Supplementary Information for "Uncertainties in global crop model frameworks: effects of cultivar distributions, crop management and soil handling on crop yield estimates"

S1 Differences in model setups and processes

S1.1 Handling of long-term simulations and implications for carry-over effects in GEPIC

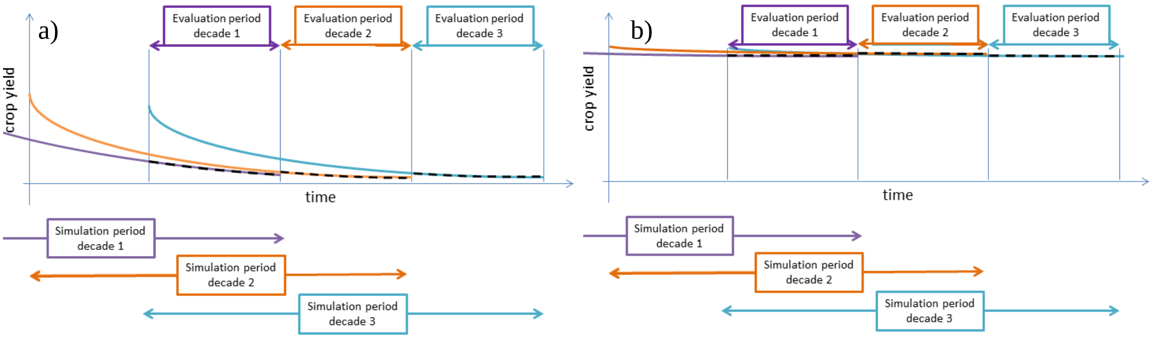

Figure S 1-1: Schematic representation of decadal GEPIC runs with dynamic soil profile and erosion for (a) high nutrient input and (b) low nutrient input conditions. Colors represent simulations for three decades with a 20 year spin-up for each decade, which is discarded. Only the last ten years are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

All model frameworks (MFWs) except GEPIC were run fully transient from a warm-up period to the end of the simulation period. Usually, these MFWs disable soil erosion and use a static soil profile, which limits soil degradation and nutrient depletion. As GEPIC is frequently used for evaluating effects of soil (nutrient) management on crop yields (Folberth et al., 2012; Folberth et al., 2013; Folberth et al., 2014), it takes soil nutrient depletion and erosion into account. The authors found that when using dynamic soil profiles, the model reproduces yields in low-input regions like sub-Saharan Africa around the year 2000 well after a spin-up of 30 years when evaluating the last 10 simulation years. Extending the simulation period with such a setup would potentially result in erosion of the whole soil profile at some point and/or complete nutrient depletion in grid cells that lack fertilizer inputs. Therefore, the model is run for each decade of the study time 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 1-1).

S1.2 Differences between EPIC model versions 0810 and 1102

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 1-2). To exact identification of drivers in differences between the models is beyond the scope of this study and will require more in-depth field-scale studies.

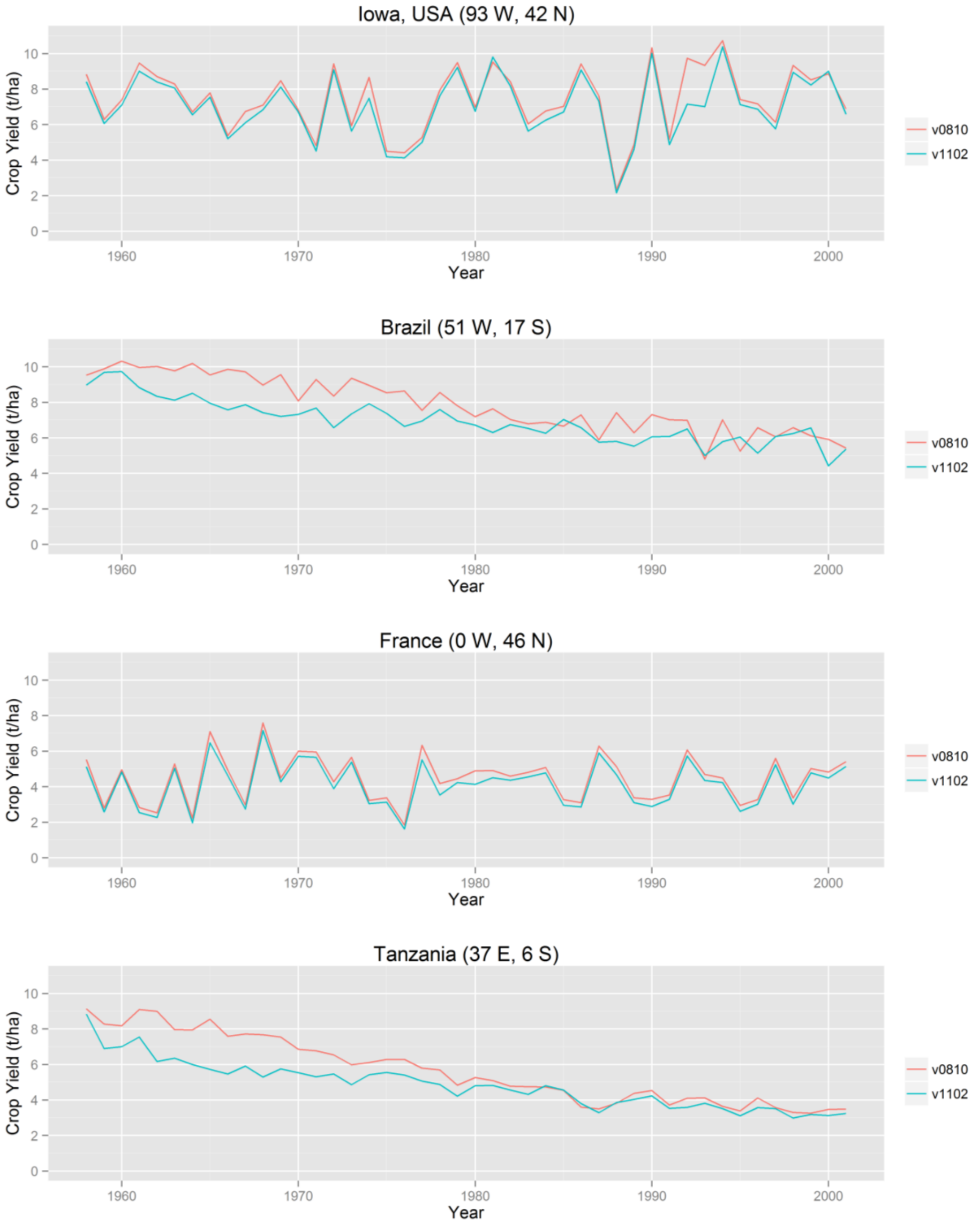

Figure S 1-2: Maize yield estimates of EPIC v0810 and EPIC v1102 for four contrasting locations.

S1.3 Parameterization of the model frameworks

S1.3.1 Parameterization of maize cultivars

Table S 1-1: Parameterization of different maize cultivars used in the MFWs as shown in Figure 1 of the manuscript. 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, TOP=optimum temperature, HI=harvest index, GSL=growing season length.

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 Supplementary results and evaluations

Table S 2-1: 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

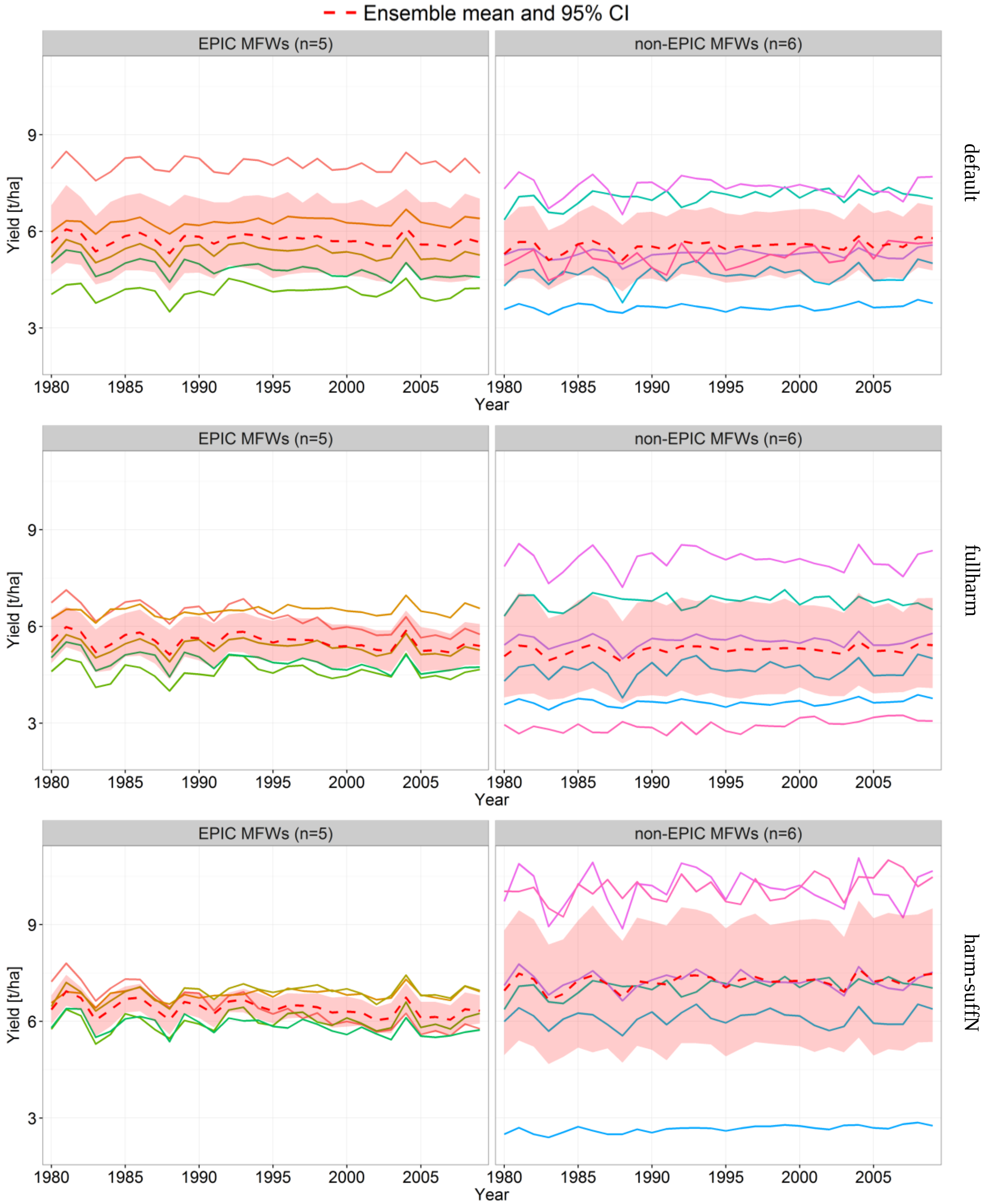

Figure S 2-1: Global average area-weighted maize yields and 95% confidence interval of the mean for EPIC-MFWs and non-EPIC-based GGCMs for three management scenarios. Solid lines show outputs from single models.

