Title: Estimating the effects of extreme weather on food security

Authors: Erik Chavez[1]*, Gordon Conway[2], Michael Ghil[3,4], Marc Sadler[5]

Affiliations:

[1] Imperial College London, Centre for Environmental Policy and Imperial College Business School, Finance Department

[2] Imperial College London, Agriculture for Impact, Centre for Environmental Policy

[3] Ecole Normale Supérieure, Paris, Geosciences Department and Environmental Research and Teaching Institute

[4] University of California, Los Angeles, Department of Atmospheric and Oceanic Sciences and Institute of Geophysics and Planetary Physics

[5] The World Bank, Agriculture and Environmental Services Department, Risk and Markets Practice

* Correspondence and requests for materials should be addressed to erik.chavez07@imperial.ac.uk

Text:

Initial paragraph: 223words

Text: 1856 words

Total: 2,079

Both governments and the private sector urgently require better estimates of the likely incidence of extreme weather events[1], their impacts on food crop production and the potential consequent social and economic losses[2]. Current assessments of climate change impacts on agriculture mostly focus on average crop yield vulnerability[3] to climate and adaptation scenarios[4,5]. Also, although new-generation climate models have improved and there has been an exponential increase in available data[6], the uncertainties in their projections over years and decades and at regional and local scale, have not decreased[7,8]. We need to understand and quantify the non-stationary, annual and decadal climate impacts using simple and communicable risk metrics[9] that will help public and private stakeholders manage the hazards to food security. Here we present an 'end-to-end' methodological construct based on weather indices and machine learning that integrates current understanding of the various interacting systems of climate, crops, and the economy to determine short to long-term risk estimates of crop production loss, in different climate and adaptation scenarios. For provinces north and south of the Yangtze River in China, we have found that risk profiles for crop yields that translate climate into economic variability follow marked regional patterns, shaped by drivers of continental-scale climate. We conclude that to be cost-effective, region-specific policies have to be tailored to optimally combine different categories of risk management instruments.

An increasing body of scientific evidence, derived from both observations and model simulations, indicates that the climate system never was, nor is it likely to ever be, statistically stationary[10]. Moreover, statistical characterization of slowly changing weather extremes is fraught with difficulties[11]. These stem partly from the potentially large effects caused by lack of stationarity and partly from the existence of complex nonlinear processes and threshold effects. The assessment and the prediction of such effects, both deterministic and stochastic, on weather extremes depend on a number of interconnected drivers. For example, changes in weather variability season-to-season and year-to-year that affects food production derive from shifts in the statistics of decade-to-decade climate processes[12,13]. Thus, changes in the large-scale climate processes that drive both regional and global climate variability affect the annual onset of rainfall in the tropics and subtropics, as well as rainfall patterns in temperate latitudes, so playing a significant role in the variability of regional rain-fed crop production[14]. The risk estimation methodology proposed here integrates large- and small-scale information, and is based on both observed and simulated data for weather, climate, crop vulnerability and economic conditions.

The overall, end-to-end methodological construct is illustrated in Fig. 1. It relies on machine learning involving weather indices that characterize the vulnerability of crops to weather variability in different technological scenarios (Fig 1a).

Figure 1 near here

We here used a stochastic "weather-within-climate" downscaling approach that quantifies the interaction of low- and high-frequency climate variability (Fig. 1b) to determine the crop loss, risk profiles (Fig. 1d) for future climate scenarios. These are then used to model the direct and indirect economic impacts subject to supply loss shock (Fig. 1e) and to determine optimum mix of risk transfer and mitigation policies in a particular region or country (Fig. 1f). We assessed the potential of this methodological construct by using data for weather, crops, and the economy in four provinces (Shandong, Hebei, Guangdong, and Guangxi) of the People's Republic of China, north and south of the Yangtze River.

Existing Integrated Assessment Models (IAMs) have attempted to provide first estimates of future possible costs of climate impacts on the economy subject to different global warming scenarios[15,16]. However, the sensitivity of these IAMs to individual economic parameters, such as the discount rate, has limited their usefulness. Taking this into account, the methodology presented in Fig. 1 focuses on the economic impacts driven by the local and regional characteristics of weather variability and climate state changes, the local response of the system considered (e.g. the crop production sector), and different scenarios of technological risk mitigation.

Weather indices were devised as proxies of physical crop response to two of the main drivers of yield variability: precipitation variability and exposure to excess-temperatures. Other hazards such as cold shocks or radiation variability are not considered here. Observed historical daily weather data and soil databases for the studied provinces are used to simulate crop yields using mechanistic crop modelling. Daily precipitation and temperature data are used to build pixel-level databases of precipitation and temperature variability indices. Each index captures exposure to deficit precipitation or excess temperature during different time intervals of crop growth.

The translation of the metrics of physical-loss risk into metrics of direct and indirect economic loss is carried out through macroeconomic modelling of exogenous, supply-side shocks. Probabilistic and scenario-based risk modelling is cascaded from climate to agricultural and finally economic loss through data clustering, by using machine learning techniques of recursive partitioning[17] and Nonhomogeneous Hidden Markov Models[18] (NHMMs), as illustrated in Supplementary Fig. 1. The joint effects of precipitation variability and excess temperature were modelled through stochastic-copula dependency; see **Methods** and Supplementary Fig. 2. Finally, complete province-level profiles of economic-loss risk were obtained by considering several technological scenarios for climate risk mitigation. While a historical climate scenario is presented here, the same methodological construct is applicable to obtain risk profiles in future climate scenarios by using simulated large scale climate driver NHMM covariates.

