Supplementary Information for "*Uncertainties in a global crop model ensemble: effects of parameterization and setup harmonization on maize yield estimates and model performance*"

S1 Background on the EPIC model, EPIC-based GGCMs, and wider ensemble included in the study

S1.1 Relevant routines of the EPIC model

This sections provides an overview of selected routines of the EPIC model relevant for this study. The descriptions are based on the original documentation of the EPIC model, which is available at http://epicapex.tamu.edu/files/2015/05/EpicModelDocumentation.pdf. Further detail on the soil organic matter routines is provided in Izaurralde et al. (2006, 2012). Parameter numbers mentioned throughout the section refer to those in Table 2 of the main article and Table S 1-2 below.

S1.1.1 Plant growth, yield formation, and associated stresses

Phenologic development of the crop follows the heat unit (HU) accumulation approach. The average temperature on each day is summed up throughout the growing season until a defined heat unit requirement (growing degree days; GDD) is met. On each day HU are calculated as

$$HU_k = \frac{T_{\max,k} - T_{\min,k}}{2} - T_b \tag{S1}$$

where HU_k is the heat units accumulated on day k [°C], $T_{max,k}$ and $T_{min,k}$ are the maximum and minimum temperatures on the day [°C], and T_b is the base temperature [°C] as specified in Table S 1-4 for the maize cultivars used in this study. Potential biomass increase ΔB_p on each day is estimated according to

$$\Delta B_p = 0.001 \times BE \times PAR \tag{S2}$$

where ΔB_p [t ha^{-1}] is biomass gain, BE [(kg ha^{-1})/(MJ m^{-2})] is a crop-specific biomass-energy-conversion coefficient, and PAR [MJ m^{-2}] is intercepted photosynthetic active radiation depending on LAI, solar radiation, and plant row width (see Table S 1-3 for examples) according to

$$PAR = 0.5 \times RA \times (1 - e^{-(0.685 - 0.209 \times RIN) \times LAI}) \tag{S3}$$

where RA [MJ m^{-2}] in incoming solar radiation on a given day, RIN is the row width at planting [-] and LAI is the leaf area index [-] on the respective day.

Biomass gain is constrained mainly by water and nutrient (N and P) deficits as well as temperature and aeration stress. Only the dominant stress on a given day regulates plant growth through a plant growth regulation factor REG ranging from 0 to 1. The sum of the daily values for each stress factor over the growing season is referred to as "stress days" [d].

Water stress (WS) is based on the concept that drought stress is proportional to the transpiration reduction. It is estimated as

$$WS_i = \frac{\sum_{l=1}^{M} u_{l,i}}{PET_i} \tag{S4}$$

where WS_i is the amount of water stress on day i [-], l is a given soil layer [-], M is the total number of soil layers [-], $u_{l,i}$ is the plant available water in layer l on day i [mm], and PET_i is the potential evapotranspiration on day i [mm]. In addition, water deficit has an impact on HI as described below. Temperature stress (TS) occurs if the average air temperature TG is above the optimum temperature TO or below the base temperature TB on a given day according to

$$TS_i = \sin\left(\frac{\pi}{2} \times \frac{TG_i - TB_i}{TO_i - TB_i}\right) \tag{S5}$$

or if the average daily temperature exceeds TO by 50%. Nutrient deficits (N stress (NS) and P stress (PS)) vary non-linearly between optimum supply and 50% of the optimum supply when the deficit reaches 100%. First, a scaling factor SNS (here for NS) on a given day i is calculated as

$$SNS_i = 200 \times \left(\frac{UN_i}{cNB_i \times B_i} - 0.5\right) \tag{S6}$$

where UN_i is the N uptake on day i [kg ha^{-1}], cNB_i is the optimum N concentration in biomass on day i [kg kg^{-1}] and B_i is the total plant biomass on day i [kg ha^{-1}]. This factor is then used in the estimation of actual nitrogen stress NS according to

$$NS_i = \frac{SNS_i}{SNS_i + \exp(3.52 - 0.026 \times SNS_i)} \tag{S7}$$

The calculation of PS follows the same pattern. Aeration stress (AS) occurs if the soil humidity approaches water saturation and depends on pore space volume, soil humidity, and a crop-specific sensitivity factor. AS occurs rarely in global simulations as depth to ground water is typically not considered. It is hence foremost limited to soils prone to water logging such as vertisiols. The plant growth regulation factor REG is finally calculated as

$$REG = \max(WS, TS, NS, PS, AS) \tag{S8}$$

and subsequently actual biomass gain ΔB_a [t ha^{-1}] as

$$\Delta B_a = \Delta B_p \times REG \tag{S9}$$

where ΔB_p [t ha^{-1}] is the potential biomass gain on a given day. At maturity, crop yield is calculated by multiplying total above-ground biomass with a water stress-adjusted harvest index (HI_a). HI_a is estimated from simulated potential HI (HI_{max}; depending on HU accumulation) and a defined minimum HI (HI_{min}) according to

$$HI_a = (HI_{\max} - HI_{\min}) \times \frac{WUR}{WUR + \exp(6.13 - 0.0883 \times WUR)} + HI_{\min} \tag{S10}$$

where WUR is the water use ratio. The fraction of the growing season when water stress starts to limit HI_a is defined in parameter 33. Values for HI_{max} and HI_{min} used in this study are provided in Table S 1-4. WUR is estimated at harvest as

$$WUR = 100 \times \frac{\sum_{i=1}^{K} U_i}{\sum_{i=1}^{K} E_{Pi}} \tag{S11}$$

where U_i [mm d^{-1}] is the actual and E_{Pi} [mm d^{-1}] the potential plant water use rate for day i. K is the total number of days of the growing season.
Biomass accumulation may in addition be limited by root growth stresses (parameter 32), which include soil strength, salinity, and aluminum toxicity. Herein, only one of the EPIC-based GGCMs (EPIC-TAMU) considers these types of stresses.

S1.1.2 Hydrology

Daily precipitation is split at the surface into runoff and infiltration. Daily runoff Q [mm] is estimated according to the USDA SCS curve number equation

$$Q = \frac{(R - 0.2s)^2}{R + 0.8s} \quad \text{for } R>0.2s \tag{S12}$$

where R is precipitation [mm] and s is the retention parameter [mm]. At R≤0.2s, no runoff occurs. Parameter s depends on the curve number (CN) according to

$$s = 254 \times \frac{100}{CN} - 1 \tag{S13}$$

The CN reflects different types of soils, landuse and management, and are adjusted for slope, surface roughness, and soil humidity in the model. Impacts of these factors on CN as partly calculated dynamically in the model, but includes also various coefficients that need to be determined exogenously. These are among others the CN number index coefficient (parameter 11), which determines the impact of ET on water retention, a CN adjustment factor for plant residue on the field (parameter 12), and the soil variable dependence of CN (parameter 10). Soil humidity is included in the estimation of the retention parameter directly, based on the fraction of field capacity (FC; parameter 9) at present soil humidity. A higher fraction of FC causes higher runoff rates. Percolation from one soil layer l to a deeper soil layer or aquifers starts when soil water storage exceeds FC according to

$$PC_l = (SW_{0,l} - FC_l) \times 1 - e^{(-\Delta t - TT_l)} \tag{S14}$$

where PC_l is the percolation rate, $SW_{0,l}$ is the initial soil water storage, FC_l is the water content at field capacity, Δt is the time interval (here 24h) and TT_l is the travel time through the soil layer. The latter is a function of porosity, field capacity, and saturated conductivity. Also upward movement of soil water may occur if soil layer of low porosity are saturated.

Actual evapotranspiration (ETa) is driven by potential evapotranspiration (PET; parameter 1) and water available for evaporation from soil and transpiration by the crop. EPIC offers 5 methods for estimating PET, namely Baier-Robertson, Hargreaves, Penman, Penman-Monteith, and Priestly-Taylor. A detailed description and evaluation of the methods in the global EPIC-based GGCM PEPIC is provided in Liu et al. (2017). Various coefficients of the PET estimation can be adjusted manually to local conditions, such as a crop canopy resistance factor for Penman-Monteith, or the linear and exponential coefficients of the Hargreaves method (parameters 2 and 3). ETa is estimated following the approach of Ritchie (1972) with plant transpiration based on leaf area index alone, except for Penman-Monteith, which also considers vapor pressure deficit and canopy resistance. Plant water use throughout the soil profile depends on water demand, root distribution (affected among others by parameters 30 and 31), water availability in each layer, and a water tension function defined by parameter 8. Soil evaporation is the remainder of ETa after transpiration demand is met. Various coefficient regulate soil evaporation, such as parameters 6 and 7, which regulate the evaporation rate with soil depth and parameters 4 and 5, which increase the albedo of soil cover and accordingly decreases the effect of solar radiation on PET estimation if considered by the selected method.

S1.1.3 Soil organic matter turnover and nutrient mineralization

Soil organic matter (SOM) and nitrogen (N) cycling follow largely the routines of the CENTURY model (Parton et al., 1996), which have been adapted for EPIC by Izaurralde et al. (2006). SOM is split into the pools standing dead residue and roots, metabolic and structural litter, slow humus, passive humus, and microbial biomass, which vary in exchange and turnover rates such as the slow to passive humus partitioning coefficient (parameter 23). Mineral and organic C, N and P may leave the system through erosion, leaching and volatilization (C and N). Fluxes between different pools depend on soil and crop management, soil hydrology, temperature, oxygen availability, stoichiometry, directly parameterized coefficients, and the selection of subroutines. The complex routines of OM turnover are laid out in Izaurralde et al. (2006) in detail.

Field capacity (FC) and wilting point (WP), which differ here most substantially between the soil data used in EPIC-IIASA and GEPIC (Figure S 1-3) are key parameters in various soil microbial routines. The soil water control factor SUT for biological processes is estimated according to

$$SUT = \sqrt{\frac{e^{6\times(DB-DB_P)}}{FC-WP}} \tag{S15}$$

where DB is the soil bulk density and DB_P is the bulk density of the plough layer. SUT is limited to a maximum of 1. SUT is part of the combined factor CS for soil microbial activity with

$$CS = \sqrt{CDG \times SUT} \times MAC \times OX \times e^{6\times(DB-DB_P)} \tag{S16}$$

where CDG is a soil temperature coefficient, MAC is the soil microbial activity coefficient (parameter 22), OX is a coefficient for oxygen availability depending on soil depth (modulated by parameters 24 and 25s), and the last term reflects the ratio of average soil bulk density and plough layer bulk density as in Eq. S17.

Mineralization from and transformation between pools of N is driven by supply and demand functions as well as transformation coefficients. Net mineralization of N is the differences between gross mineralization (sum of N mineralization from all OM pools) and N immobilized in microbial biomass, depending in turn on the C:N stoichiometry of the biomass. P mineralization is based on the PAPRAN model (Seligman and van Keulen, 1981;

Jones et al., 1983) and depends on the CS factor (Eq. S16) and P concentration in fresh biomass and humus. For both nutrients, the solving of turnover equations follows the law of conservation of mass.

The total volume of nitrification and volatilization is first estimated simultaneously based on NH_3 concentration in the profile, water saturation, pH, soil temperature, and wind sped. The volatilization coefficient is then calculated based on a user specified coefficient (parameter 26), the total nitrification and volatilization volume, and soil layer depth. Denitrification can be estimated by one of three methods. The original EPIC method is based on soil temperature and humidity, total OC, the NO_3 concentration, and a user defined soil humidity threshold and estimates N_2 production only. The more recently introduced Armen Kemanian method considers in addition a respiration factor and produces also N_2O loss estimates without user input requirements. The yet most detailed representation of denitrification processes has been introduced by Cesar Izaurralde and includes production and consumption of O_2, CO_2, and N_2O throughout the soil profile at an hourly time step based on a mechanistic electron transfer model (Izaurralde et al., 2012).

S1.1.4 Soil dynamics and degradation

Soil profiles can be handled dynamically or statically (parameter 19). In the latter case, all soil characteristics except pools of mineral nutrients are reset to initial conditions at the beginning of each year. Soil degradation is represented in the form of nutrient and OM depletion in case these constituents are not replenished in sufficient amounts. In addition, the model considers wind and water erosion (parameters 13, 14 and 16) using the WEPS model in the first case – depending on slope, wind speed and field size (parameters 17 and 18) - and providing 6 methods (USLE, RUSLE, modified RUSLE (RUSL2), MUSLE, MUST, MUSS) in the latter case, with differences in driving forces (rainfall or runoff) and assumptions concerning the watershed/field size. Water erosion can be scaled directly, using a water erosion conservation practice factor (parameter 15), which reflects assumptions on farmers practice concerning countermeasures for erosion.

S1.1.5 Crop management

The EPIC model allows for a wide range of management specifications from field preparation over planting, fertilizer, manure, pesticide, and irrigation water application, and weeding to harvest and plant residue handling. Fertilizer and irrigation water applications can follow a rigid schedule based on heat unit scheduling or fixed dates or occur automatically based on plant stresses. Triggers (parameters 27 and 29) are set to specific values of the plant growth regulating factor REG (see Sect. S1.1.1). E.g. a fertilizer application trigger of 0.9 causes fertilizer application if potential plant growth would be limited on a given day by >10%.

Examples of more detailed operation schedules for two GGCMs are provided in Table S 1-3. Besides their direct purpose, field operations are defined by their soil mixing efficiencies, effects on surface roughness, affected soil depth, and ridge height, all of which affect soil OM turnover and hydrologic processes. For planting operations, the initial row width has a direct impact on the estimation of potential plant growth (see Sect. S1.1.1).

S1.2 Differences between EPIC model versions v0810 and v1102

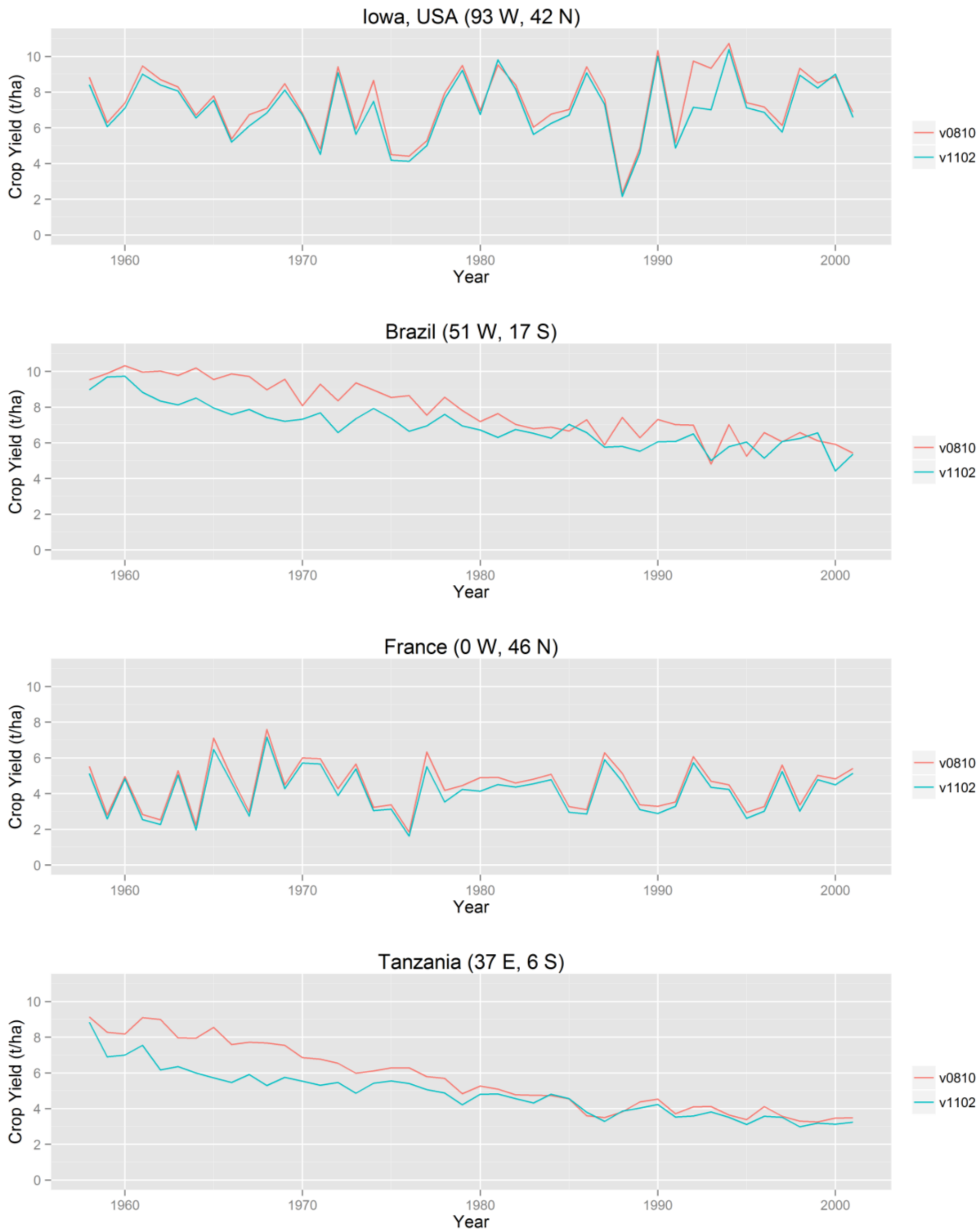

Figure S 1-1: Maize yield estimates of EPIC v0810 and EPIC v1102 for four contrasting locations.