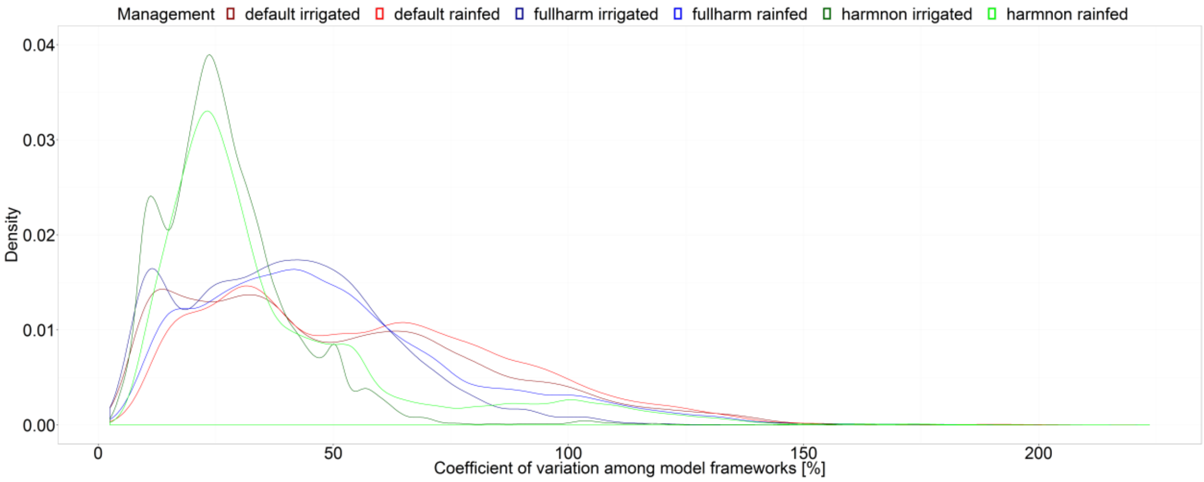

Figure S 2-2: Coefficient of variation among EPIC model frameworks depending on the setup and management scenarios (see Table 1 in main paper).

Table S 2-2: Quantiles of coefficient of variation among EPIC model frameworks for maize yield estimates depending on the setup and management scenarios (see Table 1 in paper).

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

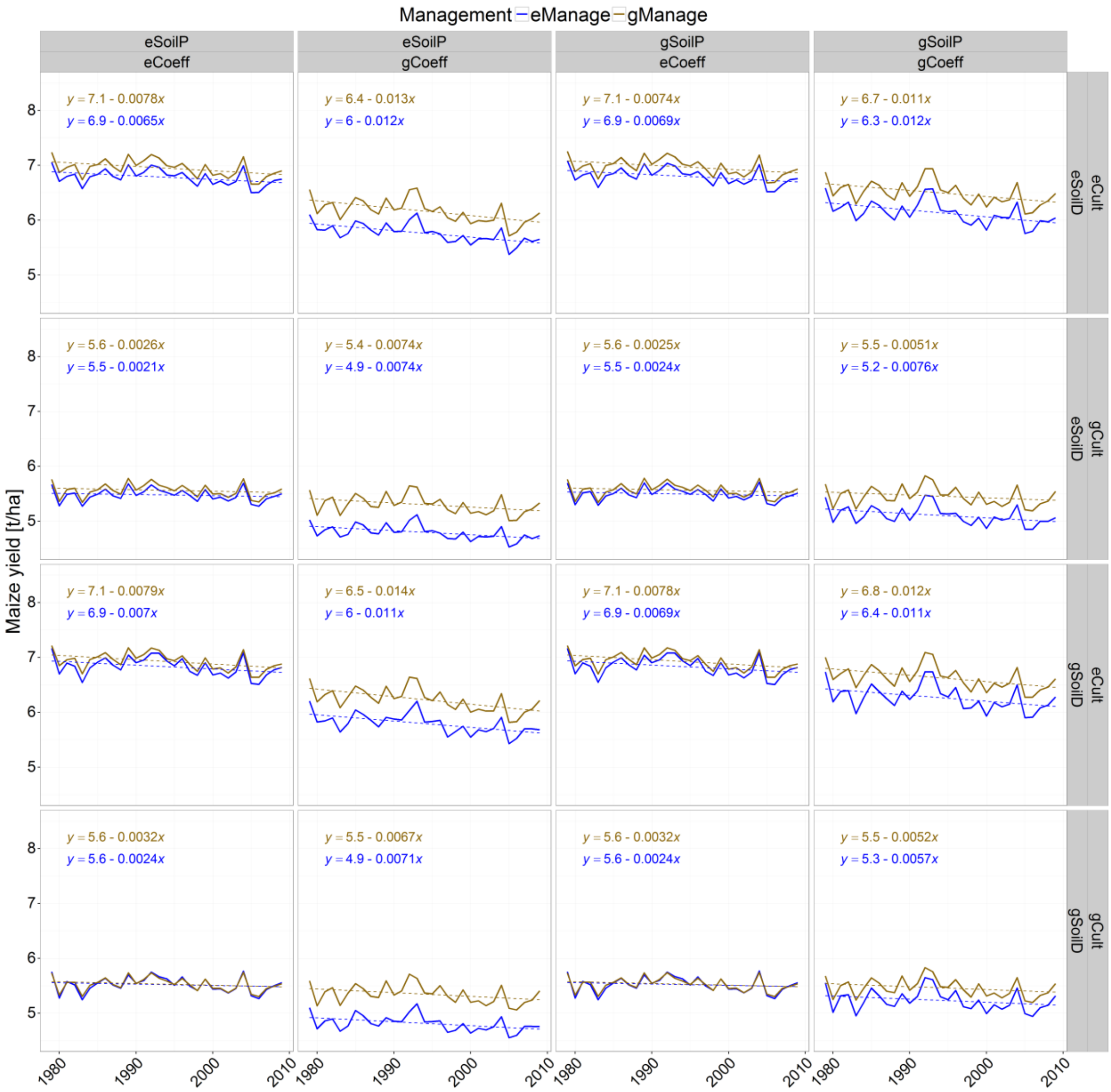

Figure S 2-3: Global average irrigated maize yields over a 31 year time period for 32 setup combinations based on the EPIC-IIASA and GEPIC setups as described in Table 2 of the manuscript. e=EPIC-IIASA, g=GEPIC, Cult=cultivar definition and distribution, SoilD=soil data, SoilP=spin-up and soil handling, Coeff=coefficients, Manage=crop management. Dashed lines show linear regressions to identify trends.