Vulnerability of crops to weather variability varies strongly over their growing period. The length of this period and of the occurrence of stages of development such as flowering and maturity is also constrained by local weather variability and environmental conditions, as well as by genetic traits. In addition to extreme weather events, slight changes in planting season and duration of weather patterns may also reduce yields[19]. The weather indices are used to capture the response of crop growth to different features of weather variability. Excess heat indices are built by counting the number of days where the maximum temperature, Tmax, surpasses a critical threshold, Tc, of 30 or 35°C – for instance the number of days with Tc > 30°C from day 10 to day 40 of crop development. Precipitation deficit indices account for cumulative rainfall during a given period of crop growth. **Supplementary Figure 3** summarizes the different periods of aggregation of weather indices and the colour code used in **Figure 2**.

The machine learning methodology applied here to select pixel-level weather indices shows that the weather indices which best capture weather-driven yield variability exhibit spatial heterogeneity relative to the portion of the growing cycle accounted in the index. For instance, the optimal indices for the effects of precipitation variability (Fig. 2a) and excess heat (Fig. 2b) on maize yield variability in the northeastern province of Shandong are heterogeneous, with several pixels spatially clustered according to different periods of the growing season (Supplementary Fig. 3) during which the crop is most sensitive to climatic effects. The spatial clustering of indices appears to follow topographical features of Shandong province. For instance, the central mountainous and the westernmost regions of the province are dominated by precipitation indices capturing vulnerability during, respectively, the middle and the end of the crop development. This spatial pattern of precipitation indices also depends on the technological scenario considered (i.e. local rain-fed variety, local irrigated variety, switched rain-fed variety), as shown in Supplementary Fig. 4. In contrast, a marked index spatial homogeneity is observed regarding the choice of critical temperature used to build heat wave indices. For each pixels, two sets of 25 heat wave indices using 30 or 35°C as critical temperature was used to determine the optimum heat wave index. 30°C appears is homogeneously selected across all Shandong province (Figure not shown).

Figure 2 near here

Heat wave–driven variability in rice yield in the Southern provinces of Guangxi and Guangdong possesses similar spatial variability; see Supplementary Figs. 5a,b. Estimated impacts of weather variability and climate change on crop production are usually based in IAMs which implies spatially homogenous hydrometeorological indicators[20]. Doing so is likely to underestimate local-to-regional yield losses. In effect, the rate of succession of phenological growth stages in crops depends on the accumulation of temporal photo-thermal units[19]; this accumulation, in turn, depends on the interaction of local environmental variables. Therefore, the use of homogenous hydrometeorological indicators may fail to systematically capture times of peak vulnerability, e.g., during reproductive stages that vary with location.

Results obtained for northern Shandong (Fig. 2) and Hebei (not shown) provinces illustrate the importance of modelling the joint impacts of precipitation variability and excess temperature stresses on rain-fed crops. Under the baseline scenario of the currently grown, rain-fed maize variety, average yield variation throughout Shandong province, subject to the stress of precipitation variability alone, produces slightly positive yield anomalies, while the joint modelling of excess temperature and precipitation variability leads to spatially homogenous negative anomalies. Supplementary Figure 4 illustrates the latter.

The nonlinearity of maize yield losses due to drought and heat stress is captured by our modelling and is consistent with agricultural field studies[22,23]. The relatively homogenous yield losses for irrigated rice subject to increasing heat wave exposure throughout the southern Guangdong (not shown) and Guangxi provinces in Supplementary Fig. 5 are consistent with existing literature[24,25] and might actually be underestimated[26].

Figure 3 near here

The results demonstrate that important variations in province-level risk profiles depend on the regional features of weather and climate variability.

To capture dependence on large-scale, low-frequency climate variability, we have constructed and applied an NHMM[18]; see **Methods** and Supplementary Fig. 6. In the northeastern provinces of Shandong (Fig. 3) and Hebei (not shown), the effect of low-frequency climate change, modelled by this NHMM, is masked by high-frequency weather variability. In fact, northeastern China is strongly affected by mid-latitude weather systems, as well as by teleconnections from the Tropical Pacific[27,28].

In contrast, for the southern Guangdong and Guangxi provinces, risk driven by weather variability depends strongly on the climate state. For a given state, the risk profiles in the southern provinces exhibit minimum variation for varying return periods of weather events, whereas drastic jumps, of 0.18 % and 1.15 % in losses of provincial gross domestic product (GDP) occur in Guangdong and Guangxi, respectively, as central-Eastern Pacific sea surface temperatures shift from a warm to a cold event, as captured by the Niño-3.4 index in our NHMM[29] and illustrated in Supplementary Figs. 6 and 7.

We have considered three different technological scenarios: (i) continuing use of a local rain-fed variety; (ii) switching to another, more drought tolerant rain-fed variety; and (iii) the use of a local irrigated variety. Their effects on the risk profiles are illustrated in Fig. 3a and Supplementary Fig. 4.

The probabilistic risk profiles of economic loss obtained by the present methodology are strongly driven by the physical-loss risk. But the different magnitudes of aggregated direct and indirect losses also reflect the shares of agriculture within each province's GDP (Figs. 3a,b).

Our results should help formulate fiscal policy and public budgeting for these extreme weather risks. Risk management instruments can be used to minimize and cap the cost of weather and climate impacts on society, government and producers.

Investments in infrastructure that increases physical resilience are effective in mitigating risk[30]. Our results indicate a maize production loss generated by a 1-in-50-year event of excess temperature and precipitation variability produces an aggregate 0.7 % loss of Shandong provincial GDP (see Fig. 3b). They also indicate that in under an irrigation scenario, production and aggregate economic losses are cancelled. As shown in Supplementary Table 1, estimations of the cost of deploying new irrigation infrastructure and restoring existing decaying structures could be performed at a cost of up to 0.73 % of Shandong GDP.

The economic efficiency of risk mitigating investments decreases, however, with the risk level considered and is only justifiable up to certain risk level[31]. In order to manage the residual risk, instruments of risk transfer and risk forecast can decrease the ex-post event costs of damage.

