*Lobell et al.*, 2013].

The simulated yield variability is often higher than the observed variability (see Fig. 4). On the one hand, a smaller variability of the observational data could be explained by ad hoc interventions of farmers reacting to climate extremes, and a generally larger diversity of management decisions, making the real world conditions more resilient to weather fluctuations than the associated model representations. On the other hand, models providing a higher standard deviation may also overestimate the yield responses to weather fluctuations. The overall standard deviation of the simulated yield fluctuations can be significantly reduced by irrigation in individual countries (see Fig. S8 and S9 of the SI).

4 Discussion

The identification of the main sources of variability is critical when thinking about measures to minimize overall yield fluctuations. Here, we estimate to what degree reported national crop-yield fluctuations can be explained by our current understanding of the complex relationship between weather

and yields, as implemented in process-based GGCMs. We show that individual GGCMs when solely accounting for weather fluctuations can explain more than 50% of the reported variability of yields in individual countries. Water stress is identified as the main driver of historical reported yield fluctuations by showing that, in most of the countries, the variance of the reported yields that is explained by the simulations is strongly reduced when water stress is avoided in the simulations.

An alternative study estimating the weather contribution to reported crop-yield variability on the global scale, used empirical models to come to the conclusion that there is a significant influence of weather on crop yields on about 70% of maize harvesting regions, 53% of rice harvesting regions, 79% of wheat harvesting regions and 67% of soybean harvesting regions [*Ray et al.*, 2015]. Averaged over these regions they found that about one third of the reported crop-yield variability was induced by weather for each of the four crops, where regionally their numbers also exceeded 75%. Since their analysis has been carried out on the finer spatial scale of 13,500 sub-national political units, instead of the national level used here, their estimates cannot be directly compared to our numbers. Even a spatial averaging of their explained variances across countries is not expected to provide the explained variances that could be derived by directly applying the statistical model on the national scale, due to additional aggregation effects [*Gornott and Wechsung*, 2016]. That is why we independently fitted different empirical models to the national yield statistics in our study.

To improve our understanding of the effect of weather influences on crop-yield variability, and to reduce the inter-model spread in explained variances, studies would benefit substantially from detailed temporally and spatially-resolved information about human management, such as fertilizer input, tillage, sowing and harvesting dates, and cultivar selection. Better knowledge of these practices would allow for a better adjustment of process based models to regional conditions [*Asseng et al.*, 2013]. In light of the current lack of detailed information about regional management and soil characteristics, the considered default simulations, which cover a wide range of different implementations, could potentially be used to inversely derive a proxy for these conditions. This may be done by comparing simulated yields to observed yields and, for example, choosing country-specific settings based on the model that provides the highest agreement with reported crop yields. Such an approach will have to be carefully evaluated. For example, whether or not the derived input data also explain a higher variance across other GGCMs would need to be tested, and data will have to be adjusted in light of new reported data.

Temporally-resolved information would even allow for testing to what degree the remaining unexplained variance can be resolved by these processes. The same applies to a reporting of the occurrence of diseases, pests and weeds that would allow for explicit estimation of the influence of these drivers, which is still hidden within the unexplained variances reported in this study. Ideally, changes in crop yields from one year to the next could be fully explained by process-based models, which account for management changes (and the occurrence in diseases, pests and weeds) and weather fluctuations. In this case, weather-induced effects could be estimated from the difference between year-to-year changes in crop yields, which are derived from model runs accounting for the combination of all effects, together with a run where weather conditions are held constant from year to year. In our sensitivity experiment, potential long-term variations in LAI and HI, as well as illustrative variations in crop densities or reported increases in irrigation areas, show only minor effects on the correlation between the reported

and simulated time series. Thus, this particular representation of potential management changes does not interact strongly with yield responses to weather fluctuations. However, the considered management influences cover by no means the full range of potential management-induced changes of crop yield responses to weather fluctuations. More research has to be done to get a better idea of the interaction between management changes and weather-induced crop-yield fluctuations. That is particularly important for dampening crop-yield fluctuations.

Given the limited information about regional management conditions and uncertainties in the representation of crop-model responses to weather fluctuations, the multi-model mean seems to be better suited to quantifying weather-induced variabilities than a single model. The multi-model ensemble provides an implicit sampling of: 1) possible real-world conditions such as temporally and spatially-resolved information about sowing and harvesting dates or fertilizer application rates; and 2) different implementation of processes across the models [*Asseng et al.*, 2013] leading to errors cancelling out. While the "ensemble" mean does not offer a physiologically-coherent explanation of observed crop-yield fluctuations as provided by individual crop models, ensemble means could still be relevant for projecting short-term national yield fluctuations across different regions if associated weather predictions were available. In addition, the analysis of the multi-crop model ensemble reveals a large potential for learning from each other by e.g. identifying the differences in model-specific implementations of national management conditions.