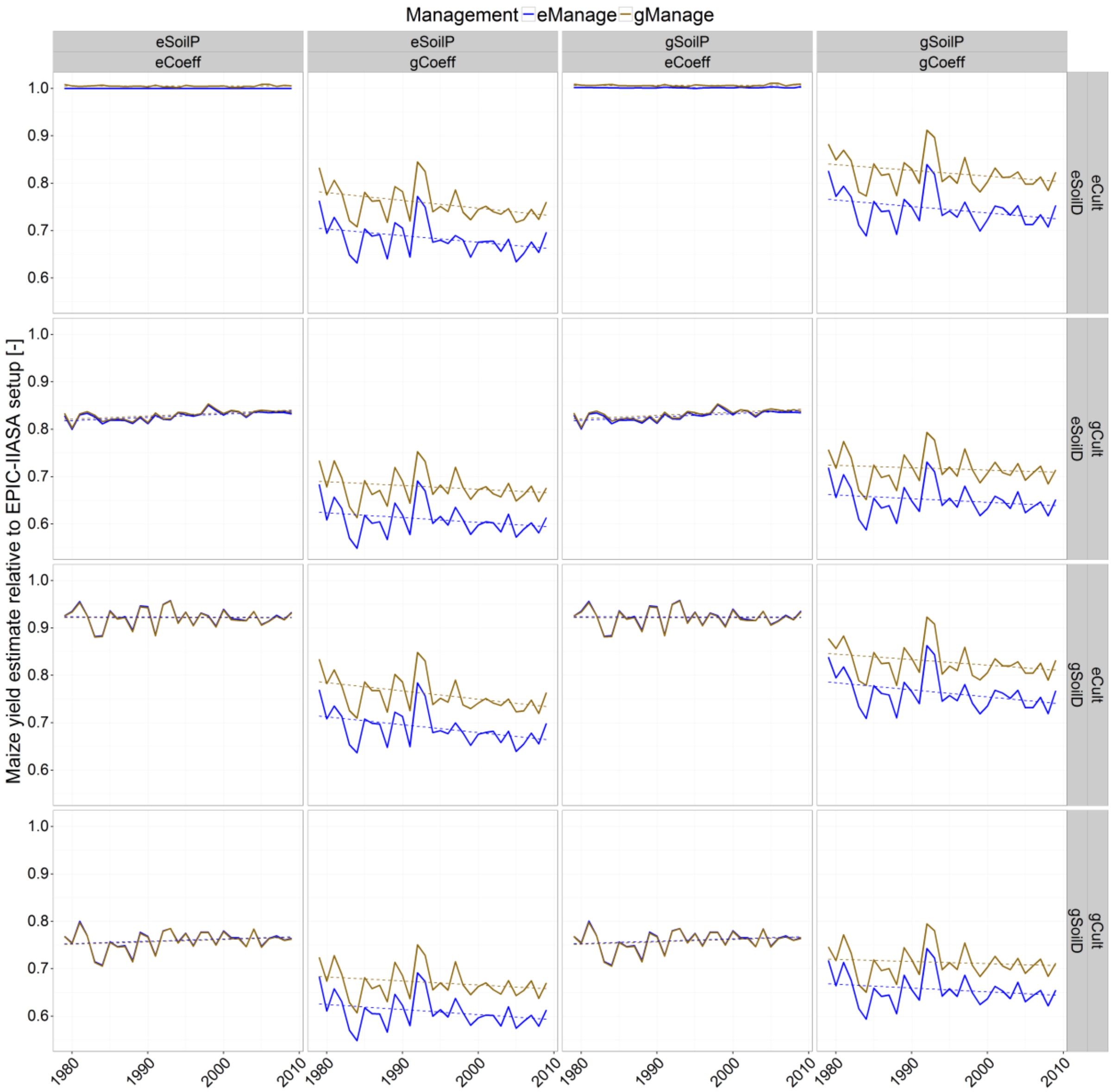

Figure S 2-4: Relative difference [-] between simulated yields for rainfed maize using the entirely EPIC-IIASA based setup (top left panel) as a reference over a 31 year time period for 32 setup combinations based on the EPIC-IIASA and GEPIC setups as described in Table 2. Absolute yields are shown in Figure 5 of the manuscript. e=EPIC-IIASA, g=GEPIC, Cult=cultivar definition and distribution, SoilD=soil data, SoilP=spin-up and soil handling (e.g. erosion), Coeff=coefficients, Manage=crop management. Dashed lines show linear regressions indicating trends.

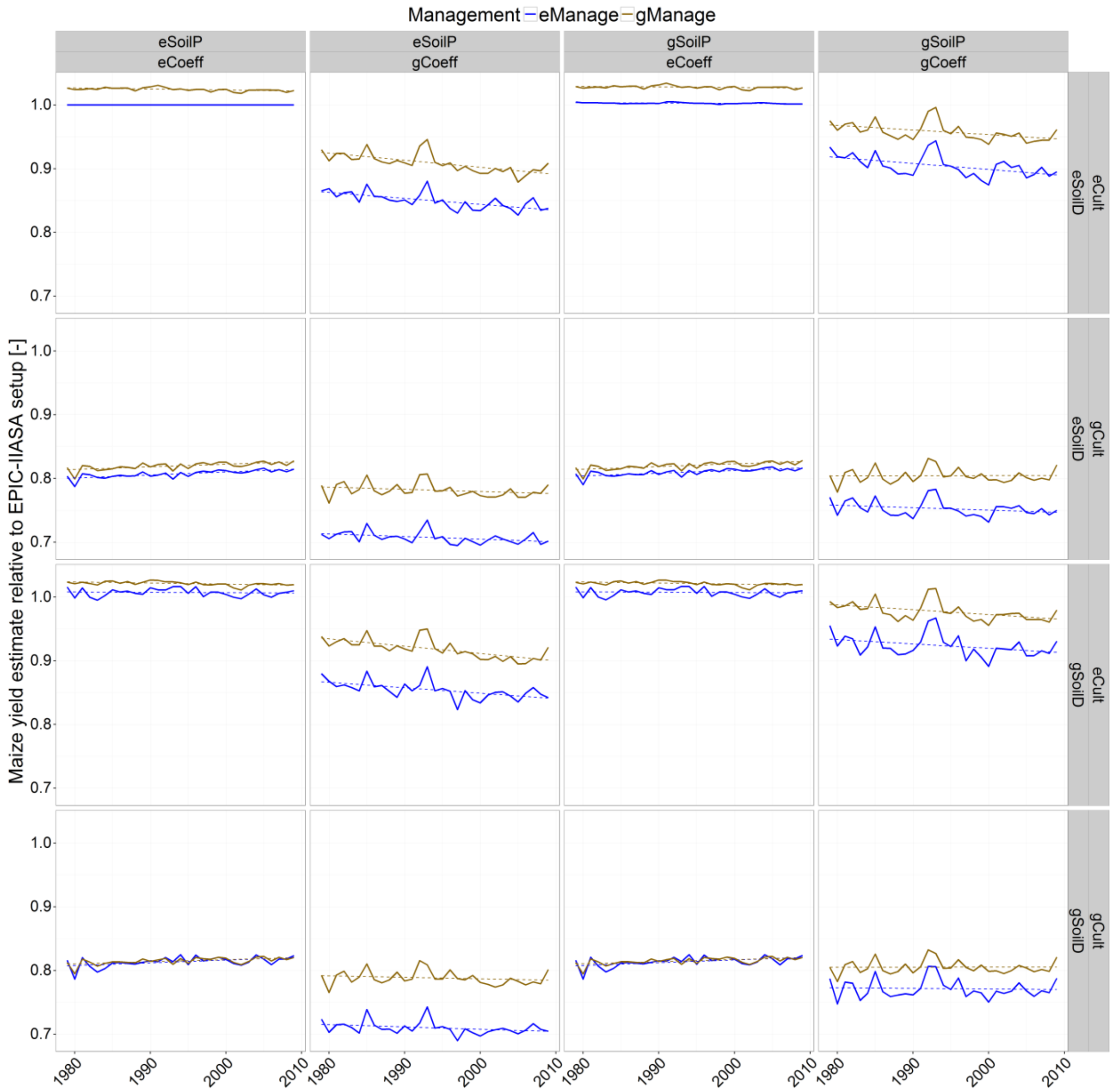

Figure S 2-5: Relative difference [-] between simulated yields for irrigated maize using the entirely EPIC-IIASA based setup (top left panel) as a reference over a 31 year time period for various GGCM setup combinations based on the EPIC-IIASA and GEPIC setups as described in Table 2. Absolute yields are shown in Figure S 2-3. e=EPIC-IIASA, g=GEPIC, Cult=cultivar definition and distribution, SoilD=soil data, SoilP=spin-up and soil handling (e.g. erosion), Coeff=coefficients, Manage=crop management. Dashed lines show linear regressions indicating trends.

Table S 2-3: Effects of selected setup options and combinations on global average maize yields as shown in Figure 5 of the manuscript.

Setup options	Observed effects
eCoeff vs gCoeff	no difference between managements yield increase over time lower inter-annual yield variability
eSoilP vs gSoilP	lower yields in combination with gCoeff (dynamic soil profile) due to nutrient mining differences in inter-annual yield variability
eSoilD vs gSoilD	impact on inter-annual yield dynamics in combination with eCoeff
eCult vs gCult	proportional upshift in yields when combined with eSoilP x eCoeff or gSoilP x gCoeff lower yield increase in eCoeff simulations but stronger decrease in gCoeff simulations
eManage vs gManage	lower yields in simulations with (partly) transient soil nutrient dynamics slight impact on inter-annual yield variability with on average lower variability in eManage