We propose a "three-pillar"–based approach for rural development and food security risk management. The three pillars are: (i) risk mitigation, (ii) risk forecast, and (iii) risk transfer instruments. These need to be tailored and combined to respond to specific climate risk profiles characterizing a given region. We believe the results of the end-to-end probabilistic risk assessment methodology presented here will be particularly effective in setting the balance of these three pillars. The implications of this work are of concern for farmers and policy makers, as well as for the whole value chain of the food-and-fibre industry, and for its long-term sustainability. The crucial importance of providing such detailed end-to-end information to stakeholders is further summarized in the **Supplementary Discussion**.

References:

1. Intergovernmental Panel on Climate Change *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (IPCC, 2013).

2. Conway, G. *One Billion Hungry: Can We Feed the World?* (Cornell University Press, Ithaca, New York, 2012)

3. Asseng, S., Ewert, F., Rosenzweig, C., Jones, J. W., Hatfield, J. L., *et al*. Uncertainty in simulating wheat yields under climate change. *Nature Clim. Change* **3**, 827–832 (2013)

4. Challinor, A. J., Watson, J., Lobell, D. B., Howden, S. M., Smith, D. R., Chhetri, N. A meta-analysis of crop yield under climate change and adaptation. *Nature Clim. Change* **4**, 287–291 (2014)

5. Moore, F. C., Lobell, D. B. Adaptation potential of European agriculture in response to climate change. *Nature Clim. Change* **4**, 610–614 (2014)

6. Overpeck, J. T., Meehl, G. A., Bony, S., Easterling, D. R., Climate data challenges in the 21st century. *Science*. **331**, 700–702 (2011)

7. Knutti, R., Sedláček, J. Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Clim. Change* **3**, 369–373 (2013)

8. Hannart, A., Ghil, M., Dufresne, J-L, Naveau, P., Disconcerting learning on climate sensitivity and the uncertain future of uncertainty. *Climatic Change* **119**, 585–601 (2013)

9. Kunreuther, H., Heal, G., Allen, M., Edenhofer, O., Field, C. B., *et al*. Risk management and climate change. *Nature Clim. Change* **3**, 447–450 (2013)

10. Martinson, D. G., Bryan, K., Ghil, M., Hall, M. M., Karl, T. R. *et al*., Eds. *Natural Climate Variability on Decade-to-Century Time Scales* (National Academy Press, Washington, D.C., 630 pp., 1995)

11. Ghil, M., Yiou, P., Hallegatte, S., Malamud, B. D., Naveau, P. *et al*., Extreme events: dynamics, statistics and prediction. *Nonlin. Processes Geophys*. **18**, 295–350 (2011)

12. Ghil, M., Robertson, A. W., “Waves” vs. “particules” in the atmosphere’s phase space: A pathway to long-range forecasting? *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2493–2500 (2002)

13. Chang, C. P., Ghil, M., Latif, M., Wallace, J. M. (Eds.), *Climate Change: Multidecadal and Beyond* (World Scientific Publ. Co./Imperial College Press, London, 2015, in press)

14. Rosenzweig, M. R., Binswanger, H. P. *Wealth, Weather and the Composition and Profitability of Agricultural Investment* (The World Bank Policy Research Working Paper, WPS 1055, 42 pp. 1992)

15. Stern, N. *The Stern Review on the Economics on Climate Change* (Cambridge University Press, 2007)

16. Nordhaus, W. D., Boyer, J. *Warming the World: Economic Models of Global Warming* (MIT Press, 2000)

17. Deloncle, A., Berk, R., D'Andrea, F., Ghil, M. Weather regime prediction using statistical learning, *J. Atmos. Sci.* **64**, 1619–1635 (2007)

18. Ihler, A.T., Kirshner, S., Ghil, M., Robertson, A. W., Smyth, P. Graphical models for statistical inference and data assimilation, *Physica D*. **230,** 72–87 (2007)

19. Läderach, P., Haggar, J., Lau, C., Eitzinger, A., Ovalle, O. *et al*. *Mesoamerican Coffee Building a Climate Change Adaptation Strategy* (International Center for Tropical Agriculture Policy Brief No. 2, February 2013)

20. Joshi, M. M., Turner, A. G., Hope, C. The use of land-sea warming contrast under climate change to improve impact metrics. *Climatic Change*. **117**, 951–960 (2013)

21. Sanchez, B., Rasmussen, A., Porter, J. Temperatures and the growth and development of maize and rice: a review. *Global Change Biol*. **20**, 408–417 (2014)

22. Lobell, D. B., Burke, M. B. Why are agricultural impacts of climate so uncertain? The importance of temperature relative to precipitation. *Environ. Res. Lett.*, **3**. 034007 (2008), doi:10.1088/1748_9326/3/3/034007

23. Lobell, D. B., Bänzinger, M., Mogorokosho, C., Bindiganavile, V. Nonlinear heat effects on African maize as evidenced by historical yield trials. *Nature Clim. Change*. **1**, 42–45 (2011) doi:10.1038/nclimate1043

24. Cheng, W., Sakai, H., Yagi, K., Hasegawa, T. Combined effects of elevated [CO_2] and high night temperature on carbon assimilation, nitrogen absorption, and the allocations of C and N by rice (Oryza sativa L.). *Agric. For. Meteorol.* **150**, 1174–1181 (2010)

25. Mohammed, A. R., Tarpley, L. High night time temperatures affect rice productivity through altered pollen germination and spikelet fertility. *Agric. For. Meteorol.* **149**, 999–1008 (2010)

26. Sheehy, J. E., Mitchell, P. L., Ferrer, A. B. Decline in rice grain yields with temperature: Models and correlations can give different estimates. *Field Crop. Res.* **98**, 151–156 (2006)

27. Chang, C.-P. (Ed.). *East Asian Monsoon* (World Scientific, 2004)

28. Ding, Y., Chan, J. C. L. The East Asian summer monsoon: an overview. *Meteorol. Atmos. Phys*. **89(1-4)**, 117–142 (2005)