Apart from reporting detailed management conditions on the regional level, our understanding of the weather signal in reported crop-yields fluctuations could significantly benefit from a separate reporting of irrigated and rain-fed crop yields. These data would allow for direct quantification of the fraction of the observed yield variability of irrigated crops that is still due to weather. Except for rice, current national crop-yield variability is dominated by yields under rain-fed conditions, with little possibility of separating the responses of irrigated crops.

References

Asseng, S. et al. (2013), Uncertainty in simulating wheat yields under climate change, *Nat. Clim. Chang.*, *3*, 827–832.

Asseng, S. et al. (2014), Rising temperatures reduce global wheat production, *Nat. Clim. Chang.*, *5*(2), 143–147, doi:10.1038/nclimate2470.

Bassu, S. et al. (2014), How do various maize crop models vary in their responses to climate change factors?, *Glob. Chang. Biol.*, *20*(7), 2301–2320, doi:10.1111/gcb.12520.

Biesbroek, G. R., R. J. Swart, T. R. Carter, C. Cowan, T. Henrichs, H. Mela, M. D. Morecroft, and D. Rey (2010), Europe adapts to climate change: Comparing National Adaptation Strategies, *Glob. Environ. Chang.*, *20*(3), 440–450, doi:10.1016/j.gloenvcha.2010.03.005.

Bondeau, A. et al. (2007), Modelling the role of agriculture for the 20th century global terrestrial carbon balance, *Glob. Chang. Biol.*, *13*(3), 679–706, doi:10.1111/j.1365-2486.2006.01305.x.

Brisson, N., P. Gate, D. Gouache, Charmet, Gilles, F.-X. Oury, and F. Huard (2010), Why are wheat yields stagnating in Europe? A comprehensive data analysis for France, *F. Crop. Res.*, *119*(1), 201–212, doi:10.1016/j.fcr.2010.07.012.

Challinor, A., K. Cochrane, M. Howden, M. M. Iqbal, D. Lobell, and M. I. Travasso (2014), Chapter 7. Food Security and Food Production Systems, in *Climate Change 2013: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, USA.

Dee, D. P. et al. (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*, 553–597, doi:10.1002/qj.828.

Deryng, D., W. J. Sacks, C. C. Barford, and N. Ramankutty (2011), Simulating the effects of climate and agricultural management practices on global crop yield, *Global Biogeochem. Cycles*, *25*(2), 1–18, doi:10.1029/2009GB003765.

Deryng, D., D. Conway, N. Ramankutty, J. Price, and R. Warren (2014), Global crop yield response to extreme heat stress under multiple climate change futures, *Environ. Res. Lett.*, *9*(3), 034011, doi:10.1088/1748-9326/9/3/034011.

Elliott, J. et al. (2014), The Global Gridded Crop Model intercomparison: data and modeling protocols for Phase 1 (v1.0), *Geosci. Model Dev. Discuss.*, *7*(4), 4383–4427, doi:10.5194/gmdd-7-4383-2014.

Expert Group, L. D. C. (2012), *National Adapation Plans - Technical guidelines for the national adaptation plan process*, Bonn, Germany.

FAO (2013), Crop production database - Food and Agriculture Organization of the United Nations (FAO)., Available from: http://faostat3.fao.org

Gornott, C., and F. Wechsung (2016), Statistical regression models for assessing climate impacts on crop yields: A validation study for winter wheat and silage maize in Germany, *Agric. For. Meteorol.*, *217*, 89–100.

Gregory, P. J., S. N. Johnson, A. C. Newton, and J. S. I. Ingram (2009), Integrating pests and pathogens into the climate change/food security debate., *J. Exp. Bot.*, *60*(10), 2827–38, doi:10.1093/jxb/erp080.

Haile, M. G., M. Kalkuhl, and J. Von Braun (2016), Worldwide Acreage and Yield Response to International Price Change and Volatility: A Dynamic Panel Data Analysis for Wheat, Rice, Corn, and Soybeans, *Am. J. Agric. Econ.*, *98*(1), 172–190, doi:10.1093/ajae/aav013.

Headey, D. (2011), Rethinking the global food crisis: The role of trade shocks, *Food Policy*, *36*(2), 136–146, doi:http://dx.doi.org/10.1016/j.foodpol.2010.10.003.