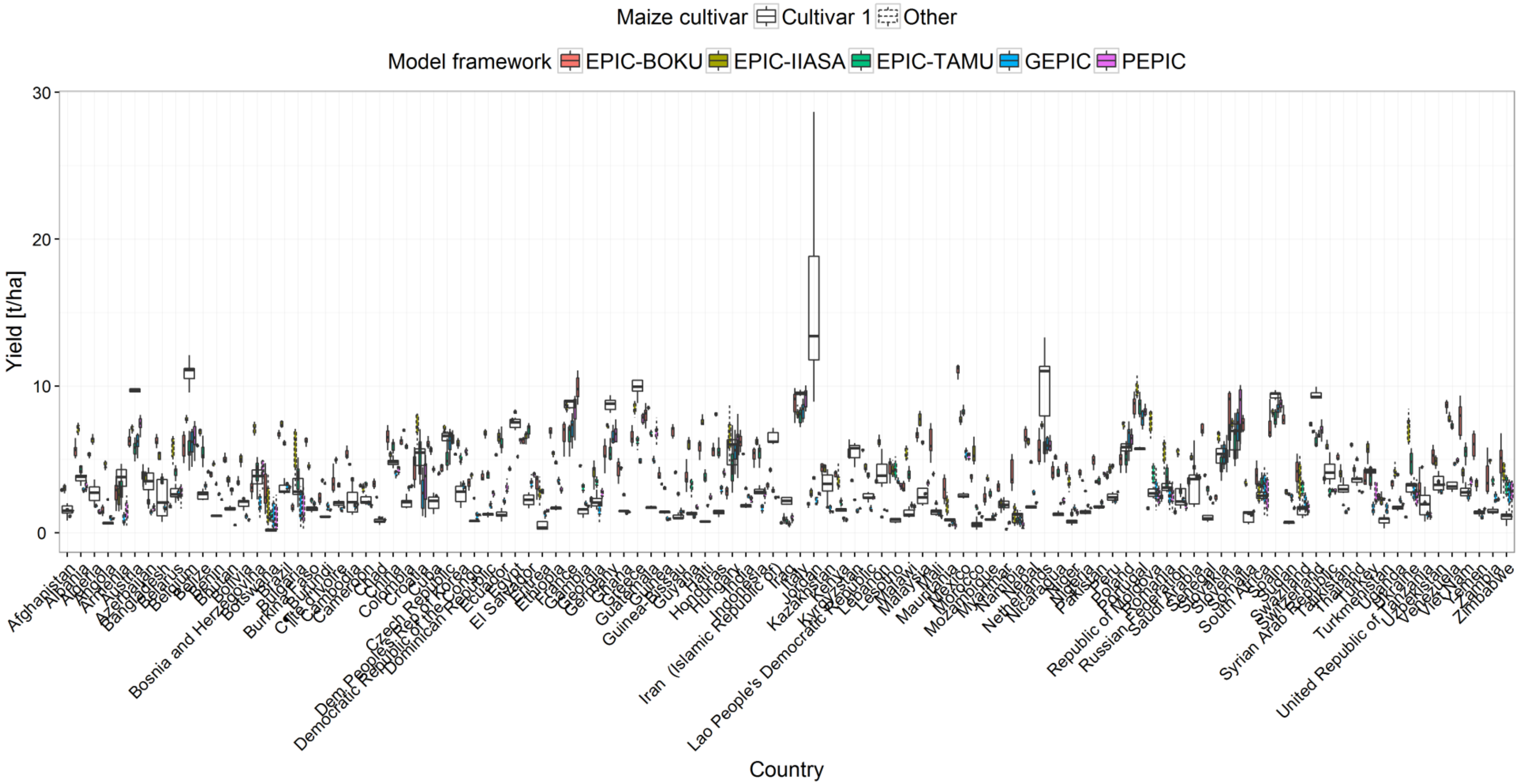

Figure S 2-6: Reported and simulated maize yields in countries, in which the attribution of cultivars (see Figure 1a-d in manuscript) differs by more than 30% between at least two GGCMs.

S3 Model framework performance in reproducing reported maize yields

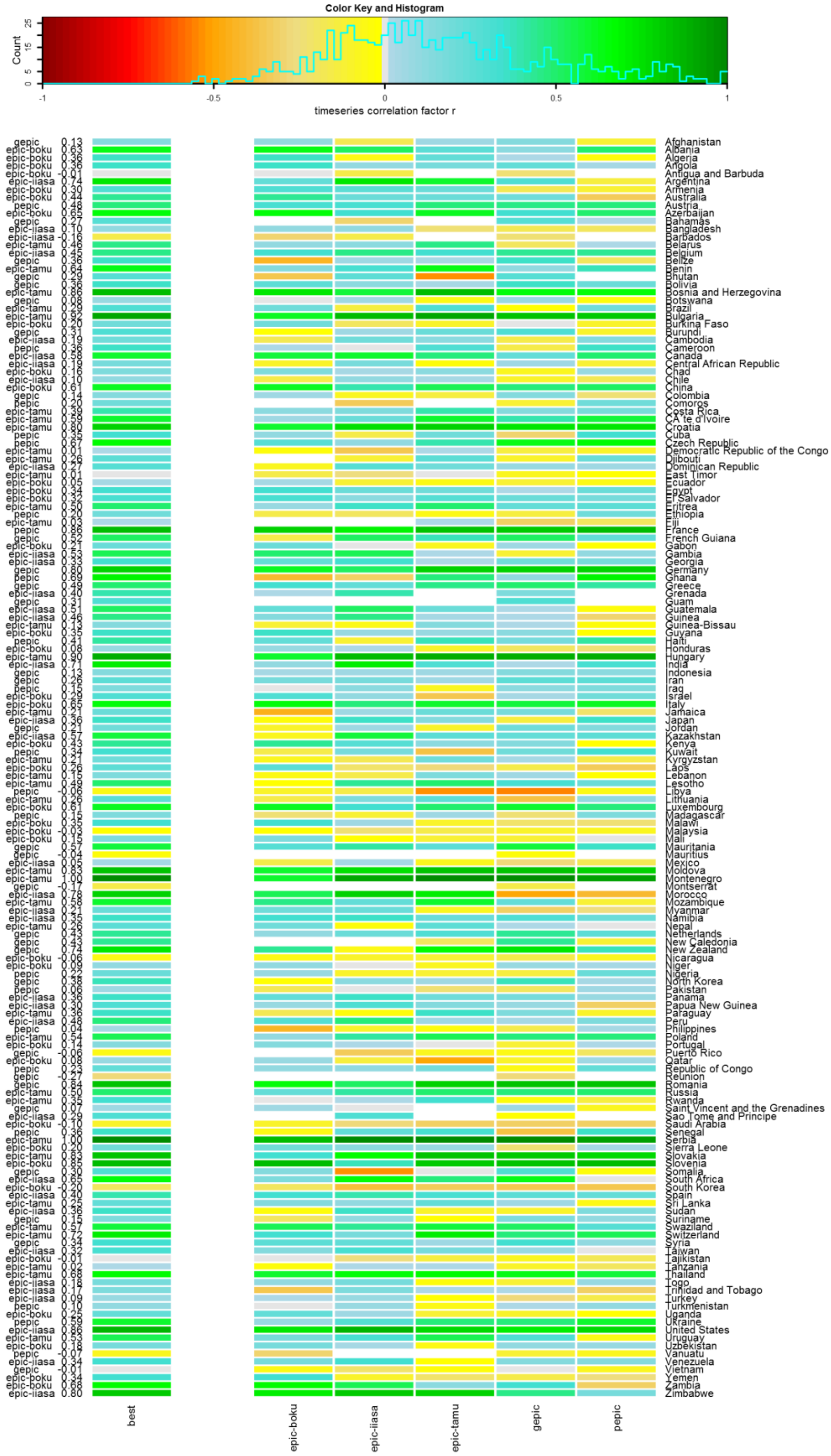

Figure S 3-1: Performance of all EPIC MFWs for estimating maize yields in each country measured as time series correlation (tscorr). The left axis depicts the MFW with the highest performance and r value for the respective country shown on the right axis.

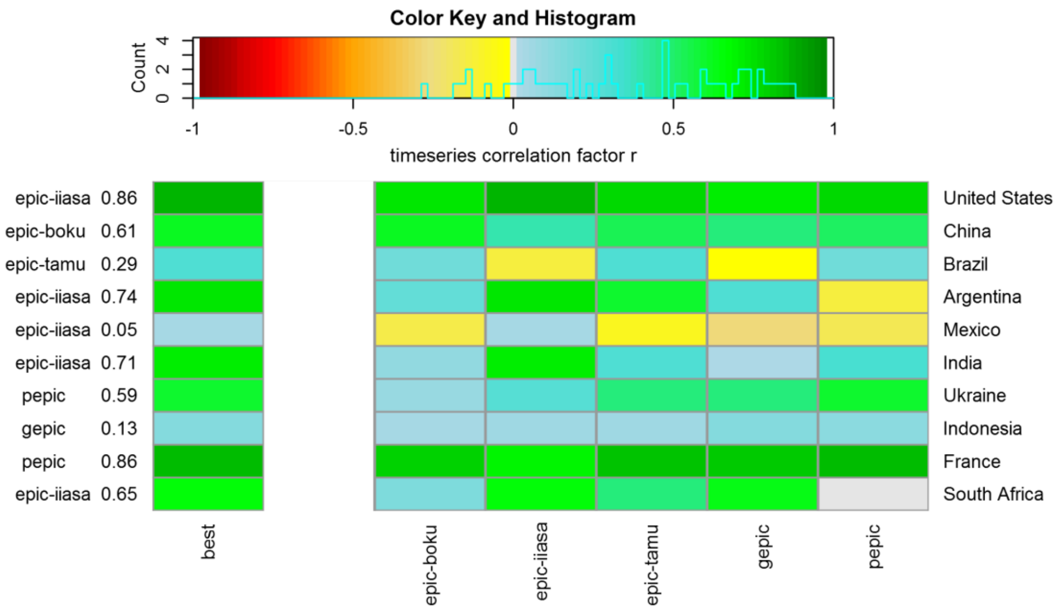

Figure S 3-2: Performance of all EPIC MFWs for estimating maize yields in the top ten maize producing countries measured as time series correlation (tscorr). The left axis depicts the MFW and setup scenario with the highest performance in the respective country shown on the right axis.