Two versions of the EPIC field-scale model were used in this study, designated as v0810 and the more recent v1102. The first is the presently publicly available version from the developers at Blackland Research Center of Texas A&M University. The latter has been modified by the developers of the global model framework EPIC-TAMU, mainly with more detailed and revised routines for soil nutrient and carbon cycling. These include gas diffusion routines, root respiration, nutrients in microbial biomass, and improved (de-)nitrification among others. Testing both field-scale models at four sites in differing climate, soil and management conditions shows that the absolute yield levels are at least after a spin-up period mostly at a comparable level and also inter-annual yield dynamics are mostly very similar (Figure S 1-1). The identification of drivers in differences between the models is beyond the scope of this study.

S1.3 Description of EPIC-based GGCMs

S1.3.1 EPIC-BOKU

EPIC-BOKU was initially developed to provide yield estimates at contrasting management intensities for land use change and agro-economic models (Havlík et al. 2011, Schneider et al. 2011; Nelson et al., 2014; Frank et al., 2015) at the European and global scales (Stolbovoy et al. 2007; Skalský et al., 2008; Elshout et al., 2015).
The spatial structure of its input data is based on a regular 5 arcmin grid, which is first aggregated to homogenous response units (HRUs) based on a classification of physical characteristics (elevation, slope, soil). The HRUs are subsequently intersected with administrative units (national borders at the global scale) that determine specific crop management parameters, which are derived from databases or socio-economic data. The field-scale model is run for each of the resulting simulation units (SimU). For comparison with GGCMs running at a 0.5°x0.5° resolution, the results from the SimUs were resampled based on the pixel-weighted 5 arcmin model outputs per 0.5°x0.5° grid. Presently, the GGCM runs two nutrient management intensities, high-input and low-input agriculture with accordingly high or low fertilizer application rates. For the default simulations, outputs from the high-input runs were submitted, corresponding to non-nutrient limited yield potential with default growing season assumptions.

S1.3.2 EPIC-IIASA

EPIC-IIASA has been developed in parallel to EPIC-BOKU, partly by the same researchers and shares in principle the same spatial data structure based on HRUs and SimUs whereas simulations can also be run at the 0.5°x0.5° grid cell level. Parameterizations and input data have been adjusted throughout research projects resulting in a substantially differing setup with the major remaining communality being the use of a static soil profile (Table 1 in main manuscript). Growing seasons have been adopted from Sacks et al. (2010) and crop-specific spatially explicit N and P application rates from Mueller et al. (2012). Focus regions of recent studies for which model setups have been adjusted are the EU (e.g. Balkovič et al., 2013) and China (Xiong et al., 2014a; Xiong et al., 2014b) besides global applications (Balkovič et al. 2014 Xiong et al. 2016). Simulations for the default setup were carried out at the SimU level and the harmonized runs were based on the 0.5° grid.

S1.3.3 EPIC-TAMU

EPIC-TAMU follows the model development and implementation of EPIC, version 1102, which accounts for C and N stocks and flows in managed terrestrial ecosystems (Izaurralde et al. 2012). As in EPIC v. 0810, the coupled C and N model in EPIC-TAMU follows the conceptual pool structure of the Century model (Izaurralde et al., 2006). Mineralization and immobilization of C and N also follows the approach in Century but a recent option has been added to describe C and N of microbial biomass following the approach used in the Phoenix model (McGill et al., 1981). The EPIC-TAMU version also contains algorithms to model the effects of biochar additions on crop productivity, soil pH, and cation exchange capacity (Lychuk et al., 2015). Other developments include a mechanistic model to describe microbial denitrification and the corresponding feedback on decomposition (Izaurralde et al., 2012).
EPIC-TAMU has primarily been used for field-scale and regional-scale simulations (Gelfand et al., 2013; Zhang et al., 2015). It has been adapted with minimal changes for use as part of the AgMIP GGCMI project, and has otherwise not been previously used for global simulations. As a result, no default simulations (see Sect. 2.2 in main manuscript) were produced. To keep the number of EPIC-based GGCMs in evaluations across management scenarios constant, the fully harmonized setup was also used as default.

S1.3.4 GEPIC

The GEPIC GGCM was originally developed for studies of global crop-water relations (Liu et al., 2007). In its present version, it uses input data for planting dates, growing season length, P fertilizer application rates, and cultivar distributions besides the original input data elevation, slope, country/region, N fertilizer application rates, and irrigation water management (Folberth et al., 2012). Default fertilizer inputs are mainly based on the FertiStat database (FAO, 2007) and have been extrapolated for maize based on the human development index (HDI) for countries lacking data.

More recently, the GGCM has been setup for applications in sub-Saharan Africa based on a regional calibration (Gaiser et al., 2010) and thoroughly evaluated in various studies (e.g. Folberth et al., 2012; Folberth et al., 2014; van der Velde et al., 2014). The authors found that when using dynamic soil profiles in the setup, the model reproduces yields around the year 2000 well after a spin-up of 30 years (Folberth et al., 2012). Extending the simulation period may result in erosion of the whole soil profile at some point or complete nutrient depletion in grid cells that lack fertilizer inputs. To avoid resulting detrimental effects on crop yields and unrealistically long monocultures, the model is run for each decade of the study period separately, which aims at mimicking fallow rotation with an average cultivation period of 40 years and complete recovery of the soil profile afterwards (see Figure S 1-2).

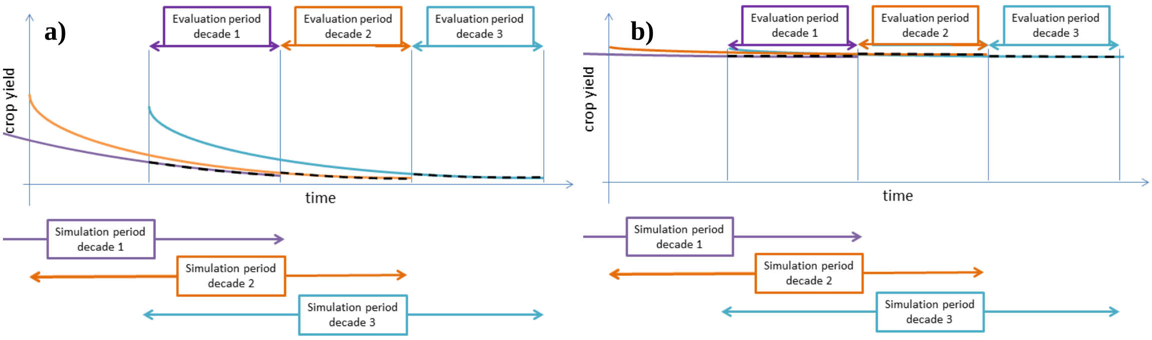

Figure S 1-2: Schematic representation of decadal GEPIC runs with dynamic soil profile and erosion for (a) high nutrient input and (b) low nutrient input conditions. Colors represent simulations for three decades with a 20 year spin-up for each decade, which is discarded. Only the last ten years are part of the evaluation as indicated by the dashed black lines.

S1.3.5 PEPIC

PEPIC is a global EPIC-based GGCM developed at the Swiss Federal Institute of Aquatic Science and Technology (Eawag) initially based on GEPIC, which had been developed at the same institute. Hence, the two GGCMs have similar features in software design and default input data. However, one of the main purposes of PEPIC was to develop a fully free tool, which can be used without any software license. Therefore, unlike GEPIC, PEPIC was compiled by a free computer language, Python. In addition, the parameterization and setup has been adjusted in large parts (Table 1 in main manuscript) to match focus research purposes. It was initially developed to investigate the impacts of different PET methods on crop-water relations (Liu et al., 2016a). Presently, applications of PEPIC focus on assessing trade-offs between crop yields and nutrient losses, e.g. N, in the context of global agricultural intensification (Liu et al., 2016b). For such assessments, N was applied three times during the whole season following a fixed schedule (Liu et al., 2016b). Besides the parameterization, this is a single major difference compared to the other four EPIC-based GGCMs, which used automatic N fertilization based on plant nutrient requirements.

Table S 1-1: Relevant characteristics of GGCMs in this study based on Müller et al. (2017) and amended by stress and fertilizer handling. More information is provided in the reference.

GGCM	Type[1]	Leaf area development[2]	Light utilisation[3]	Evapotrans-piration[4]	Growth stresses[5]	Yield formation[6]	Soil input data[7]	Soil layers[8]	Soil C and nutrient models[9]	Crop residue handling	Spin-up[10]
CLM-crop	Eco	DA	P-R	TF	W (S),N,H	Prt	IGBP Global Soil Data Task 2000	10	C N	To litter pool	NA
EPIC-BOKU	Site (EPIC)	PS	RUE	PM	W (E), T, H, A, N, P	HIws Prt B	ISRIC-WISE; ROSETTA; AWC; Albedo D; HYD USDA	10	C N B(1) P(6)	To litter pool	Soil OM, C, NH3, NO3, H2O, P (1)
EPIC-IIASA	Site (EPIC)	PS	RUE	HG	W (E), T, H, A, N, P	HIws Prt B	ISRIC-WISE; ROSETTA; AWC; HYD USDA	10	C N B(1) P(6)	To litter pool	Soil OM, C, NH3, NO3, H2O, P (50)
EPIC-TAMU	Site (EPIC)	PS	RUE	PM	W (E), T, H, A, N, P	HIws Prt B	ISRIC-WISE	3	C N B(1) P(6)	To litter pool	Soil OM, C, NH3, NO3, H2O, P, CR (10)
GEPIC	Site (EPIC)	PS	RUE	HG	W (E), T, H, A, N, P	HIws Prt B	ISRIC-WISE	5	C N B(1) P(6)	80% removed 20% litter	Soil OM, C, NH3, NO3, H2O, P, CR (30)
LPJ-GUESS	Eco	DA	P-R	PT	W (S), T	HIws	HWSD; STC; HYD C; THM L	2	NA	Removed, no effect on yield	H2O (30)
LPJmL	Eco	PS	P-R	PT	W (S), T	HIws	HWSD; STC; HYD C; THM L	5	NA	Removed, no effect on yield	H2O, Tsoil (200)
ORCHIDEE-crop	Eco	DA	P-R	PT	W (S),T,N	Prt	NA	11	NA	Removed, no effect on yield	H2O (1)
pAPSIM	Site	DA	RUE	TE	W (E,S), T, H, A, N	Gn Prt	HWSD	5	C,N,P,B(3)	NA	NA
pDSSAT	Site	PS	RUE/P-R	PT/PM	W (E), T, H, A, N	Gn	HWSD	4	C N P(3)	Removed, no effect on yield	Soil OM, C, NH3, NO3, H2O (1)
PEGASUS	Eco	DA	RUE	PT	W (E), T, H, N, P, K	Prt	ISRIC-WISE (AWC)	3	NA	NA	H2O (4)
PEPIC	Site (EPIC)	PS	RUE	PM	W (E), T, H, A, N, P	HIws Prt B	ISRIC-WISE	5	C N B(1) P(6)	Yes	Soil OM, C, NH3, NO3, H2O, P, CR (20)

[1] Site: site-base crop model; Eco: ecosystem model
[2] DA: Dynamic simulation based on development and growth processes; PS: prescribed shape of LAI curve as function of phenology, modified by water stress & low productivity
[3] RUE: Simple (descriptive) radiation use efficiency approach; P-R: Detailed (explanatory) gross photosynthesis-respiration
[4] TF: Turbulent Flux, PM: Penman-Montieth, HG: Hargreaves, PT: Priestly-Taylor, TE: Transpiration Efficiency
[5] W: water stress with (S)=water available in root zone and (E)=ratio of supply to demand; T: temperature stress; H: specific heat stress; A: aeration stress; N: nitrogen stress; P: phosphorus stress; K: potassium stress; BD: bulk density; AL: aluminum stress
[6] Yield formation depending on: HI: fixed harvest index; B: total (above ground) biomass; Gn: number of grains and grain growth rate; Prt: partitioning during reproductive stages; HIws: harvest index modified by water stress
[7] Major source of soil property input data and methods for manipulation to derive parameters required by the model; Albedo D: Albedo according to Dobos, 2006; AWC: Available Water Capacity (Van Genuchten et al., 1992) ; HYD USDA: hydraulic soil parameters according to USDA and NRCS, 2015; HYD C: hydraulic soil parameters according to Cosby et al., 1984; THM L: thermal parameters according to Lawrence and Slater, 2008; HWSD: Harmonized world soil database (Fischer et al., 2008); STC: soil texture classification based on the USDA soil texture classification (http://ufdc.ufl.edu/IR00003107/00001); ISRIC-WISE (Batjes, 2006) ; ROSETTA (Schaap and Bouten, 1996)
[8] Number of soil layers considered in simulation
[9] C: carbon/organic matter model; N: nutrient cycling model; P(x): x number of organic matter pools; B(x): x number of microbial biomass pools
[10] Modules affected by spin-up: OM: organic matter, C: carbon; NH3: ammonia; NO3: nitrate; H2O: soil water; P: phosphorus; CR: crop residues; (X) number of spin up years

Table S 1-2: Legend for Table 2 in main paper with brief explanation for parameters differing among EPIC-based GGCMs. Numbers in round braces refer to the number of subroutines available for estimating the respective model output. Routines considering the parameters are laid out in Section S1.1.