29. NOAA. *Historical El Niño/La Niña episodes (1950–present)* (http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml)

30. Michel-Kerjan, E., Hochrainer-Stigler, S., Kunreuther, H., Linnerooth-Bayer, J., Melcher, R. *et al*. Catastrophe risk models for evaluating disaster risk reduction in investments in developing countries. *Risk Anal.* **33(6)**, 984–998 (2013)

31. Mechler, R., Islam, N. in *The Economic Impacts of Natural Disasters*, D. Guha-Sapir, I. Santos, A. Borde, Eds. (Oxford University Press, USA, 2013) pp.80–106

Acknowledgments:

Gridded weather data were shared by Prof Jiang Tong, National Climate Centre of the China Meteorological Administration. Provincial Input-Output tables were shared by Prof Yang Cuihong, Chinese Academy of Mathematics and System Sciences. Simulated crop data were shared by Prof Xiong Wei, Chinese Academy of Agricultural Sciences. The work of EC was supported by the CONACyT, Mexico, and Grantham Institute for Climate Change of Imperial College London. MG acknowledges support from the U.S. National Science Foundation, grant DMS-1049253.

Author contributions:

EC, GC and MG designed the study. EC obtained the data and carried out the calculations. MS provided further insights into the application of risk profiles to market practice. All four authors contributed to the writing.

Competing interest declaration

None of the authors declare any competing financial interest.

Figure Legends and Tables

Figure 1. **Schematic diagram of the end-to-end methodology for deriving crop production and economic-risk profiles**. Panel (c) uses input from panels (a) and (b) to produce grid-to-province PDFs of yield loss captured by weather indices, conditional on large-scale interannual climate processes. Panel (d) uses panel (c) grid-level yield loss PDFs and yield response functions subject to GHG and technological scenarios to derive regional-level risk profiles of production loss. If the region matches an economic administrative unit (e.g. province, country), panel (e) uses (d) to derive distributions of province-level economic losses. Panel (f) uses panel (d) and/or, if relevant, panel (e), to determine optimum combinations of risk mitigation and transfer instruments to minimize risk of climate-driven losses.

Words: 114

Figure 2. **Results of weather index–based modelling of maize yield in Shandong province**. (a, b) Maps of indices selected to best capture on a 0.25° × 0.25° longitude-latitude grid (a) deficit precipitation, and (b) excess temperature–driven yield variability. Color scale (see Supplementary Fig. 3) indicates the phase of crop growth in which the selected index captures highest sensitivity. (c) Map of 10-year return period production (see **Methods**) of ~200 to ~1,400 tons/pixel. Panels (d)–(g) present computations for a heat wave index. (d) Mixed univariate distributions of the index, subject to each NHMM state. (e) Viterbi-weighted sum of each distribution. The convolution of (f), the response function of yield to

heat wave, with (e) allows obtaining the distribution of yield (g). Results shown for a single local maize variety rainfed technological scenario.

Words: 132

Figure 3. **Risk profiles of province-level physical production and aggregate economic loss in China's northeast Shandong province**. (a) Risk profiles of maize provincial production loss, driven by the joint impacts of excess temperature and precipitation variability, subject to three different technological scenarios: (i) continuous line – local rain-fed variety; (ii) dotted line – switched rain-fed variety; and (iii) long-dashed line – local irrigated variety. (b) Risk profiles of direct and indirect aggregate economic loss expressed as percentage of provincial gross domestic product (GDP_{2008}): (i) black bars – local rain-fed variety; (ii) yellow bars – switched rain-fed variety; and (iii) red bars – local irrigated variety.

Words: 97

Methods

Word Count: 840

Data sources

Daily observed weather data on precipitation, radiation, and maximum and minimum temperatures were used. The data set was provided by the National Climate Centre (NCC) of the China Meteorological Administration (CMA) on a 0.25° × 0.25° longitude-latitude grid, available from 1961 to 2012; it covered the two northeastern provinces of Shandong and Hebei, and the two southern provinces of Guangxi and Guangdong. Grid-level maize and rice yields were simulated in those northeastern and southern provinces, respectively, using a mechanistic crop model called DSSAR-CERES.

Random forest-based indices selection

We selected the most effective pixel-level pairs of indices to capture the effects of deficit precipitation and excess temperature on yield variability by a random-forest algorithm. This algorithm uses ensemble-based recursive partitioning and thus permits one to circumvent the issues of cross-correlation between indices and of a large number of variables vs. a small sample size.

Extreme value multivariate modelling

Robust stochastic characterization of the interannual variability of the optimum grid-level weather indices was carried out using univariate distributions of mixed, exponential–Generalized Pareto Distribution (GPD) type. The latter allows one to accurately estimate the risk of occurrence of events that are both rare and extreme, within a modified Generalized Pareto Distribution framework across the whole gridded domain studied. The stochastic dependence of deficit precipitation and excess temperature is characterized by coupling their univariate mixed distributions F_X and F_Y within a Gumbel-Hougaard copula model, as described in the equations (1) and (2) below.

$$F(X,Y) = C_\theta(F_X, F_Y) \tag{1}$$

Here C_θ is the Gumbel-Hougaard Archimedean extreme value copula,

$$C_\theta = \left\{-\left(\left((-\log(u_X))^\theta + (-\log(u_Y))^\theta\right)^{-1/\theta}\right)\right\} \tag{2}$$

The coefficient of dependence is $\theta \geq 1$, where $\theta = 1$ characterizes independence of the uniform transforms u_X and u_Y of the mixed univariate F_X and F_Y distributions of precipitation and heat wave grid-level indices, respectively.

The Gumbel-Hougaard Archimedean copula enables us to characterize dependence in both the upper and lower tails without assuming independence of extreme-value occurrences, as is the case in Gaussian copulas. An example of stochastic dependence of two weather indices, at the same location and subject to a technological scenario, is presented in Supplementary Fig. 2.