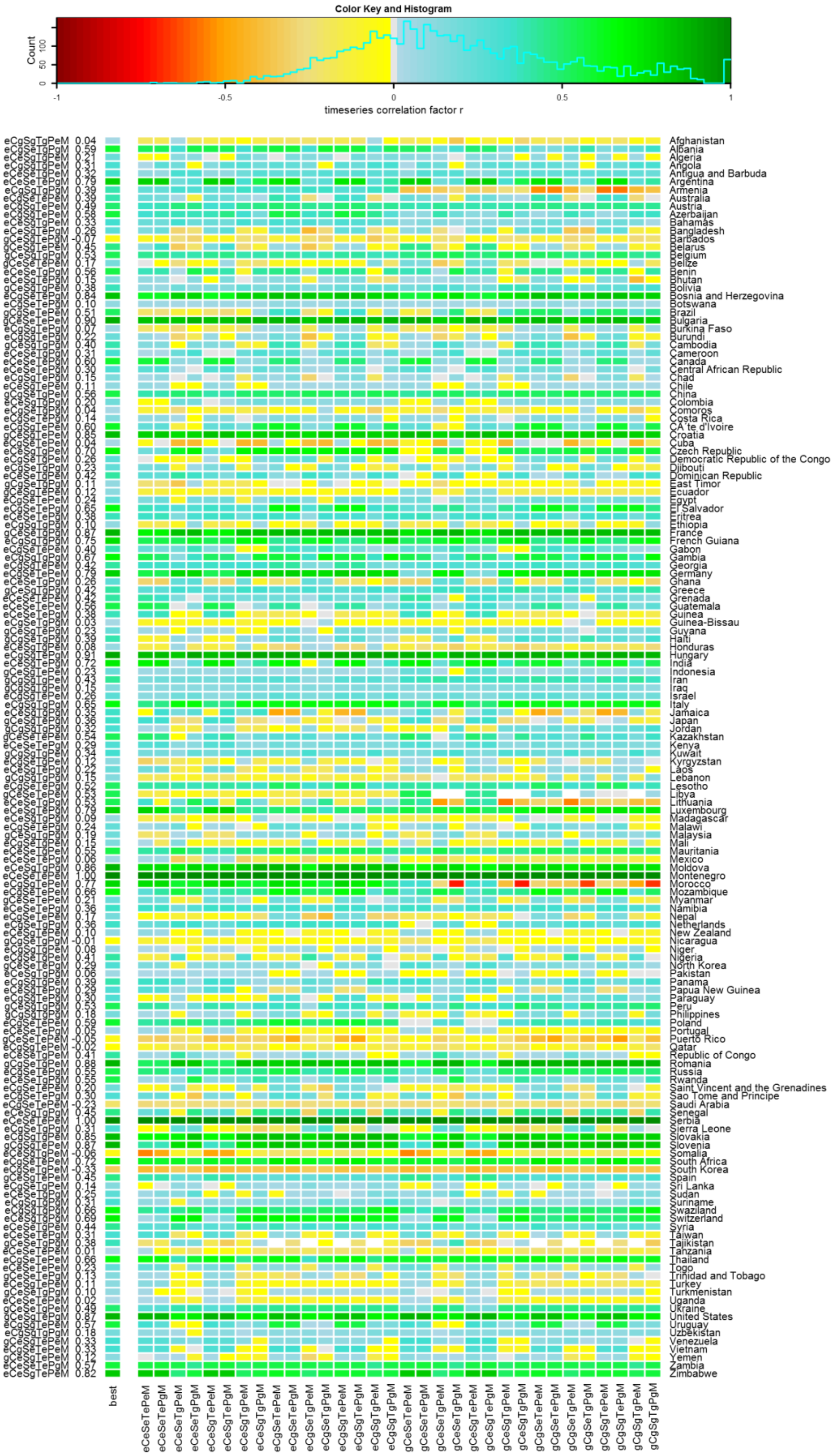

Figure S 3-3: Performance of all EPIC-IIASA and GEPIC setup permutations for estimating maize yields in each country measured as time series correlation (tscorr). The left axis depicts the setup scenario with the highest performance and r value for the respective country shown on the right axis. g/e=GEPIC/EPIC-IIASA, C=cultivar, S=soil data, T=soil handling, P=coefficients, M=management.

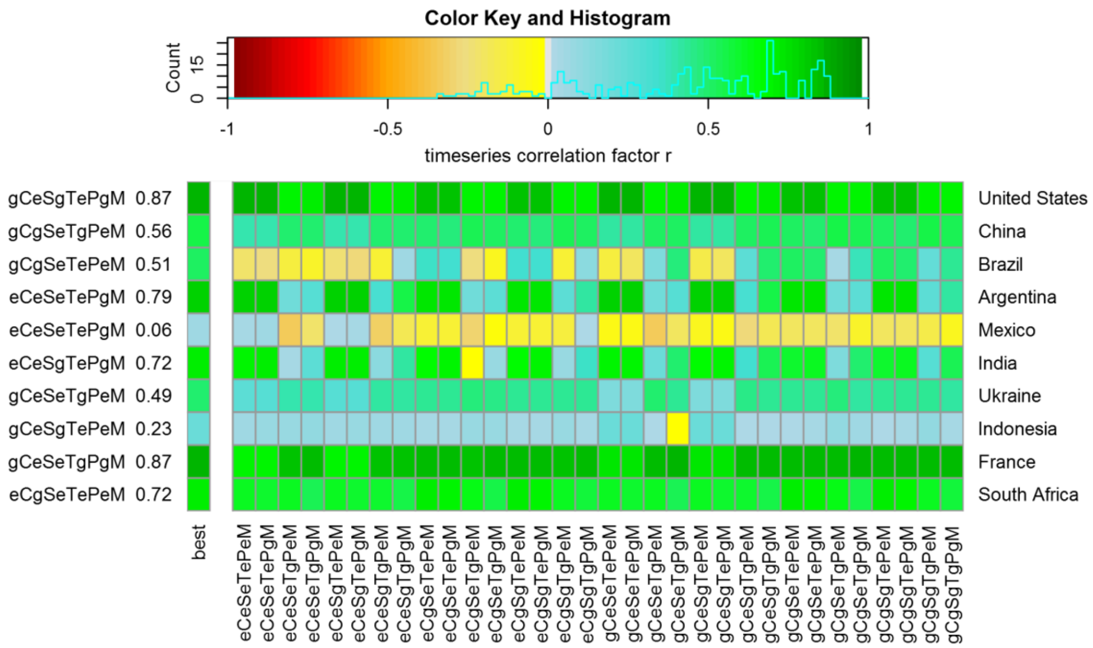

Figure S 3-4: Performance of all EPIC-IIASA and GEPIC setup permutations for estimating maize yields in the top ten maize producing countries measured as time series correlation (tscorr). The left axis depicts the setup scenario with the highest performance and r value for the respective country shown on the right axis. g/e=GEPIC/EPIC-IIASA, C=cultivar, S=soil data, T=soil handling, P=coefficients, M=management.

S4 Complementary evaluation of differences among model frameworks in wheat simulations

S4.1 Distribution of wheat types

All MFWs follow established rules for the distribution of wheat types (Figure S 4-1). Winter wheat is generally assumed to be grown successfully at latitudes above 30°N and below 27°S (Nuttonson, 1955; Curtis, 2002). Similarly, EPIC-IIASA plants winter wheat above 25°N and below 25°S in its default setup, except for parts of the Arab peninsula (Figure S 4-1a). The distribution in EPIC-BOKU is based on growing season data with planting of winter wheat if the growing season length is >180 days (Figure S 4-1b). Another approach is the use of a thermal envelope based on the temperature range required for vernalization of winter wheat and the threshold for frost damage. This is provided if the mean temperature in the coldest month of the year is within the range of -10°C - +5°C (Stehfest et al., 2007). This distribution was used in the harmonized setups of EPIC-IIASA (Figure S 4-1c) and all setups of EPIC-TAMU, GEPIC and PEPIC (Figure S 4-1d).

Although the approach of Stehfest et al. (2007) shows the best agreement with a set of representative sites used in the field-scale comparison of various wheat models (Asseng et al., 2013), the limited number and regional concentration of data points does not allow for a thorough validation. On the other hand, only the approach of EPIC-BOKU to distribute wheat types according to the growing season calendar allows for consistency between reported growing seasons and wheat types in the harmonized setups.

The fact that both spring and winter wheat may be grown in close proximity is not taken into account by the mutually exclusive zoning in all EPIC-MFWs (and other GGCMs). Collecting further data will be required to develop databases of global wheat type distributions for winter and spring wheat separately. A first such dataset has recently been published by Gbegbelegbe et al. (2016) presenting global distributions of wheat cultivars for agro-climatologic zones. Producing model outputs for both spring and winter wheat in in regions in which both may be grown will then allow for distributing the different wheat types *ex-post*. To allow for this, GGCM outputs will be collected separately for spring and winter wheat each covering the whole globe in phase two of GGCMI.

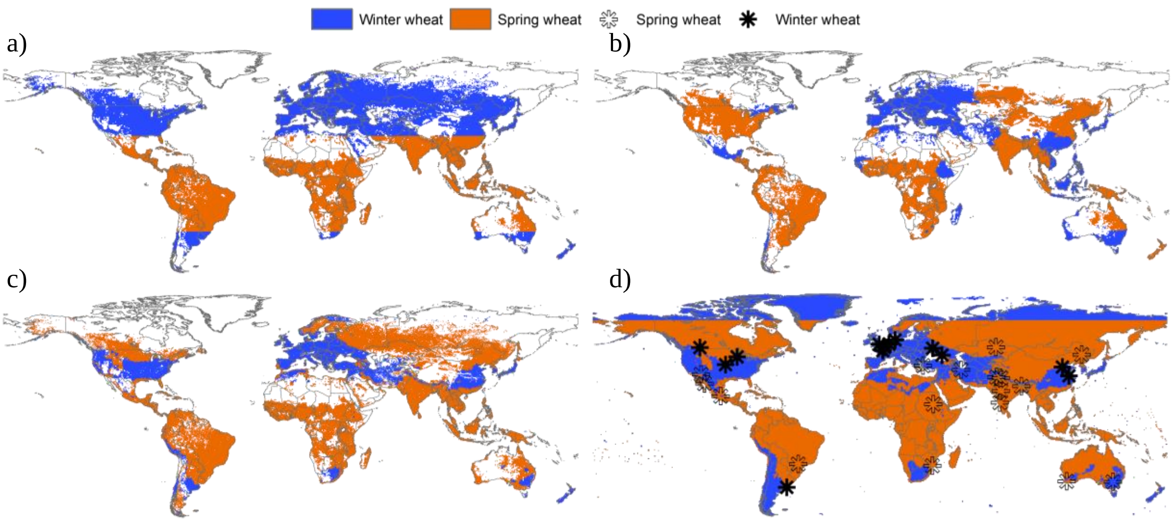

Figure S 4-1: Distributions of wheat types (spring and winter) in the model frameworks for (a) EPIC-IIASA default scenario, (b) EPIC-BOKU, (c) EPIC-IIASA fullharm and harm-suffN scenarios, and (d) GEPIC, PEPIC, and EPIC-TAMU (based on Stehfest et al., 2007). Asterisks in (d) indicate the distribution of winter and spring wheat for the AgMIP wheat pilot sites as reported by Asseng et al. (2014).