	No	Parameter	Description/Effect
Hydrology	1	PET estimation method (5)	Affects PET and hence water stress and hydrology
	2	Hargreaves exp. coefficient	Higher values increase importance of diurnal temperature range in Hargreaves equation
	3	Hargreaves linear coefficient	Linearly scales PET estimation in Hargreaves eq.
	4	Soil evaporation-cover coefficient	Affects evaporation from soil through surface albedo
	5	Soil cover-temperature function	Defines two points on a sigmoid curve describing soil cover effect on soil temperature
	6	Soil evaporation coefficient	Higher values increase soil evaporation from top 0.2m in exponential term
	7	Soil evaporation-depth function	Defines two points on a sigmoid curve describing evaporation from soil with depth
	8	Plant water use-soil water tension function	Defines two points on a sigmoid curve describing plant water use in relation to soil water tension
	9	Field capacity and wilting point estimation method (11)	Affects wide range of soil hydrologic and microbial processes by defining soil water holding capacity and saturation and a given water content
	10	Soil variable dependence of daily curve number (CN) estimate (5)	Daily CN estimates may depend on soil water with or without weighting by depth or be static
	11	CN number index coefficient	Regulates impact of PET in runoff retention parameter. Higher values increase runoff
	12	CN adjustment for standing dead residue	Higher values increase runoff if residue<1 t ha^{-1} and increase if >1 t ha^{-1}
Soil degradation	13	Wind erosion considered	Loss of top soil and total soil volume
	14	Water erosion considered	Loss of top soil and total soil volume
	15	Water erosion conservation practice	Lower values proportionally limit water erosion
	16	Water erosion estimation method (6)	Magnitude of soil erosion and differences in driving factors (e.g. rainfall force vs. runoff)
	17	Field length for wind erosion	Higher values increase wind erosion
	18	Field width for wind erosion	Higher values increase wind erosion
	19	Soil profile handling (static/dynamic)	Static soil profile is reinitialized at the beginning of each year, except for mineral nutrient pools
	20	Simulation continuity (transient/decadal)	Simulations are carried in a fully transient way or as in the GEPIC model separately for each decade
Organic matter and nutrient cycling	21	Denitrification method (3)	Affects magnitude and dynamics of denitrification
	22	Microbial decay rate	Higher value increases microbial turn-over off labile C pool
	23	Slow to passive humus coefficient	High values allocate more slow to passive humus, resulting in slower nutrient mobilization from OM
	24	Oxygen content-soil depth function	Defines two points on a sigmoid curve describing oxygen distribution with soil depth together with parameter 20 below
	25	Oxygen coefficient for microbial activity	Higher values decrease O_2 avail. with soil depth
	26	N volatilization coefficient	Fraction of potential nitrification and volatilization of NH_3 allocated to volatilization
Management	27	Automatic irrigation trigger	Potential biomass reduction factor on a given day that triggers application of irrigation water (e.g. 0.9 corresponds to reduction by ≥10%)
	28	Maximum single water application [mm]	Max. volume of water applied per irrigation event
	29	Automatic fertilizer application trigger[4)]	Value of potential biomass reduction factor on a given day that triggers application of fertilizer
Growth	30	Coefficient for lin. or exp. root growth	Allocates root growth to lin. and exp. functions
	31	Coefficient for root growth dist. by depth	Higher values increase root growth with soil depth
	32	Root growth stress considered	Higher values render root growth less sensitive to bulk density constraints (≥2 eliminates constraint)
	33	Fraction of growing season from which HI_{min} affects yield formation	Fraction of growing season in terms of PHU from which on water availability affects yield formation

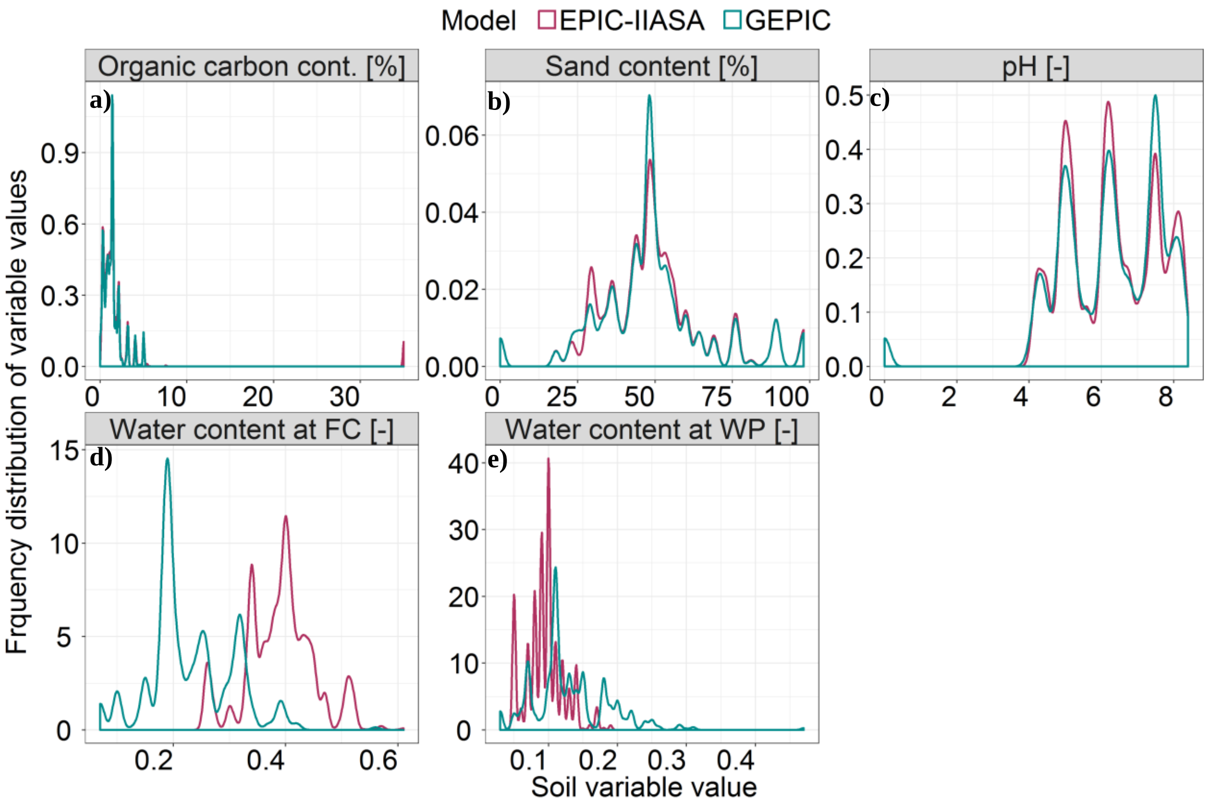

Figure S 1-3: Density distributions of key soil parameters in the original ISRIC WISE dataset used in GEPIC and the processed soil data used in EPIC-IIASA based on WISE. Water contents [-] at field capacity (FC) and wilting point (WP) are not provided in the original dataset and were estimated by different methods as specified in Table 2 (parameter 6) of the main paper. Very low pH values for WISE indicate the coincidence of water bodies with the land mask. Organic soil such as histosols do not exist in the original soil data and were parameterized for EPIC-IIASA only as indicated by the high OC values occurring in this dataset.

Table S 1-3: Crop management operations of EPIC-IASA and GEPIC

Operation	EPIC-IIASA	GEPIC
Fertilizer application[1]	Fixed P	Fixed P
Ploughing	Moldboard plough Tillage depth=150 mm Surface roughness=30 mm Mixing efficiency=99%	Tandem disk Tillage depth=40 mm Surface roughness=50 mm Mixing efficiency=75%
Planting	Regular planter Tillage depth=40 mm Surface roughness=10 mm Mixing efficiency=10% Row spacing=1.0 m Ridge height=75 mm	Regular planter Tillage depth=20 mm Surface roughness=10 mm Mixing efficiency=10% Row spacing=0.5 m Ridge height=75 mm
In-season operations		2x row cultivation (at 10% and 25% growing season) Tillage depth=25 mm Surface roughness=15 mm Mixing efficiency=25%
Harvest	Combined harvester	Combined harvester
Residue handling	No residue removal	Removal of approx. 80% stover

[1]N fertilizer is applied automatically in both models based on plant stress as specified in Table 2 of the main article

Table S 1-4: Parameterization of different maize cultivars used in the GGCMs as shown in Figure 1 of the manuscript with corresponding coloring of column headings. Cultivar 1 is the default in the EPIC model and corresponds to a high-yielding variety. Cultivar 2 has been calibrated for applications in Europe (Cabelguenne et al., 1999). Cultivar 3 is a faster maturing version of Cultivar 1. Cultivar 4 has been parameterized for West Africa and North-Eastern Brazil (Gaiser et al., 2010). TBS=base temperature for plant growth, TOP=optimum temperature, HI_{max}=maximum harvest index without water stress, HI_{min}=minimum harvest index under water stress. The temperature requirement (GDD), commonly used for distinguishing cultivars, is here prescribed by the growing season length of the spatially explicit planting and harvest dates.

Parameter	Cutivar1	Cultivar1b	Cutivar2	Cutivar3	Cutivar4
TBS [°C]	8	8	6.5	8	8
TOP [°C]	25	25	22.5	25	25
HI_{max} [-]	0.5	0.55	0.5	0.5	0.35
HI_{min} [-]	0.4	0.4	0.4	0.4	0.01

S2 Input and evaluation data

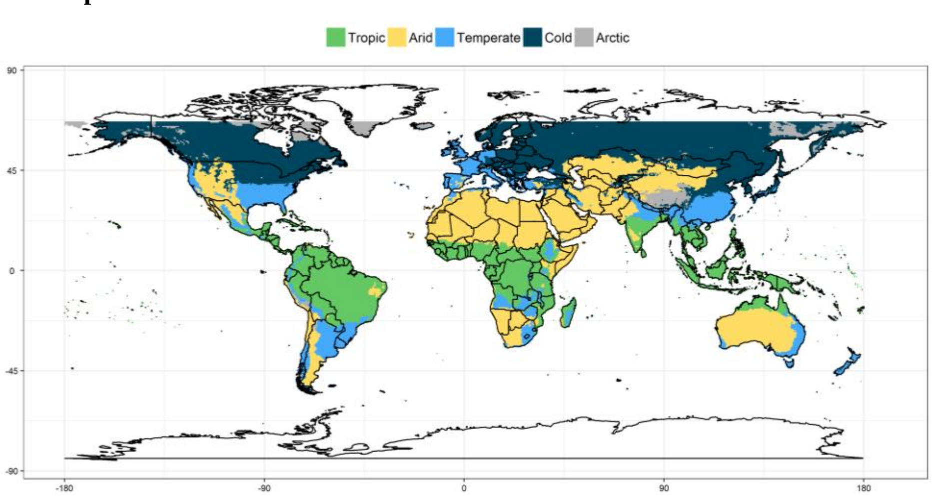

Figure S 2-1: Major Koeppen-Geiger climate regions according to (Peel et al., 2007) based on the climate data used in this study. Climate data were available only up to 67°N above which no harvest areas for maize or wheat are reported.

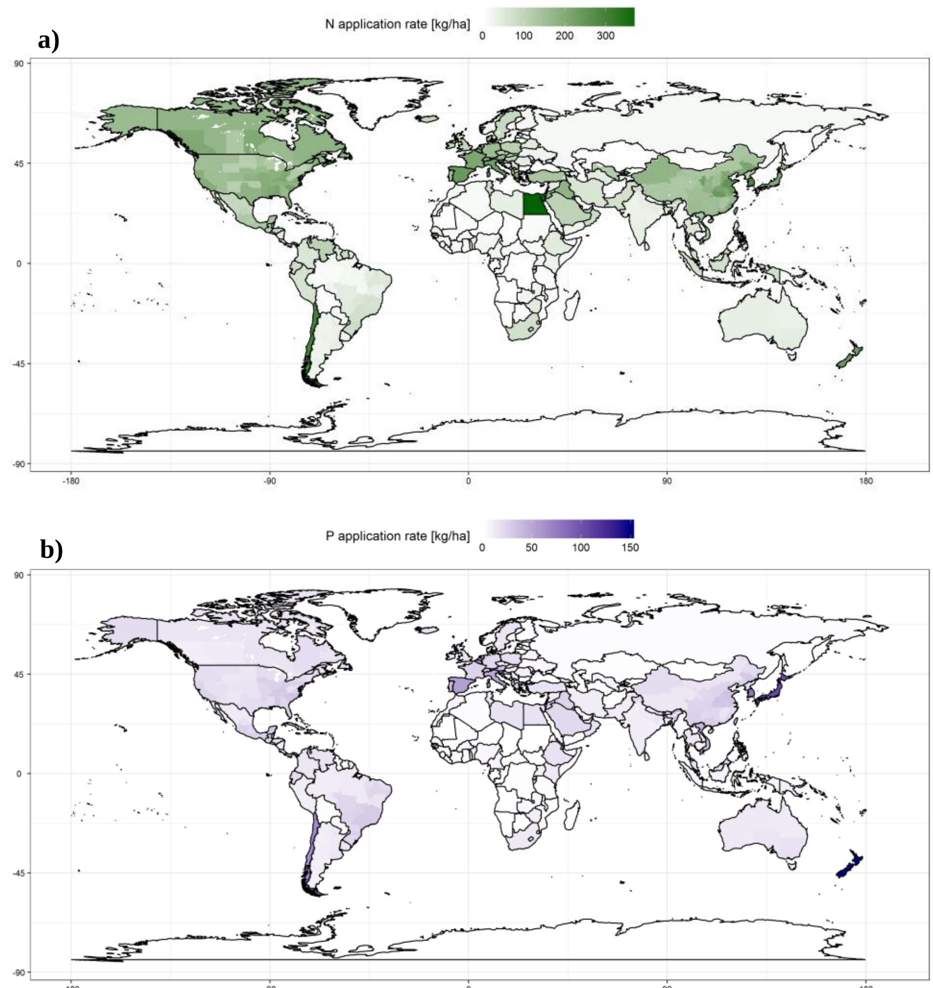

Figure S 2-2: (a) Nitrogen and (b) phosphorus fertilizer application rates used in the fullharm setup (Elliott et al., 2015)

.

S3 Supplementary results

Table S 3-1: Relative spread of maize yield estimates measured as yields of the highest estimate in relation to yields of the lowest estimate in Figure 2 of the main paper. See Table 1 of the main paper for management scenarios

Management	Relative range of maize yield estimates	
	Maximum [%]	Mean [%]
default	124	95
fullharm	55	41
harm-suffN	26	18

Table S 3-2: Statistical coefficients for linear regressions of yield estimates over time in Figure 2 of the main article and in Figure S 3-1 and mean error [t ha^{-1}] compared to reported yields.

Model	Intercept	Slope	R^2	p	Mean error	Scenario and panel in Figure 2 and Figure S 3-1
EPIC-BOKU	8.068	-0.001	0.000	0.909	3.676	default panel a
EPIC-IIASA	6.178	0.006	0.114	0.068	1.891	
EPIC-TAMU	5.454	-0.005	0.045	0.258	0.987	
GEPIC	4.128	0.001	0.001	0.901	-0.246	
PEPIC	5.100	-0.018	0.376	0.000	0.439	
EPIC-BOKU	6.861	-0.039	0.660	0.000	1.880	fullharm panel b
EPIC-IIASA	6.399	0.005	0.073	0.150	2.096	
EPIC-TAMU	5.454	-0.005	0.045	0.258	0.987	
GEPIC	4.611	-0.001	0.001	0.896	0.216	
PEPIC	5.162	-0.017	0.314	0.001	0.508	
EPIC-BOKU	7.409	-0.063	0.823	0.000	2.049	harm-suffN panel c
EPIC-IIASA	6.747	0.006	0.075	0.143	2.450	
EPIC-TAMU	6.773	0.007	0.064	0.178	2.501	
GEPIC	5.895	0.005	0.026	0.395	1.596	
PEPIC	6.070	-0.015	0.218	0.009	1.459	

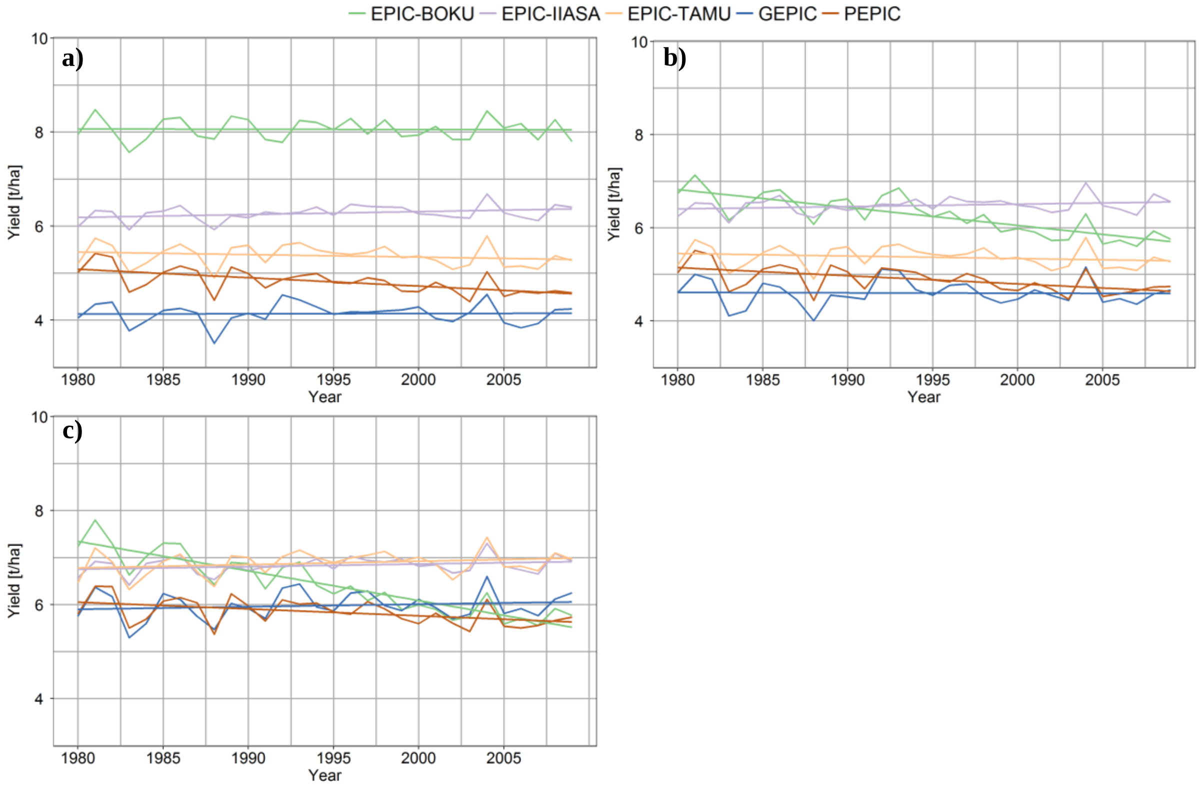

Figure S 3-1: Same as Figure 2 in the main manuscript but with linear regressions included but without ensemble mean and reported yields. Regression coefficients are presented in Table S 3-2 for better readability.