Iizumi, T., H. Sakuma, M. Yokozawa, J.-J. Luo, A. J. Challinor, M. E. Brown, G. Sakurai, and T. Yamagata (2013), Prediction of seasonal climate-induced variations in global food production, *Nat. Clim. Chang.*, *3*(7), 1–5, doi:10.1038/nclimate1945.

Ivanic, M., and W. Martin (2008), Implications of higher global food prices for poverty in low-income countries, *Agric. Econ.*, *39*, 405–416, doi:10.1111/j.1574-0862.2008.00347.x.

Jones, J. ., G. Hoogenboom, C. . Porter, K. . Boote, W. . Batchelor, L. . Hunt, P. . Wilkens, U. Singh, a. . Gijsman, and J. . Ritchie (2003), *The DSSAT cropping system model*.

Keating, B. . et al. (2003), An overview of APSIM, a model designed for farming systems simulation, *Eur. J. Agron.*, *18*(3-4), 267–288, doi:10.1016/S1161-0301(02)00108-9.

Klein Goldewijk, K., and G. van Drecht (2006), HYDE 3: current and historical population and land cover. Integrated modelling of global environmental change. An overview of IMAGE 2.4, edited by A. F. Bouwman, T. Kram, and K. K. Goldewijk, pp. 93–111, Netherlands Environmental Assessment Agency, Bilthoven, The Netherlands.

Lesk, C., P. Rowhani, and N. Ramankutty (2016), Influence of extreme weather disasters on global crop production, *Nature*, *529*(7584), 84–87, doi:10.1038/nature16467.

Lindeskog, M., A. Arneth, A. Bondeau, K. Waha, G. Schurgers, S. Olin, and B. Smith (2013), Effects of crop phenology and management on the terrestrial carbon cycle: case study for Africa, *Earth Syst. Dynam. Discuss.*, *4*(2), 235–278.

Lobell, D. B., and C. B. Field (2007), Global scale climate–crop yield relationships and the impacts of recent warming, *Environ. Res. Lett.*, *2*(1), 014002, doi:10.1088/1748-9326/2/1/014002.

Lobell, D. B., W. Schlenker, and J. Costa-Roberts (2011a), Climate trends and global crop production since 1980., *Science*, *333*(6042), 616–20, doi:10.1126/science.1204531.

Lobell, D. B., M. Bänziger, C. Magorokosho, and B. Vivek (2011b), Nonlinear heat effects on African maize as evidenced by historical yield trials, *Nat. Clim. Chang.*, *1*(1), 42–45, doi:10.1038/nclimate1043.

Lobell, D. B., A. Sibley, and J. Ivan Ortiz-Monasterio (2012), Extreme heat effects on wheat senescence in India, *Nat. Clim. Chang.*, *2*(3), 186–189, doi:10.1038/nclimate1356.

Lobell, D. B., G. L. Hammer, G. McLean, C. Messina, M. J. Roberts, and W. Schlenker (2013), The critical role of extreme heat for maize production in the United States, *Nat. Clim. Chang.*, *3*(5), 497–501, doi:10.1038/nclimate1832.

Lucas-Picher, P., F. Boberg, J. H. Christensen, and P. Berg (2013), Dynamical Downscaling with Reinitializations: A Method to Generate Finescale Climate Datasets Suitable for Impact Studies, *J. Hydrometeorol.*, *14*(4), 1159–1174, doi:10.1175/JHM-D-12-063.1.

Maltais-landry, G., and D. B. Lobell (1998), Evaluating the Contribution of Weather to Maize and Wheat Yield Trends in 12 U.S. Counties, , 301–311, doi:10.2134/agronj2011.0220.

Müller, C. et al. (2016), Global Gridded Crop Model evaluation: benchmarking, skills, deficiencies and implications, *Geosci. Model Dev. Discuss.*, doi:10.5194/gmd-2016-207.

Osborne, T. M., and T. R. Wheeler (2013), Evidence for a climate signal in trends of global crop yield variability over the past 50 years, *Environ. Res. Lett.*, *8*(2), 024001, doi:10.1088/1748-9326/8/2/024001.

Portmann, F. T., S. Siebert, and P. Döll (2010), MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling, *Global Biogeochem. Cycles*, *24*(1), GB1011, doi:10.1029/2008GB003435.

Ray, D. K., J. S. Gerber, G. K. MacDonald, and P. C. West (2015), Climate variation explains a third of global crop yield variability, *Nat. Commun.*, *6*, 5989, doi:10.1038/ncomms6989.