No clear pattern can be identified for differences in wheat type distribution and yield estimates (Figure S2-8). Although only EPIC-BOKU has a substantially differing pattern of wheat type distribution in the harmonized setups, EPIC-BOKU and EPIC-IIASA exhibit similar yield levels in many countries in which wheat types differ by more than 30% of the harvested area among at least two models as do EPIC-TAMU, GEPIC and PEPIC. Differences in other setup domains such as soil handling and nutrient supply may outplay differences in wheat type distributions.

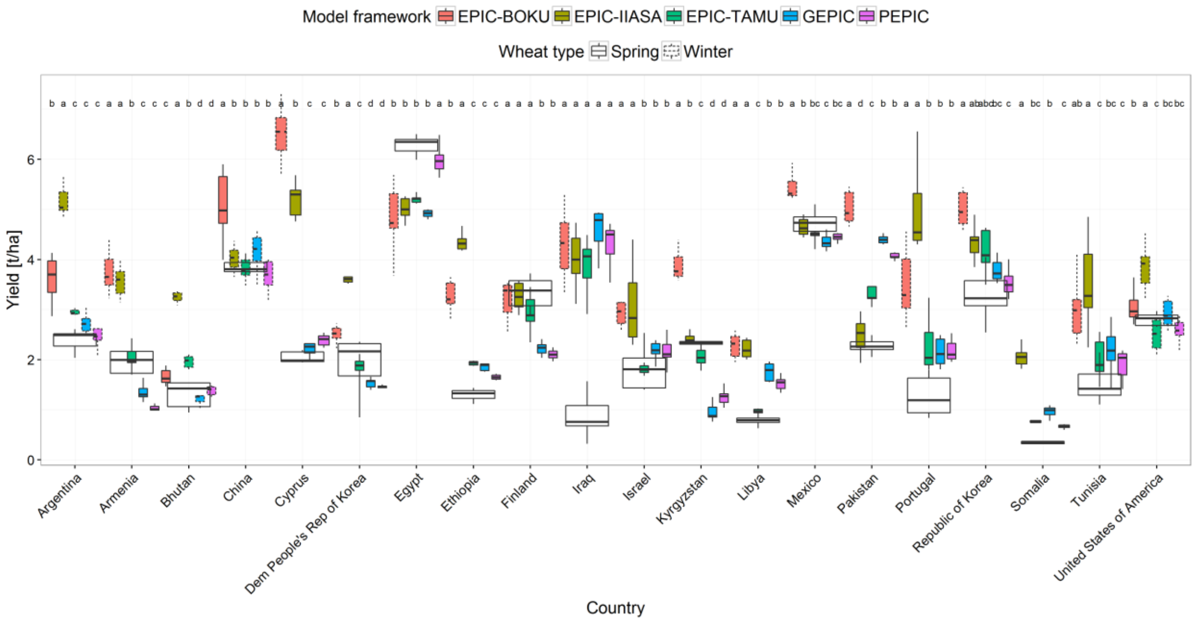

Figure S 4-2: Reported and simulated wheat yields in countries, in which the attribution of wheat types (see Figure 1e-h in manuscript) differs by more than 30% between at least two GGCMs.

S4.2 Global average wheat yields

With default setups, global average simulated wheat yields have a relative range of up to 119% annually for wheat (mean 82%; Figure S 4-3a; Table S 4-1). EPIC-BOKU and EPIC-IIASA show very high yields at 3.0-3.5 t ha^{-1}, while the other EPIC-MFWs have yield estimates around 2.0-2.5 t ha^{-1}. Yield estimates from PEPIC decrease further over time. The yield range decreases to 52% if harmonized planting dates and fertilizer application rates are used (Figure S 4-3b) and further to 32% with sufficient nutrient supply (Figure S 4-3c). In addition, the order of EPIC-MFWs mean biases changes: While EPIC-IISA provides the highest wheat yield estimates in the default runs followed by EPIC-BOKU, GEPIC produces the highest yields in the harm-suffN management and EPIC-BOKU the lowest.

As for maize, the continuous decrease in the relative range among EPIC-MFWs with increasing level of harmonization and elimination of nutrient limitations is contrasted by an increasing range for the non-EPIC-based MFWs (Figure S 4-4). This is mainly driven by on the one hand very high yield estimates by four non-EPIC-based GGCMs with non-nutrient limited wheat yield potential twice as high as the EPIC ensemble. On the other hand, one non-EPIC-based GGCM estimates very low yields, and two are at a level similar to the EPIC-MFWs.

EPIC-BOKU and PEPIC exhibit declining yield trends over the simulation period, whereas the other EPIC-MFWs show fairly stable fluctuating yields. For EPIC-BOKU this is most pronounced in the harm-suffN scenario and for PEPIC in the default setup. The inter-annual yield variability appears largely similar among all MFWs, but yield dynamics can be contrasting in certain years, especially in the second half of the study period. The whole EPIC-MFW ensemble indicates a peak in global average yield in 1993, which is picked up as well by the non-EPIC-based GGCMs (Figure S 4-4), but is not apparent in the reported data.

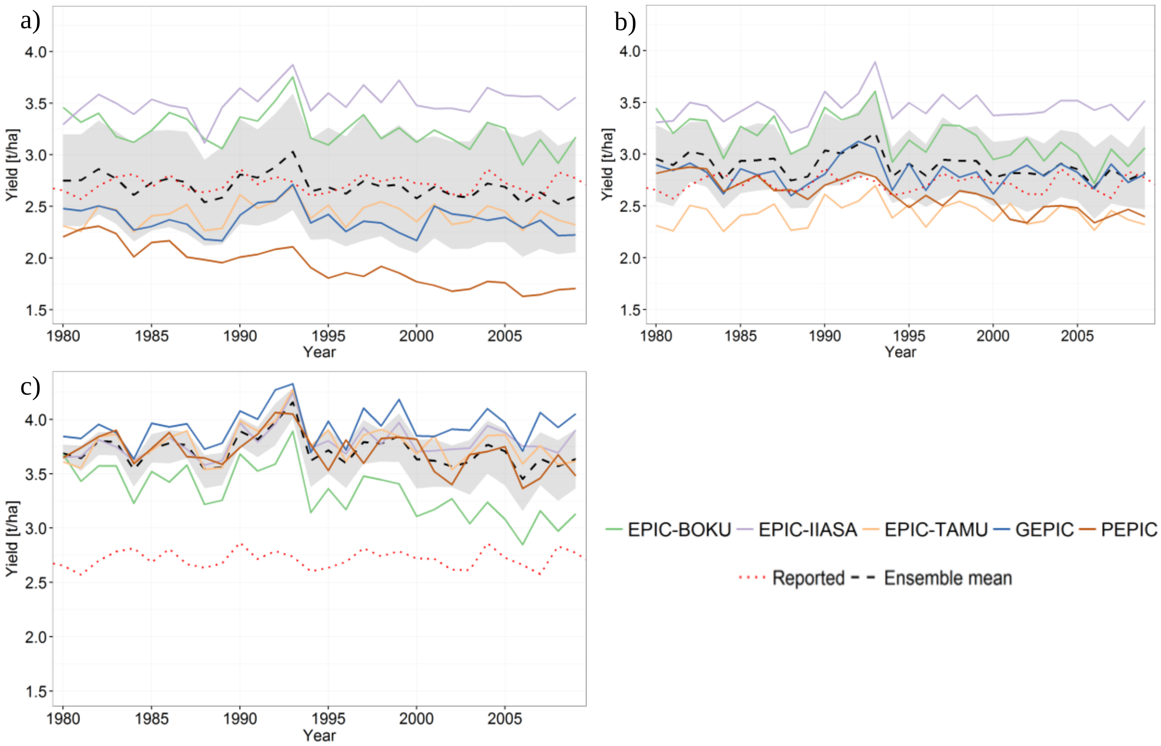

Figure S 4-3: Global average area-weighted wheat yield estimates of five EPIC-MFWs for the (a) default, (b) fully harmonized (fullharm), and (c) fully harmonized scenario with sufficient nutrient supply (harm-suffN) management scenario (Table 1 in manuscript). Reported yields are based on FAOSTAT (FAO, 2014) and have been detrended with a seven-year moving average (Elliott et al. 2015). The black dashed line represents the ensemble mean. The grey ribbon shows the 95% confidence interval of the mean. For EPIC-TAMU, outputs from the fully harmonized (fullharm) simulations were used as a substitute for missing default outputs to keep the number of EPIC-MFWs across management scenarios constant.

