Supplementary Material

1 Climate Data Selection and Processing

1.1 Historical Climate Data

CFSR data (Climate Forecast System Reanalysis; Saha et al., 2010) is taken as historical reference climate data for a baseline period, from 1996-2005. To ensure the accuracy of the baseline data set, the CFSR data is bias-corrected using climate records of three available local climate stations, located in and in close proximity to the study area (namely: Faisalabad, Lahore and Sialkot). Non-parametric quantile mapping is used as statistical fitting method between observed and simulated data over the time period 1979-2005. The bias correction is conducted using the R-package "Qmap" (Gudmundsson et al., 2012).

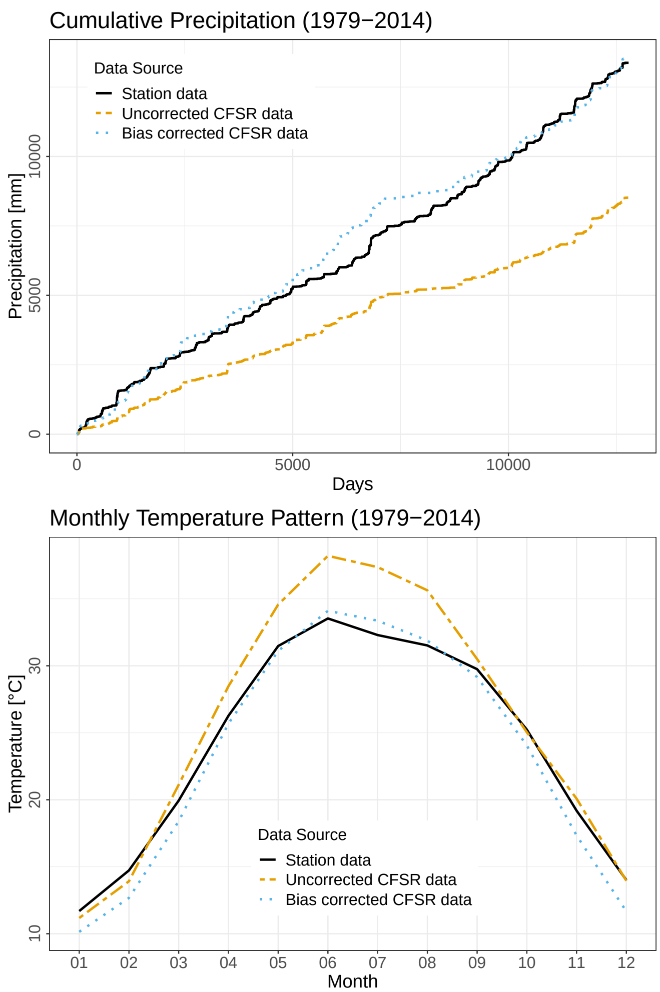

Figure S1: Results of bias correction procedure. Station data vs. uncorrected and bias corrected CFSR data

Figure S4 shows the fit of the reanalysis data before and after the bias correction, for precipitation and mean temperature. It shows, that the bias correction procedure was successful in removing the negative bias of original CFSR precipitation data as well as the positive bias of summer temperatures. Winter temperatures are "over corrected" and are now slightly underestimating temperatures from November until March. As the study focuses on climate impacts during the summer period (May-October), where a clear improvement of the fit between observed and bias corrected CFSR data can be achieved, we accept these results despite the winter-deviation from the observed data.

The same correction procedure (i.e. quantile mapping) was used to correct the remaining climate variables, namely Relative Humidity, Solar Radiation and Wind Speed, which are used by SWAT to calculate evapotranspiration rates according to Penman-Monteith (results not shown).

1.2 Climate Change Data

Climate change data is taken from the Coordinated Regional Climate Downscaling Experiment (CORDEX; www.cordex.org), which provides a suite of regional climate projections based on Global Climate Models of the Coupled Model Intercomparison Project, Phase 5 (CMIP5; Taylor et al., 2012).

We first select daily CORDEX climate projections, of 15 climate models, of the South-Asian CORDEX domain, at a resolution of 0.44 degrees x 0.44 degrees. CORDEX hindcast data of the same 15 GCM-RCM model combinations, is used to test the fit of CORDEX model outputs with the chosen baseline climate data from the CFSR-data set. Here, we use the maximum overlapping time period of the data products (1979-2005) to test their agreement. The performance of single CORDEX GCM-RCM combinations with respect to the baseline climate data is displayed in Figure S2, which shows their goodness of fit in terms of standard deviation, correlation and RMSE, with respect to each climate variable (i.e. precipitation, temperature, solar radiation, relative humidity, and wind speed).