Table S 3-3: Quantiles of coefficient of variation [%] among EPIC-GGCMs for grid-wise maize yield estimates (Figure 3 in main article) depending on the setup and management scenarios (see Table 1 in main article).

Management scenario	Irrigation regime	0%	25%	50%	75%	100%
default	irrigated	5.14	24.27	44.21	70.95	223.61
default	rainfed	3.85	30.66	52.51	77.33	223.61
fullharm	irrigated	2.49	23.28	39.01	53.89	223.61
fullharm	rainfed	2.89	28.82	44.52	63.63	223.61
harm-suffN	irrigated	3.22	18.26	25.23	33.94	223.61
harm-suffN	rainfed	4.38	20.82	28.51	46.73	223.61

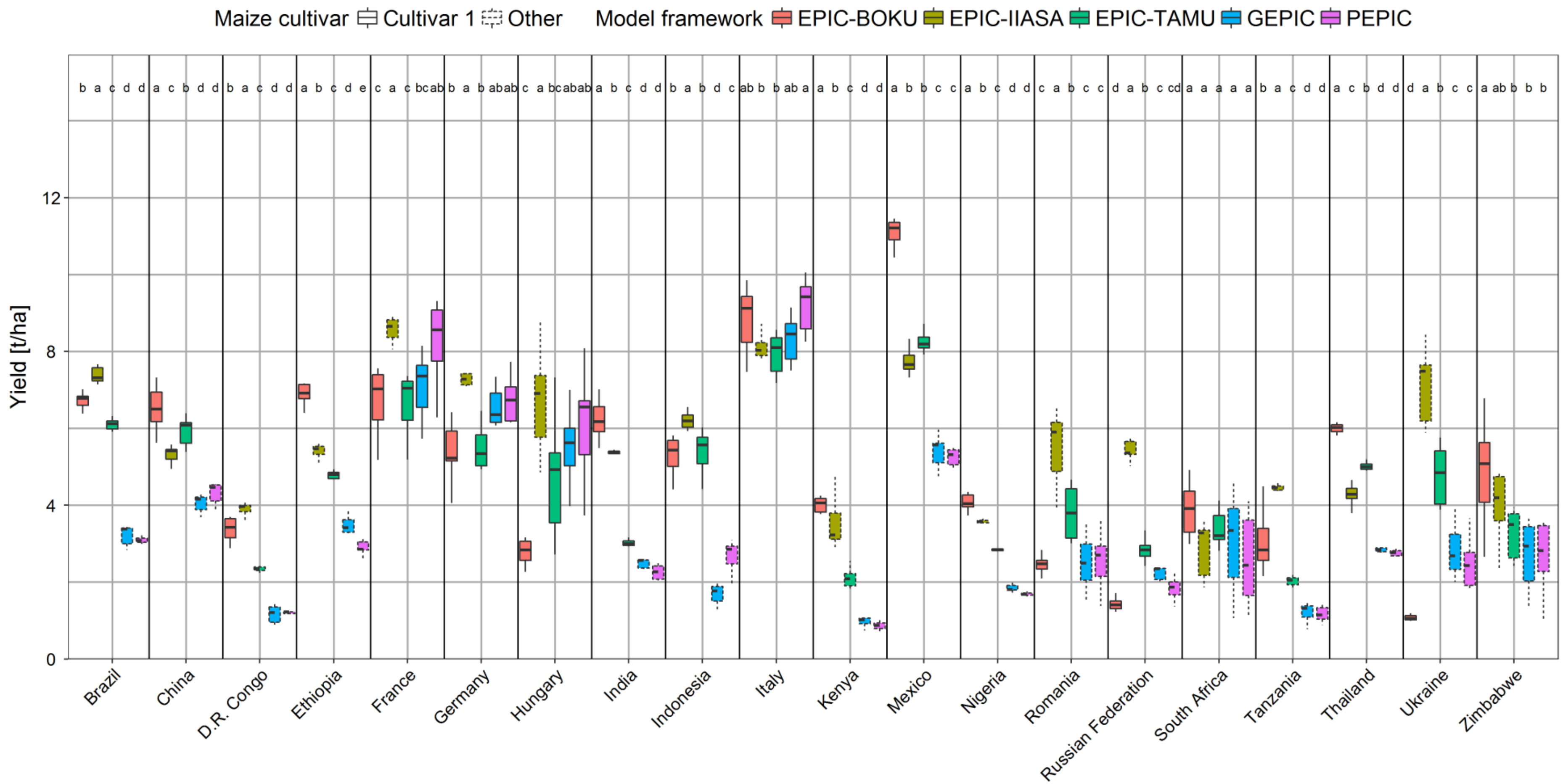

Figure S 3-2: Box-and-whisker plots of national annual maize yield estimates throughout the study period for the fullharm scenario in countries, in which the attribution of cultivars (see Figure 1a-d in main paper) differs by >30% among at least two EPIC-based GGCMs. Only the 20 countries with the largest harvest areas for maize and the specified differences in maize cultivar attribution are displayed. The cultivar "Other" refers to cultivar 2 for EPIC-IIASA in European countries and Russia (Figure 1a). Otherwise the cultivar "Other" refers to cultivar 4. Letters indicate significantly different groups according to Tukey's HSD test.

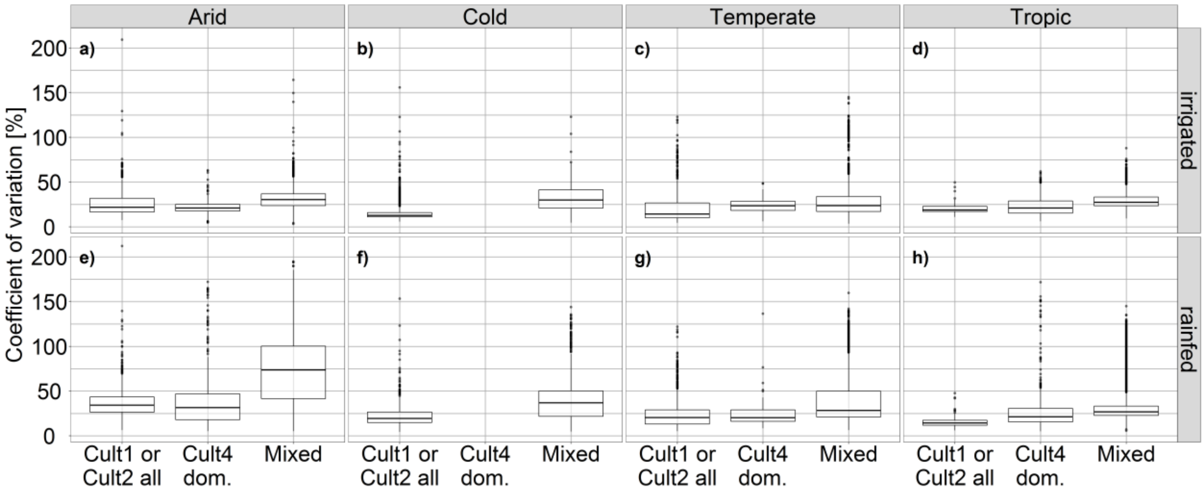

Figure S 3-3: Coefficient of variation (CV_{av}) among maize yield estimates in the harm-suffN scenario in grid cells in which either all GGCMs plant the high-yielding cultivars 1 or 2 (Figure 1 in main paper) or in which at least four GGCMs plant the low-yielding drought-sensitive cultivar 4 or in which cultivar types are mostly mixed.

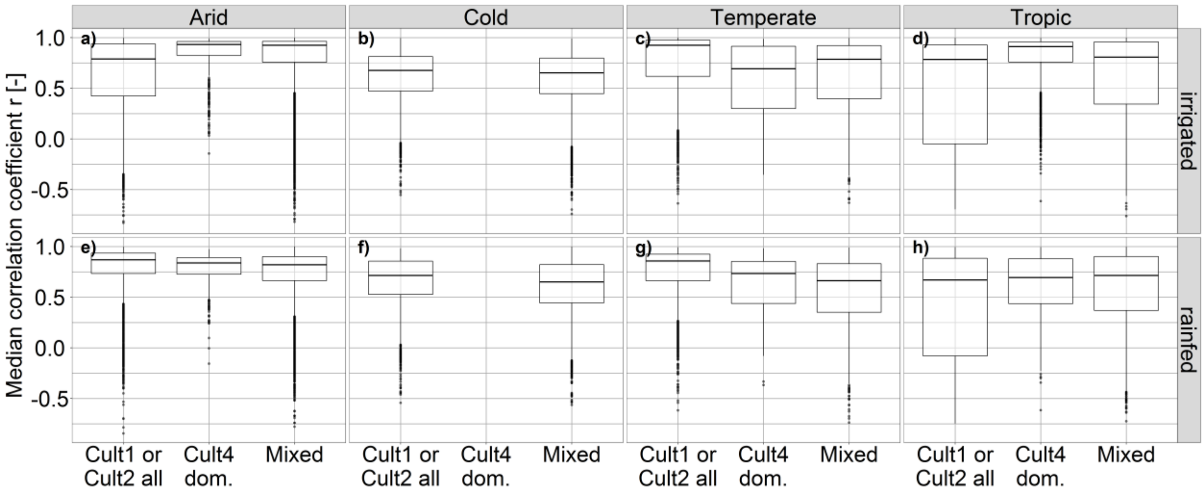

Figure S 3-4: Median time-series correlation coefficient in the harm-suffN scenario in grid cells in which either all GGCMs plant the high-yielding cultivars 1 or 2 (Figure 1 in main paper and Table S 1-4) or in which at least four GGCMs plant the low-yielding drought-sensitive cultivar 4 or in which cultivar types are mostly mixed.

Table S 3-4: Fractions of grid cells [%] in which the median time series correlation among the EPIC-based GGCMs (Figure 4 in main article) fulfils a certain level of significance.

Management scenario	Water supply	p<0.1	p<0.05	p<0.01
default	irrigated	16.38	10.41	3.14
default	rainfed	41.43	32.96	17.11
fullharm	irrigated	24.38	17.37	8.58
fullharm	rainfed	53.59	45.40	29.26
harm-suffN	irrigated	69.53	63.92	52.51
harm-suffN	rainfed	69.18	61.57	44.85

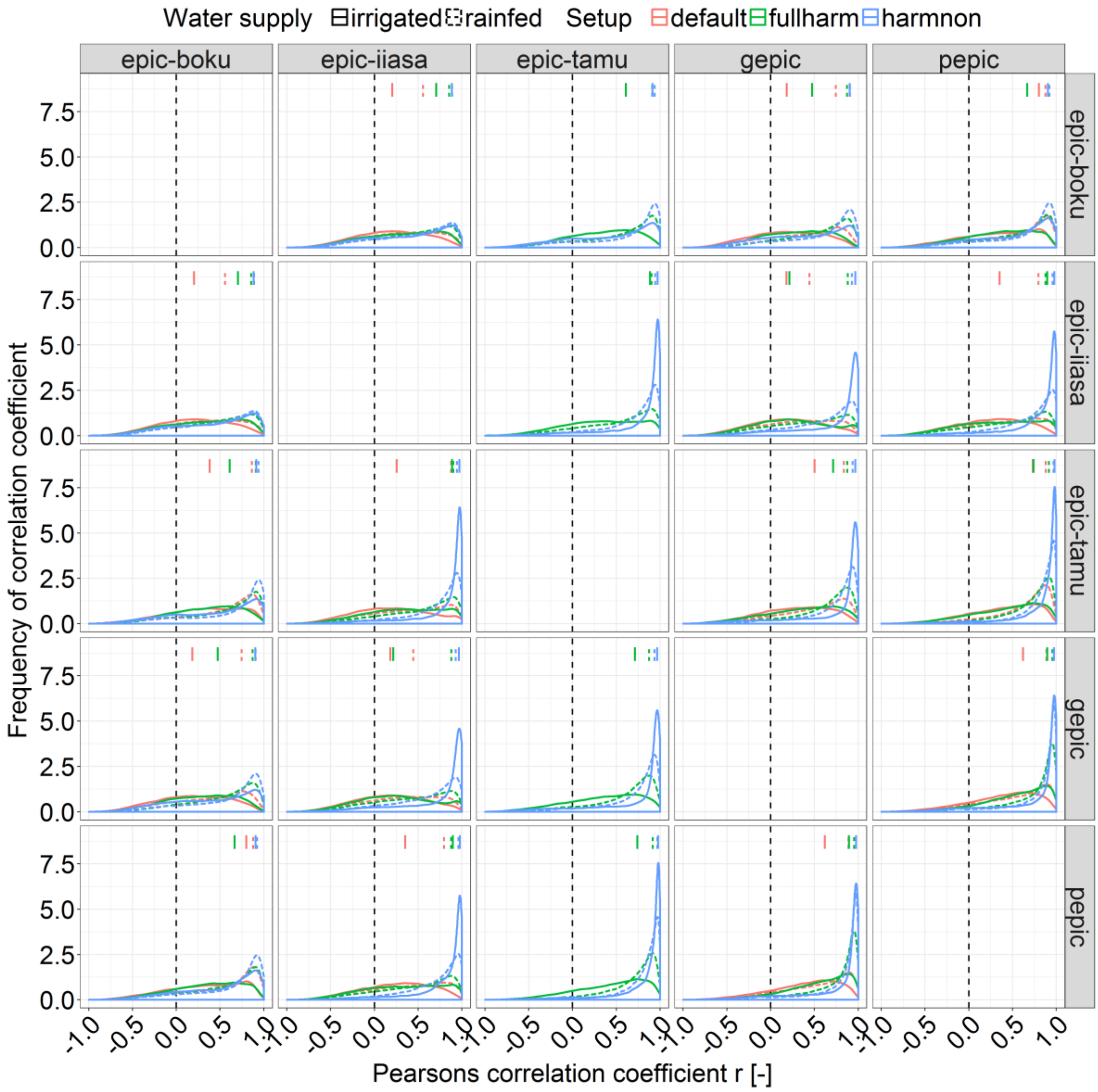

Figure S 3-5: Frequency distribution of time-series correlation coefficients among EPIC-based GGCMs in each grid cell and for each management scenario.

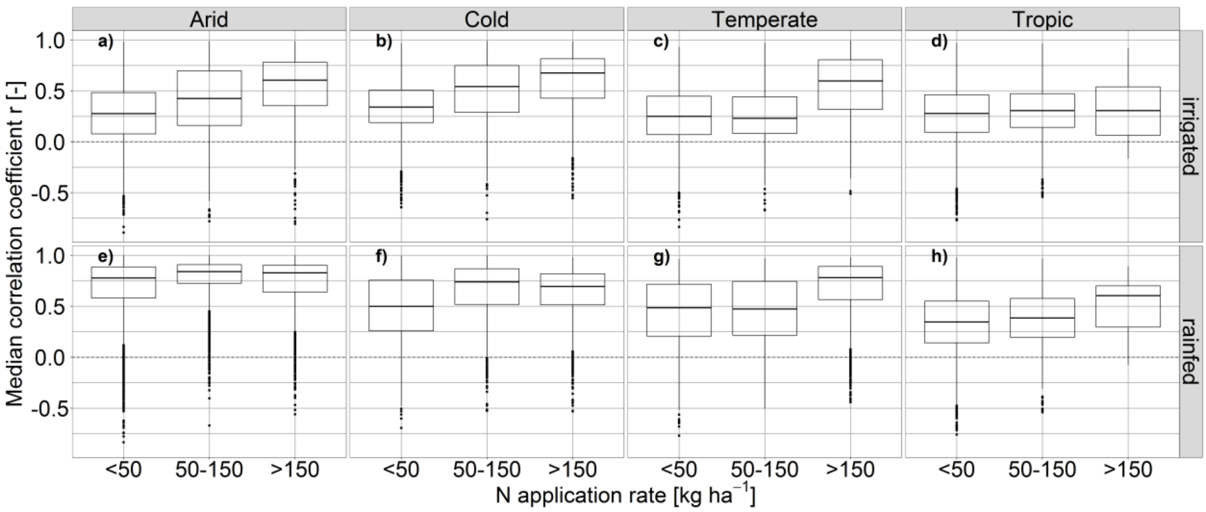

Figure S 3-6: Median time-series correlation coefficient r for maize yields among EPIC-based GGCMs compared to binned fertilizer application rates in the fully harmonized management scenario (fullharm) with sufficiently irrigated (a-d) or rainfed (e-h) water supply in each grid cell of four major climate regions.

