

Seventh Framework Programme
Theme 6
Environment



Project: 603864 – HELIX

Full project title:
High-End cLimate Impacts and eXtremes

Deliverable: 4.3

**Report on the sensitivities of impact models to climate drivers as well as their
ability to represent effects of extreme events**

Version 2.0

Original Due date of deliverable: 31/10/16

Actual date of submission: 31/10/16

Title	Report on the sensitivities of impact models to climate drivers as well as their ability to represent effects of extreme events
Creator	Richard Betts
Editor	Richard Betts
Brief Description	This report assesses the performance of a number of models in simulating biophysical processes in response to climate variability and extremes. This includes hydrological models, simulating flood and drought conditions, land biosphere models, and global crop models.
Publisher	
Contributors	Lorenzo Alfieri, Hylke Beck, Gustavo Naumann, Philippe Ciais, Jingfeng Chang, Katja Frieler, Anselmo Garcia Cantu Ros, Fang Zhao, Catherine Morfopoulos,
Format	Report
Creation date	7/10/2015
Version number	0
Version date	04/08/2016
Last modified by	GN
Rights	
Audience	<input type="checkbox"/> internal <input checked="" type="checkbox"/> public <input type="checkbox"/> restricted, access granted to: EU Commission
Action requested	<input checked="" type="checkbox"/> to be revised by Partners involved in the preparation of the deliverable <input type="checkbox"/> for approval of the WP Manager <input type="checkbox"/> for approval of the Internal Reviewer (if required) <input type="checkbox"/> for approval of the Project Co-ordinator
Deadline for approval	

Report on the sensitivities of impact models to climate drivers as well as their ability to represent effects of extreme events

October 31

2016

Contents

PART A: LISFLOOD HYDROLOGICAL MODEL AND RELATED WORK

1. LISFLOOD hydrological model.....	5
1.1 Lisflood	5
1.2 LISVAP	6
1.3 Model performance	6
1.4 Model Calibration and Evaluation.....	8
2. Characterization of droughts using standardized indicators	10
2.1 Projection of occurrence of extreme dry events Using the Standardized Precipitation Index and Standardized Precipitation Evaporation Index.....	10
2.2 Standardized precipitation index (SPI) and Standardized Precipitation-Evaporation Index (SPEI).....	10
2.3 The relative SPI / SPEI	14
3. APPENDIX.....	21
3.1 Calibration score cards	21
4. References	40

PART B. EVALUATING AND IMPROVING THE SIMULATION OF PEAK RIVER DISCHARGE BY GLOBAL HYDROLOGICAL MODELS

42

PART C. MEAN AND LOW HYDROLOGICAL STATES IN JULES

43

PART D. ABILITY OF BIOME MODELS TO REPRODUCE THE MAGNITUDE OF CARBON FLUX CHANGES FOR EXTREME CLIMATE ANOMALIES PROVIDED BY LA NINA AND EL NINO OCCURRENCES OVER THE THE LAST 40 YEARS

59

PART E. THE EXPONENTIAL RATE OF CHANGE IN GLOBAL BIOMASS STOCKS: PLANT MORTALITY, EL NINO ANOMALIES

63

PART F. UNDERSTANDING THE WEATHER-SIGNAL IN NATIONAL CROP-YIELD VARIABILITY

71

PART G. REPRESENTATION OF HEAT EFFECTS ON CROP YIELDS IN GLOBAL GRIDDED CROP MODELS

73

1. LISFLOOD HYDROLOGICAL MODEL

1.1 LISFLOOD

The hydrological simulation of large river basins forced by a variety of meteorological input and of climate projections have become feasible through the revolutionary increase in computing and network power over the past few decades and the advanced Geographical Information System (GIS) analysis tools that are now available. Traditionally, GIS was associated with static data such as topography, forest areas, urban areas or infrastructure, occasionally satellite images, often in a rather static way. The incorporation of dynamic data sets such as remotely sensed data (vegetation cover, snow cover), gridded meteorological data (precipitation, temperature, evaporation, either observed or forecasted), and weather station observations are becoming increasingly used also in a GIS environment to form the basis for effective hydrological monitoring on a variety of spatial and temporal scales.

The hydrological model described here is called Lisflood. It was developed at the Institute for Environment and Sustainability (IES) of the European Commission - Joint Research Centre (JRC) and its main features are described by van der Knijff et al. (2010) and Burek et al. (2013a). Lisflood is a hybrid between a conceptual and a physically-based rainfall-runoff model combined with a routing module for transporting the water through river channels. It has been specifically designed for large river catchments (De Roo et al. 2001) but has also been applied to smaller catchments (e.g., Alfieri et al. 2012; Younis et al. 2008). Recent works using different model versions include the comparison of various satellite precipitation datasets over Africa (Thiemig et al. 2013), a global setup for streamflow estimation, flood forecasting and early warning system (Alfieri et al., 2013; Beck et al., 2015), and the simulation of climate scenarios over Europe on multi-decadal time scales for floods (Alfieri et al., 2015; Rojas et al., 2012) and droughts (Forzieri et al. 2014) impact assessment under a changing climate.

Whenever possible parameters in Lisflood are linked to physical properties, e.g. soil or land use properties. Although for all model parameters default values are proposed, a number of calibration works have been performed on different model domains for better model performance (e.g., see Alfieri et al., 2014; Revilla-Romero et al., 2015). Analysis of model parameter uncertainty and its impact on streamflow simulated by the Lisflood model is presented in Feyen et al. (2007). The JRC stores all maps required to set up Lisflood at the global and European scale, with grid resolution as fine as 0.1 degrees and 1 km, respectively. A global setup at 0.5 degree grid resolution was recently set up to enable transient daily simulations over the time range of a century, and for facilitating the data sharing with international initiatives such as ISI-MIP (Warszawski et al., 2014). In the context of HELIX, in the WP4, global hydrological simulations with Lisflood are foreseen to be run at 0.1 and 0.5 degree grid resolution. In particular, the higher resolution runs (0.1 deg) are used to validate the model in the quantitative estimation of the streamflow characteristics, while the 0.5 degree runs will be dedicated to hydrological projections of the future global streamflow, focusing on the expected changes as compared to a consistent baseline scenario. This appears as a good tradeoff between computational feasibility and high enough level of detail to reproduce accurate streamflow statistics for a wide range of river basins. Available global static maps at both resolutions include maps related to topography (i.e., digital elevation model, local drain direction, slope gradient, elevation range), land use (i.e., land use classes, forest fraction, fraction of urban area), soil (i.e., soil texture classes,

soil depth), and channel geometry (i.e., channel gradient, Manning's roughness, bankfull channel depth, channel length, bottom width and side slope).

Monthly climatic mean Leaf Area Index (LAI) maps were derived from SPOT-VGT data (http://wdc.dlr.de/data_products/SURFACE/LAI/). Meteorological data needed as input from the Lisflood model are precipitation, average daily temperature, and daily potential evaporation rate (from free water surface, bare soil and reference crop). These maps are to be provided for each time step of simulation (i.e., daily or sub daily), although for temperature and evaporation, daily or longer term averages are required by the model.

1.2 LISVAP

When potential evaporation maps are not directly available for use in Lisflood, they can be estimated from other meteorological variables from different potential evaporation formulations. To this aim, a software named Lisvap (Burek et al. 2013) was developed at the JRC-IES. Lisvap is a pre-processor that calculates potential evaporation from gridded meteorological observations that can subsequently be used as input to Lisflood. The approach is based on the Penman-Monteith equation or the Hargreaves equation. The calculation of potential evaporation is confounded somewhat by the fact that the different datasets that are usually available are rather heterogeneous. For instance, incoming solar radiation can be estimated from sunshine duration or cloud cover data. Some data suppliers do not offer this kind of information, but provide pre-calculated grids of components of the radiation balance instead. Wind speed is sometimes provided in the form of U and V components. Vapour pressure is sometimes used in place of dew point temperature. Lisvap has been designed to handle this heterogeneity in a straightforward way. It contains various options that allow the user to select the data to use, and combinations of different data sources (e.g. vapour pressure and dew point temperature) can be combined within one Lisvap run. Similar to Lisflood, Lisvap is implemented in the PCRaster Environmental Modelling language (Wesseling et al. 1996), wrapped in a Python based interface. It runs on any operating system for which Python and PCRaster are available. Currently, these include 32 and 64-bits Windows (e.g. Windows XP, Vista, 7, 8) and a number of Linux distributions.

1.3 MODEL PERFORMANCE

Observed streamflow (Q) data from 966 small catchments (1000–5000 km²) around the globe were used to evaluate the Q estimates (1979–2012) of LISFLOOD driven by the WATCH Forcing Data methodology applied to ERA-Interim reanalysis (WFDEI, Weedon et al., 2014). The model output is a set of hydrological variables in the form of daily maps in raster format, at 0.1x0.1 deg resolution, though for the validation work, simulated streamflow was resampled to a 0.5x0.5 deg grid. We used small catchments to minimize the influences of channel routing and human activity. In total 966 catchments (median size 1970 km²) were found to be suitable for the analysis, of which 641 catchments had daily Q data and 325 catchments (mainly located in Russia) had only monthly Q data. The median record length during 1979–2012 was 19 y. For the evaluation, we computed Pearson linear correlation coefficients (r) between daily, log-transformed daily, 5-day, monthly, monthly climatologic, and annual time series of simulated and observed Q. The r values range from -1 to 1, with higher values corresponding to better model performance. In terms of daily correlation (see Figure 1), good results were obtained for the Rocky Mountains and the northern Appalachian

Highlands in the USA and the UK. Conversely, relatively poor results were obtained for arid regions such as southern USA and south-eastern Africa, due to the highly non-linear response under these conditions. When computed on monthly streamflow, the linear correlation tends to improve, thanks to the averaging effect and consequent decreasing influence of mismatch in the peaks timing (Figure 2).

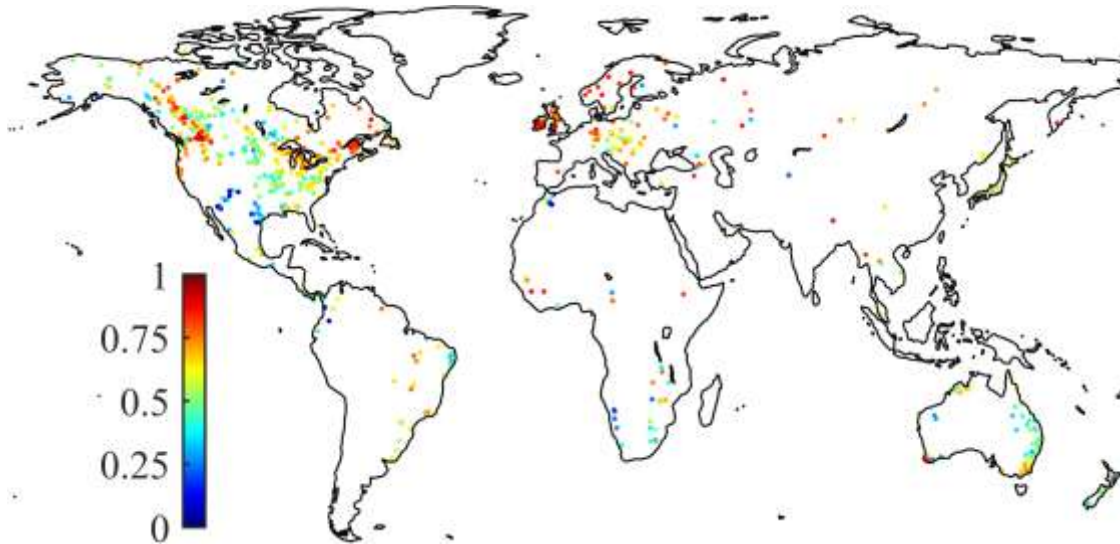


Figure 1: Pearson correlation coefficient computed between simulated and observed daily streamflow. Each data point represents a catchment ($n=641$).