The unsatisfactory performance of all models which employ the Regional Climate Model “RegCM4-4” for downscaling purposes (Figure S2, triangles), leads to the exclusion of these models. Based on the goodness of fit between CORDEX and CFSR data, we select the data sets of all 9 models which use “RCA4” Regional Climate Models for downscaling (Figure S.2, dots).

To minimize any further bias, which the data sets of 9 selected RCMs still present, we apply a commonly used linear scaling correction approach (Teutschbein and Seibert, 2012) to each ensemble member and to each climate variable. Here, the relative changes, based on 10-year monthly mean difference between CORDEX hindcasts (1996-2005) and CORDEX future forecasts (2021-2030 and 2041-2050) are used as correction factors to create new daily future climate times series. The final future climate time series consist of the daily CFSR baseline data plus the added difference between CORDEX hindcasts and CORDEX future data, defined by the respective monthly correction factors.

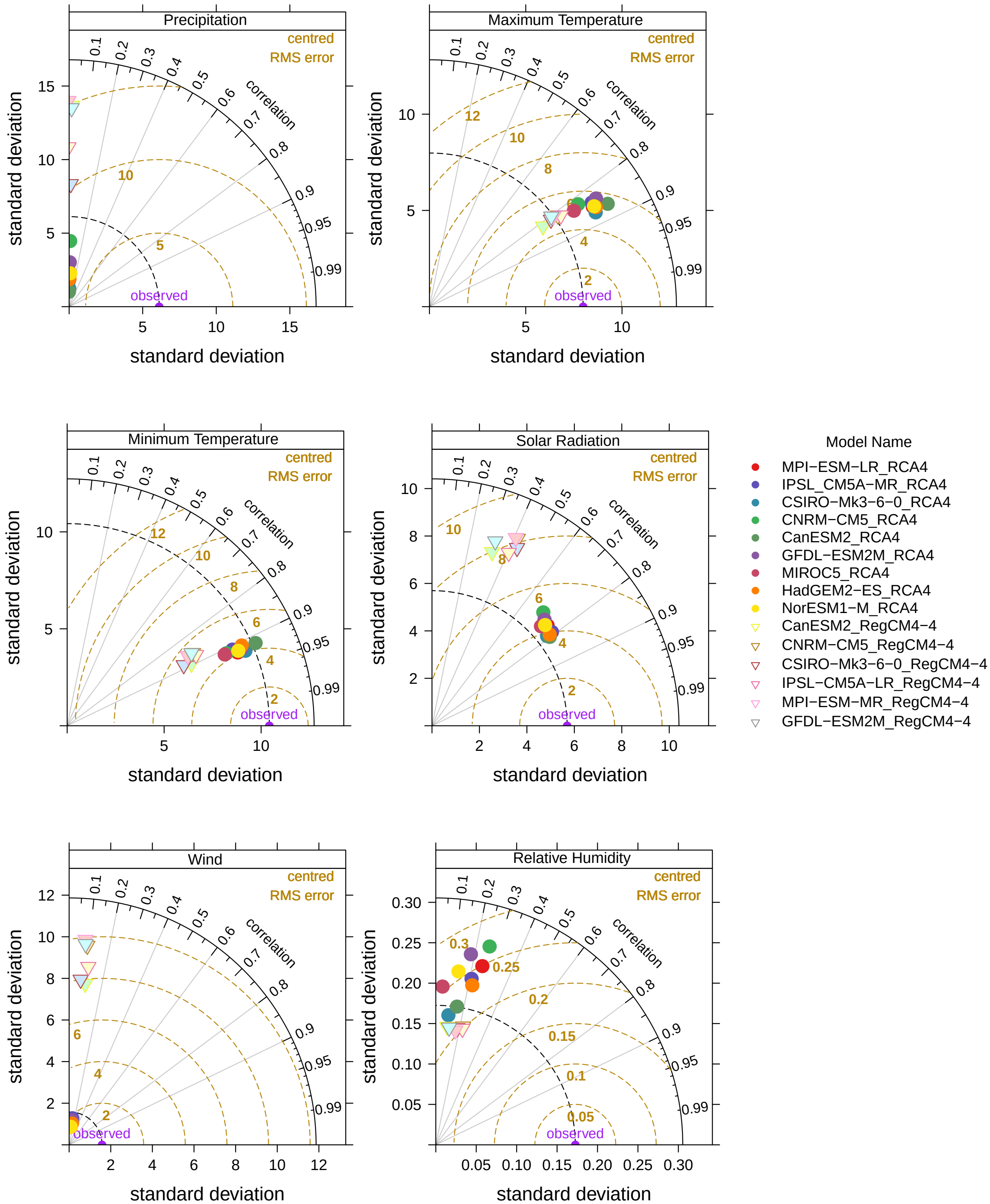

Figure S2: Goodness-of-Fit between CFSR Reanalysis Data and CORDEX GCM-RCM models