Table S 4-1: Relative spread of wheat yield estimates measured as yields of the highest estimate in relation to yields of the lowest estimate in Figure S 4-3. See Table 1 of the main paper for management scenarios

Management	**Relative range of wheat yield estimates**	
	Maximum [%]	Mean [%]
default	119	82
fullharm	52	42
harm-suffN	32	18

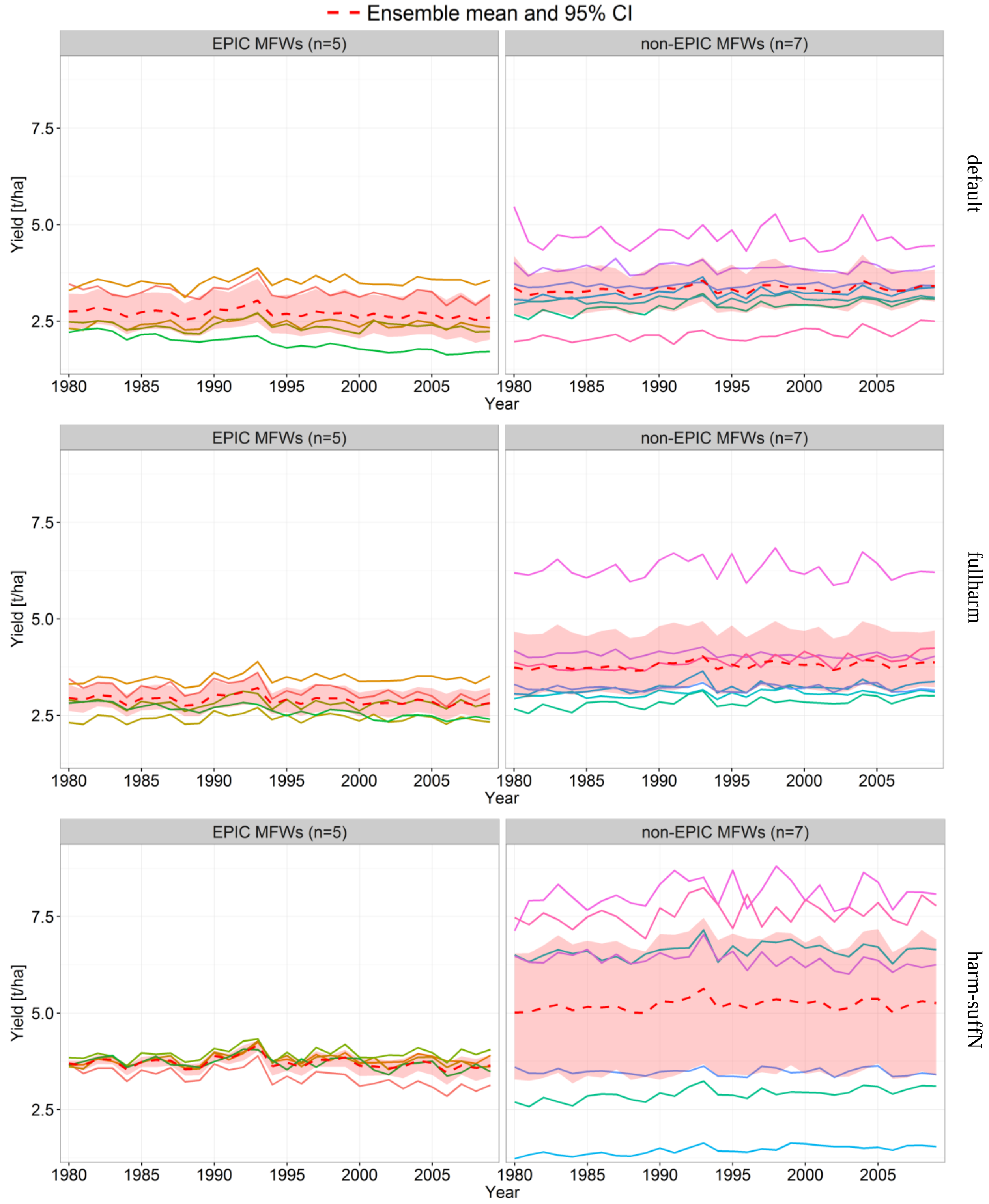

Figure S 4-4: Global average area-weighted wheat yields and 95% confidence interval of the mean for EPIC-MFWs and non-EPIC-based GGCMs for three management scenarios. Solid lines show outputs from single models.

S4.3 Spatial patterns of differences in maize yields among model frameworks

The evaluations of spatial differences in wheat yield estimates follow mostly the same pattern as those for maize. A major difference is that the CV among wheat yield estimates (Figure S 4-5) is in most regions lower than that for maize (Figure 3 in the manuscript) in all management scenarios. With sufficient nutrient supply, the largest differences remain in southern Africa, South America, India and especially at the edge of the Himalaya under rainfed cropping. The same is the case under irrigated conditions but with lower CV.

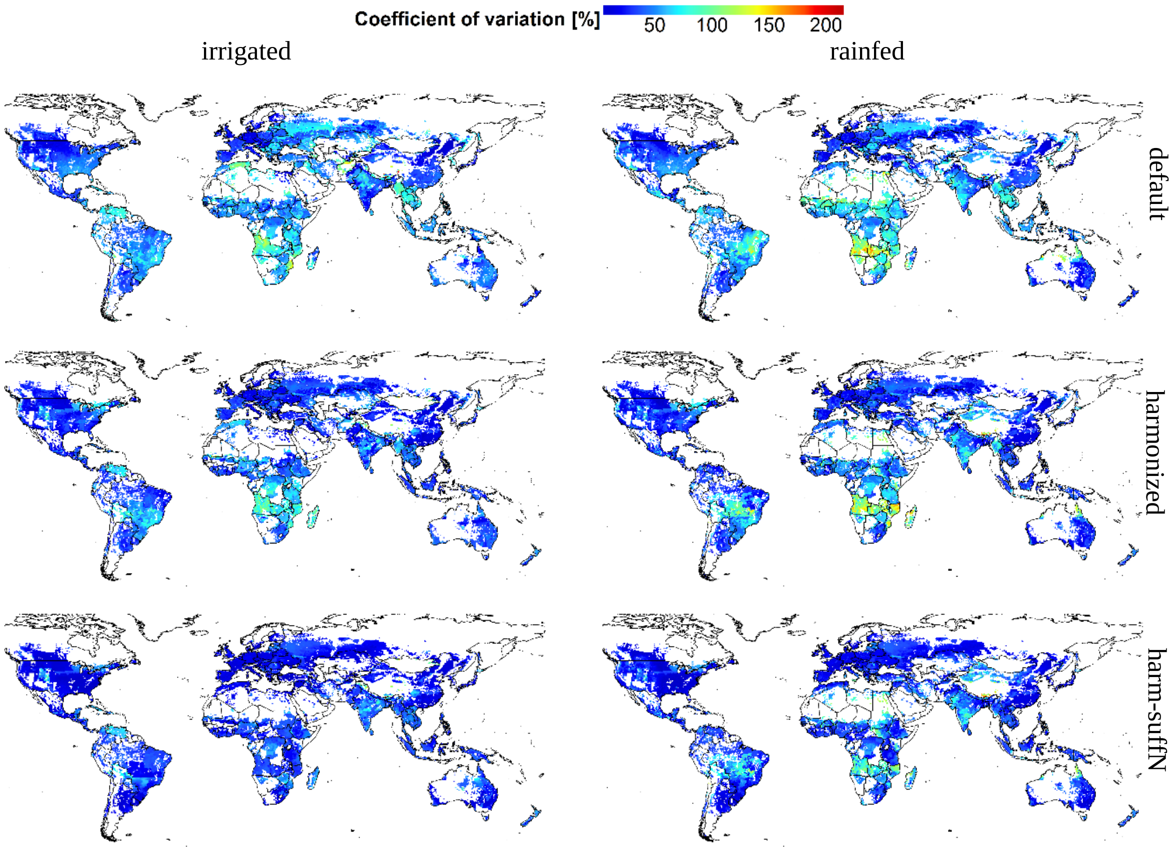

Figure S 4-5: Coefficient of variation for wheat yield estimates among EPIC-MFWs for each of the six crop management scenarios defined in Table 1 of the manuscript. For EPIC-TAMU, outputs from the fully harmonized (fullharm) simulations were used as a substitute for missing default outputs to keep the number of EPIC-MFWs across management scenarios constant.

S4.4 Impact of fertilizer supply on wheat yield deviations

The relationship between differences in yield estimates and fertilizer application rates follows for wheat (Figure S2-9) a similar pattern as for maize (Figure 3 in the manuscript). The CV decreases under irrigated conditions in all climates with increasing fertilizer application rates. This relationship is less pronounced than in the case of maize. Larger deviations remain at high fertilizer application rates in temperate and especially tropic zones. Under rainfed conditions, there is no relationship between nutrient application and model deviation in arid regions and only a weak one in cold regions. Temperate and tropic regions show a similar picture as with sufficient irrigation but with a wider range of yield deviations especially at low application rates.

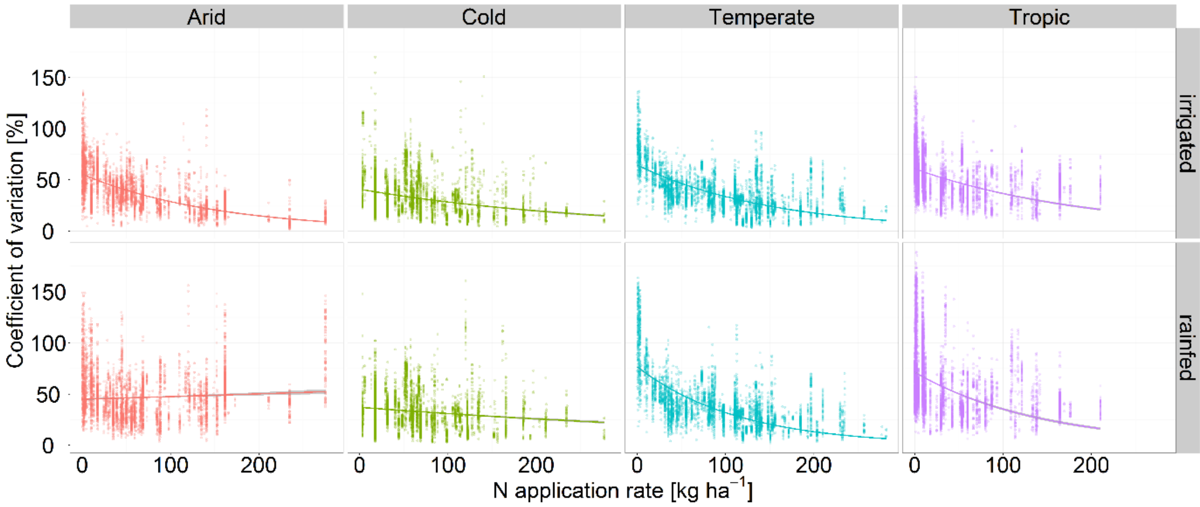

Figure S 4-6: Coefficient of variation for wheat yields among EPIC-MFWs compared to fertilizer application rates in the fully harmonized (fullharm) management scenario with sufficiently irrigated (upper row) or rainfed (lower row) water supply in each grid cell of four major climate regions.