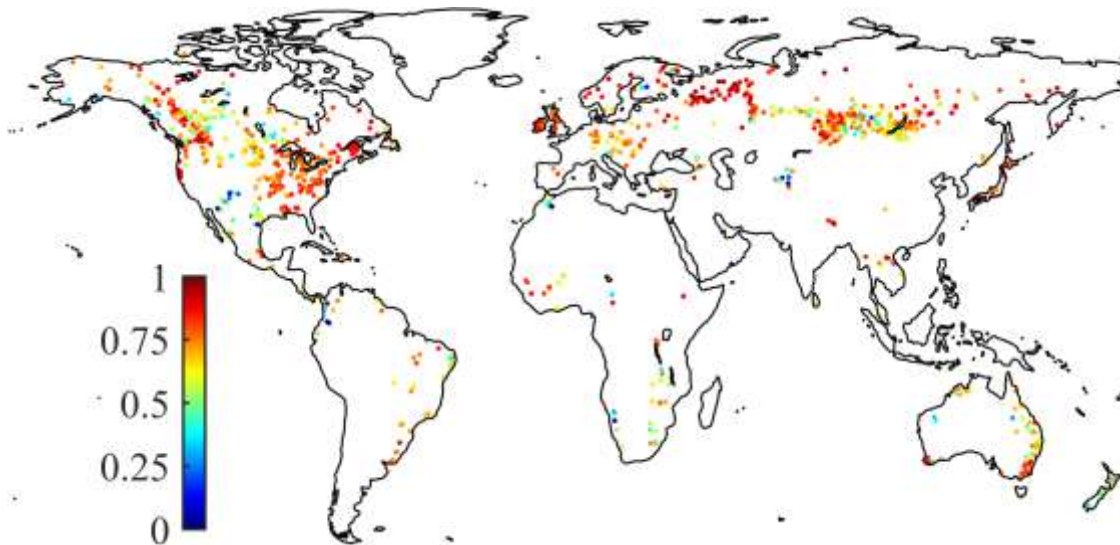


Figure 2: Pearson correlation coefficient computed between simulated and observed monthly streamflow. Each data point represents a catchment ($n=966$).

The magnitude of the peak flows (i.e., extreme events) simulated by LISFLOOD was evaluated by computing the spatial Spearman rank correlation coefficient between simulated and observed values of the 1st percentile streamflow (termed Q1; Figure 3). The Q1 values were square-root transformed to give more weight to small values. Blue indicates regions where the model underestimates the peak flow magnitude, whereas yellow and red indicate regions where the model overestimates the peak flow magnitude. Overall, the differences appear to be quite small, suggesting that the model performs well. However, the model generally appears to slightly

underestimate the peak flow magnitude for most catchments, particularly over the Rocky Mountains, in the temperate eastern USA, in Japan, and in New Zealand.

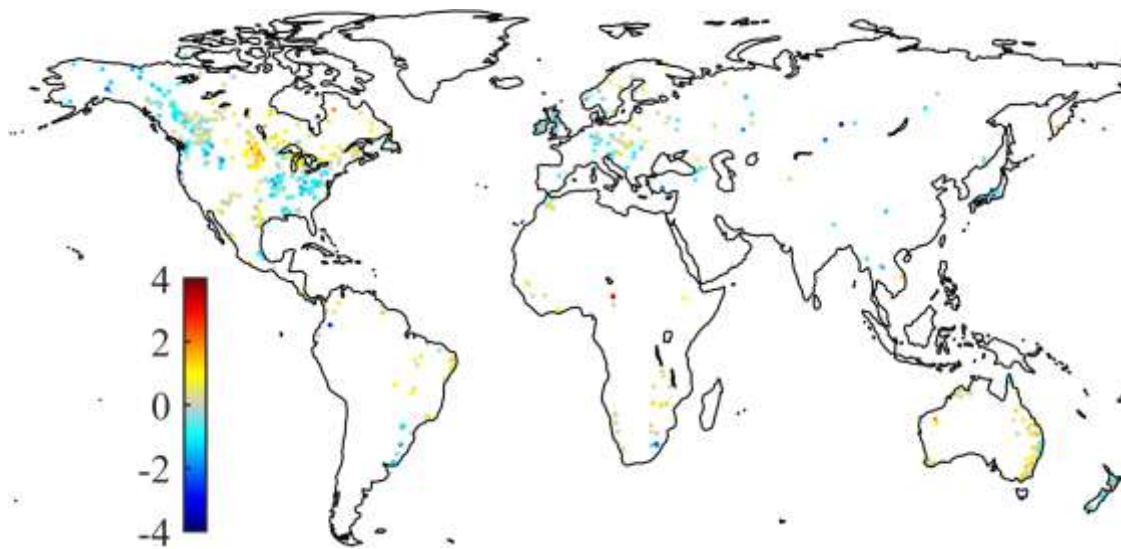


Figure 3: Square-root transformed 1st percentile exceedance flow.

1.4 MODEL CALIBRATION AND EVALUATION

The LISFLOOD model was calibrated for 24 large catchments around the globe, ranging in size from 84,230 to 4,680,000 km², against daily observed streamflow. The most recent half of the record of simultaneous observed streamflow and forcing data was used for calibration, while the other half was used for validation. As objective function we used an aggregate objective function incorporating the correlation coefficient, the Nash-Sutcliffe Efficiency computed from log-transformed flows, and the percentage bias. An example of the calibration results is shown in Figure 4 for the Amazon river basin at Obidos – Linografo. The figure compares the simulation and observed streamflow time series for the (a) calibration and (b) validation periods. Panels (c) and (d) compare the monthly climatology, while panels (e) to (g) provide statistics of the calibration procedure. Calibration results for the remaining catchments are provided in the Appendix.

The discharge time series in the appendix demonstrate that Lisflood is able to reproduce skillfully the streamflow variability and the extreme values in most river basins. In cold climates, where the timing and the dynamics of snowmelt play a key role, the model shows a tendency to underestimate the peak flow, as can be seen for the rivers Lena, Yenisey and MacKenzie. The weakest model performance was found in arid climates (Niger, Zambezi and Orange rivers), where simulated streamflow was considerably overestimated and NSE values were negative.

G1156: AMAZONAS at OBIDOS - LINIGRAFO

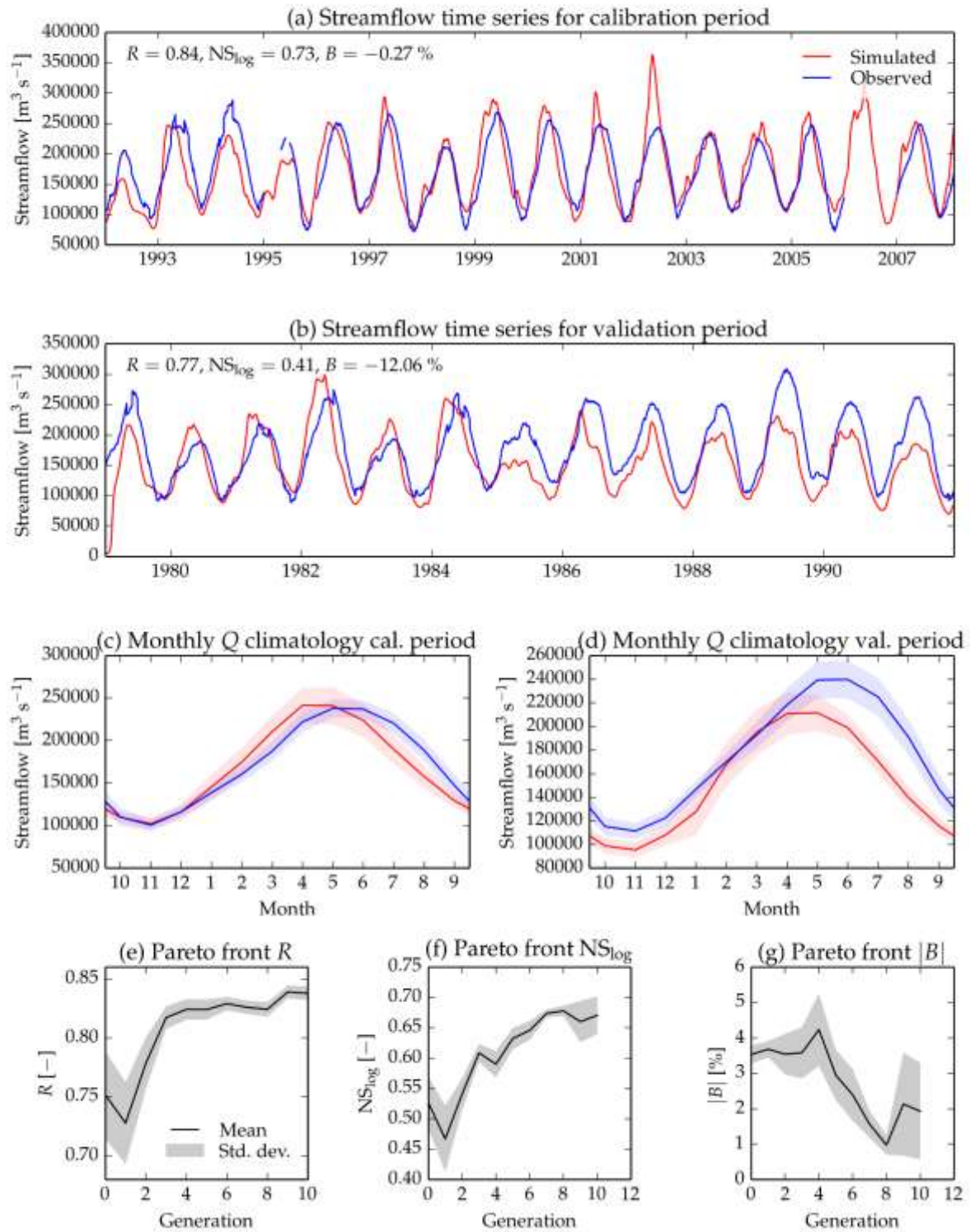


Figure 4: Score-card for the calibration of the Amazon river basin at Obidos – Linigrafo.

2. CHARACTERIZATION OF DROUGHTS USING STANDARDIZED INDICATORS

Due to its complexity, there exists no single definition of drought and meteorological, agricultural, hydrological, ground-water, and socio-economic droughts are often distinguished (Heim Jr., 2002). However, a meteorological drought (i.e. a prolonged rainfall deficit) is the primary cause of a drought while the other types of drought describe secondary effects on specific ecological and economic compartments (e.g., soil moisture, river flows, reservoirs, and economic sectors). Meteorological drought, which can be seen as a prolonged deficit in precipitation over a defined region and period of time as compared to climatological average values. Also temperature (and therefore evapotranspiration) is used in some of the indicators used in this study since meteorological drought conditions can be aggravated by high temperatures, low relative humidity, or strong and persisting desiccating winds, for example (Zampieri et al., 2009; Lei and Duan, 2011). Precipitation provides a direct measurement of water supply conditions over different time scales, and several commonly used drought indicators rely on precipitation measurements only or on simplified water balance models that include evapotranspiration (Heim 2002; Vicente-Serrano, 2010). The longer and the more spatially extensive the lack of precipitation, the more likely different types of droughts will occur.

2.1 PROJECTION OF OCCURRENCE OF EXTREME DRY EVENTS USING THE STANDARDIZED PRECIPITATION INDEX AND STANDARDIZED PRECIPITATION EVAPORATION INDEX

The changes in the probability of occurrence of extreme droughts was assessed globally using the Standardized Precipitation Index (SPI) and Standardized Precipitation-Evaporation Index (SPEI). Both indicators were computed using precipitation and potential evapotranspiration (LISVAP) simulations over the period 1970-2100. Derived variables like changes in frequency, duration and severity of droughts were used for drought characterization. The changes between present climate and projections were assessed by computing the differences between drought characteristics derived from the baseline defined as the 1970-2005 period.