2 SWAT and APSIM

2.1 Model description and model differences

In this study, we use two models from two different research communities. On the one hand we use the hydrological Soil & Water Assessment Tool (SWAT, Version SWAT2012 rev 664; Arnold et al., 2012) and on the other hand the biophysical-crop modelling framework Agricultural Production Simulator (APSIM, Version APSIM classic; Holzworth et al., 2014). Based on their model structure, both models have strengths and weaknesses in predicting future yields and plant water requirements. Calculation procedures for central variables discussed in the main paper (i.e. biomass production, LAI and yield) are as follows:

2.1.1 SWAT

Potential biomass production (biom) calculated in SWAT is based on plant specific radiation-use efficiency (RUE), which defines the conversion of intercepted light at the leaf surface into biomass, and the amount of intercepted light available for photosythesis ($H_{photosyn}$).

$$\Delta biom = RUE * H_{photosyn} \tag{1}$$

While the radiation-use efficiency itself is assumed to be independent of the plant growth stage, the amount of intercepted light depends on the plant leaf area development (Neitsch et al 2009).

$$H_{photosyn} = 0.5 * H_{day} * (1 - exp(-k_l * LAI_{act})) \tag{2}$$

With H_{day} being the incident total solar radiation, k_l being the light extinction coefficient and LAI being the leaf area index.

The leaf area index, which controls the amount of intercepted light, is simulated dynamically based on the concept of potential heat units (PHUs). The heat unit theory assumes plant specific temperature requirements for the different phenological stages of plant maturation, denominated as "heat units", or more commonly known as "growing degree days". Heat units are accumulated over time and control the leaf and plant growth until a maximum LAI and plant maturity is reached, after which leaf senescence begins and LAI declines.

To estimate the actual plant growth, the actual LAI (LAI_{act}) and the actual biomass growth ($biom_{act}$) are reduced according to the stress experienced by plants due to extreme temperature stress (tstrs). Water stress (wstrs) and nutrient stresses (nstrs and pstrs) are reduced to a minimum in our study by assuring constant irrigation and sufficient fertilization.

$$\Delta LAI_{act} = \Delta LAI * \sqrt{1 - max(wstrs, tstrs, nstrs, pstrs)} \tag{3}$$

$$\Delta biom_{act} = \Delta biom * \{1 - max(wstrs, tstrs, nstrs, pstrs)\} \tag{4}$$

Finally, yield is estimated in SWAT by multiplying the actual biomass, produced at the time of harvest, by a plant specific harvest index (HI).

$$Yield = biom_{act} * HI_{act} \tag{5}$$

The harvest index depends on the fraction of accumulated potential heat units (fr_{PHU}) and is defined as

$$HI = HI_{opt}(\frac{100 * fr_{PHU}}{100 * fr_{PHU} + exp[11.1 - 10 * fr_{PHU}]}). \tag{6}$$

The actual harvest index (HI_{act} in eqn. 5) is a reduced HI, depending on the impact of water deficit stresses. In our study water deficit is close to zero and HI = HI_{act}, due to constant and demand based irrigation. Optimal harvest indices (HI_{opt}) used in this study are 40% for cotton and 50% for maize and rice, according to Awan et al. (2016).

This procedure shows that SWAT simulations of leaf area development, biomass and ultimately yield production are highly dependent on one single dominating environmental stress factor. In our study this results in a strong sensitivity of plant productivity to temperature stress, which leads to significant reductions of LAI, biomass and yield with increasing temperatures.

2.1.2 APSIM

APSIM is a modelling framework with separate crop modules for each plant type. It uses the Ozcot Model for cotton simulations (Hearn, 1994), the CERES-based maize model for maize simulations (Jones and Kiniry, 1986) and the Oryza2000 model for rice simulations (Bouman et al., 2001). Each model accounts for crop specific physiologies such as plant phenology, photosynthesis, plant stresses, nutrient cycling and carbon allocation. Details on the model structures and calculation steps for each separate model can be found the the above cited references.