Nonhomegenous Hidden Markov Model "weather-within-climate" modelling

Historical univariate or multivariate distributions of weather indices are derived by adopting a "weather-within-climate" modelling framework. The distributions are modelled conditionally on hidden regional weather states, S_t that capture seasonal variability. These states are conditioned themselves on observed or simulated continental and planetary-scale climate drivers that capture interannual modes of variability. A Nonhomogenous Hidden Markov Model (NHMM) is used to achieve this two-step conditioning and enable the introduction of non-stationarity, as illustrated in Supplementary Figure 1 across a gridded domain and equation (3) below.

The weather index distributions, $P(O_{1:T}, S_{1:T} | \lambda, z_{1:T})$, thus use continental-scale climate variables, $z_{1:T}$, observed or, potentially, simulated by high-end general circulation models, subject future greenhouse gas scenarios[45]. The non-stationary univariate distributions of pixel-level precipitation and excess heat, $O_{1:T}$, follow the mixed GPD-exponential univariate framework presented above. The copula-characterized stochastic dependency between marginal is considered stationary across weather states.

Here $1961 \leq t \leq 2012$ while S_t are the hidden states of the two-states Markov chain, z_t is the non-stationary NINO3.4 index acting as covariate, and $\lambda = \{a_i, \pi_i\}_{i=\{1,2\}}$ contains the transition parameters a_i, and initial probabilities π_i, of the NHMM, and b_{S_t} the distribution of the observed weather indices at time t, depending on the state S_t as follows:

$$P(O_{1:T}, S_{1:T}|\lambda, z_{1:T}) = \pi_i(z_1) b_{S_1}(O_1|z_1) \prod_{t=1}^{T-1} a_{ij}(z_t) b_{S_t}(O_{t+1}|z_{t+1}) \quad (3)$$

And where

- $a_{ij}(z_t)$ is the transition probability from state i at time t to j at time $t+1$ of a first-order Markov chain as a function of the non-stationary covariate z_t;
- $\pi_i(z_1)$ is the probability that the initial hidden state at $t = 1$ is i , $S_1 = i$; and
- $b_{S_t}(O_{t+1}|z_{t+1})$ is a component of the vector of observed weather indices characterized by mixed densities F_X and F_Y cited above, and dependent on the value of the non-stationary covariate z_{t+1}.

Generalized Additive Mixed crop response modelling

In order to model the vulnerability functions of crop yield to the combined or individual effects of precipitation variability and excess temperature exposure, Generalized Additive Mixed Models (GAMMs) are used. The use of a GAMM $g(\mu_i)$ enables capturing non-linear response of crop yield μ_i to the varying values of a single or several weather indices, cf. Fig. 2 (f),

$$g(\mu_i) = X_i\theta + f_1(x_{1i}) + f_2(x_{2i}) + \ldots \quad (4)$$

Here $\mu_i \equiv E(Y_i)$, with Y_i the rice or maize yield response variable following an exponential-family probability distribution function with and y_i is the i^{th} observation of the rice or maize yield variable, X_i is the i^{th} row of the model matrix with its corresponding θ parameter vecto

Also, in order to model the univariate model of rice or maize yield response to heat waves or deficit precipitation, a smoothing basis composed of natural cubic splines is used. Ultimately, the convolution of the GAMM-based yield response function with the distribution of the corresponding grid-level indices results in the distribution of yield loss as a function of indices values.

Input-Output-based economic impact modelling

An Input-Output modelling approach is used to assess direct and indirect Province-level economic impacts due to weather-driven maize production shortfall. Additional details concerning the methodology can be found in the **Supplementary Information** section.

Title: Estimating the effects of extreme weather on food security

NCLIM-14111599 revised manuscript

Authors: Erik Chavez[1]*, Gordon Conway[2], Michael Ghil[3,4], Marc Sadler[5]

Supplementary Information

Words: 1,413

Supplementary Methods

Data sources

The datasets contain less than 0.1 % data gaps[1]. The quality of the datasets was controlled by the CMA following Qian and Lin (2005)[2]. The temperature data homogeneity was controlled by CMA using the method of standard homogeneity test[3], the moving *t*-test[4], and departure accumulating method[5]. Precipitation datasets are not adjusted, while temperature datasets were homogeneity-adjusted[1].

The mechanistic crop model used has been calibrated with observed yield and soil data at the Chinese Academy of Agricultural Sciences[6,7,8].

Random forest-based indices selection

In each pixel, two sets of databases of (i) 25 precipitation indices, and (ii) 50 excess heat indices are built based on the pixel-specific date of planting, The 25 different periods of aggregation of the weather indices across the crop growing period are represented in **Supplementary Figure 3**. Datasets of excess temperature indices are computed in each pixel using two critical temperatures $T_c = 35°C$ or $T_c = 30°C$, and accounting for the numbers of days with $T_{max} > T_c$ during each of the aggregation periods. Precipitation indices are built by computing cumulative precipitation during the same 25 different aggregation periods.

We selected the most effective pixel-level pairs of indices to capture the effects of deficit precipitation and excess temperature on yield variability by a random-forest algorithm[9]. Therefore, for each pixel one precipitation index and one excess heat index are selected from the population of 25 precipitation 50 and excess heat potential indices.

For each set of 25 precipitation or 50 excess heat pixel-level indices, the Random Forest algorithm was programmed to extract a subset of 5 indices randomly for 5,000 times to compute regression trees. The indices importance measure is obtained after computation of the average of the 5,000 initial trees.

Extreme value multivariate modelling

A dynamic mixture model[10] was used to enable unsupervised threshold setting for the fitting of the GPD distribution. The stochastic dependence of the exponential-GPD mixed distributions of precipitation and excess heat indices is subsequently characterized using a Gumbel-Hougaard copula framework[11].