Rienecker, M. M. et al. (2011), MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, *J. Clim.*, *24*, 3624–3648, doi:10.1175/JCLI-D-11-00015.1.

Rosenzweig, C. et al. (2014), Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison, *Proc. Natl. Acad. Sci.*, (14), 1–6, doi:10.1073/pnas.1222463110.

Ruane, A. C., R. Goldberg, and J. Chryssanthacopoulos (2015), Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation, *Agric. For. Meteorol.*, *200*, 233–248, doi:10.1016/j.agrformet.2014.09.016.

Saha, S. et al. (2010), The NCEP Climate Forecast System Reanalysis, *Bull. Amer. Meteorol. Soc.*, *91*, 1015–1057, doi:10.1175/2010BAMS3001.1.

Schauberger, B. et al. (2017), Consistent negative response of US crops to high temperatures in observations and crop models, *Nat. Commun.*, *8*(13931), doi:10.1038/ncomms13931.

Schewe, J., C. Otto, and K. Frieler (2017), The role of storage dynamics in annual wheat prices, *Environ. Res. Lett.*

Schlenker, W., and M. J. Roberts (2009), Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change., *Proc. Natl. Acad. Sci. U. S. A.*, *106*(37), 15594–8.

Weedon, G. P., G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo (2014), The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, *Water Resour. Res.*, *50*, 7505–7514, doi:10.1002/2014WR015638.

Williams, J. R. (1995), The EPIC Model. In: V.P. Singh (eds). Computer Models of Watershed Hydrology., *Water Resour. Publ. Highl. Ranch, Color.*, 909–1000.

Williams, J. R., C. A. Jones, J. R. Kiniry, and D. A. Spanel (1989), The EPIC crop growth model, *Trans ASAE*, *32*, 497–511.
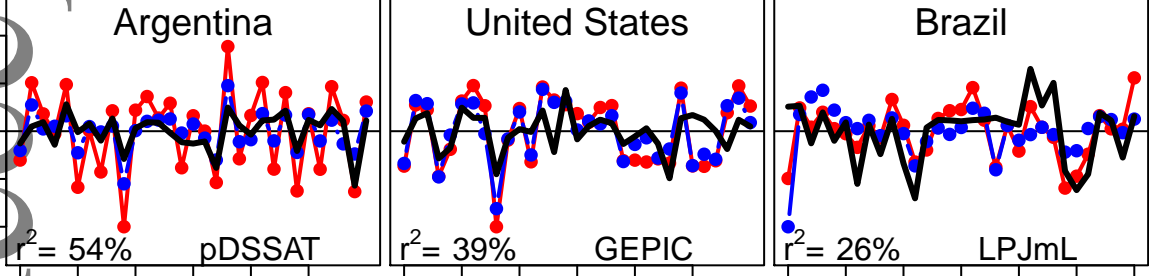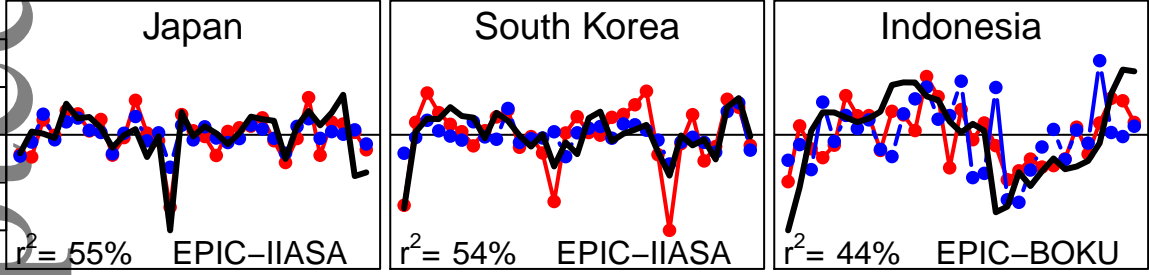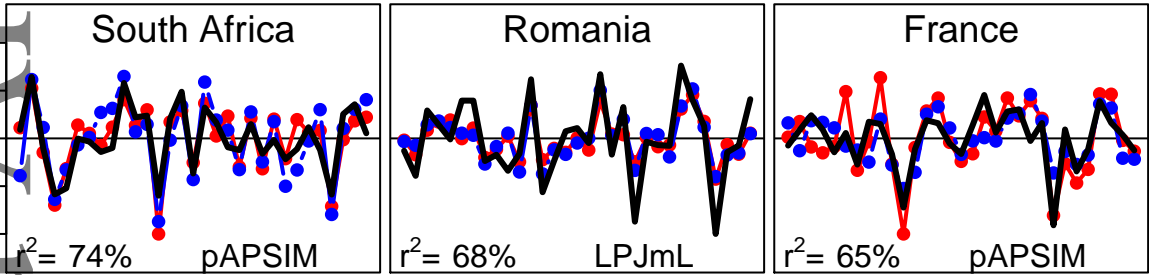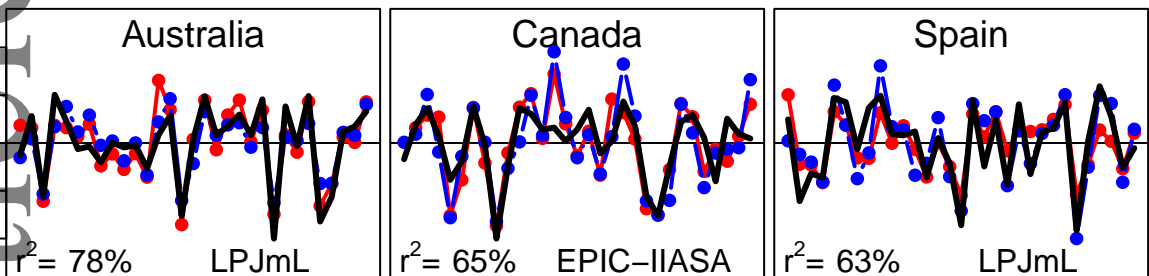
Figure 1: Reproduction of recorded annual yield variations by global gridded crop models. Black curves: De-trended recorded time series of national yields. Red curves: De-trended simulated yield variations ΔY_{sim} for the process-based crop model that explains the highest country-specific fraction of the variance of the FAO time series (model name and r^2 given in lower part of the panel). Blue curves: De-trended time series of multi-model mean yields. Results are shown for the three countries with the highest explained variance provided by an individual process-based model for each crop. Simulated and reported time series were scaled by the maximum of the absolute values of all three time series.