The main difference to SWAT is the strength in accounting for more detailed bio-physical processes. The cotton model for example, includes a plant respiration factor (Resp) in the photosynthesis calculation (adopted from Hearn 1994).

$$H_{photosyn} = 2.391 + H_{day}(1.374 - (0.0005414 H_{day})) * (1 - exp(-k_l * LAI)) - Resp \tag{7}$$

The respiration factor is affected by temperature and water stress. Thus the photosynthesis and carbon assimilation part already accounts for the direct impact of stress factors and not only indirectly through LAI (eqn. 2). The maize model uses a similar concept to SWAT for the estimation of the light driven biomass production but directly includes temperature, phosphorus and nitrogen stresses into the calculation (adopted from Jin et al 2016).

$$\Delta biom = H_{day} * RUE * min\{tstrs, nstrs, pstrs\} \tag{8}$$

Accounting for all stresses (i.e. water, temperature and nutrients) in different phenological stages, makes the APSIM models less sensitive to one single dominating stress (= temperature stress in SWAT). Yet, some stages are lacking the impact of temperature stress, which might lead to an overestimation of plant productivity in environments, where the remaining stresses are low. The cotton model for example does not account for heat stress in the initial parts of leaf area development. In the leaf area formation, temperature stress is not considered until the first square event. After the first square event, it is indirectly included through vapor pressure deficit (VPD).

$$\Delta LAI_{act} = \sqrt{0.1847 - 0.1165 * SMI - 1.514 * VPD + 1.984 * SMI * VPD} \qquad (9)$$

When the soil moisture index (SMI) and VPD are low (as in the case of our intensively irrigated and from monsoon rainfall impacted region), LAI development will hardly be affected by environmental stresses. This is one reason for the strong differences between SWAT and APSIM results regarding their LAI estimations.

Yield estimation of the APSIM models are based on crop specific fruiting dynamics rather than on the more stringent harvest index (HI) method in SWAT. Yield is estimated based on grain number and grain filling (maize and rice) or ball growth rates (cotton), which makes it less dependent on LAI and dry matter production. This explains, why yield declines are projected even under strengthening leaf area development.

2.2 Management parameters

The following table lists the settings of both models regarding management parameters, which define irrigation settings, planting, growing and harvesting decisions as well as plant type specific information on optimal growing conditions (i.e. optimal temperature ranges of single cultivars).

Table S1: Management parameters

Parameter	APSIM	SWAT
Irrigation frequency	demand based	demand based
Irrigation efficiency [%]	0.7	0.7
Irrigation trigger	soil water deficit	soil water deficit
Fraction of available soil water below which irrigation is applied	0.9	0.9
Fertilizer application	on sowing day and each 14 days	with irrigation
Fertilizer type	Urea_N	Urea
Sowing rule	fixed date	fixed date
Cultivar type - cotton	S71BR	not specified (from SWAT data base)
$T_{opt1;2}$ cotton [°C]	20;30	30
T_{base} cotton [°C]	8	15
Cultivar type - maize	Pioneer_3153	not specified (from SWAT data base)
$T_{opt1;2}$ maize [°C]	15;30	25
T_{base} maize [°C]	8	8
Cultivar type - rice	BR3	not specified (from SWAT data base)
T_{opt} rice [°C]	30	25
T_{base} rice [°C]	8	10

While APSIM offers the possibility to select specific cultivar types (e.g. S71BR, Pioneer_3153 and BR3), SWAT only facilitates the selection of one generic crop type (e.g. cotton, rice and maize). Due to missing information on regional specific cultivar selections, the APSIM cultivar types were chosen according to the fit between simulated APSIM yield levels and observed yield levels documented in Agricultural Statistics of Pakistan (http://www.finance.gov.pk/survey/chapters_18/02-Agriculture.pdf).

2.3 Soil parameters

SWAT is able to account for spatially distributed soil information and distinguishes between five different soil type in the study area. Their local names are: Buchiana, Chuharkana, Farida, Jhang and Nokhar soils. The soil characteristics were adopted from the study by Awan et al. (2016) and thoroughly calibrated in a complex calibration procedure described in Becker et al. (2019). In addition, a field campaign was conducted and soil samples were taken. Laboratory soil test results were used to validate and improve the soil information.