Nonhomegenous Hidden Markov Model "weather-within-climate" modelling

Using the best fit test of Aikake Information Criteria[12], a two-state Hidden Markov Model (HMM) with S_t=1 and S_t=*2* was fit in each pixel on the observed time series variables of weather indices $O_{1:T.}$ The two pixel-level states capture seasonal patterns of indices variability. For instance, at time *t* with S_t = 1, the distribution $P(O_t|S_t = 1)$ of precipitation or excess heat indices, O_t, corresponds to a characteristic distribution observed during a "less dry" and "less warm" season. In contrast, distribution of indices when S_t= 2, $P(O_t|S_t = 2)$, corresponding to a "dryer" and "warmer" state 2 type season. The different state-dependent indices distribution during "wet-mild" (blue pdf) or "dry-warm" (red pdf) states is illustrated in **Figure 2(d)**. Within the Non-homogenous Hidden Markov Model (NHMM) the sequence of weather states $S_{1:T}$ is dependent upon the sequence $z_{1:T}$ of large scale climate driver covariates (i.e. Niño3.4 index). These covariates can be observed or, potentially, simulated by high-end general circulation models, subject future greenhouse gas scenarios[13] Supplementary Figure 7 illustrates the parameters of a two-state NHMM fitted in one of Shandong province 280 pixels with Niño3.4 index used as non-stationary covariate.

While the El Niño Southern Oscillation is known to be amongst the main drivers of the Asian Summer Monsoon[14,15] the seasonal and interannual variability of the summer Monsoon in North East China is also associated with other drivers that are not taken into account in the model used here where only one non-stationary covariate is included. Other drivers such as the snow cover conditions in Eurasia and the Tibetan Plateau[16,17], the Indian Ocean Dipole interannual oscillation[18], and tropospheric cooling over Northern latitudes of China[18,19] have also been shown to exert an influence on the summer Monsoon variability in North East China, in conjunction with ENSO. Given the demonstrative nature of the manuscript in illustrating the methodological construct developed, only ENSO, the main driver of the Asian Summer Monsoon was used, and a more detailed study would allow characterization of the relative influences and interactions of the various climate drivers cited here on the Northeast China Summer Monsoon variability.

Furthermore, the interaction of global climate forcing, derived from increased emissions of greenhouse gases, with regional climate forcings[20,21], which result from tropospheric pollution and natural climate variability, amplify the uncertainty of projections of local weather variability in climate models. In particular, the prediction of local precipitation variability, both seasonal and interannual, such as the dates of rainfall season onset, is uncertain and represents a persistent barrier to robust forecasting of the impacts of weather variability on food supply. Furthermore, the uncertainty of future tropospheric pollution and the negative sensitivity of crop production to solar dimming increases the uncertainty of future food production in regions such as northeast China[22].

Generalized Additive Mixed crop response modelling

Within the GAMM[23] described, f_i are smooth functions of the x_i covariates that can be defined using a basis function that can be expressed linearly as follows with $b_j(x)$ the j^{th} element of the basis function and β_j scalar parameter values:

$$f(x) = \sum_{j=1}^{n} b_j(x)\beta_j \tag{5}$$

Here a spline basis due to the ability provided to estimate the properties of f over a large domain of the response variables. Cubic splines are used as smooth functions within the GAMM. Cubic splines can be described as portions of cubic polynomials joined together at specified *knots* in the response domain. The knots are located at specific quantiles values of the response variable. Given the locations of the knots $\{x_i^*: 1, \ldots, q-2\}$ the i^{th} row of the $y = \beta X + \varepsilon$ model matrix can be written using a cubic spline as:

$$X_i = \left[1, x_i, R(x_i, x_1), \ldots, R\left(x_i, x_{q-2}\right)\right] \tag{6}$$

Input-Output-based economic impact modelling

Supply-side shock is simulated using a Gosh model formulation[24] of province-level Input-Output tables[25] as detailed in equation (5) below. The crops considered are singled out from the rest of the economic network in order to model both direct and indirect economic losses derived from supply shortages Δv. Input-Output tables were obtained from the National Bureau of Statistics[26] repository and province-level maize and rice grain production used to single out these sectors in the tables were retrieved from Provincial Agricultural Statistical Records[27,28,29,30], and Δx below is the vector of changes in final supply for each sector represented.

$$\Delta x = G'\Delta v \tag{7}$$

Here G is the Gosh inverse, Δx the vector of changes in final demands and productions of each of the $n = 47$ represented sectors of the provinces economies, subsequent to a change in supply of $\Delta v = (0,0,\Delta_{crop}, \dots ,0)'$ of supply in maize in Shandong and Hebei or rice in Guangdong and Guangxi. The elements of the Gosh inverse coefficients reflect the total value of production δx_j coming about in sector $j \in [|1, n|]$ per unit of primary input δv_i in sector $1 \leq i \leq n$.

Supplementary Discussion

More frequent and broad-spread crop failures resulting from extreme weather conditions require new sources and types of financial products. Here the main driver is ensuring the sustainability of product sourcing by minimizing and smoothing in time, the costs caused by climate and weather hazards to farmers, the food-and-fiber industry, and society. Developing-country farmers are vulnerable to climate change and to the impacts of extreme events. Lack of resources reduces their ability to cope with these conditions. Moreover, the occurrence of natural disasters frequently forces their governments to divert planned investments to immediate post-catastrophe aid and reconstruction.

The role of improved modelling of future agricultural production loss risk on food stocks at both the national and international levels is becoming critically important. At the national level, the ability to base policy, procurement and safety net decisions on reliable data is vital. The previous existence of global food stocks and surpluses meant that shortfalls at the national level could be managed through access to international markets. With the reduction in global stocks and the fact that the majority of these are not liquid — as they are situated in countries unlikely to allow their export — the ability of national governments to purchase internationally has decreased[35].