2.2 STANDARDIZED PRECIPITATION INDEX (SPI) AND STANDARDIZED PRECIPITATION-EVAPORATION INDEX (SPEI)

- The Standardized Precipitation Index (SPI) was developed by McKee et al. (1993) to provide a spatially and temporally invariant measure of the precipitation deficit (or surplus) for any accumulation timescale. It is computed by fitting a parametric Cumulative Distribution Function (CDF) to a homogenized precipitation time-series and applying an equi-probability transformation to the standard normal variable. This gives the SPI in units of number of standard deviations from the median. Typically, the gamma distribution is the parametric CDF chosen to represent the precipitation time-series (e.g. McKee et al. 1993, 1995; Lloyd-Hughes and Saunders 2002) since it has the advantage of being bounded on the left at zero and positively skewed (Wilks 2011). In this study the Maximum-Likelihood Estimation (MLE) method was used to estimate the parameters of the gamma distribution.

- The Standardized Precipitation Evapotranspiration Index (SPEI, Vicente-Serrano et al., 2010) is based on precipitation and potential evaporation data, and it has the advantage of combining different time dimensions (like the SPI) with the capacity to include the effects of atmospheric humidity demand. The calculation combines a climatic water balance, the accumulation of a water deficit/surplus at different timescales, and an adjustment to a log-logistic probability distribution. Reference values for potential evapotranspiration and evaporation were estimated using the Penman-Monteith equation (Supit et al., 1994, Supit & Van Der Goot, 2003) using LISVAP. LISVAP is the Evaporation Pre-Processor for the LISFLOOD Water Balance (Burek et al., 2013).

A reduction in precipitation with respect to the normal precipitation amount is the primary driver of drought, resulting in a successive shortage of water for different natural and human needs. An increased demand of moisture from the atmosphere represented by increased evapotranspiration would potentially aggravate the water shortage.

As both drought indicators are given in units of standard deviation from the standardized mean, negative values correspond to drier periods than normal and positive values correspond to wetter periods than normal. The magnitude of the departure from the mean is a probabilistic measure of the severity of a wet or dry event (Table 1). Moreover, since the SPI and SPEI can be calculated over different rainfall accumulation periods, different SPIs allow for estimating different potential impacts of a meteorological drought:

- SPI/SPEI for short accumulation periods (e.g., 1 to 3 months) are indicators for immediate impacts such as reduced soil moisture, snowpack, and flow in smaller creeks;
- SPI/SPEI for medium accumulation periods (e.g., 3 to 12 months) are indicators for reduced stream flow and reservoir storage; and
- SPI/SPEI for long accumulation periods (12 to 48 months) are indicators for reduced reservoir and groundwater recharge, for example.

The exact relationship between accumulation period and impact depends on the natural environment (e.g., geology, soils) and the human interference (e.g., existence of irrigation schemes). In order to get a full picture of the potential impacts of a drought, SPI/SPEIs of different accumulation periods should be compared.

Definition of SPI classes and their probability of occurrence (in percent)

$SPI \leq -2$	Extremely dry	2.3 %
$-2 < SPI \leq -1.5$	Severely dry	4.4 %
$-1.5 < SPI \leq -1$	Moderately dry	9.2 %
$-1 < SPI \leq 1$	Near normal	68.2 %
$1 < SPI \leq 1.5$	Moderately wet	9.2 %
$1.5 < SPI \leq 2$	Severely wet	4.4 %
$SPI \geq 2$	Extremely wet	2.3 %

Table 1 Definition of Standard Precipitation Index (SPI) classes according McKee et al (1995)

Whenever a drought event is detected, a set of statistical parameters are derived as depicted in Table 2:

VARIABLE	DESCRIPTION
ONSET	first month with the indicator below the threshold
END	last month with the indicator below threshold
DURATION	number of months of the event
FREQUENCY	number of drought events per period as defined above
SEVERITY	a severity score is computed as the sum of the differences, in absolute values, between the indicator values and the threshold as defined below (S)
MAGNITUDE	severity divided by duration
PEAK	month with the lowest value of the indicator during the drought event; this parameter is associated together with the severity value corresponding to that month

Table 2 Definitions of drought parameters derived from standardised indicators

Each period with continuous negative SPEI is defined as a drought event. Cumulative SPEI values of a dry event are used to measure the severity of drought event and according with McKee (1993) is defined as:

$$S_i = - \sum_{i=1}^D \text{SPEI}_i < -1$$

Where S is the severity and D represents the drought duration in months of each dry event. Figure 5 below shows a schematic representation of the S and D and the inter-arrival time between two dry periods. As SPEI and SPEI are dimensionless Drought Severity is dimensionless.

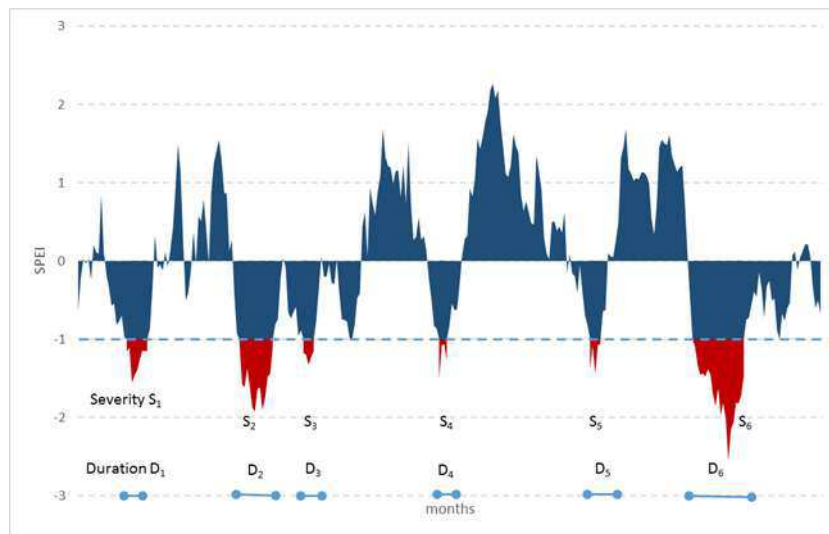


Figure 5 Definition of Standard Precipitation Index (SPI) classes according McKee et al (1995)

Estimation of return periods and return levels was assessed through the fitting of the univariate extreme value distribution functions (GEV) to drought severity data. The extreme value distributions (EVDs) have theoretical support for analysing extreme values of a stochastic process. In particular, the generalized extreme value (GEV) is appropriate for modelling block maxima (for large blocks, such as annual maxima).

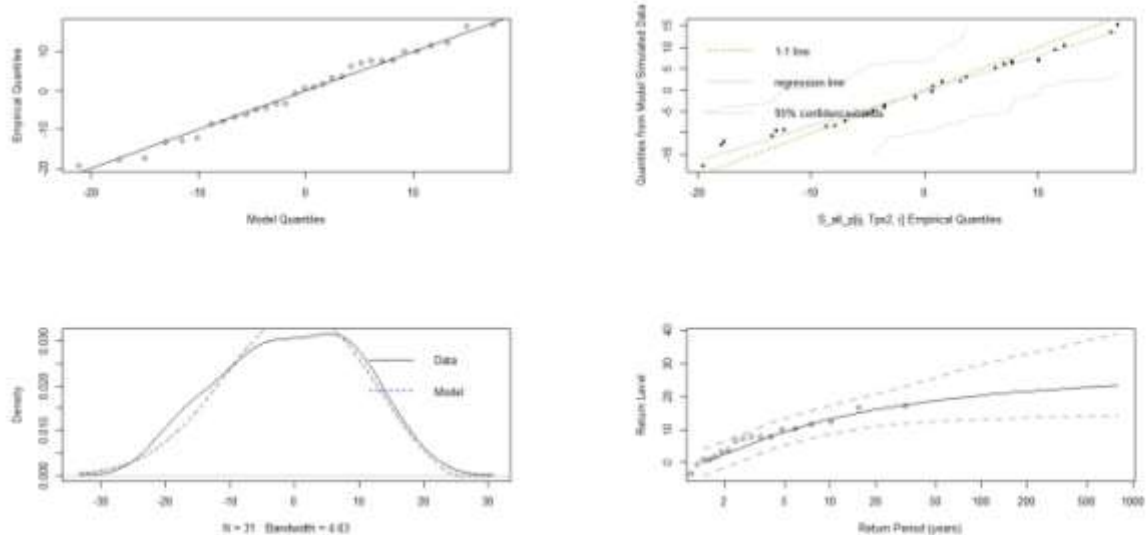


Figure 6 Estimation of the parameters of the Generalized Extreme Value analysis for annual series of drought severity.

The GEV distribution function is often associated to a family of distribution functions because it encompasses the three types of EVDs: Gumbel (shape = 0, light tail), Frechet (shape > 0, heavy tail) and the reverse Weibull (shape < 0, bounded upper tail at location - scale/shape). It was first found by R. von Mises (1936) and also independently noted later by meteorologist A. F. Jenkins (1955). It enjoys theoretical support for modelling maxima taken over large blocks of a series of data. Figure 6 shows the estimated GEV function compared with the empirical quantiles.

2.3 THE RELATIVE SPI / SPEI

The relative SPI is defined with respect to a reference period, in this case the 1970-2000. To compute the relative SPI, the Gamma distribution is estimated for the reference and for the future period. The two curves are then used to calculate the probability difference between future and present SPI values. Meteorological droughts were computed using SPI and SPEI (P-PET for the present and future conditions). Comparison and changes of different drought characteristics: frequency, duration, onset and end, peak and month of the peak.

As an example, the black, red and green curves in Figure 7 are the CDFs fitted to the precipitation data of the present and future periods (2041-2070 and 2071-2100 respectively), which is wetter (right shifting and widening of the PDF) in a point in the Limpopo Basin and drier (left shifting of the PDF) in a point in east Angola. These two points have been selected because they experience different future changes. The black curves represent the CDFs fitted to the precipitation data of the reference period (1971-2005). For the present climate, we found a value of 981 mm corresponding to a standard normal value of -2 (extremely dry according to Table 2) in Angola. In the future

climate (green line), this same amount of rain corresponds to a probability of $\approx 8\%$ or, equivalently, to a standard normal value of ≈ -1.2 (moderately dry). The probability of an extreme dry year (according to present-day standards) has, therefore, increased more than 3 times.

A similar procedure holds for the wet end of the distribution. For instance, for the point analysed in the Limpopo Basin (Figure 7), the precipitation that corresponds to a SPI of 2 (extremely wet) in the present climate is 1540 mm per year, while for the future, the same amount of precipitation corresponds to a probability of $\approx 10\%$.