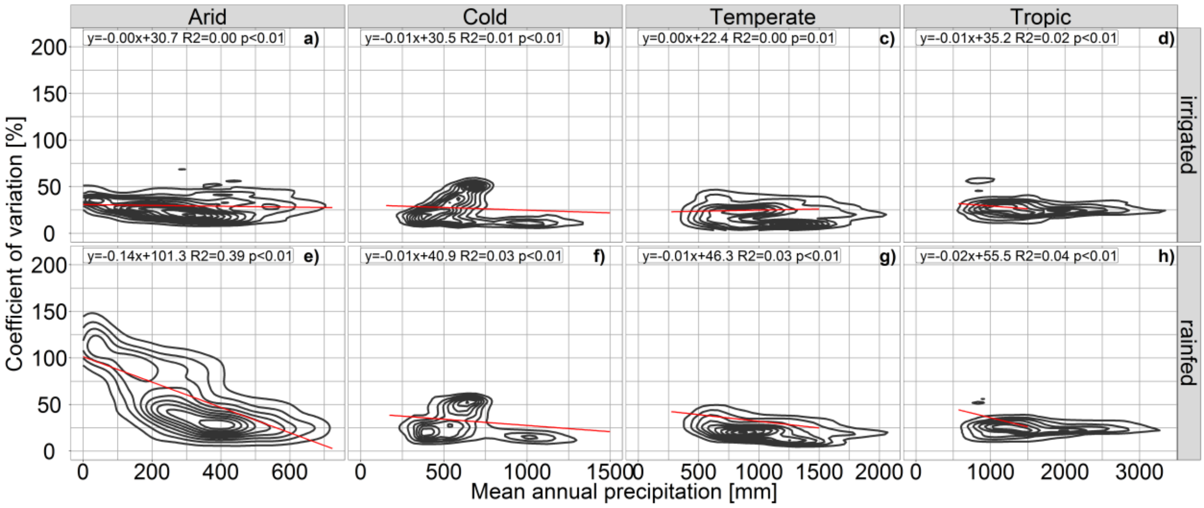

Figure S 3-7: Coefficient of variation for maize yields among EPIC-based GGCMs compared to mean annual precipitation (MAP) in the harmonized management scenario with sufficient nutrient supply (harm-suffN) with sufficiently irrigated (a-d) or rainfed (e-h) water supply in each grid cell of four major climate regions. Linear regressions are limited to MAP≤1500 mm which commonly corresponds to sufficient water supply.

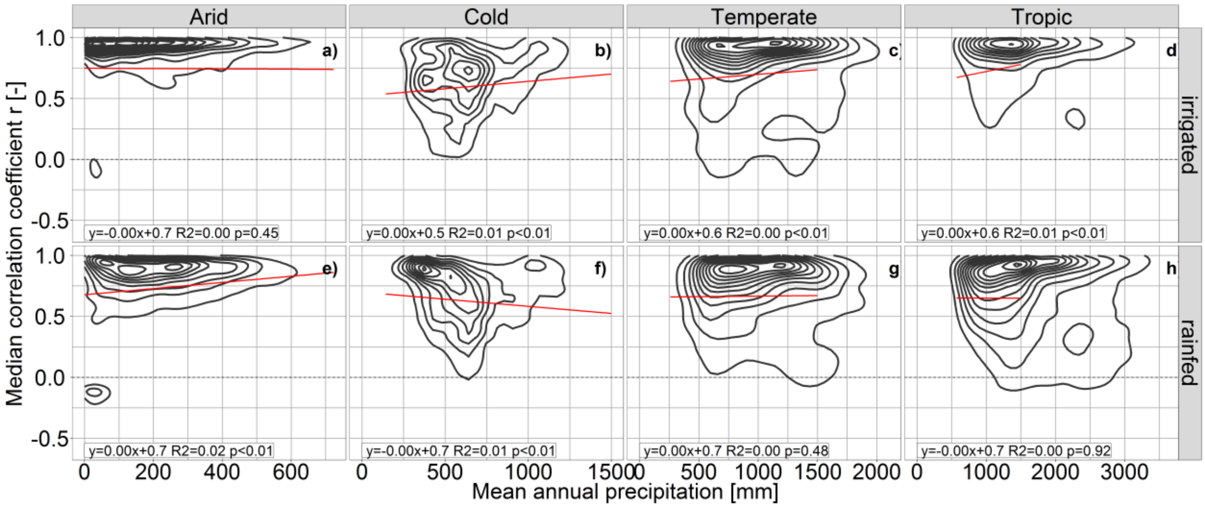

Figure S 3-8: Median time-series correlation coefficient r for maize yields among EPIC-based GGCMs compared to mean annual precipitation (MAP) in the harmonized management scenario with sufficient nutrient supply (harm-suffN) with sufficiently irrigated (a-d) or rainfed (e-h) water supply in each grid cell of four major climate regions. Linear regressions are limited to MAP≤1500 mm which commonly corresponds to sufficient water supply.

Table S 3-5: Time-series correlation coefficient r for each GGCM in the ten major maize producing countries for the fullharm and harm-suffN scenarios (Table 1 in main paper) and annual N fertilizer application rates for maize in each country.

Country	Best	EPIC-BOKU	EPIC-IIASA	GEPIC	EPIC-TAMU	PEPIC	Country median	N rate [kg ha^{-1}]	
United States	0.878	0.737	0.878	0.707	0.757	0.772	0.757	163	fullharm
China	0.711	0.711	0.388	0.549	0.594	0.563	0.563	186	
Brazil	0.331	0.293	0.224	0.331	0.247	0.036	0.247	61	
Argentina	0.796	0.234	0.796	0.349	0.564	0.567	0.564	31	
Mexico	0.328	0.085	0.067	0.216	0.302	0.328	0.216	120	
India	0.777	0.246	0.777	0.395	0.411	0.368	0.395	35	
Ukraine	0.550	0.142	0.308	0.413	0.463	0.550	0.413	22	
Indonesia	0.111	0.019	0.111	0.056	0.082	0.021	0.056	55	
France	0.848	0.786	0.716	0.782	0.832	0.848	0.786	201	
South Africa	0.622	0.513	0.622	0.609	0.508	0.507	0.513	70	
GGCM median	**0.666**	**0.269**	**0.505**	**0.404**	**0.486**	**0.528**	-	-	
United States	0.870	0.729	0.870	0.724	0.767	0.769	0.767	163	harm-suffN
China	0.710	0.710	0.391	0.549	0.582	0.564	0.564	186	
Brazil	0.381	0.278	0.315	0.351	0.381	0.238	0.315	61	
Argentina	0.787	0.409	0.787	0.672	0.756	0.647	0.672	31	
Mexico	0.270	0.270	0.099	0.192	0.153	0.232	0.192	120	
India	0.809	0.338	0.809	0.550	0.544	0.594	0.550	35	
Ukraine	0.490	0.135	0.316	0.408	0.388	0.490	0.388	22	
Indonesia	0.097	0.063	0.087	0.002	0.049	0.097	0.063	55	
France	0.849	0.789	0.720	0.826	0.832	0.849	0.826	201	
South Africa	0.690	0.432	0.690	0.678	0.664	0.557	0.664	70	
GGCM median	**0.700**	**0.373**	**0.540**	**0.550**	**0.563**	**0.560**	-	-	

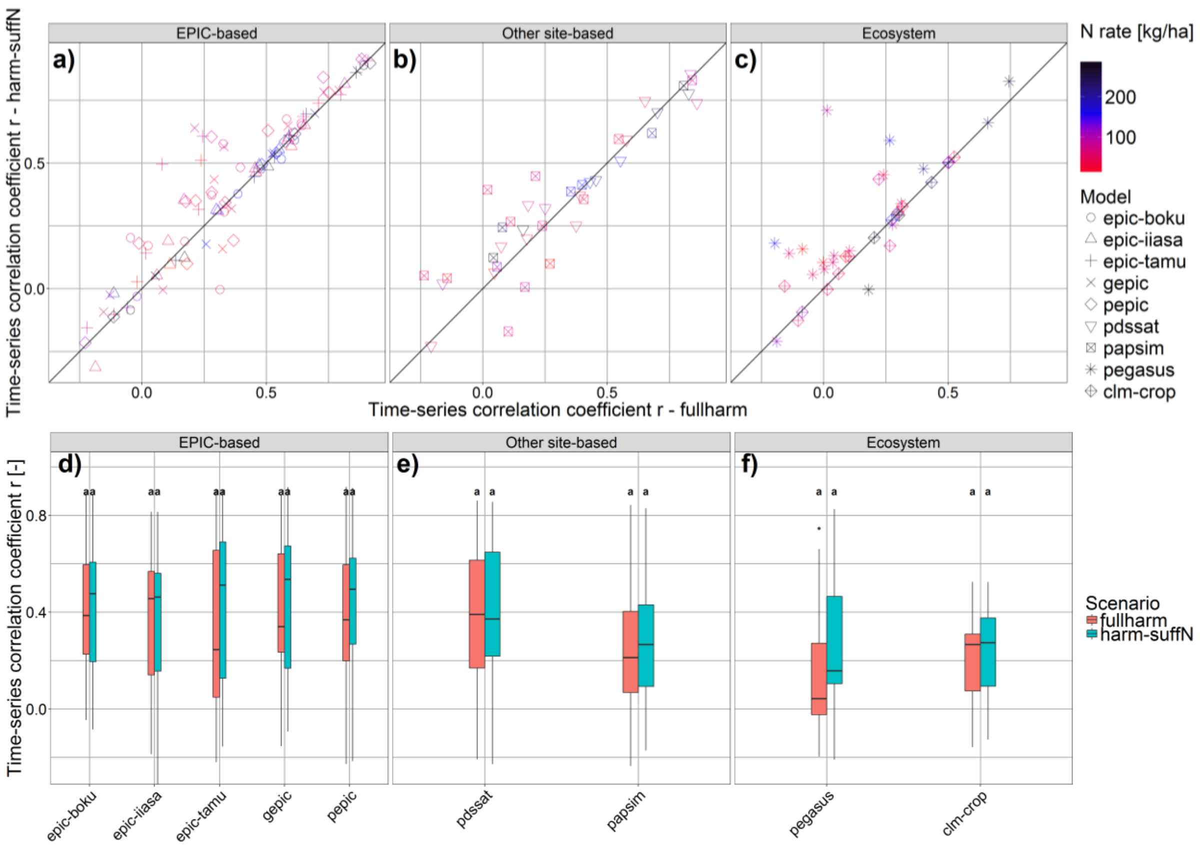

Figure S 3-9: Differences in model performance between the fullharm and harm-suffN scenarios per GGCM and GGCM type (Table S 1-1) for 19 maize producing countries for which production data have not been estimated by FAO and harvest area did not change by >100% during the study period. (a)-(c): Direct comparison of time-series correlation coefficients with different shapes for each model and color scale representing national N application rates. (d)-(f): Boxplots for each GGCM and setup scenario spanning the 19 correlation coefficients in each scenario. Letters above boxplots indicate a significant difference between the results of the fullharm and harm-suffN scenarios based on ANOVA/Tukey's HSD test.

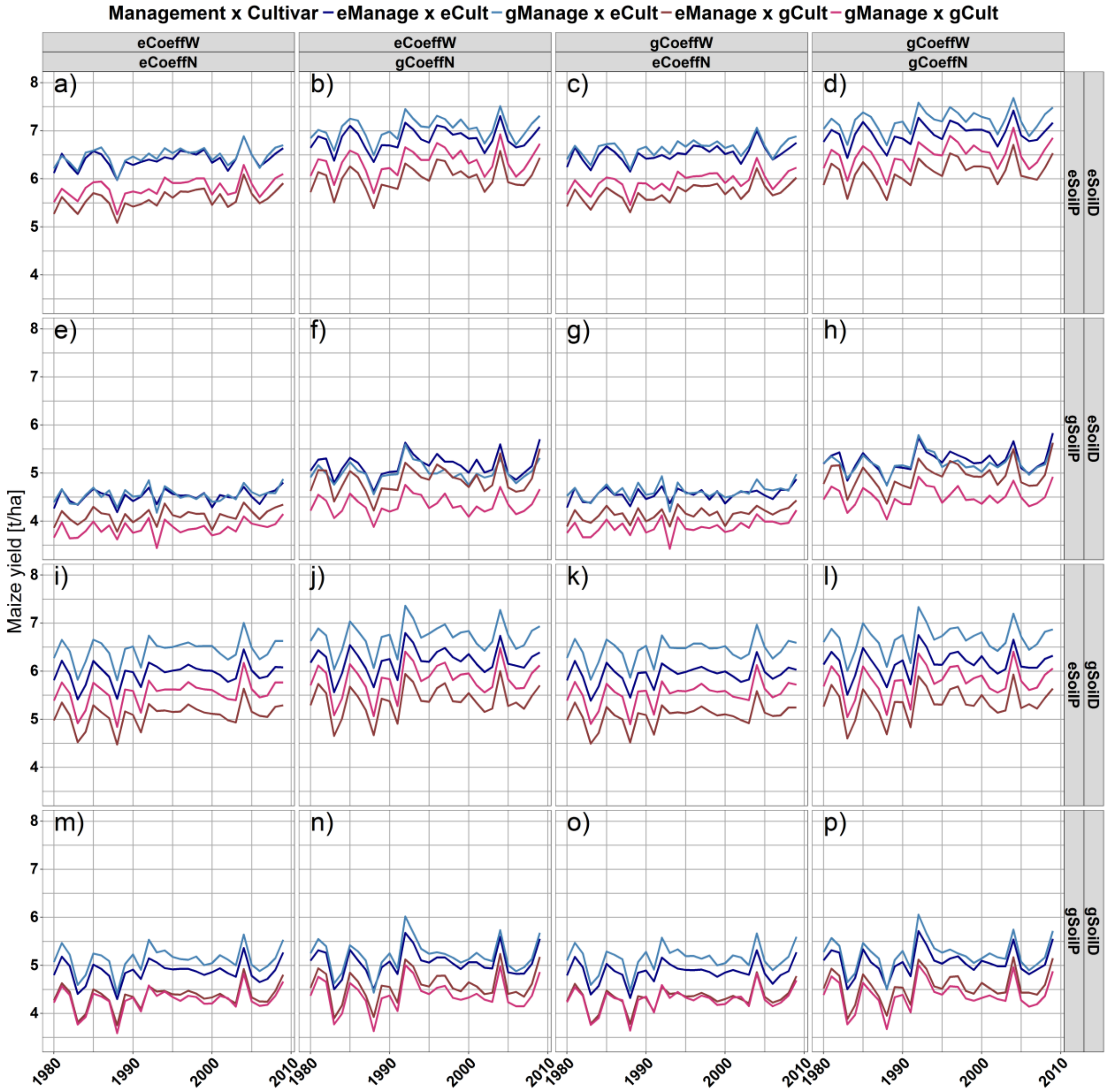

Figure S 3-10: Global average rainfed maize yields over a 29 year time period for 64 setup combinations based on the EPIC-IIASA and GEPIC setups (Table 3 in main article). e=EPIC-IIASA, g=GEPIC, Cult=cultivar definition and distribution, SoilD=soil data, SoilP=spin-up and soil handling, CoeffN=organic matter and nutrient cycling coefficients, CoeffW=hydrologic coefficients, Manage=crop management.