With the increase of long-term investment funds in the equity markets and closer financial controls resultant from the 2008 financial crisis, equity analysts are increasingly interested in long-term sustainability plans of publicly listed companies, including food purchasers and retailers. This will ultimately result in share price differentiation between those companies who are, and those who are not, building long-term variables — such as climate change — into their business models and practices.

At the international level, the use of more accurate temporal and spatial modelling of future production would enable the humanitarian-aid architecture to be better planned and resourced. Such accurate modelling would also enable multi-country policy dialogue to occur in the case of shocks to the global food system, reducing the likelihood of volatile, "beggar-thy-neighbour" policy changes. Initiatives in this direction include the Agriculture Market Information System (AMIS) and Global Agricultural Monitoring (GEOGLAM) project. Much more remains to be done and will require the establishment of innovative collaborations between different disciplines and actors, including physical, agricultural and economic researchers and institutes.

References:

1. Fischer, T., Gemmer, M., Liu, L. L., Su, B. Temperature and precipitation trends and dryness/wetness pattern in the Zhujiang River Basin, South China, 1961–2007. *Quatern. Int*. **224**, 138–148 (2011)

2. Qian, W., Lin, X. Regional trends in recent temperature indices in China. *Meteorol. Atmos. Phys*. **90**, 193–207 (2005)

3. Alexandersson, H. A homogeneity test applied to precipitation data. *J. Climatol.* **6**, 661–675 (1986)

4. Peterson, T. C. Easterling, D. R., Karl, T. R. Homogeneity adjustments of in situ atmospheric climate data: a review. *Int. J. Climatol.* **18**, 1493–1517 (1998)

5. Buishand, T. A. Some methods for testing the homogeneity of rainfall records. *J. Hydrol.* **58** 11–27 (1982)

6. Xiong, W., Matthews, R., Holman, I., Lin, E. D., Xu, Y. L. Modelling china's potential maize production at regional scale under climate change. *Clim. Change.* **85**, 433–451 (2007)

7. Xiong, W., Holman, I., Conway, D., Erda, E. D., Li., Y. A crop model cross-calibration for use in regional climate impacts studies. *Ecol. Model*. **213**, 365–380 (2008)

8. Xiong, W., Conway, D., Lin, E. D., Xu, Y. L., Ju, H., *et al.* Future cereal production in china: The interaction of climate change, water availability and socio-economic scenarios. *Glob. Environ. Chang.* **19(1)**, 34–44 (2009)

9. Breiman, L. Bagging predictors. *Machine Learning* **26**,123–140 (1996)

10. Frigessi, A., Haug, O., Rue, H. A dynamic mizture model for unsupervised tail estimation without threshold selection. *Extremes*. **5**, 219–235 (2002)

11. Gumbel, E. J. Distributions à plusieurs variables dont les marges sont données. *C. R. Acad. Sci. Paris*. **246**, 2717–2719 (1958)

12. Aikake, H. A new look at the statistical model identification. *IEEE Trans. on Autom. Control*. **19**, 716–723 (1974)

13. IPCC. *Special Report on Emissions Scenarios* (Cambridge: Cambridge University Press, 2000)

14. Chang, C.-P. (Ed.). *East Asian Monsoon* (World Scientific, 2004)

15. Ding, Y., Chan, J. C. L. The East Asian summer monsoon: an overview. *Meteorol. Atmos. Phys*. **89(1-4)**, 117–142 (2005)

16. Zhang, Y. S., Li, T., Wang, B. Decadal change of the spring snow depth over the Tibetan Plateau: The associated circulation and influence on the East Asian summer monsoon. *J Climate*. **17**, 2780–2793 (2004)

17. Wu, Z. W., Li, J. P., Jiang, Z. H., Ma, T. T. Modulation of the Tibetan Plateau Snow Cover on the ENSO Teleconnections: From the East Asian Summer Monsoon Perspective. *J. Climate*. **25**, 2481–2489 (2012)

18. Ding, Y. H., Wang, Z. Y., Sun, Y. Inter-decadal variation of the summer precipitation in East China and its association with decreasing Asian Summer Monsoon. Part I: Observed evidences. *Int. J. Climatol.* **28**, 1139–1161 (2008)

19. Ding, Y. H., Sun, Y., Wang, Z., Zhu, Y. X., Song, Y. F. Inter-decadal variation of the summer precipitation in china and its association with decreasing Asian Summer Monsoon part II: Possible causes. *Int. J. Climatol*. **29**, 1926–1944 (2009)

20. Ramanathan, V., Chung, C., Kim, D., Bettge, T., Buja, L. *et al.* Atmospheric brown clouds: Impacts on South Asian climate and hydrological cycle. *Proc. Natl. Acad. Sci. USA*, **102**, 5326–5333 (2005)

21. Bollasina, M. A., Ming, Y., Ramaswamy, V. Anthropogenic aerosols and the weakening of the South Asian summer monsoon. *Science*, **334,** 502–505 (2011)

22. Chen, C., Baethgen, W., Robertson, A. W. Assessing the individual contributions of variations in temperature, solar radiation and precipitation to crop yield in the North China Plain, 1961-2003. *Clim. Change* (2013) doi: 10.1007/s10584-012-0509-2

23. Hastie, T., Tibshirani, R. Generalized additive models (with discussion). *Stat. Sci.* **1**, 297–318 (1986)

24. Ghosh, A. Input-output approach to an allocative system, *Economica*. **25**, 58–64 (1958)

25. Leontief, W. Input–output economics. *Sci. Am*. **185** (1951)

26. National Bureau of Statistics. *China statistical yearbook.*(China Statistics Press, Beijing, 2008)

27. Department of Rural Surveys, Shandong Province. *Shandong Agricultural Statisitics Yearbook.* (China Statistics Press, Beijing, 2008)

28. Department of Rural Surveys, Hebei Province. *Hebei Agricultural Statisitics Yearbook.* (China Statistics Press, Beijing, 2008)

29. Department of Rural Surveys, Guangdong Province. *Guangdong Agricultural Statisitics Yearbook.* (China Statistics Press, Beijing, 2008)