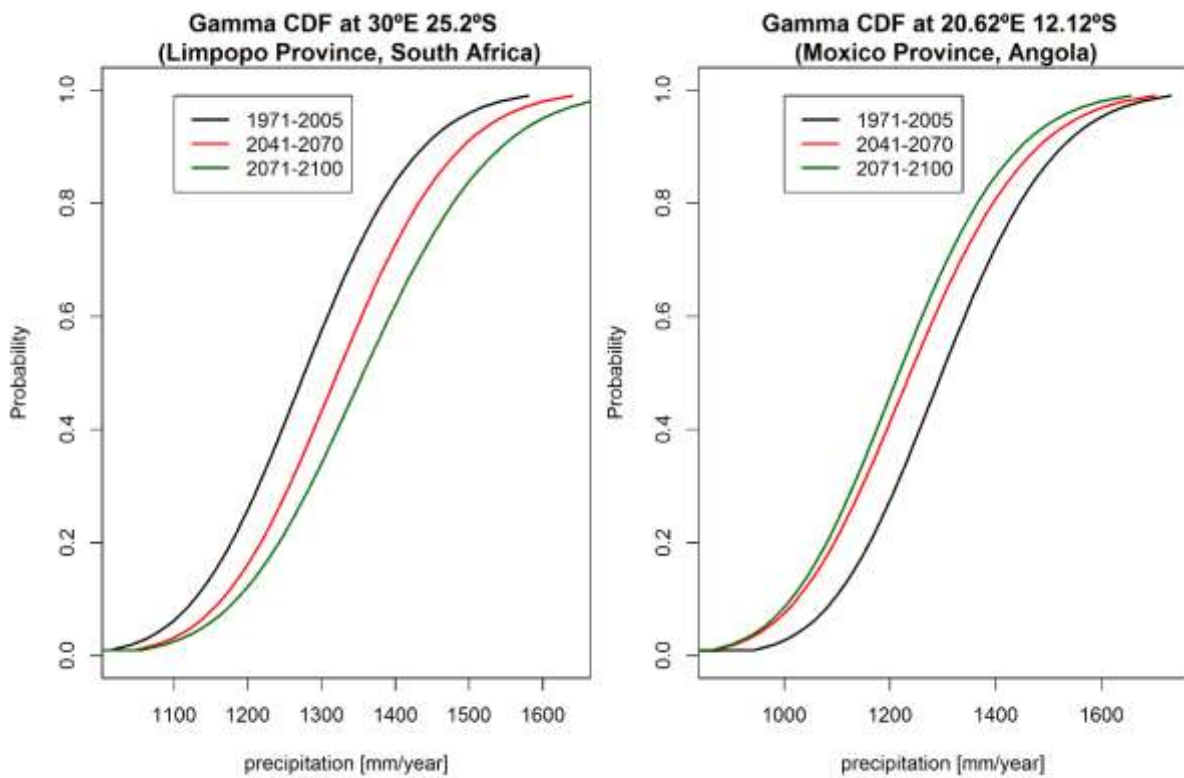


Figure 7 Gamma distribution at two points in South Africa and Angola respectively.

Estimations of the related drought variables derived from SPI and SPEI were computed for the baseline using different datasets. Two independent datasets CRU v.3.23 (Harris et al., 2014) and WATCH-Forcing-Data-ERA-Interim (WFDEI, Weedon et al., 2014) were used to evaluate the accuracy of the models in representing the different drought features.

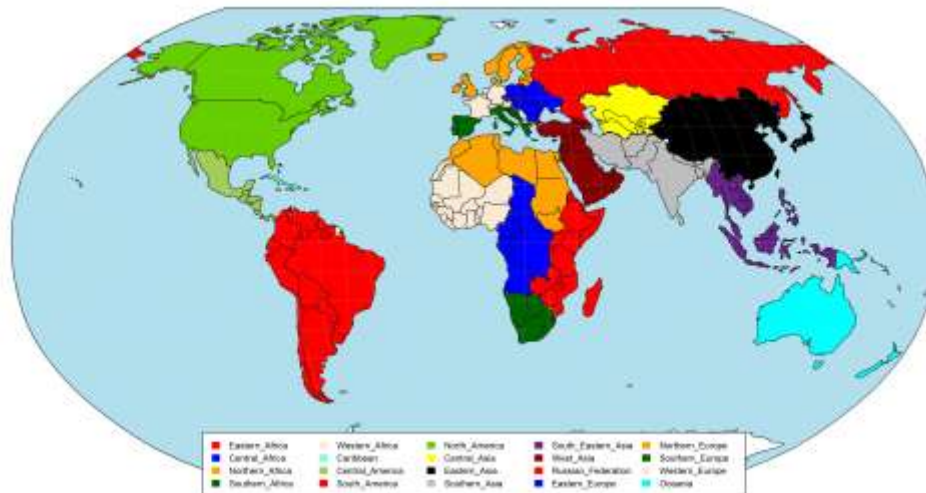


Figure 8 Composition of macro-regions as defined by UN.

Figure 9 shows the percentage of land fraction area affected by droughts in 20 different global macro-regions as defined by UN and represented in Figure 8.

In most of the areas the models are able to represent the present drought trends, however there are some regions where there are also big discrepancies between the “observed” datasets. This is the case for Southern Africa, North America and the Russian Federation. In North America and Russia the model ensemble tends to follow better the trends observed by the CRU datasets, while in Southern Africa it seems that model runs are more in agreement with the trends showed by the WFDEI dataset.

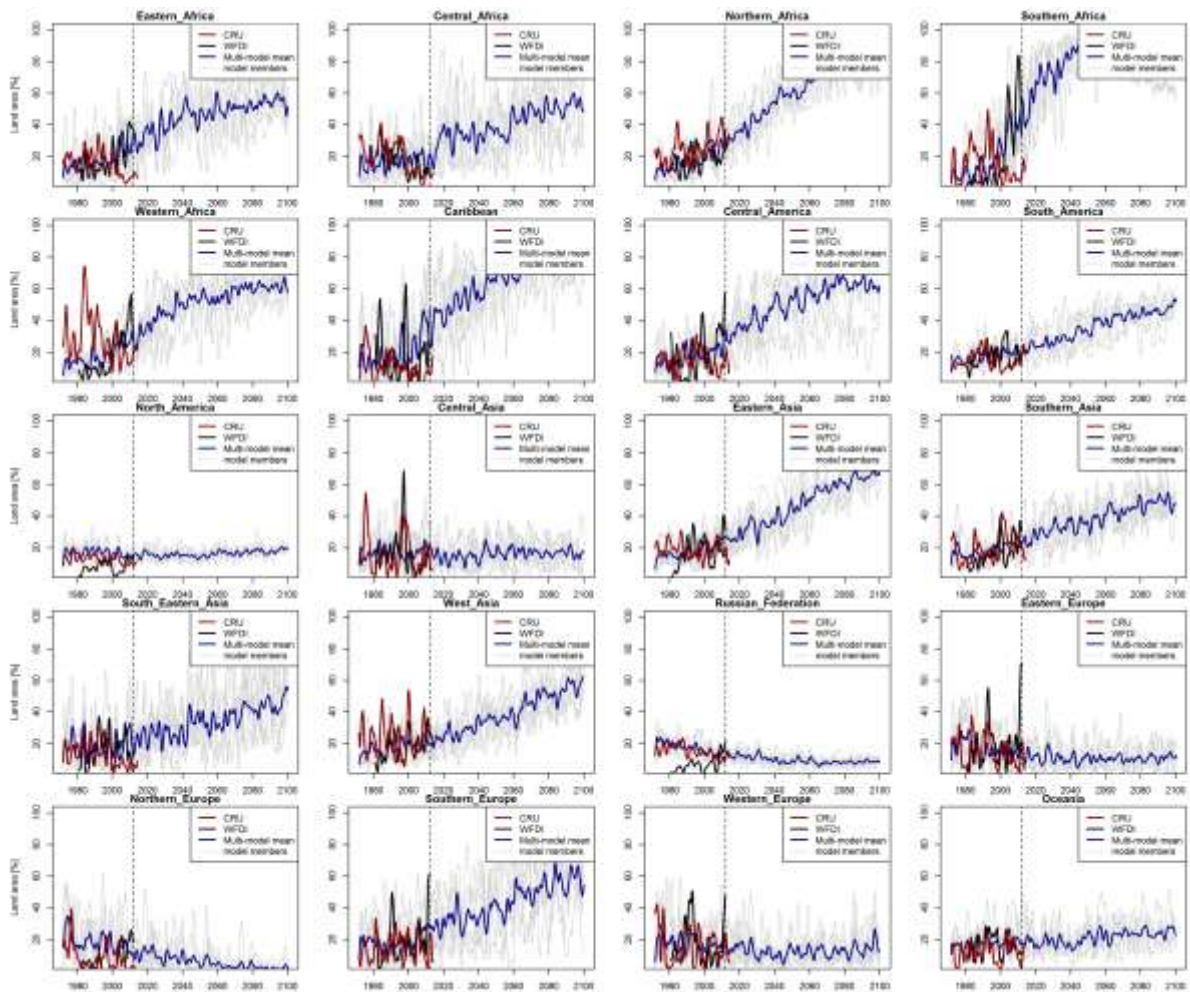


Figure 9 Time evolution of percentage of land area affected by droughts derived from SPEI-12 for different macro-regions calculated using the CRU dataset (red), WFDEI (black) and the model ensembles (grey) and mean (blue)

For the drought severity, mean climate and trends are generally well represented by the models estimations but also are evident some differences between observed datasets that represent the current conditions (Figure 10). As detected for the total affected area, Southern Africa, North America and the Russian Federation show the biggest discrepancies. North America, North Europa and the Russian Federation seems to behave more similarly for the estimations computed using the CRU datasets, while in Central and Southern Africa it seems that the simulations better agree with the WFDEI estimations. Model spread tends to be higher in tropical regions and regions that are most likely to be simultaneously covered by droughts (land areas > 60%) like Central Africa, South East Asia, Caribbean and Central America but also this is true for Southern Europe.

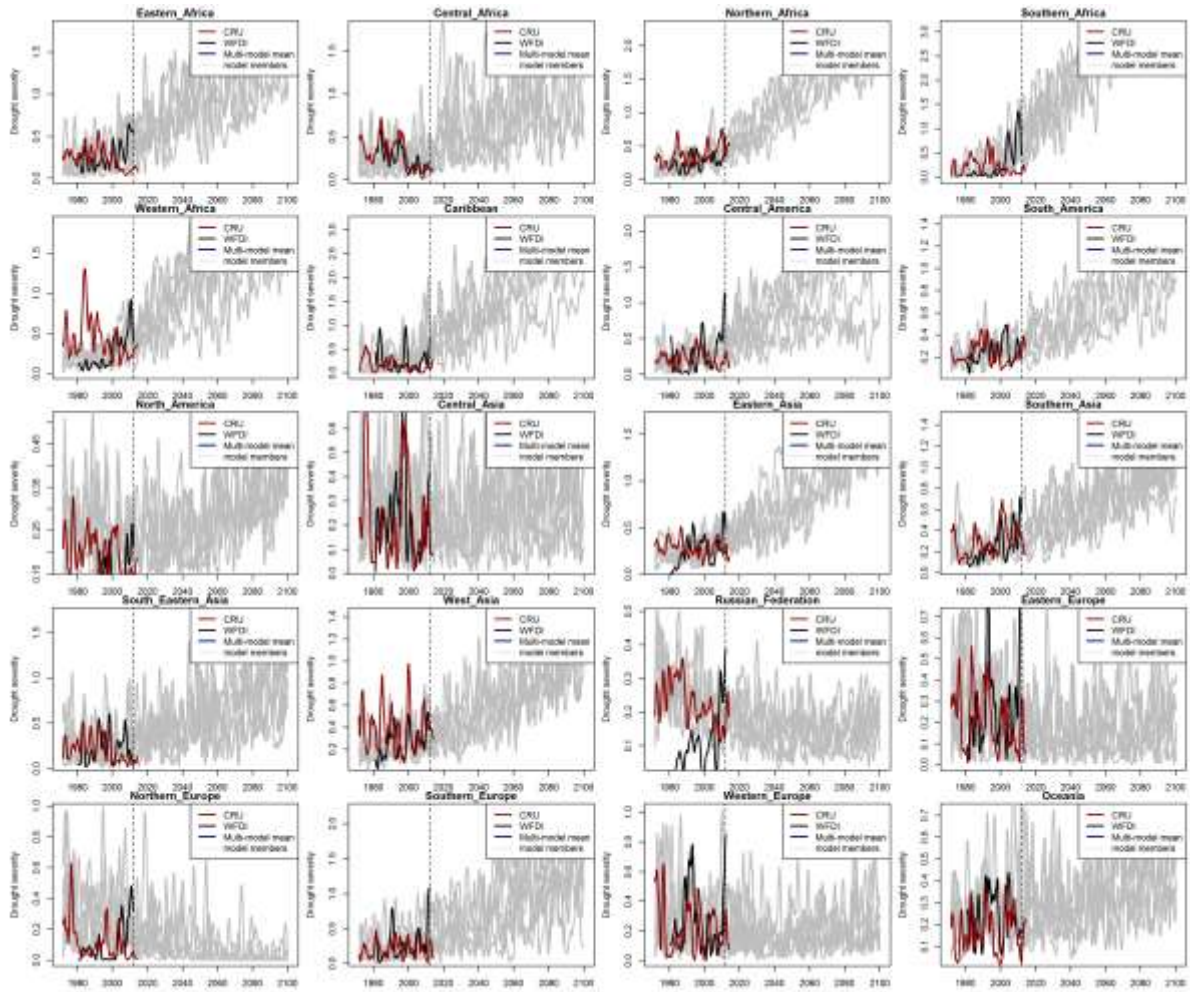


Figure 10 Time evolution of drought severity derived from SPEI-12 for different macro-regions calculated using the CRU dataset (red), WFDEI (black) and the model ensembles (grey) and mean (blue)

Relative differences between model estimations and WFDEI (Figure 11) shows mainly underestimation of most of drought duration and severity in Central Africa, Sahel, Eastern Europe, Western North America South Eastern South America, south east Australia but usually the relative differences are around or below 50 %. There are also some regions that presents overestimation of duration and severity mainly observed in north east Brazil, central Australia.