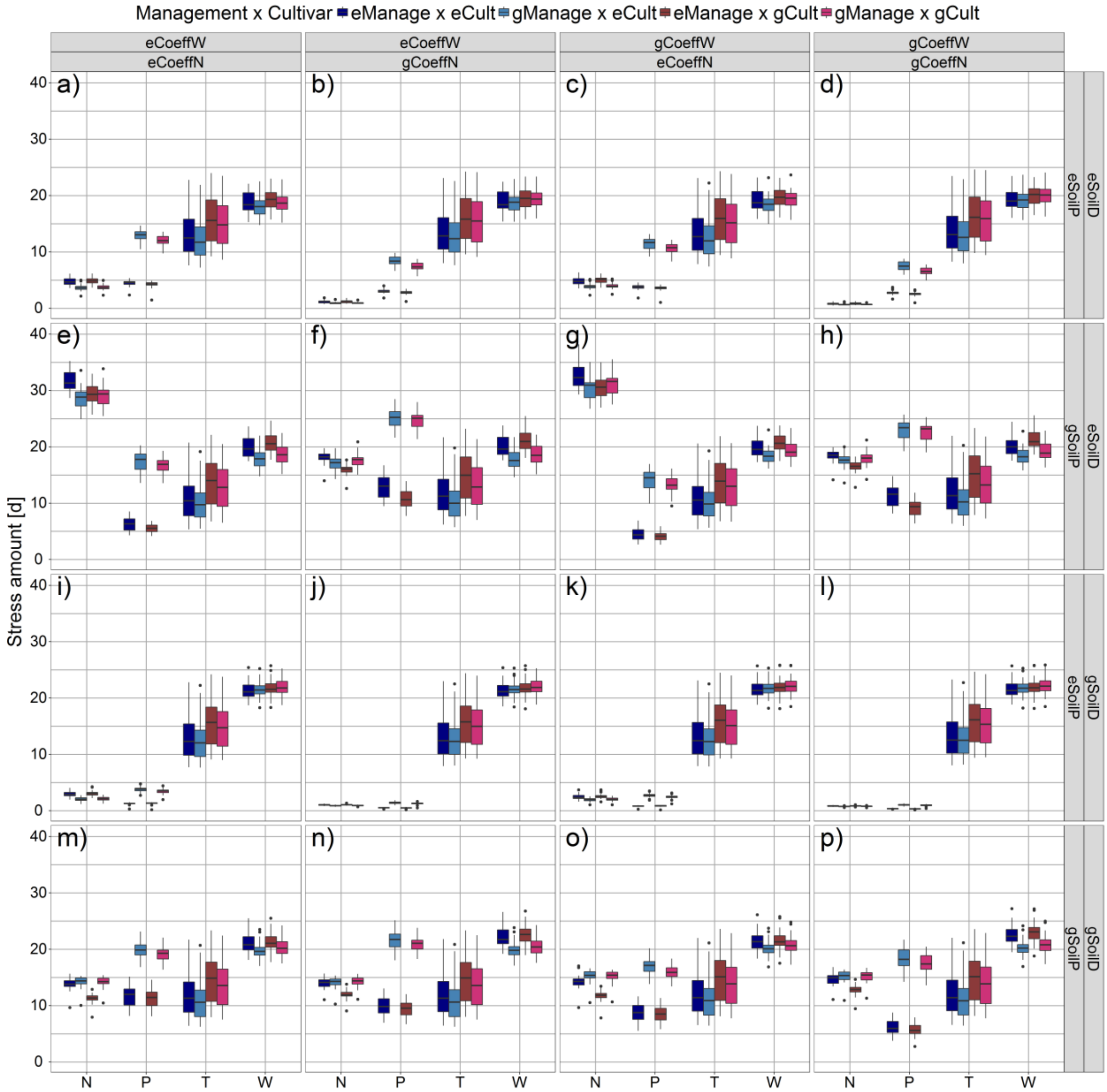

Figure S 3-11: Box-and-whisker plots of global averaged growth stresses over a 29 year time period for 64 setup combinations based on the EPIC-IIASA and GEPIC setups under rainfed conditions (Table 3 in main paper). e=EPIC-IIASA, g=GEPIC, Cult=cultivar definition and distribution, SoilD=soil data, SoilP=spin-up and soil handling, CoeffN=organic matter and nutrient cycling coefficients, CoeffW=hydrologic coefficients, Manage=crop management. The corresponding yields are shown in Figure S 3-10 and their relative difference from the EPIC-IIASA setup in Figure 7 in the main paper.

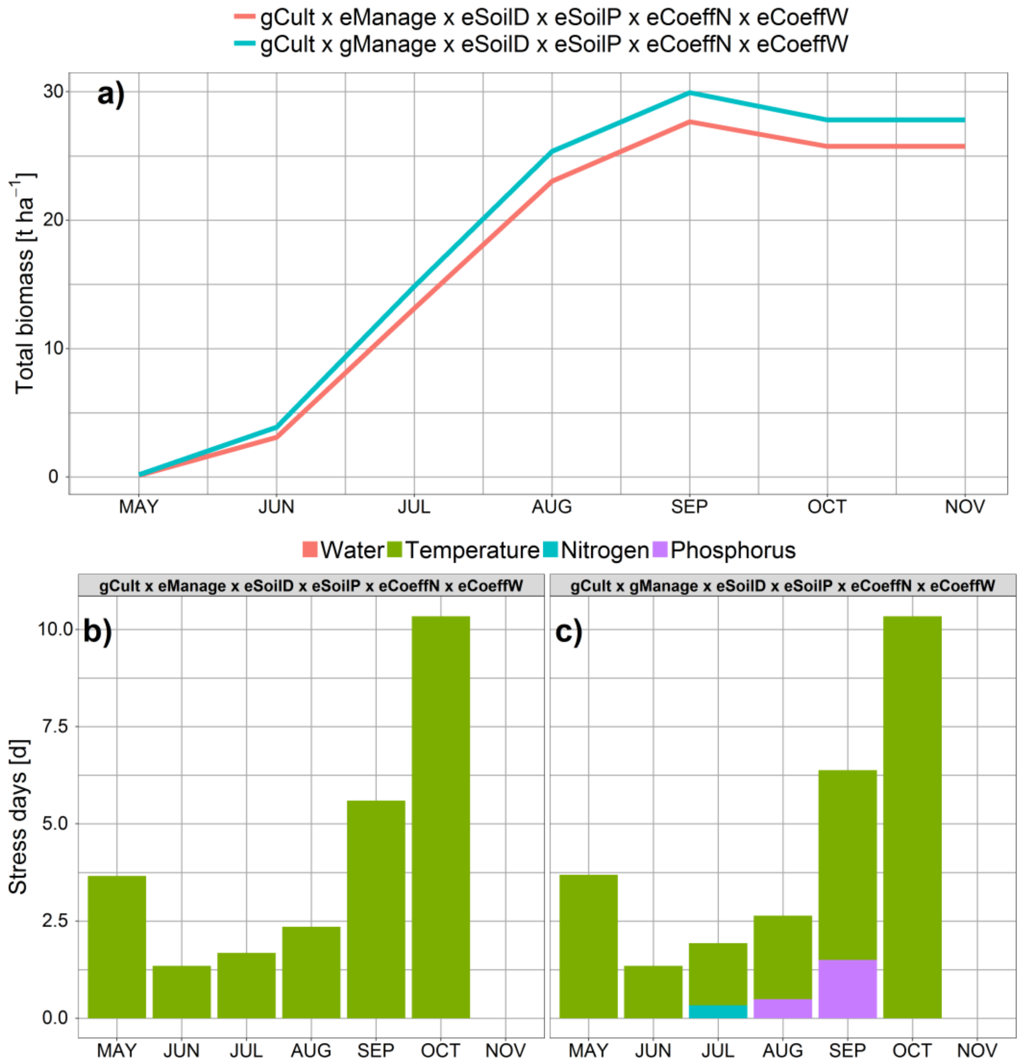

Figure S 3-12: (a) Monthly total biomass and (b-c) stress occurrence for a single year in a randomly sampled grid cell of the US Corn Belt differing the managements of EPIC-IIASA and GEPIC (eManage/gManage) with otherwise identical setups (included in Figure 9a of the main article) and a static soil profile (eSoilP). The GEPIC management results in higher biomass at any point of time despite higher stress occurrence (panel c) due to narrower row spacing than in EPIC-IIASA, which increases potential biomass estimation. Nutrient-related stresses occur only for the GEPIC management, which includes plant residue removal after harvest (Table S 1-3). This management hence causes a two-fold impact on nutrient stress with overall lower availability of nutrients in the long-run under low-input conditions and higher biomass accumulation - resulting in stronger nutrient mining - early in the growing season.

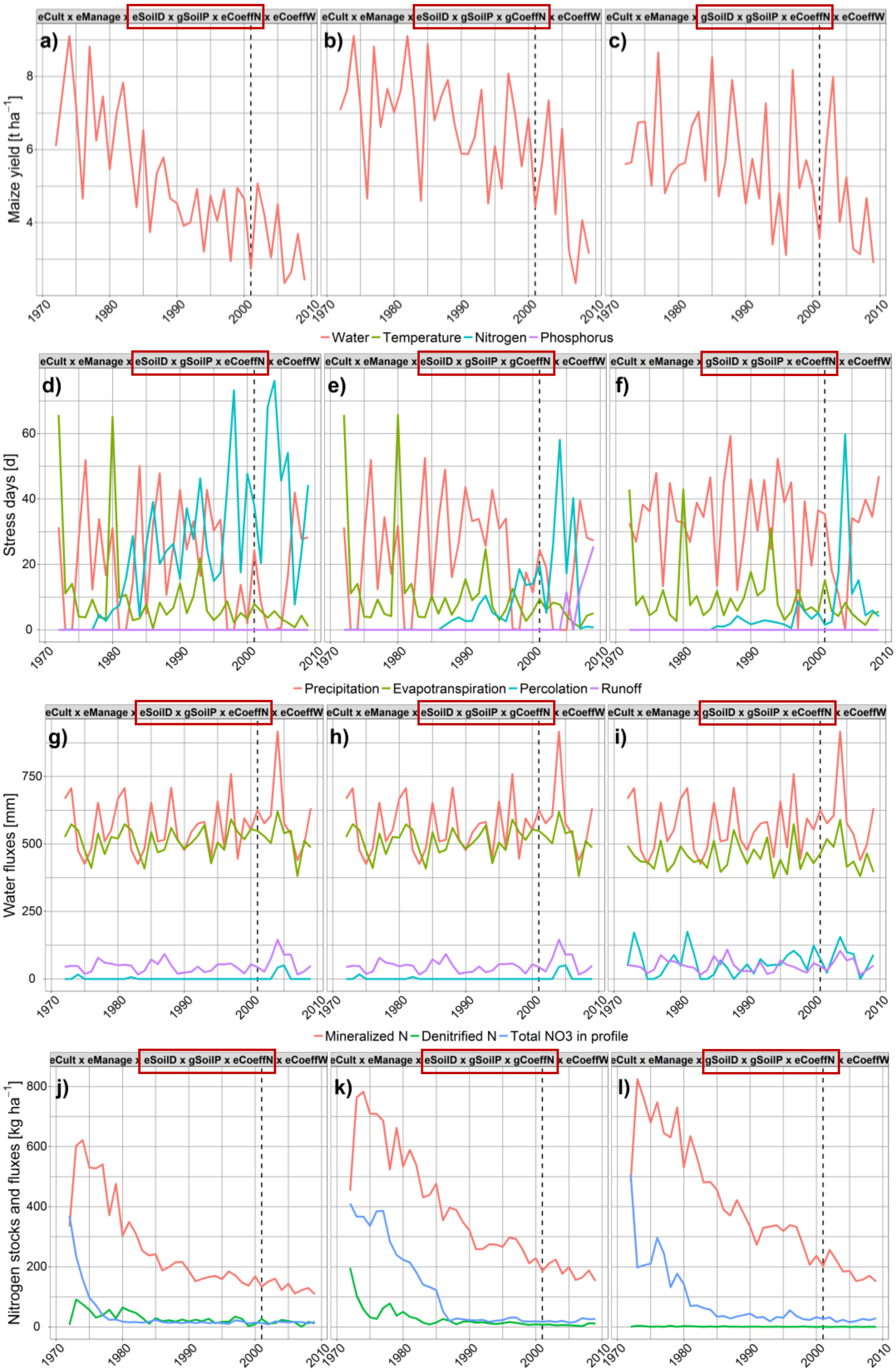

Figure S 3-13: Annual (a-c) yield estimates, (d-f) stresses, (g-i) water fluxes, and (j-l) nitrogen fluxes for a randomly sampled grid in Eastern Ukraine with low fertilizer application for three EPIC-GGCM setups differing in soil data (SoilD) and OM/nutrient cycling parameterization (CoeffN) using dynamic soil profile handling (gSoilP). Scenarios correspond to Figure 9e (left column of panels), Figure 9f (center column of panels), and Figure 9m (right column of panels) in the main article. The vertical dashed line indicates the end of the spin-up period (Figure S 1-2). The setup in the left column causes low N mineralization and early and rapidly increasing in N stress, which is lower if the nutrient coefficients of GEPIC (center column) are used due to more rapid turnover of OM. If the soil data of GEPIC are used (right vs. left panel column), water stress is

constantly high due to lower field capacity (Figure S 1-3), which in turn also triggers earlier the mineralization of OM and causes lower nutrient stresses in the model.

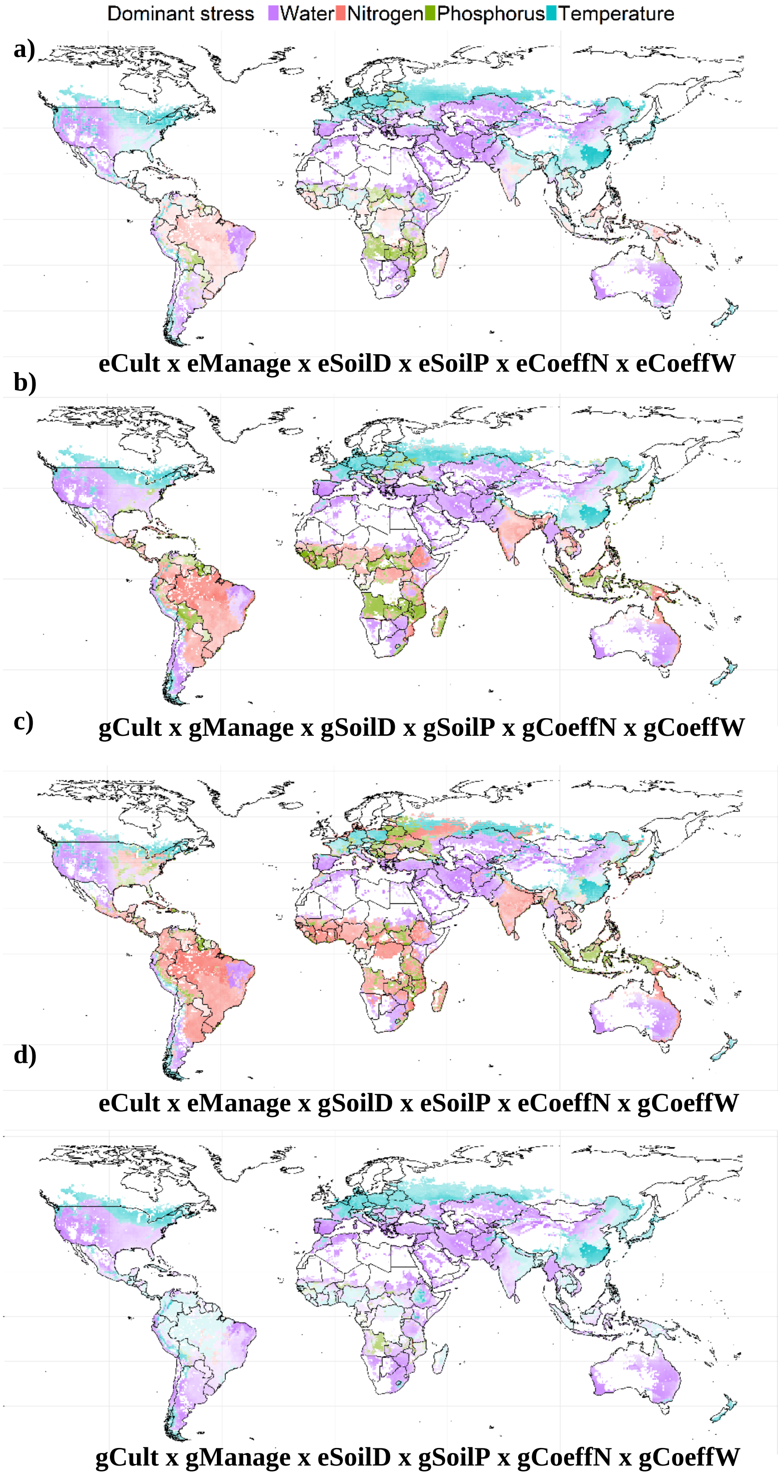

Figure S 3-14: Dominant stress per grid cell averaged over the simulation period in four selected setups shown underneath each panel. The transparency of each grid is proportional to the relative magnitude of the stress within the setup. The EPIC-IIASA setup (a) causes a dominance of climate-related stresses in most parts of the world, except for parts of the tropics. With the GEPIC setup (b) nutrient-related stresses are higher and have more coverage throughout the tropics and stretch out to parts of temperate low-input regions while water stress stretches to higher latitudes in the US. The most extensive nutrient stress occurs if the nutrient cycling coefficient of EPIC-IIASA are combined with the dynamic soil handling and management of GEPIC (c). If the GEPIC setup is run with static soil handling of EPIC-IIASA (d), hardly any nutrient-related stresses occur. The corresponding panels in Figure 9 of the main article are a ➔ a, b ➔ p, c ➔ k, d ➔ h.