The most prominent soil type is the so called "Jhang" soil type - a sandy-loam which covers approx. 63% of the study area. The second most abundant soil type (approx. 24% of the study area) is called "Farida" soil, with similar grain size distribution and soil characteristics as the "Jhang" soil. A sensitivity check of APSIM yield results with respect to soil parameters was conducted. Using "Farida" as well as "Jang" soil characteristics, APSIM results revealed that yield levels do not show any significant variation with respect to differences in these two soils (yield difference = 1.5%) . We therefore select the prominent "Jhang" soil characteristics for the APSIM soil module, while maintaining all characteristics of the five different soils for SWAT. Detailed characteristics of Jhang soils are given below and details on the remaining soil types and soil layers can be made available on request or found in Becker et al. (2019).

Table S2: Soil parameters of Jhang soils

Depth	Bulk density	AirDry	LL15 (Wilting Point)*	DUL (Field Capacity)	SAT**	KS	AWC
cm	g/cm3	mm/mm	mm/mm	mm/mm	mm/mm	mm/day	mm/mm
0-15	1.45	0.102	0.203	0.380	0.403	11280	0.177
15-30	1.61	0.180	0.225	0.310	0.342	11280	0.085
30-60	1.61	0.225	0.225	0.310	0.342	11280	0.085
60-90	1.59	0.223	0.223	0.300	0.350	11280	0.077
90-120	1.59	0.223	0.223	0.300	0.350	2064	0.077
120-150	1.59	0.223	0.223	0.300	0.350	2064	0.077

* calculated according to SWAT manual (Wilting Point = Field Capacity - Available Water Capacity (AWC))
** calculated according to APSIM Soil manual (SAT = 1-(Bulk Density/2.65)-0.05)

Table S3: Soil grain size distribution for Jhang soils

Depth	Clay	Sand	Silt
cm	%	%	%
0-15	3.5	65.5	31
15-30	3.5	65.5	31
30-60	3.5	65.5	31
60-90	3.5	65.5	31
90-120	6.5	84.4	9.1
120-150	6.5	84.4	9.1

2.4 Regional data sets used in SWAT and APSIM

Table S4: Regional data sets

Variable	Spatial Resolution	Temporal Resolution	Source
DEM	90x90 m	-	SRTM, NASA
Soil Map	500x500 m	-	Water and Soil Investigation division (WASID), Pakistan
Land-use classes	250x250 m	-	Awan and Ismaeel (2014)
Meteorological Station Data	Point data (3 Stations)	daily	Pakistan Meteorological Department (PMD)
Meteorological Reanalysis Data	approx. 25x25 km (gridded)	daily	globalweather.tamu.edu
Climate Change Projections	approx. 50x50 km (gridded)	daily	www.cordex.org

2.5 Model validation

Validating models such as SWAT and APSIM using observed yield data is challenging. Closely matching observed yield levels taken for example from national agricultural statistics is difficult, mainly due to model limitations in representing the actual environmental conditions, short-term changes in management strategies, famer's capabilities to react flexibly to environmental dynamics, potential plant diseases, technical standards in irrigation or harvesting, etc.. Using data from national statistics, which are often based on data at national or provincial level might further add to the uncertainty in yield validation of models which operate on regional scales.

Yet, to prove that crop model results show reasonable yield predictions a validation of yield estimates is necessary and a comparison with observed data is a common procedure. To support the validity of the models used in this study, we compare their yield predictions with observed yield data from Agricultural Statistics of Pakistan, published by the Ministry of National Food Security & Research (http://www.mnfsr.gov.pk/frmDetails.aspx, last accessed: 12/22/2020). Yield data for cotton, rice and maize from the province of Punjab was taken for the years 2009-2013 and compared to simulated yield levels by SWAT and APSIM. The time period 2009-2013 corresponds to the last available 5 years of the CFSR climate data set, which is taken as reference climate data input in this study. The models were run for the same time period with the respective reference climate data and, due to the above mentioned difficulties in comparing the results with observed data, their average yield (mean of 5 years) was compared with the 5-year-average of the observed yield levels.

The validation results are shown in Fig. S3, where "AGRIStats" corresponds to the data from Agricultural Statistics of Pakistan. Uncertainty bars show +/- one standard deviation. Deviations from estimated yield levels by APSIM are 15% for cotton and 19% for maize and rice. SWAT shows a stronger deviation (33% for cotton, -21% for maize and -34% for rice), but at the same time higher uncertainties in yield estimation.