Figure 2: Quantification of lower bounds of weather-induced yield variability. Colored symbols: Fraction of the variance of the reported year-to-year variability that is explained by individual crop models. Red diamonds: Highest variance explained by simple climate indicators (= CIs; see Fig S4 for the individual results of the three applied regression approaches). Black diamonds: Variance of reported yield fluctuations explained by the multi-model mean of simulated yields. Countries are ordered according to the highest variance explained by individual crop models. Grey polygons show the range from zero to the highest explained variance provided by the process-based crop models for each country.

Figure 3: Similar to Fig 2, but only showing the highest explained variances per country (colored bars identical to Fig 2) and the explained variances of the same model but assuming full irrigation (colored circles). Numbers indicate the "present-day" irrigation fractions (MIRCA2000) underlying the "default" simulations represented by the colored bars. Explained variances with full irrigation are not shown if the underlying correlation is negative.

Figure 4: Comparison of the magnitude of observed and simulated yield fluctuations. Light blue columns: Standard deviation of the de-trended observational time series (FAO). Colored bars: Standard deviation of the de-trended simulated time series. Black bars: standard deviation of the de-trended time series of the multi-model mean. We only show model results that provide an explained variance > 10%.

Normalized ΔY

Australia — r^2= 78% — LPJmL	Canada — r^2= 65% — EPIC−IIASA	Spain — r^2= 63% — LPJmL
South Africa — r^2= 74% — pAPSIM	Romania — r^2= 68% — LPJmL	France — r^2= 65% — pAPSIM
Japan — r^2= 55% — EPIC−IIASA	South Korea — r^2= 54% — EPIC−IIASA	Indonesia — r^2= 44% — EPIC−BOKU
Argentina — r^2= 54% — pDSSAT	United States — r^2= 39% — GEPIC	Brazil — r^2= 26% — LPJmL

Wheat

Maize

Rice

Soy

Years

Wheat

Maize

Rice

Soy

Standard Deviations (t/ha)

EPIC−BOKU
EPIC−IIASA
GEPIC
LPJ−GUESS
LPJmL
pAPSIM
pDSSAT
PEGASUS
Observed
Ens. mean

Australia, Canada, Spain, Hungary, Romania, Syria, Iran, Poland, Morocco, USA, S. Africa, Romania, France, USA, Hungary, Germany, Italy, Argentina, China, Canada

Japan, S. Korea, Indonesia, Brazil, USA, China, Egypt, Vietnam, Bangladesh, India, Argentina, USA, Brazil, China, India