Drought frequency shows some regional consistencies between biases if compared with the other drought variables. For instance, more frequent and severe droughts in North East Brazil and Northern Russia are estimated. However, also a positive bias is observed with higher drought frequencies in Central Africa, Greenland, South East Australia and South West South America but associated with an underestimation of the drought severity and duration.

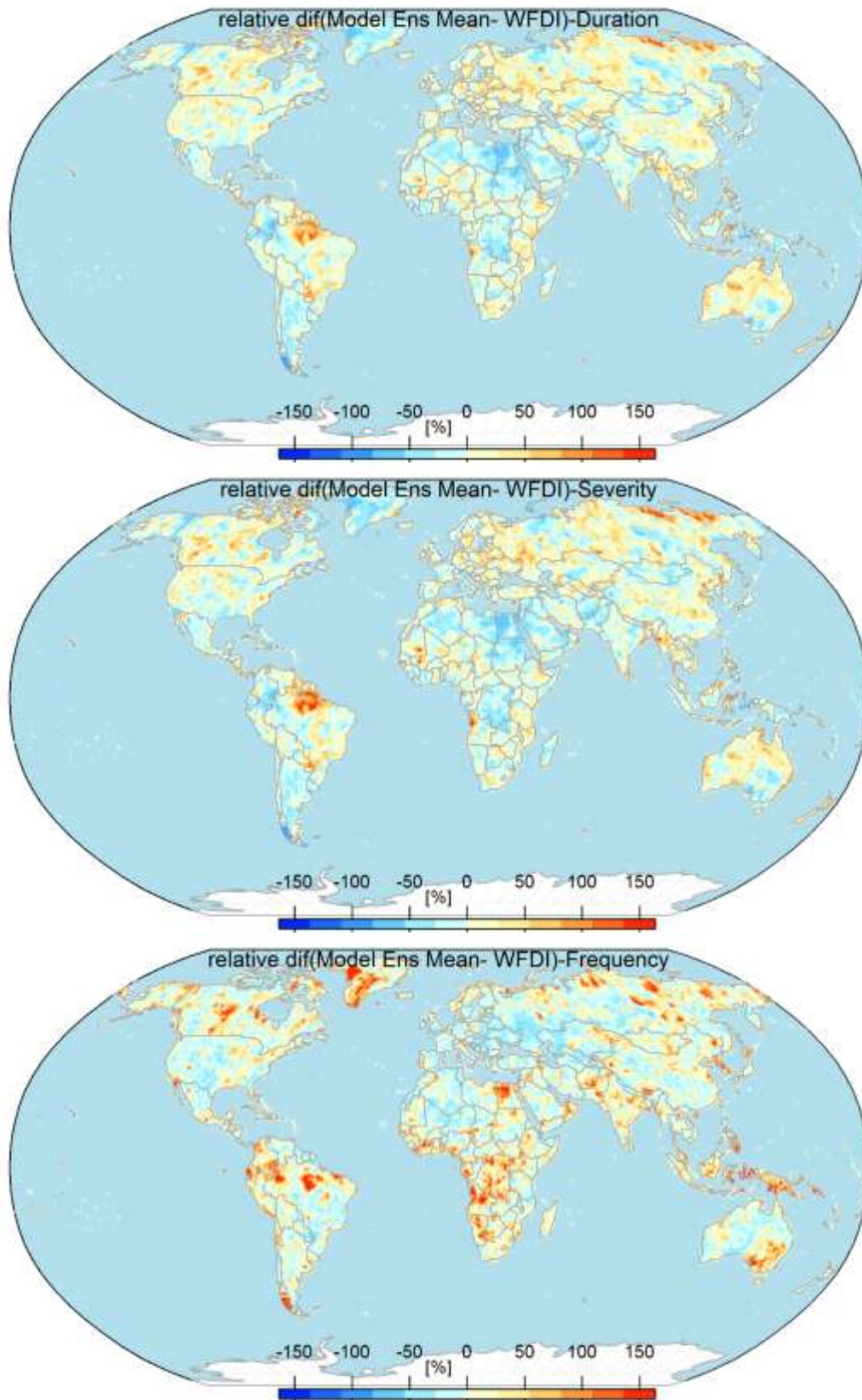


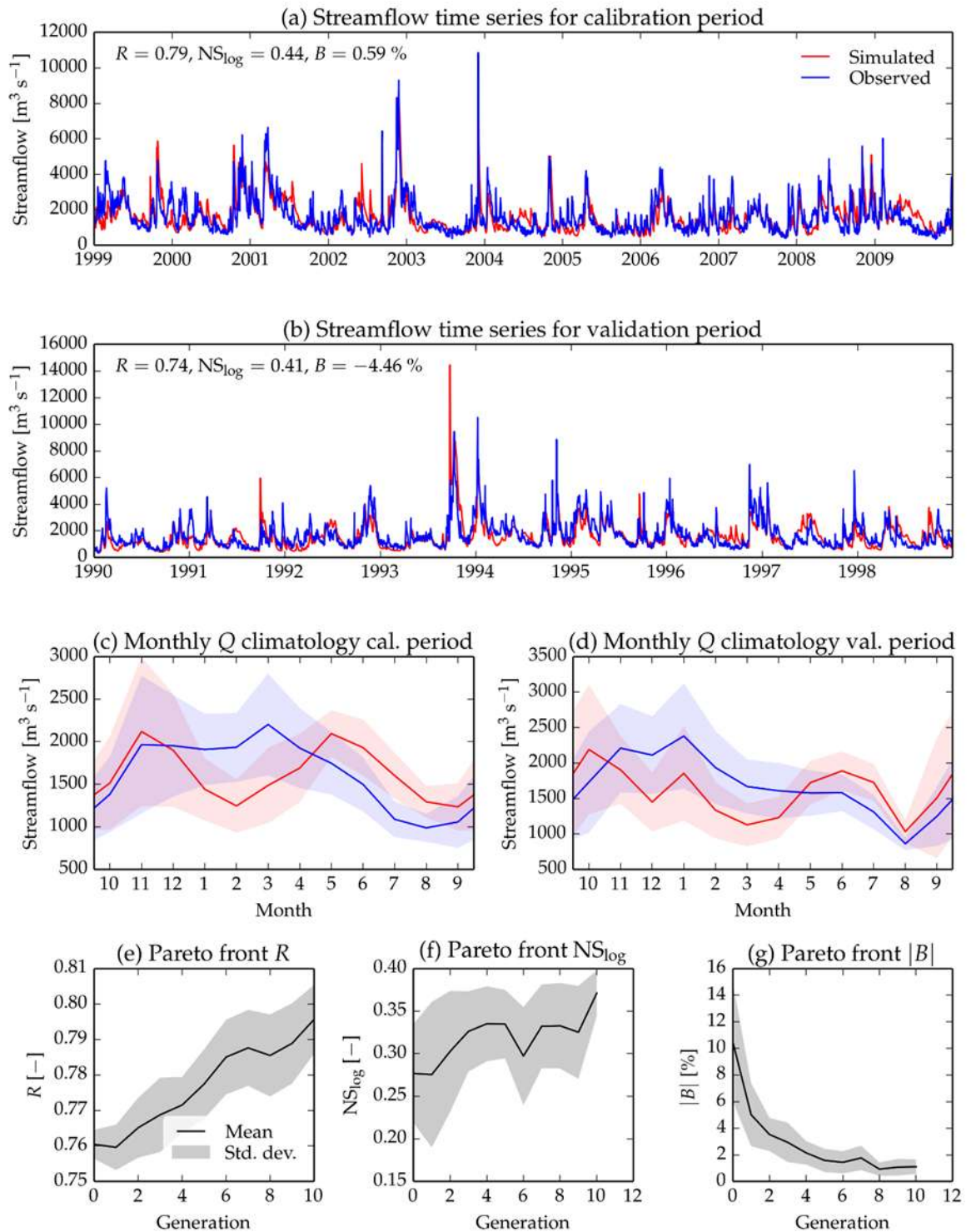
Figure 11 Relative difference between models ensemble mean and WFDI estimation of drought duration severity, duration and frequency derived from SPEI-12.

Overall, the model inter-comparison and verification using independent datasets (CRU and WFDEI) shows that the set of modelled drought parameters are able to reproduce drought variability and intensity of extreme values in most of the areas for the baseline. However, a weakest representation can be found in arctic regions like Greenland and northern Russia as well as in some tropical regions like north east Brazil and Central Africa.