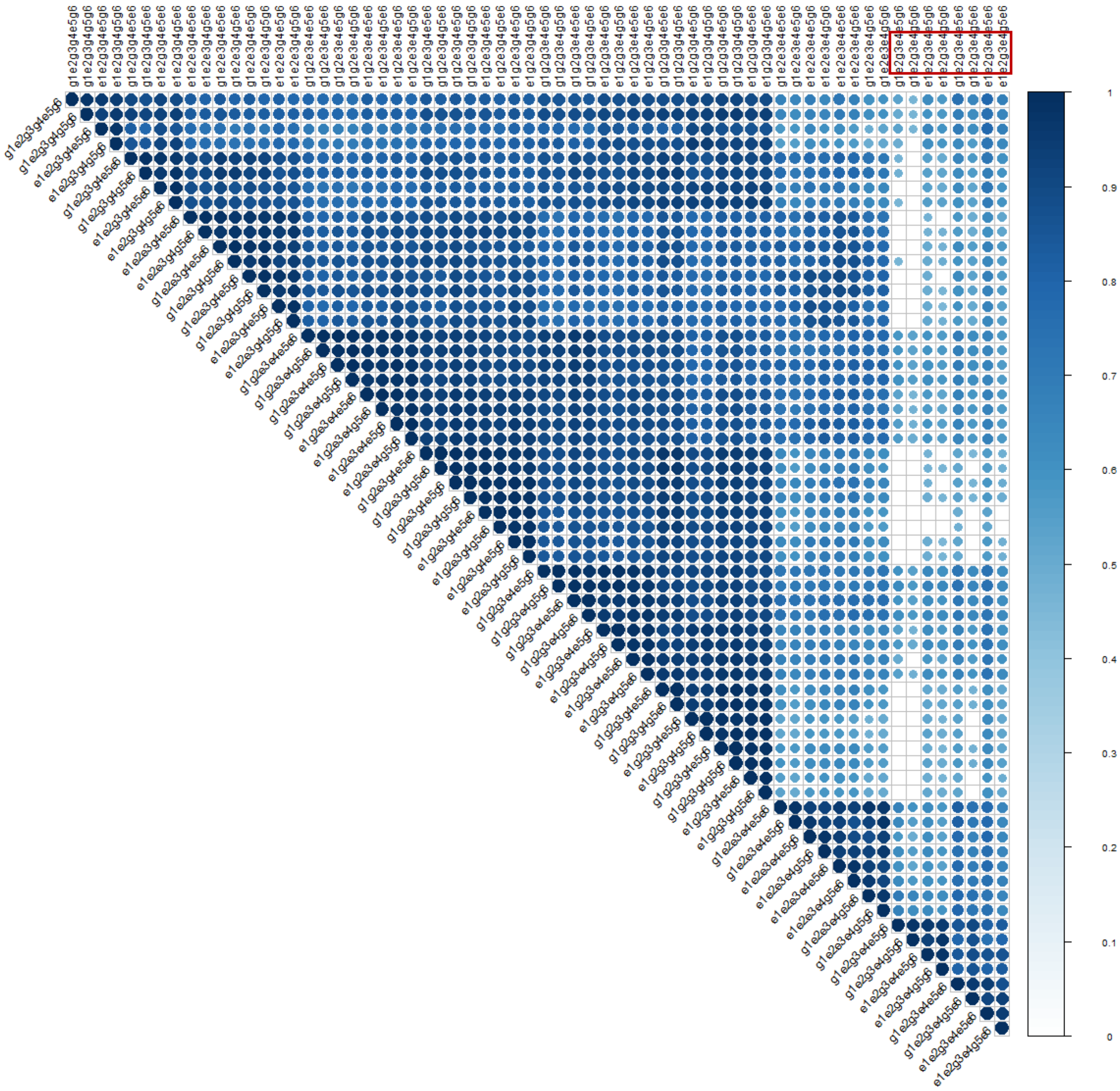

Figure S 3-15: Correlation matrix for the 64 setup permutation of EPIC-IIASA and GEPIC. Colour indicates the correlation coefficient r as shown on the right scale, circle sizes represent the level of significance. Insignificant correlations are excluded. The setups are ordered based on hierarchical clustering (Wei and Simko, 2016). The red box indicates the setup combinations showing the least agreement with most the majority of other setups. 1= cultivar distribution, 2= soil data, 3=soil handling, 4=nutrient cycling coefficient, 5=hydrologic coefficients, and 5=management. e=EPIC-IIASA and g=GEPIC.

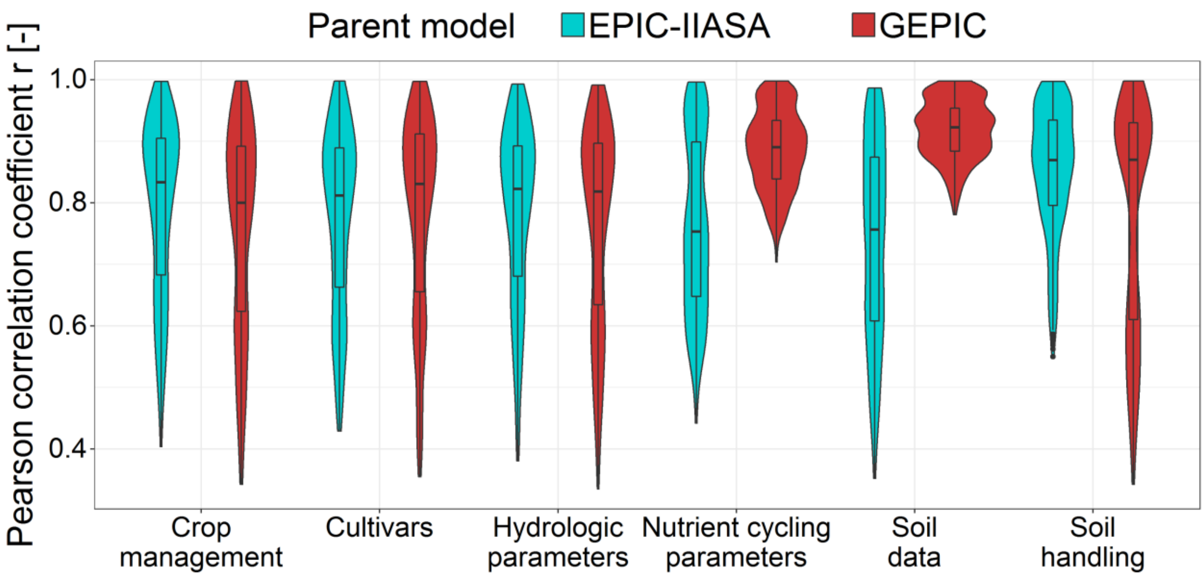

Figure S 3-16: Distributions (violins) and box-and-whisker plots of correlation coefficients among all setup combinations of EPIC-IIASA and GEPIC (Table 3 in main article) aggregated by setup domains.

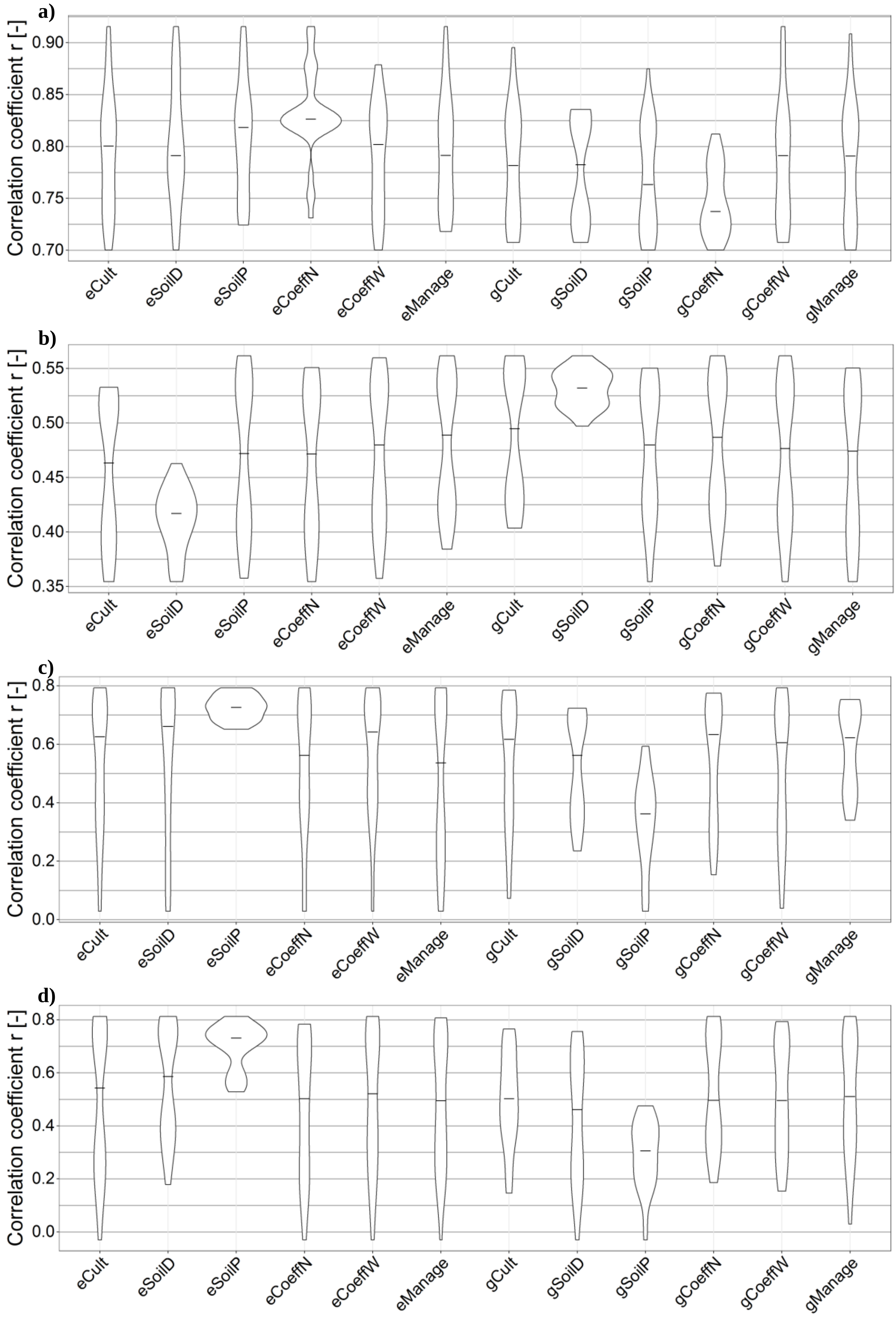

Figure S 3-17: Distributions (violins) and medians (horizontal lines) of the time series correlation coefficient r of simulated and reported yields for each setup domain (Table 3 in main article) in (a) USA, (b) China, (c) Argentina, and (d) India. Corresponding performance of each setup combination are shown in Figure 10 of the main article

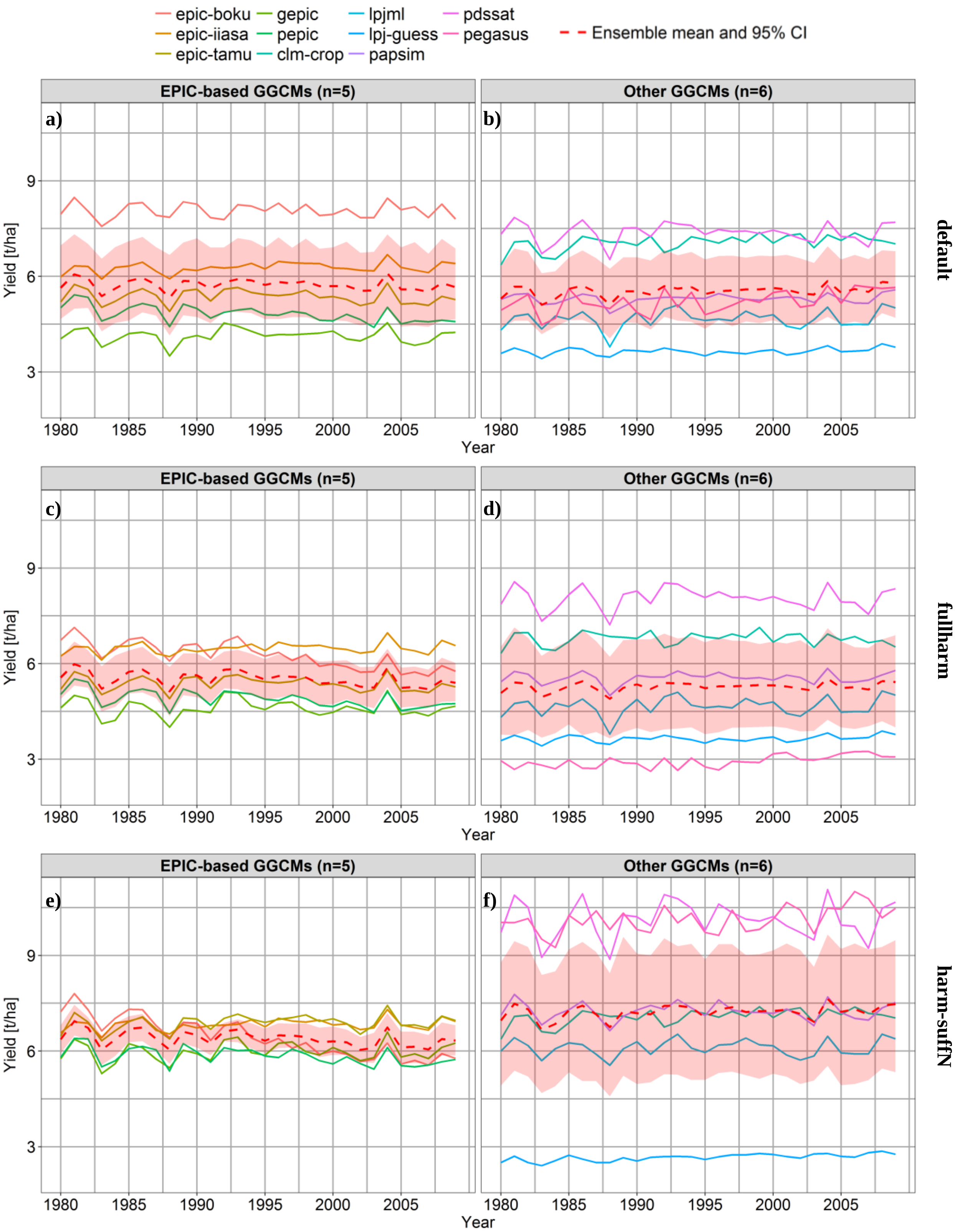

Figure S 3-18: Global average area-weighted maize yields and 95% confidence interval of the mean for EPIC-GGCMs and non-EPIC-based GGCMs for three management scenarios. Solid lines show outputs from single models.

Table S 3-6: Statistical coefficients for linear regressions of yield estimates (not shown) corresponding to global average yields in Figure S 3-18 and mean error [t ha^{-1}] compared to reported yields.