30. Department of Rural Surveys, Guangxi Province. *Guangxi Agricultural Statisitics Yearbook.* (China Statistics Press, Beijing, 2008)

31. Sadler, M., Magnan, N. Grain import dependency in the MENA region: risk management options. *Food Sec.*, **3(Suppl 1)**, S77–S89 (2011)

32. Viterbi, A. J. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE T. Inform. Theory*. **13**(2), 260–269 (1967)

33. US Department of Agriculture. *China agricultural and economic data*. http://www.ers.usda.gov/data-products/china-agricultural-and-economic-data.aspx

34. Food and Agriculture Organization AquaStat. *Global map of irrigation areas*. http://www.fao.org/nr/water/aquastat/irrigationmap/index.stm

35. Food and Agriculture Organization. *Database on investment costs in irrigation*. http://www.fao.org/nr/water/aquastat/investment/index.stm

Supplementary Figures and Tables Legends

Supplementary Figure 1. **Schematic of the "weather-within-climate" and index-based, local-to-regional weather risk modelling framework**. The bottom of the figure shows grid-level columns (i.e. databases) of *N* potential weather indices acting as proxies of weather-driven crop yield loss (colour coding of indices as described in Supplementary Figure 3). The most effective index is selected using an ensemble-based recursive partitioning algorithm resulting in a mosaic of weather indices that capture the sensitivity of crop yield to daily variability of one or more weather variables. Each selected weather index is modeled (downward dashed arrows) conditionally on latent, regional-level variables capturing intraseasonal weather variability in each region. The set of homogenous latent variables is itself modeled conditionally on observed or simulated, time varying, large-scale variables that capture interannual climate variability. The latter variables are used to project the regional set of selected indices into different climate scenarios.

Supplementary Figure 2. **Bivariate distributions of the indices for heat wave and precipitation variability, associated with a single pixel in Shandong province, subject to rain-fed local variety scenarios.** (a) Joint cumulative distribution of dependence between heat wave and precipitation variability indices, using a Gumbel copula model. (b) Return period of joint occurrence of the indices for heat wave and precipitation variability.

Supplementary Figure 3. **Weather index color code, as illustrated by sample building for a cumulative weather index, namely the deficit rainfall for the 135 daylong growing period of a given crop.** The color code of each weather index is calibrated on the yellow-red-blue color scale located above all the indices. If an index recording deficit precipitation is at the beginning of the crop growth cycle (i.e. during the first third of the 135- day period) its color is yellow and tends to green. For deficit precipitation at the middle (i.e. during the second third of the growth cycle, including the reproductive stages) the color is red. Finally, for deficit precipitation during the last third of the crop cycle the color is blue. Overlapping deficit precipitation indices capture periods are indicated by corresponding colour proportions Brown colored “NaN” is used to encode lack of data.

Supplementary Figure 4. **Matrix of the impact of weather conditions and technological scenario on the maize yield in northeast China’s Shandong province**. Rows indicate the technological scenario, while the columns indicate the individual, (a)-(f), or combined, (g)-(i), weather hazards (i.e. precipitation, heat or both).

Supplementary Figure 5. **Results of weather index–based modeling of rice yield response to excess temperature in South China's Guangxi province**. (a) Map of Guangxi with indices selected at pixel level to best capture rice yield variability driven by excess temperature. The colour scale is fully displayed in Supplementary Fig. 3 and indicates the phase of the crop growth cycle during which the selected weather index captures most significantly higher sensitivity to excess heat: beginning – green-yellow; middle – red-purple; end – purple/blue-dark blue; and grey: whole season. (b) Map of the 10-year return period for rice production, derived from the pixel-level distributions of weather indices for rice yield response, and pixel-level sown area; light-yellow–to–dark-orange scale from ~200 to ~1,400 tons/pixel. Results shown for a single local irrigated technological scenario, of local rain-fed rice.

Supplementary Figure 6. **Schematic diagram of the Nonhomogenous Hidden Markov Model (NHMM) used**. **R1** and **R2** represent the observed uni- or multivariate distributions of the weather indices. **S1** and **S2** are hidden variables that describe regional weather variability on intraseasonal scales, while $X(t)$ is a time-varying covariate that captures interannual climate variability. The vertical arrows represent conditional dependence, while the horizontal arrows linking **S1** and **S2** represent transition probabilities between the two latent variables; self-transition probabilities are represented by circular arrows. The Niño-3.4 index, based on Tropical Pacific sea surface temperatures, is used as $X(t)$, while **S1** and **S2** are derived from the observed **R1** and **R2** weather indices.

Supplementary Figure 7. **Schematic diagram of the two-state NHMM for a grid point in Guangxi province**. (a) Transition probabilities for the two states of the NHMM conditioned on central-eastern Pacific sea surface temperatures, as captured by the Niño-3.4 index. (b)

Most probable sequence of states on the same grid point as decoded using the Viterbi algorithm[32].

Supplementary Table 1. **Cost estimates for new development and renovation of irrigation infrastructure in Shandong province, expressed in millions of 2008 USD (USD×10^6) and percentage of 2008 aggregate provincial GDP (% GDP)**; the latter amounted to 3.09 trillion Yuan in 2008 (i.e. 0.46 trillion 2008 USD). Sown area figures are extracted from the USDA ERS statistical database[33]. Irrigated land areas are extracted from the National Bureau of Statistics[26] (NBS) and FAO's AquaStat[34], respectively, lower and upper bound estimates for the year 2001. The FAO irrigation infrastructure cost database is used to access potential costs of deployment and renovation/modernization of irrigation infrastructure in Shandong province used here[35]: (i) average cost of new infrastructure for underground pumped water irrigation in Asia of 550 USD/ha; and (ii) average cost of rehabilitating and modernizing underground pumped water irrigation projects in China of 1,670 USD/ha.