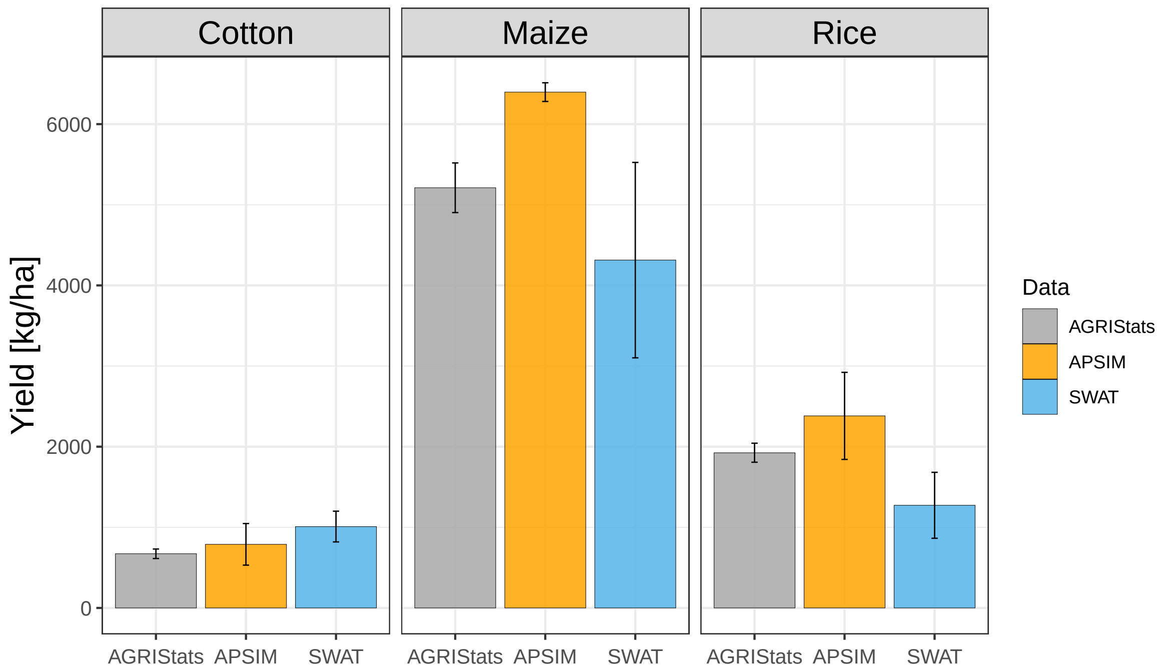

Figure S3: Validation of crop yield estimations

3 Crop specific results

In the main text we mention that our deductions on crop responses to climate change are based on the average trends estimated for maize, cotton, and rice crops. Yet, the three selected crops are reacting differently to climate change and plant specific reactions should be taken into account, when impacts on individual crop types are the focus. The following figures give detailed information on the reaction of each crop type to the climate change scenarios of this study.

Separating the changes of each crop type reveals that even though the magnitudes in crop reactions to climate change are different the crops show similar responses (e.g. strong improving trend in yield and biomass with increasing CO_2; decreasing Irrigation demand with increasing temperatures for SWAT, less effect for APSIM; clearly decreasing SWAT-LAI for all crops and significant positive CO_2 effects on APSIM-LAI). With the exception of APSIM-maize, it can be stated that the difference in plant reaction to climate change is larger between the models than between the crop types. Which again hints to the importance of considering model structural differences.