S4.5 Differences in the performance of model frameworks in reproducing reported wheat yields

The difference in EPIC-MFW performance in reproducing reported wheat yields (Figure S 4-7) are less pronounced than those for maize (Figure 6 in the manuscript). The number of countries with top performance, number of countries with top performance and tscorr with r>0.5 and all countries with tscorr with r>0.5 for each EPIC-MFW are EPIC-IIASA (36,8,21), EPIC-TAMU (29,12,22), GEPIC (18,8,25), EPIC-BOKU (16,5,19), and PEPIC (11,3,13). High r values can mainly be found in Eurasia, south-eastern Africa, Australia, and Canada, whereas four of the five MFWs have r>0.5 in each of the latter two countries (Figure S 4-7d; Figure S 4-8). There is a high correlation with r>0.5 in five of the ten major wheat producing countries (Figure S 4-9). All MFWs exhibit a high tscorr in Russia, Canada, and Turkey.

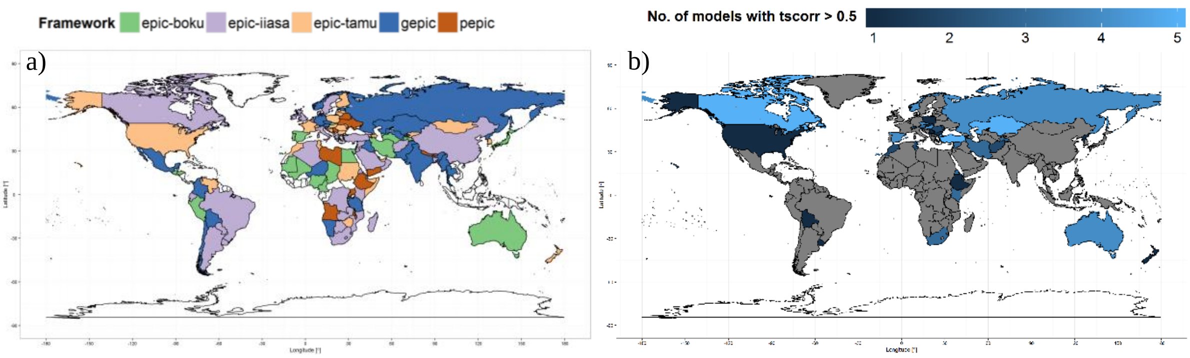

Figure S 4-7: EPIC-MFWs and setup scenarios showing the best performance in each country regarding time-series correlation (tscorr) factor r. (a) EPIC-MFWs with best performance for wheat in each country and (b) number of EPIC-MFWs in each country with r>0.5. Model outputs were post-processed by moving average detrending and mean-scale correction.

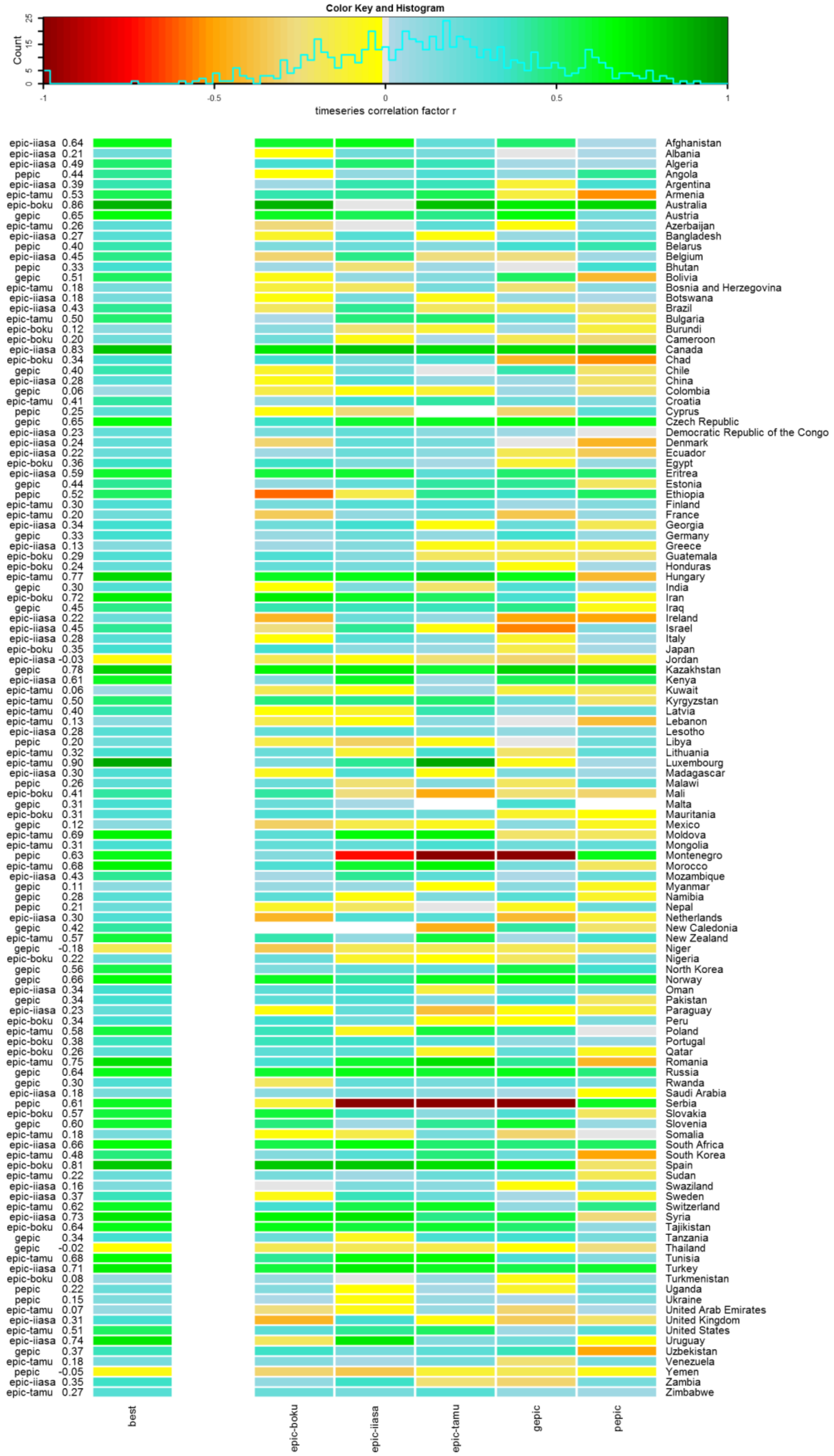

Figure S 4-8: Performance of all EPIC MFWs for estimating wheat yields in each country measured as time series correlation (tscorr). The left axis depicts the MFW with the highest performance and r value for the respective country shown on the right axis.

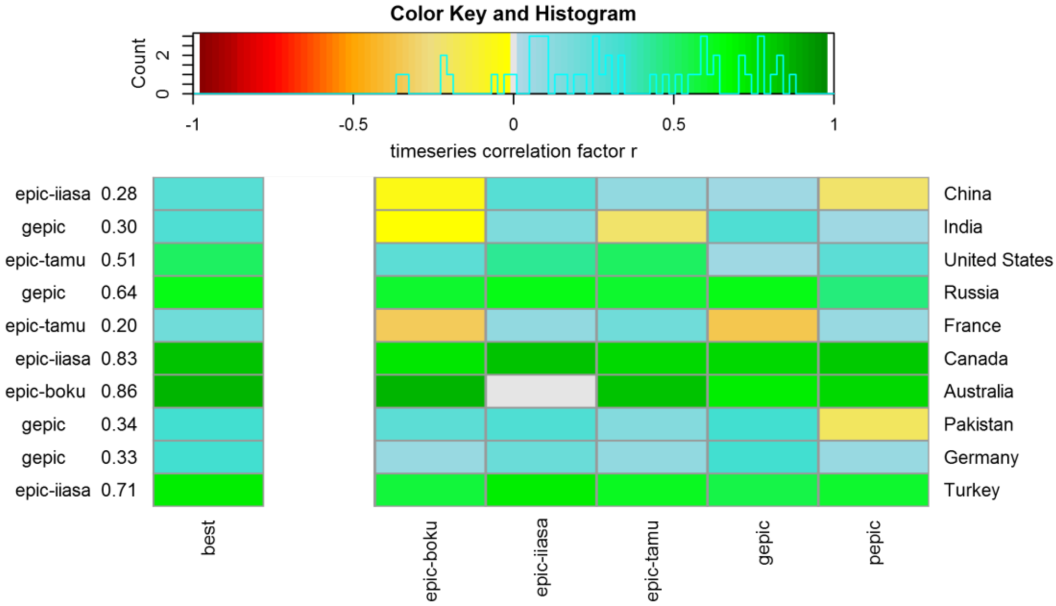

Figure S 4-9: Performance of all EPIC MFWs for estimating wheat yields in the top ten wheat producing countries measured as time series correlation (tscorr). The left axis depicts the MFW with the highest performance and r value for the respective country shown on the right axis.