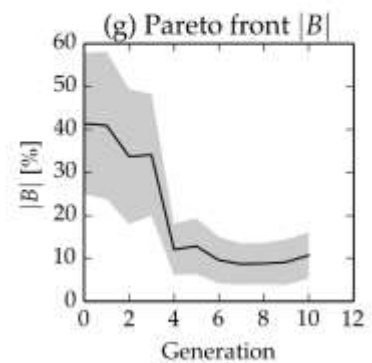
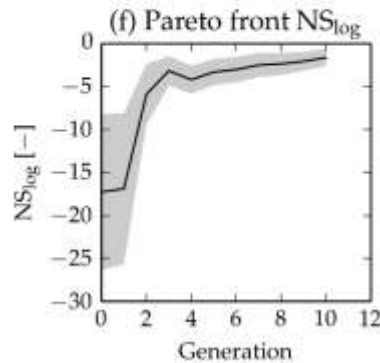
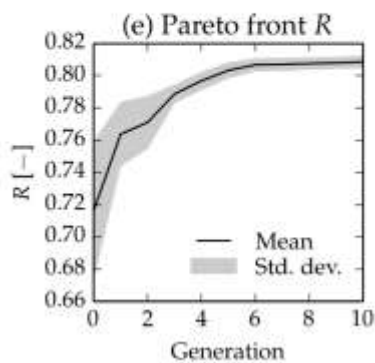
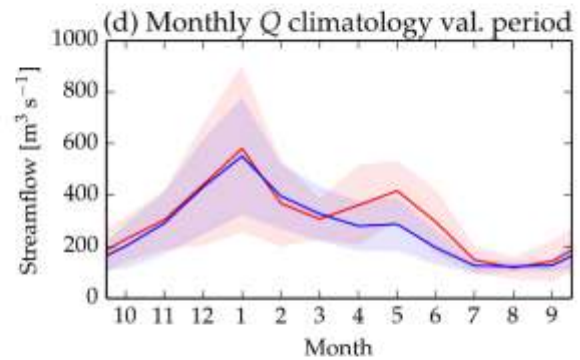
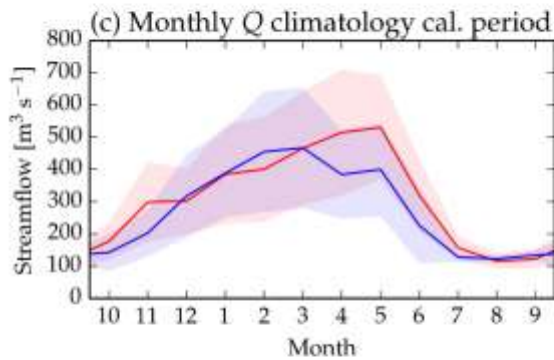
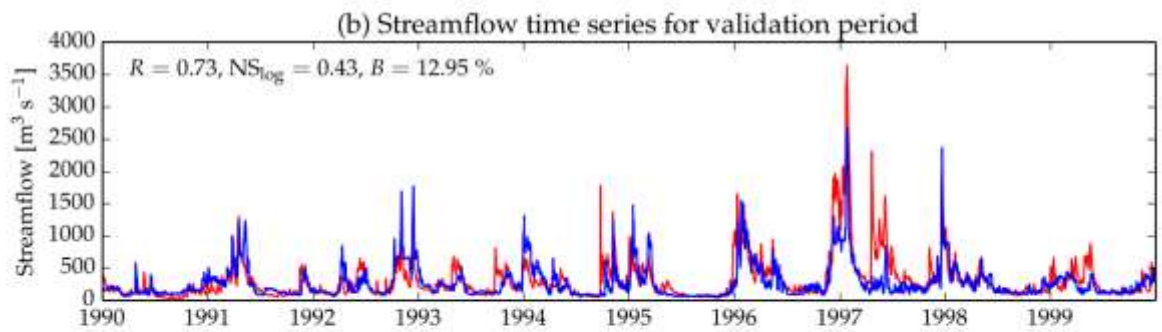
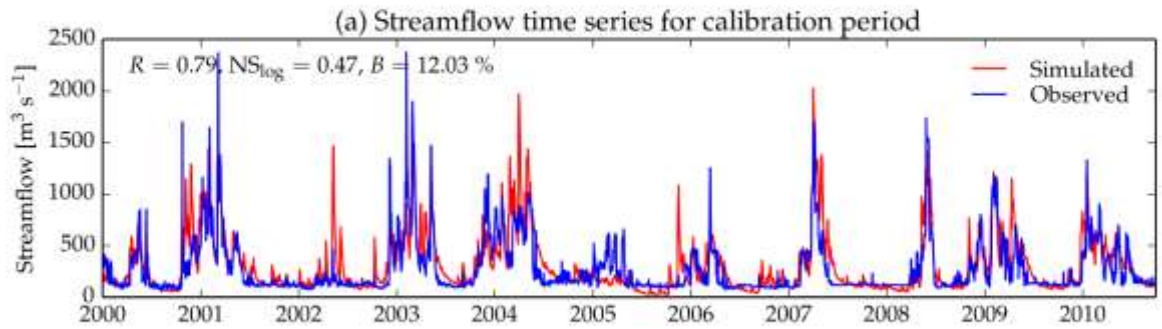
3. APPENDIX TO PART A

3.1 CALIBRATION SCORE CARDS

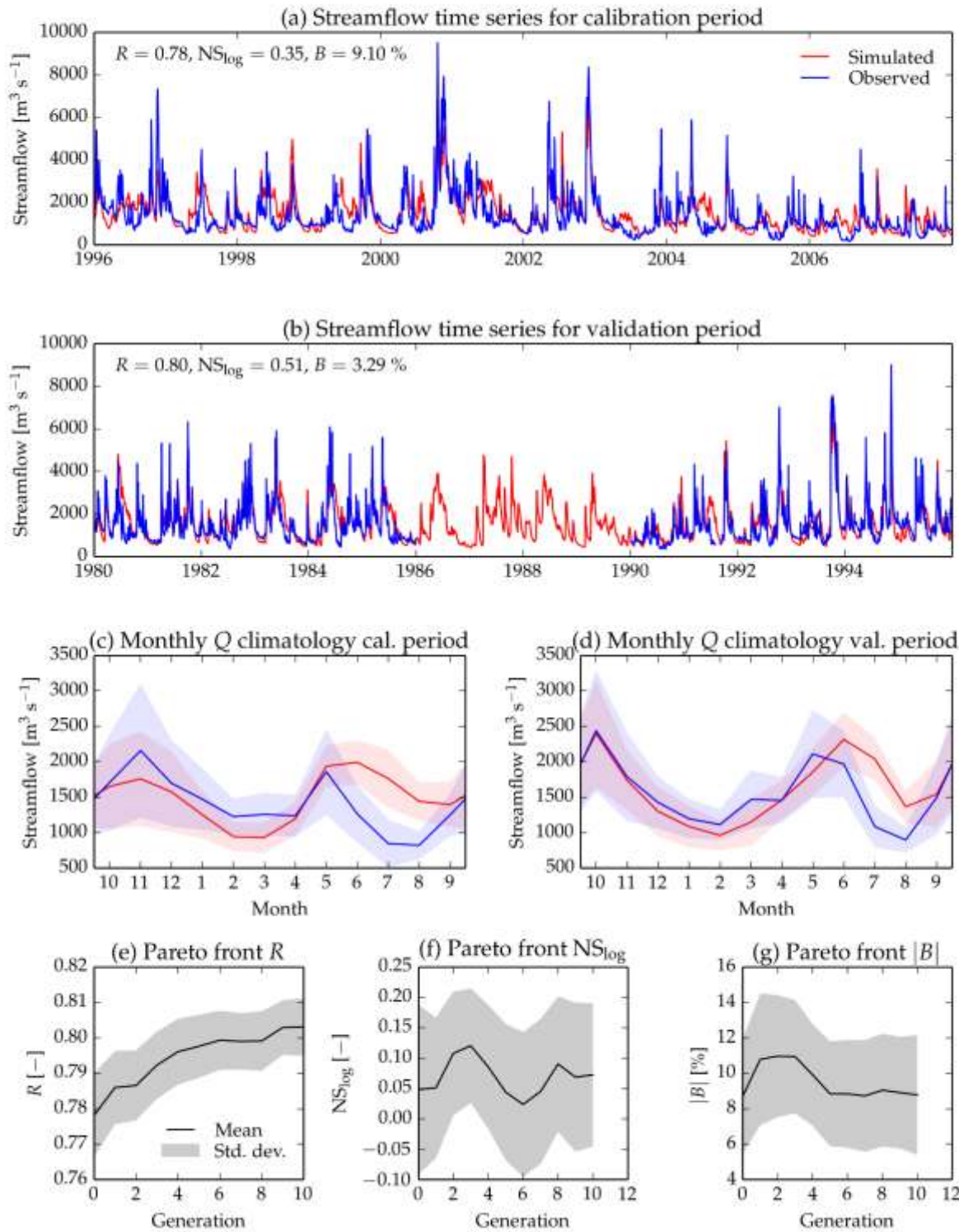
C098: Rhone at Le Rhone a Beaucaire



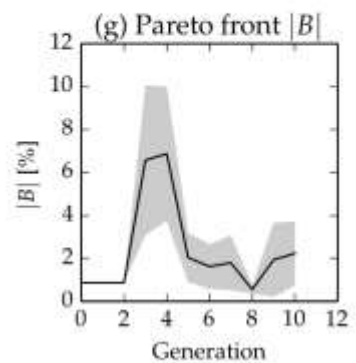
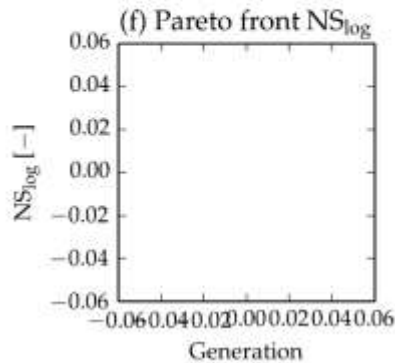
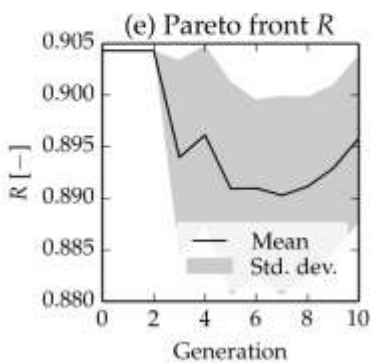
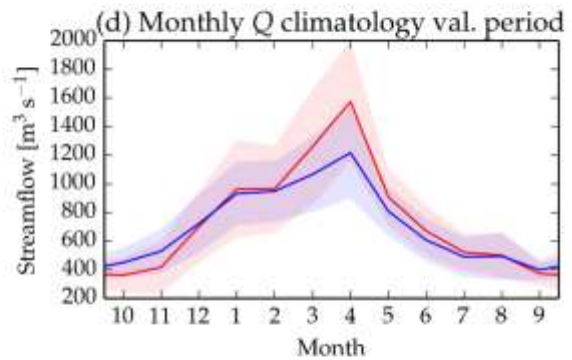
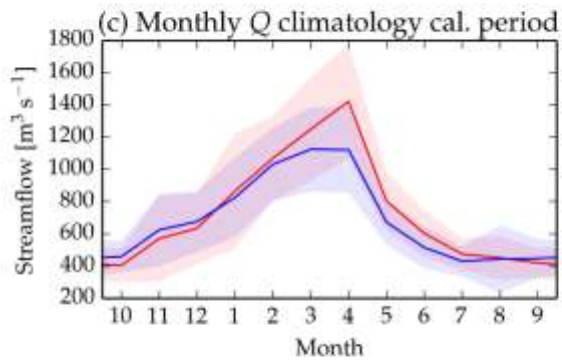
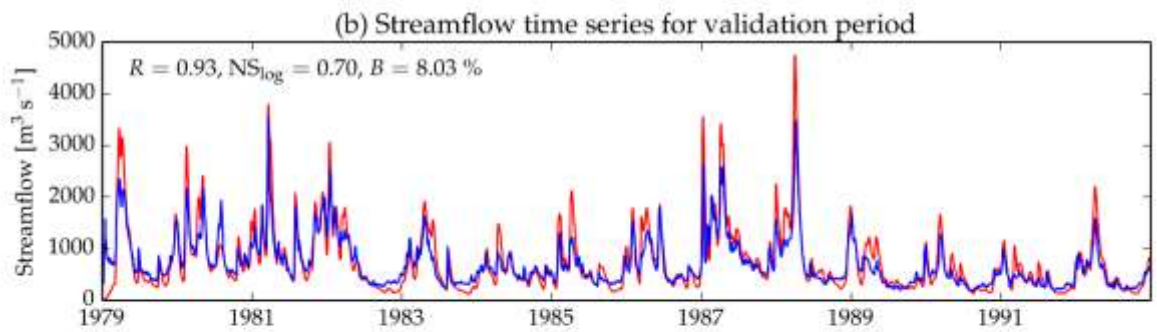
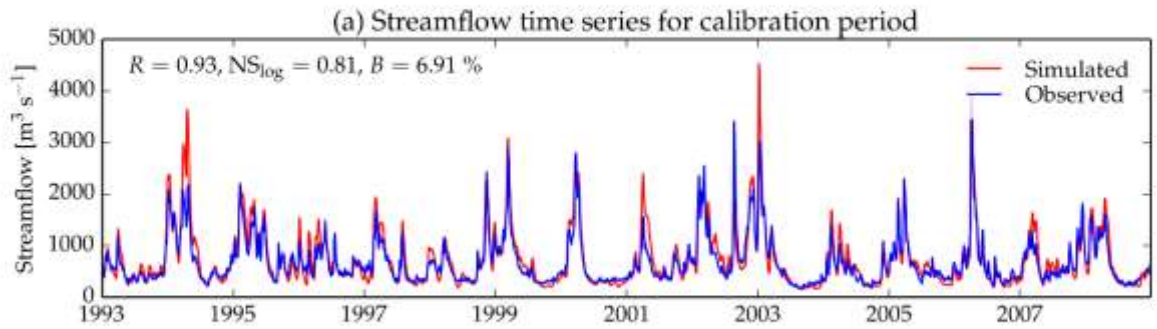
C118: Ebro at EbroEnTortosa