Model	Intercept	Slope	R^2	p	Mean error	Scenario and panel in Figure 2 and Figure S 3-18
EPIC-BOKU	8.068	-0.001	0.000	0.909	3.676	default Panel a
EPIC-IIASA	6.178	0.006	0.114	0.068	1.891	
EPIC-TAMU	5.454	-0.005	0.045	0.258	0.987	
GEPIC	4.128	0.001	0.001	0.901	-0.246	
PEPIC	5.100	-0.018	0.376	0.000	0.439	
CLM-crop	6.839	0.014	0.260	0.004	2.675	
LPJmL	4.557	0.007	0.044	0.266	0.277	
LPJ-GUESS	3.585	0.004	0.110	0.073	-0.738	
pAPSIM	5.245	0.003	0.040	0.286	0.915	
pDSSAT	7.358	0.001	0.001	0.864	2.993	
PEGASUS	4.908	0.020	0.254	0.004	0.834	
EPIC-BOKU	6.861	-0.039	0.660	0.000	1.880	fullharm Panel b
EPIC-IIASA	6.399	0.005	0.073	0.150	2.096	
EPIC-TAMU	5.454	-0.005	0.045	0.258	0.987	
GEPIC	4.611	-0.001	0.001	0.896	0.216	
PEPIC	5.162	-0.017	0.314	0.001	0.508	
CLM-crop	6.758	0.001	0.003	0.786	2.393	
LPJmL	4.557	0.007	0.044	0.266	0.277	
LPJ-GUESS	3.585	0.004	0.110	0.073	-0.738	
pAPSIM	5.511	0.003	0.018	0.476	1.170	
pDSSAT	8.044	0.001	0.000	0.912	3.674	
PEGASUS	2.724	0.013	0.392	0.000	-1.454	
EPIC-BOKU	7.409	-0.063	0.823	0.000	2.049	harm-suffN Panel c
EPIC-IIASA	6.747	0.006	0.075	0.143	2.450	
EPIC-TAMU	6.773	0.007	0.064	0.178	2.501	
GEPIC	5.895	0.005	0.026	0.395	1.596	
PEPIC	6.070	-0.015	0.218	0.009	1.459	
CLM-crop	6.850	0.014	0.263	0.004	2.689	
LPJmL	6.063	0.003	0.008	0.634	1.720	
LPJ-GUESS	2.523	0.009	0.538	0.000	-1.719	
pAPSIM	7.261	-0.001	0.000	0.920	2.869	
pDSSAT	10.064	0.003	0.003	0.782	5.733	
PEGASUS	9.783	0.022	0.226	0.008	5.741	

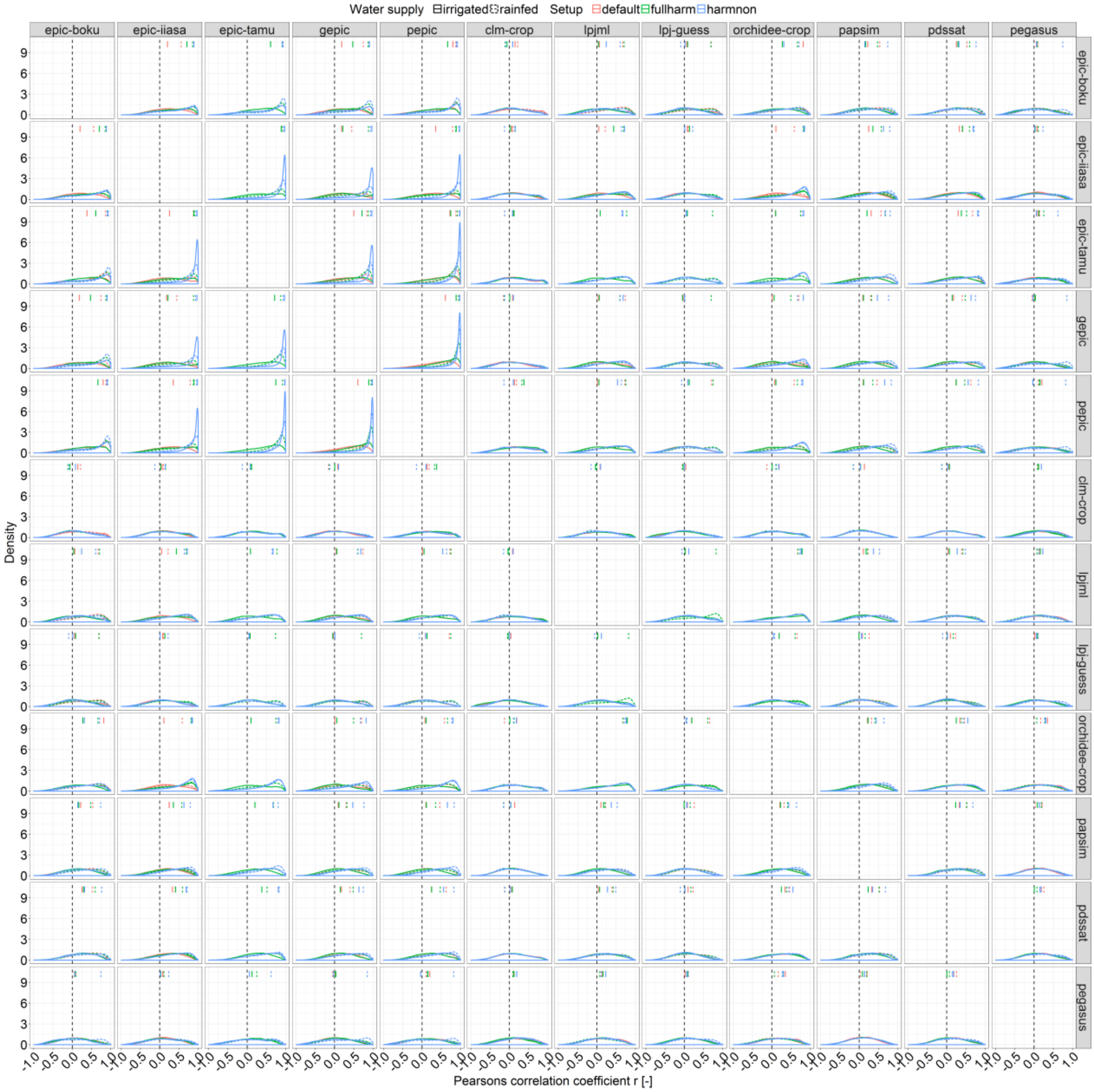

Figure S 3-19: Frequency distributions of time-series correlation coefficients in each grid cell for all GGCMs and setup scenarios (Table 1 in main article). Solid and dashed lines at the top of each panel indicated the location of the major peak in the distribution for rainfed (dashed) or sufficiently irrigated (solid) simulations of each management scenario (Table 1 in main article).

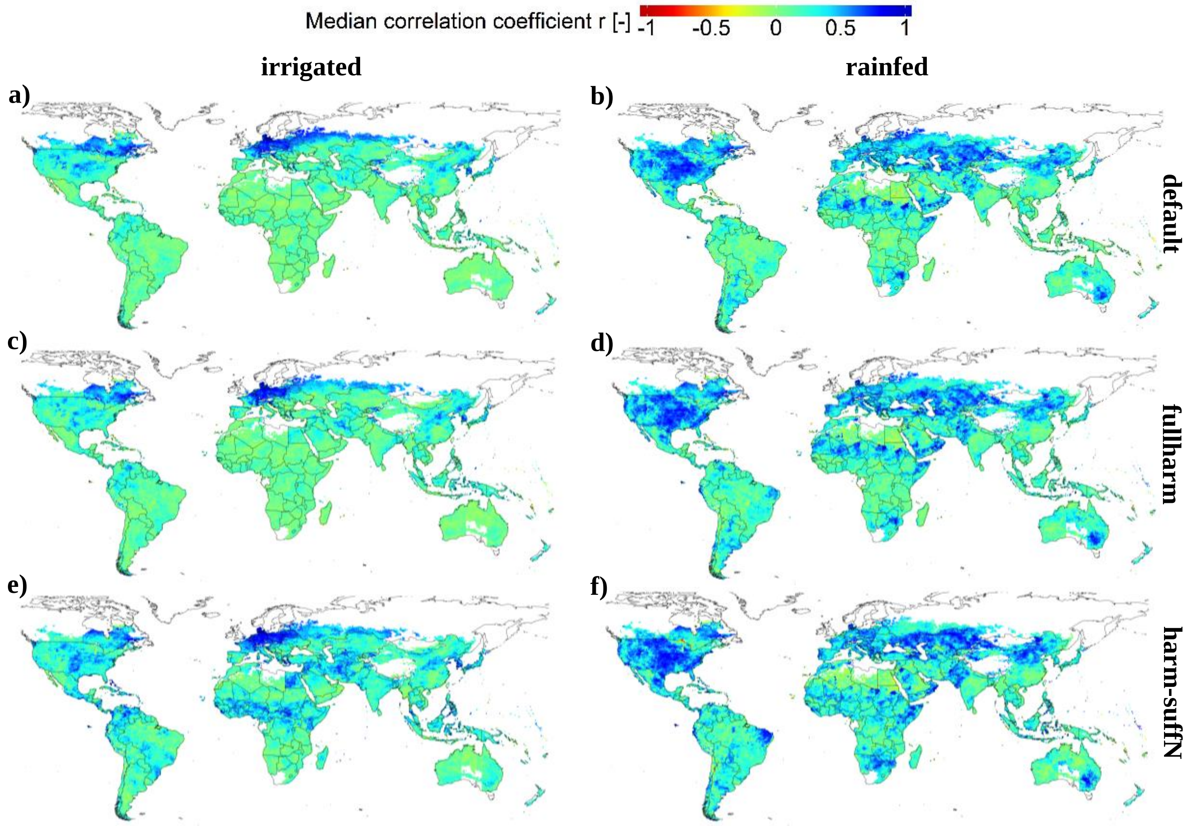

Figure S 3-20: Median of time-series correlation coefficient r for maize yield estimates among the whole GGCM ensemble for each of the six crop management scenarios defined in Table 1 of the main article. For EPIC-TAMU outputs from the fully harmonized (fullharm) simulations were used also as default and vice versa for LPJ-GUESS and LPJmL default outputs also as fullharm. ORCHIDEE-crop is not included in the harm-suffN scenario.

Table S 3-7: Fractions of grid cells [%] in which the median time series correlation among all GGCMs (Figure S 3-20) fulfils a certain level of significance.

Management scenario	Water supply	p<0.1	p<0.05	p<0.01
default	irrigated	5.22	2.60	0.41
default	rainfed	9.82	4.63	0.54
fullharm	irrigated	3.99	2.15	0.42
fullharm	rainfed	11.60	5.82	0.70
harm-suffN	irrigated	7.36	3.77	0.83
harm-suffN	rainfed	14.89	8.14	1.41

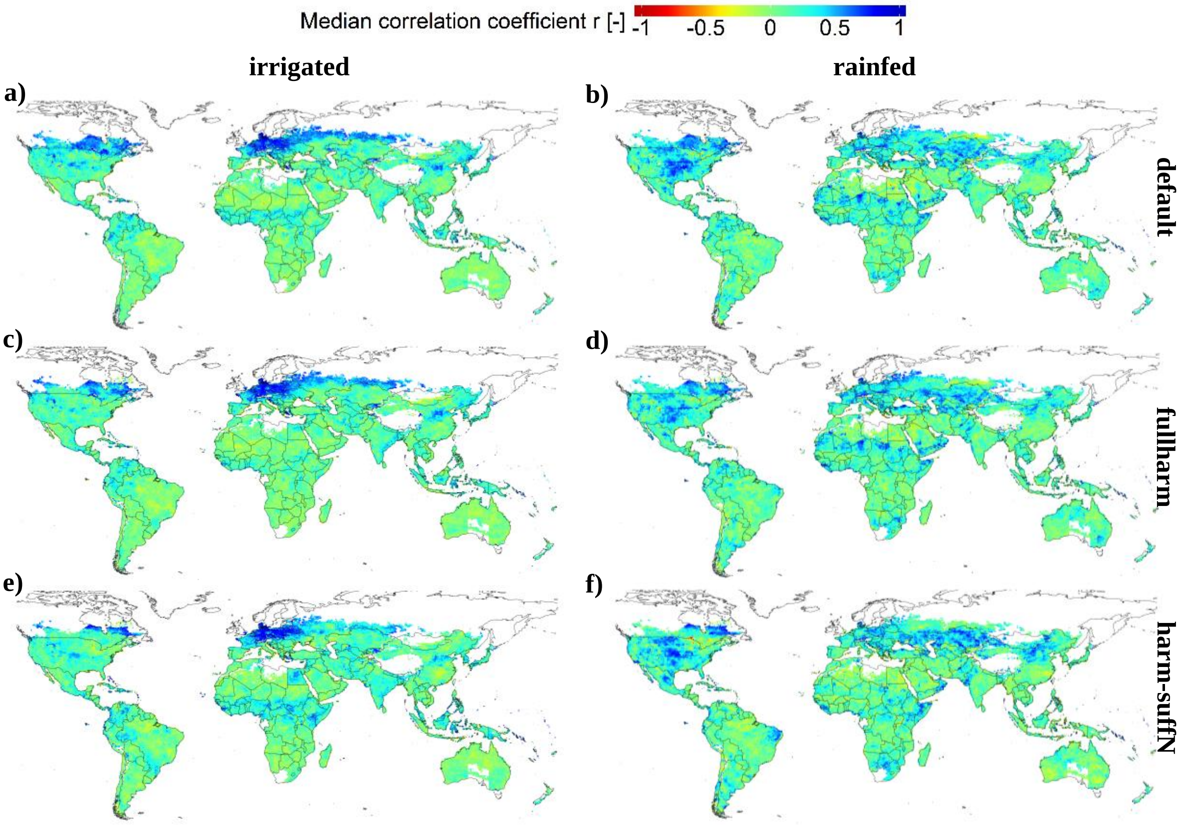

Figure S 3-21: Median of time-series correlation coefficient r for maize yield estimates among the GGCM ensemble excluding the EPIC-based GGCMs for each of the six crop management scenarios defined in Table 1 of the main article. For LPJ-GUESS and LPJmL outputs from the default simulations were used also as fullharm. ORCHIDEE-crop is not included in the harm-suffN scenario.

Table S 3-8: Fractions of grid cells [%] in which the median time series correlation among the GGCMs, excluding the EPIC-based ones, (Figure S 3-21) fulfils a certain level of significance.

Management scenario	Water supply	p<0.1	p<0.05	p<0.01
default	irrigated	5.31	2.81	0.61
default	rainfed	4.52	1.99	0.37
fullharm	irrigated	4.41	2.40	0.56
fullharm	rainfed	4.14	1.68	0.27
harm-suffN	irrigated	3.66	2.29	0.74
harm-suffN	rainfed	4.70	2.19	0.38

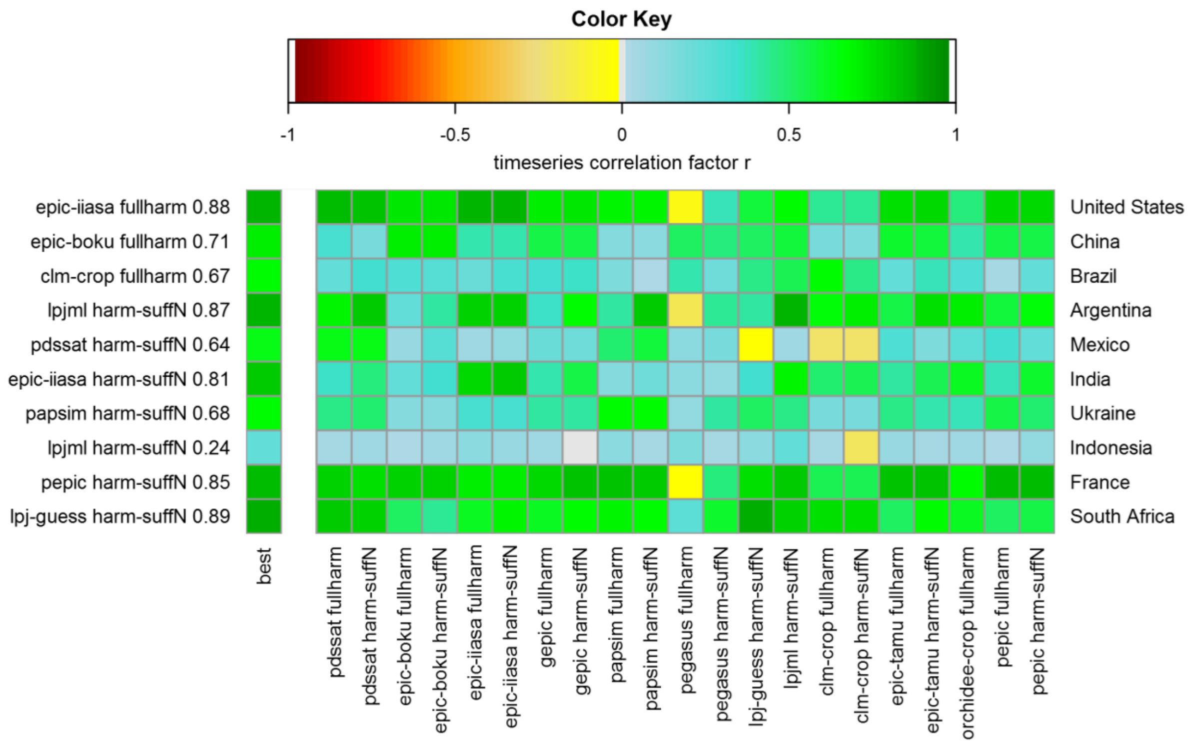

Figure S 3-22: Time-series correlation coefficients for all GGCMs with the fullharm and harm-suffN scenarios (x-axis) in the top ten maize producing countries (right y-axis) and the best performing GGCM/setup combination including the r value (left y-axis).