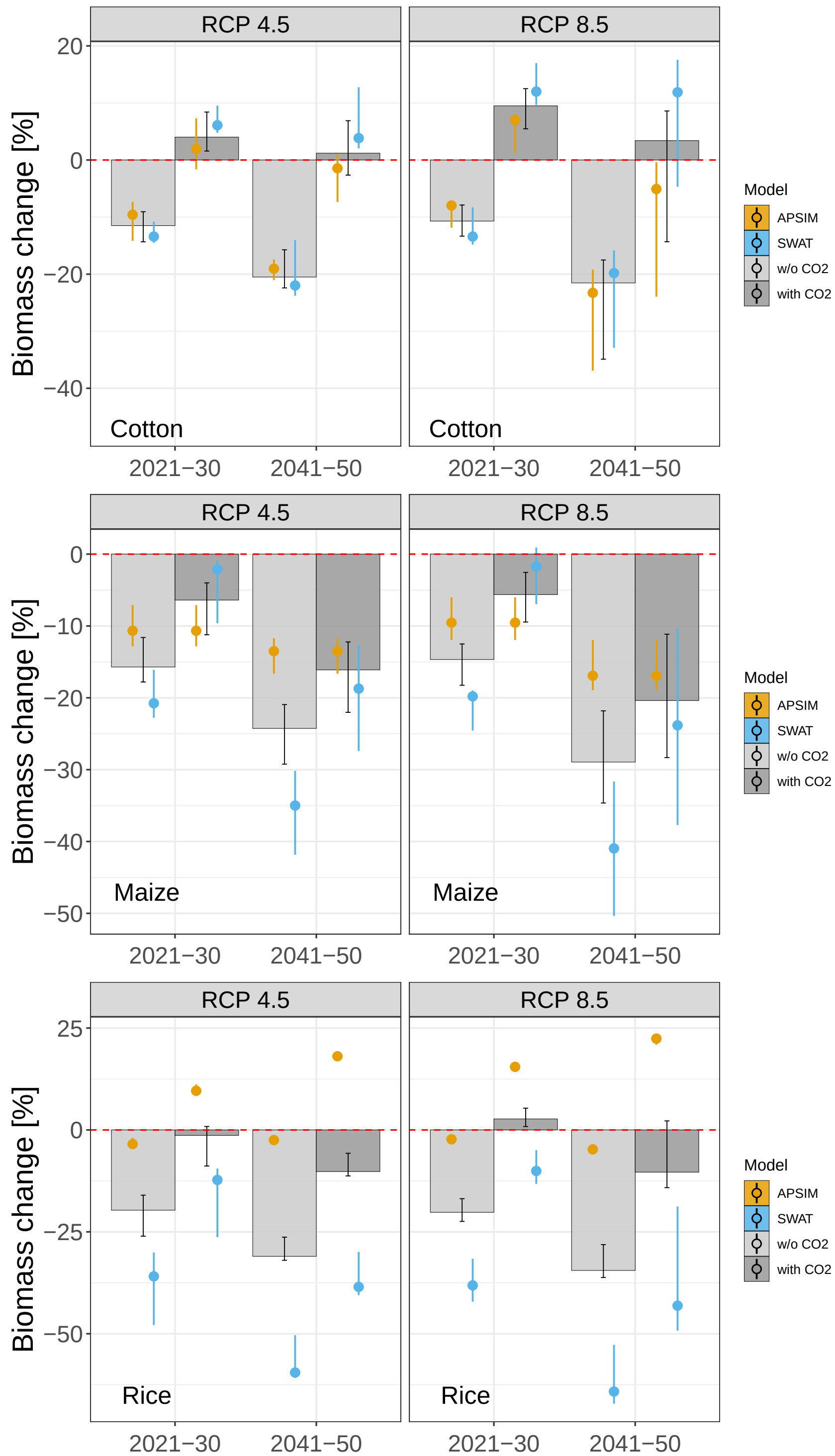

Figure S4: Simulated biomass changes for cotton, maize and rice

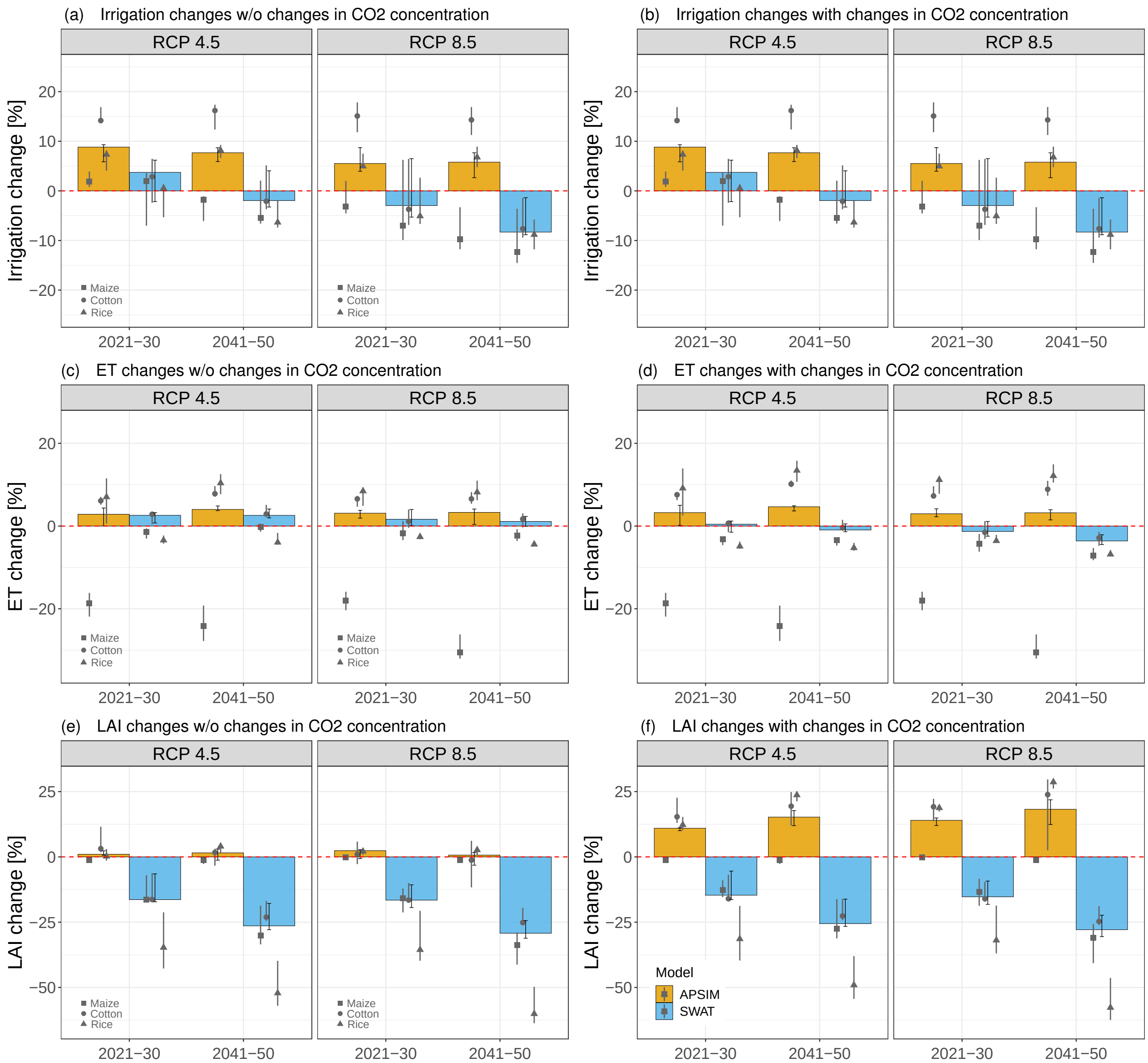

Figure S5: Simulated changes in irrigation demand, ET and LAI showing mean changes per model and for each crop type

References

Arnold, J. G., D. N. Moriasi, P. W. Gassman, K. C. Abbaspour, M. J. White, R. Srinivasan, C. Santhi, R. D. Harmel, A. van Griensven, M. W. Van Liew, N. Kannan, and M. K. Jha (2012). SWAT: Model Use, Calibration, and Validation. *Transactions of the ASABE*, 55(4):1491–1508.

Awan, U. K., Liaqat, U. W., Choi, M., and Ismaeel, A. (2016). A SWAT modeling approach to assess the impact of climate change on consumptive water use in Lower Chenab Canal area of Indus basin. *Hydrology Research*, 47(5):1025–1037.

Becker, R., Koppa, A., Schulz, S., Usman, M., aus der Beek, T., and Schüth, C. (2019). Spatially distributed model calibration of a highly managed hydrological system using remote sensing-derived ET data. *Journal of Hydrology*, 577:123944.

Bouman, B., Kropff, M., Toung, T., Wopereis, M., Ten Berge, H., and Van Laar, H. (2001). *ORYZA2000: Modeling Lowland Rice*. International Rice Research Institute, and Wageningen University and Research Centre, Los Banos, Philippines and Wageningen, Netherlands.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T. (2012). Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations. A comparison of methods. *Hydrology and Earth System Sciences*, 16(9):3383–3390.

Hearn, A. B. (1994). OZCOT: A simulation model for cotton crop management. *Agricultural Systems*, 44(3):257 – 299.

Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore, A. D., Brown, H., Whish, J. P., Verrall, S., Fainges, J., Bell, L. W., Peake, A. S., Poulton, P. L., Hochman, Z., Thorburn, P. J., Gaydon, D. S., Dalgliesh, N. P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp, J., Cichota, R., Vogeler, I., Li, F. Y., Wang, E., Hammer, G. L., Robertson, M. J., Dimes, J. P., Whitbread, A. M., Hunt, J., van Rees, H., McClelland, T., Carberry, P. S., Hargreaves, J. N., MacLeod, N., McDonald, C., Harsdorf, J., Wedgwood, S., and Keating, B. A. (2014). APSIM – Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software*, 62:327–350.

Jones, C. and Kiniry, J. (1986). *CERES-Maize: A simulation model of maize growth and development*. Texas A&M University Press, College Station.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M. (2010). The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, 91(8):1015–1058.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, 93(4):485–498.

Teutschbein, C. and Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, 456-457:12–29.