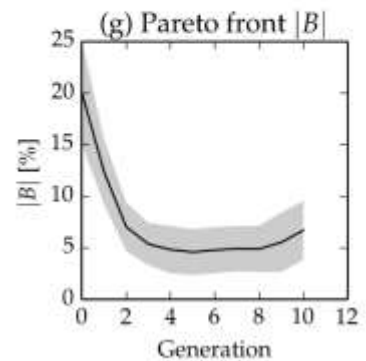
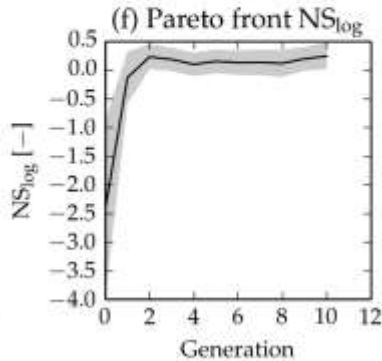
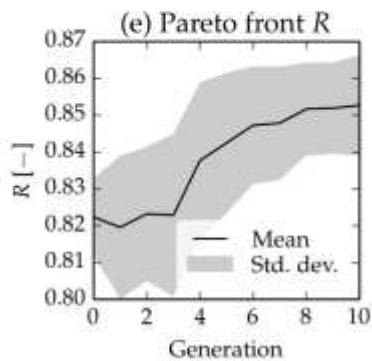
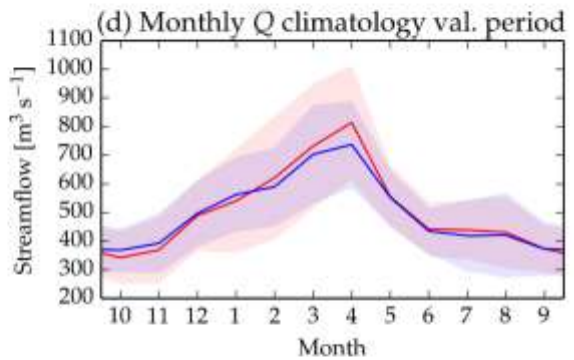
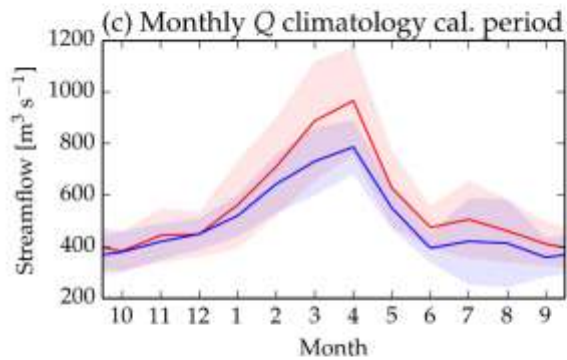
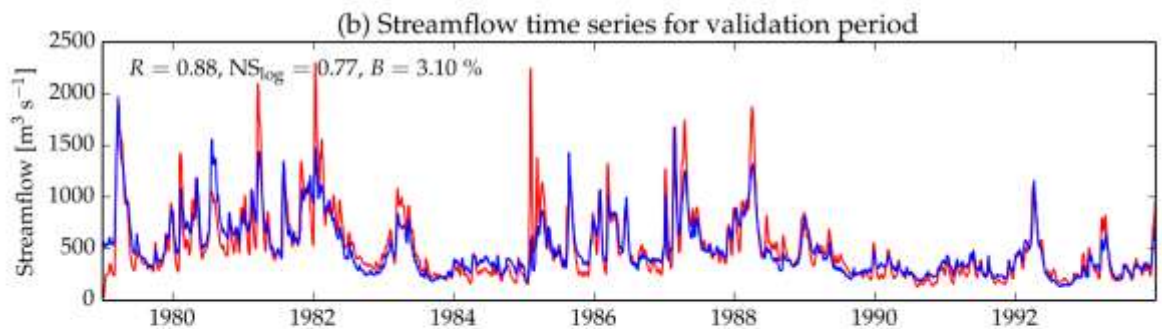
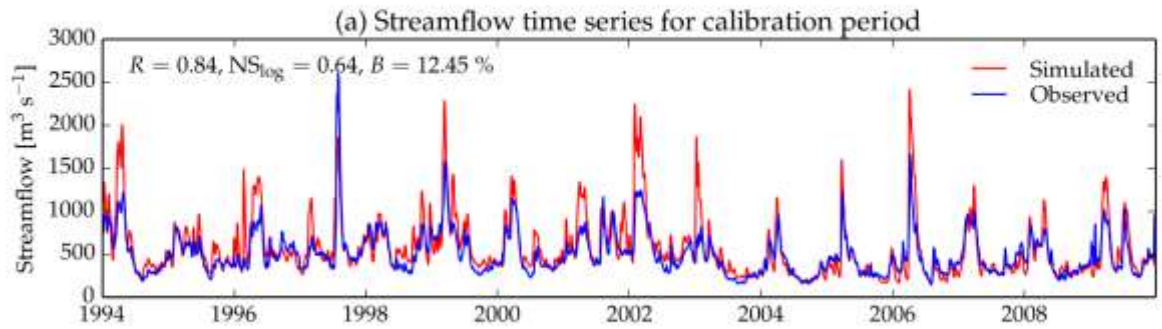
C140: Po at Lagoscuro



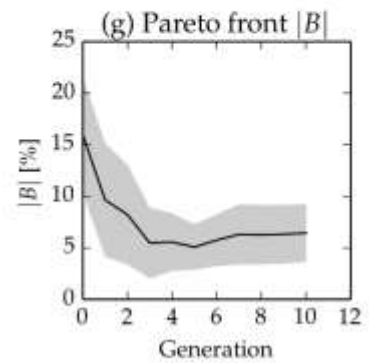
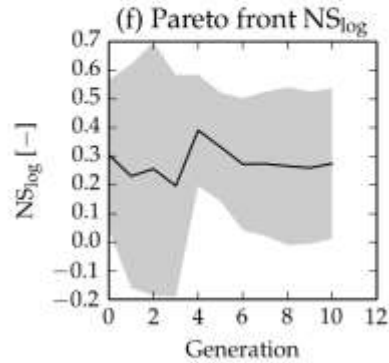
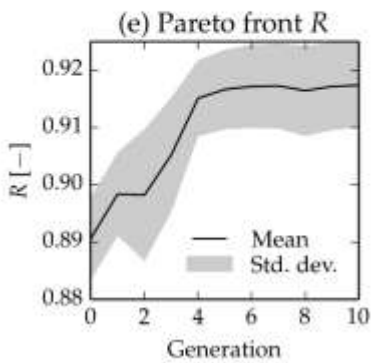
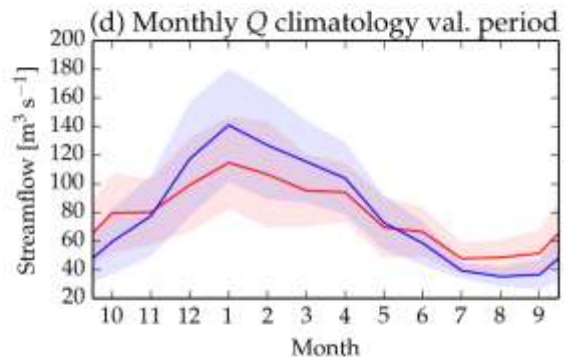
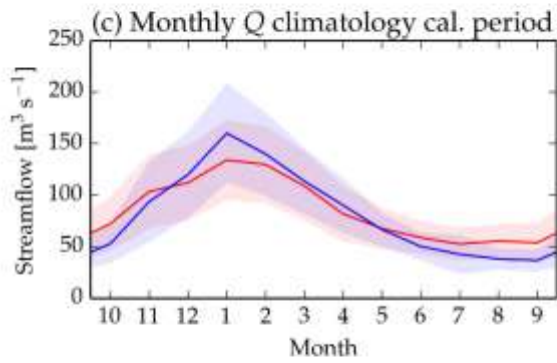
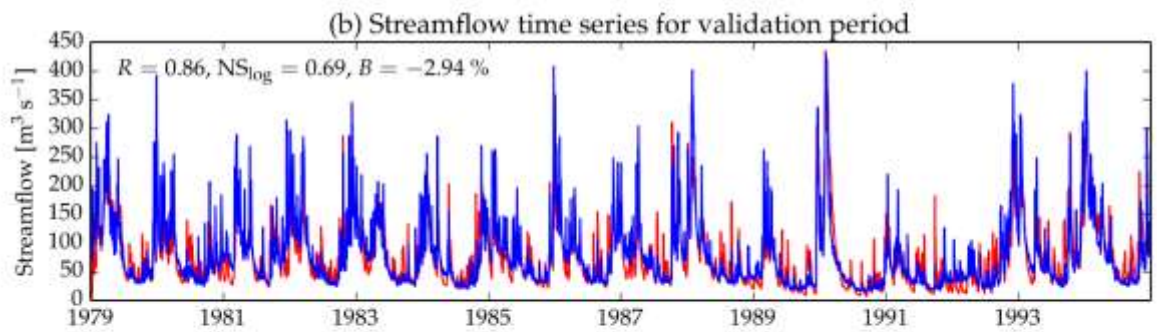
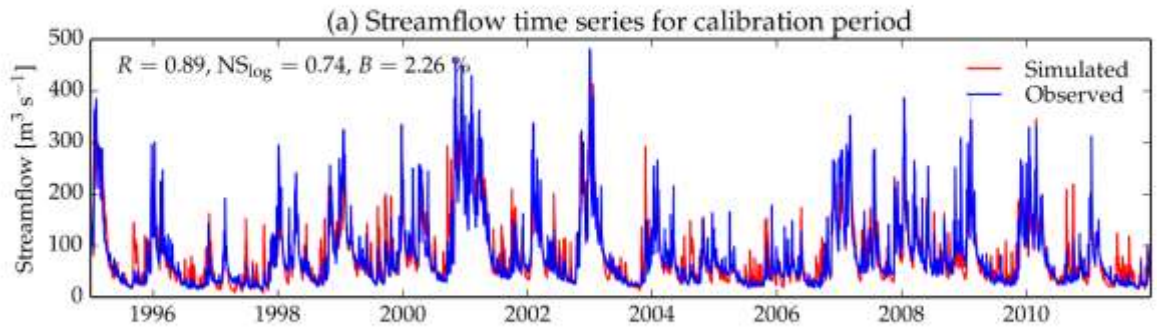
C172: Elbe at NeuDarchau



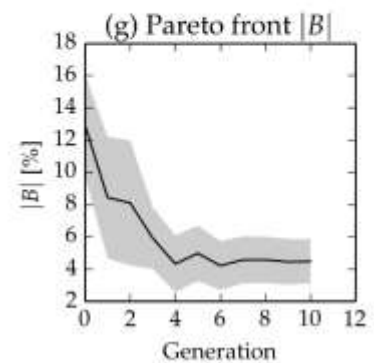
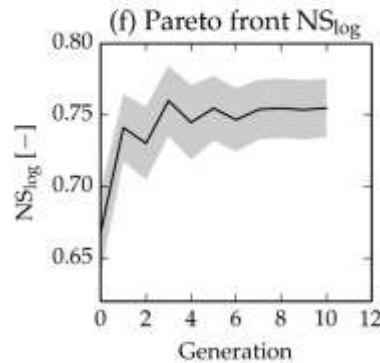
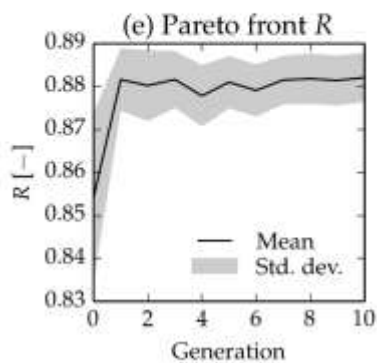
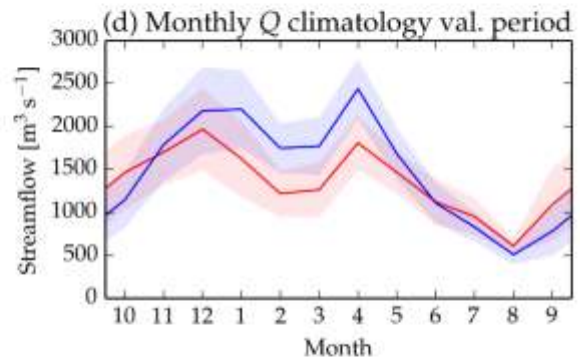
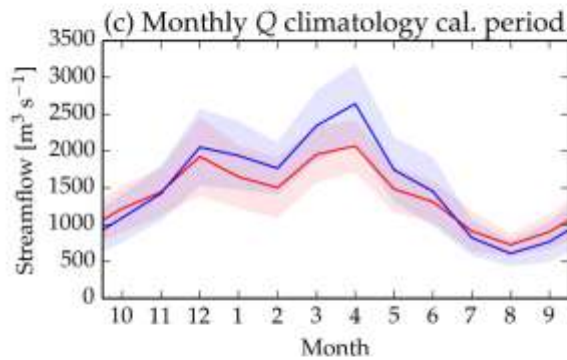
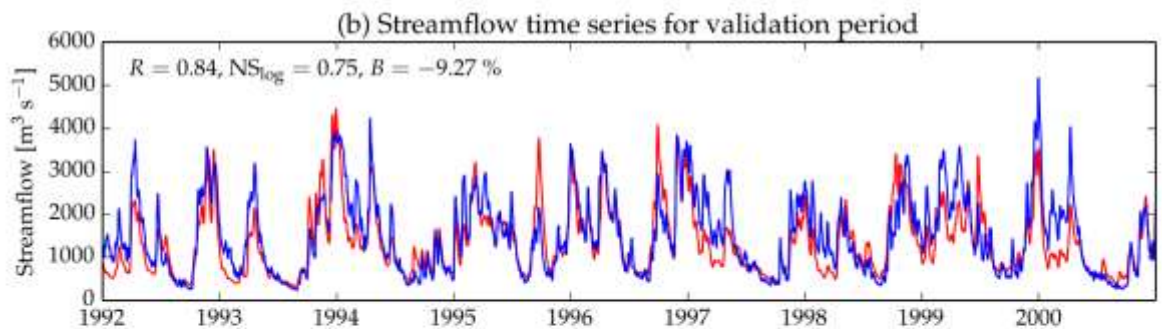
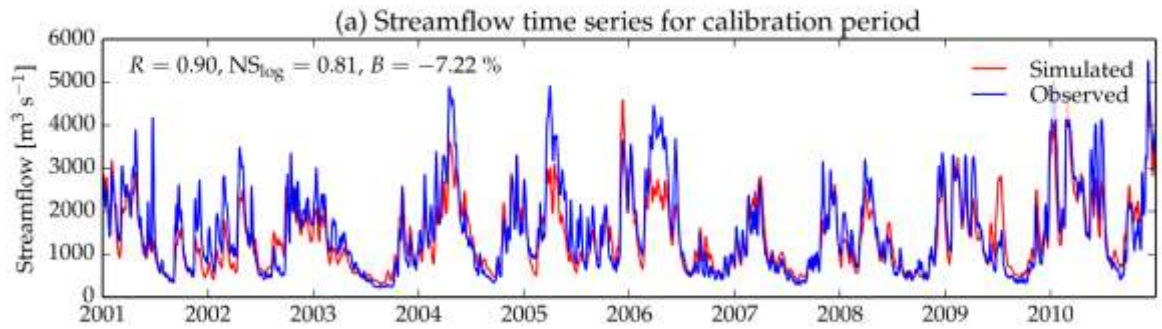
C178: Oder at Hohensaaten



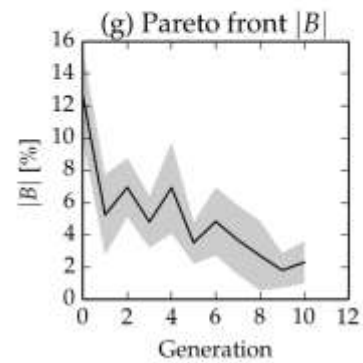
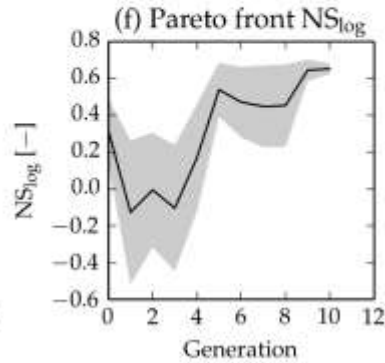
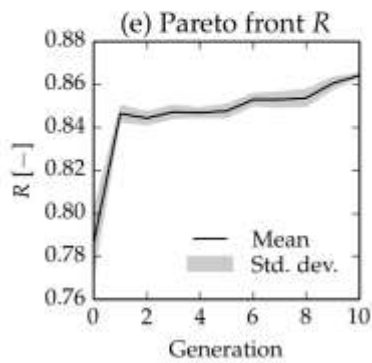
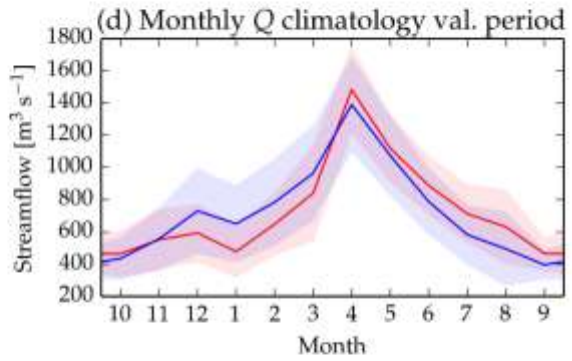
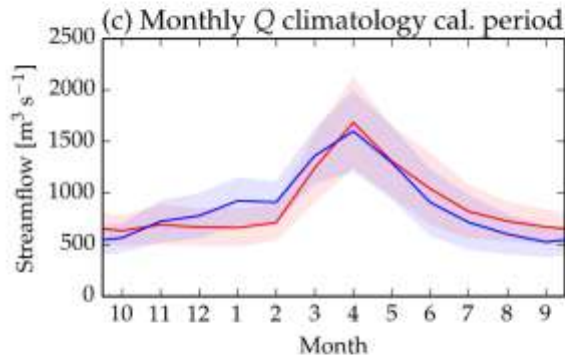
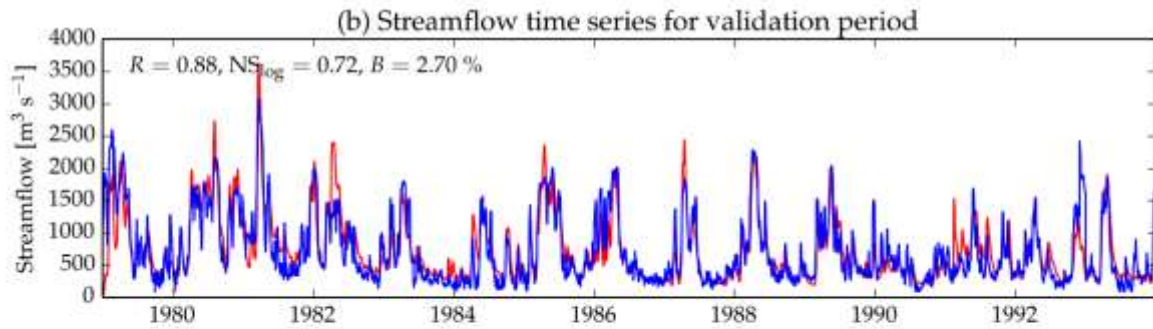
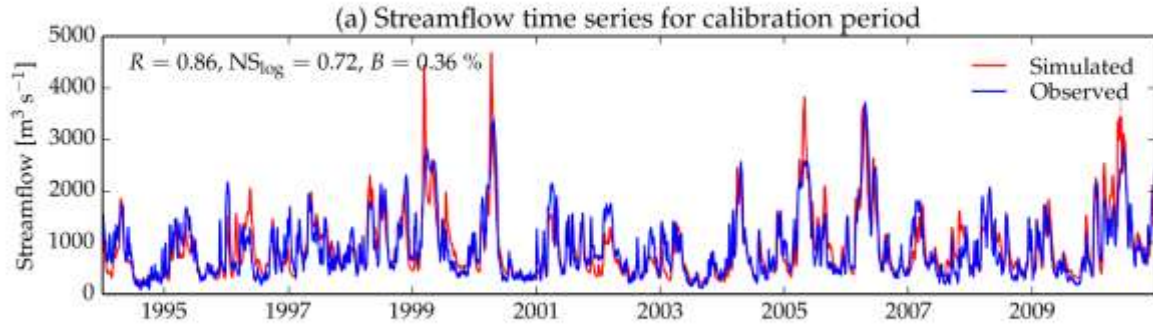
C653: nan at KINGSTON_ON_THAMES



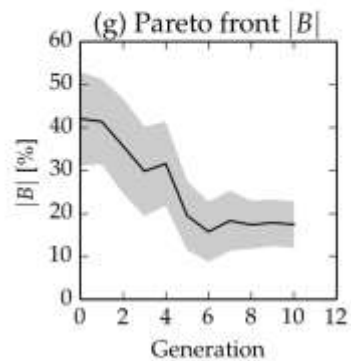
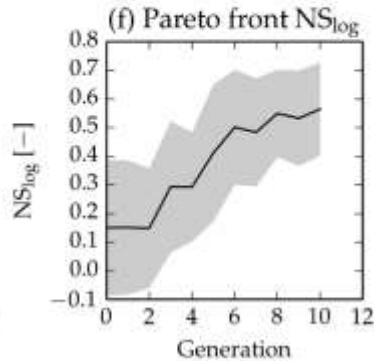
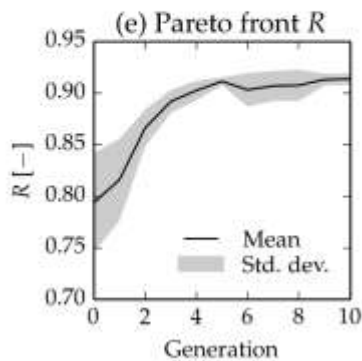
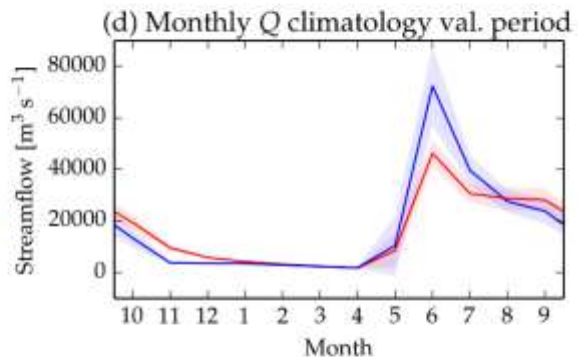
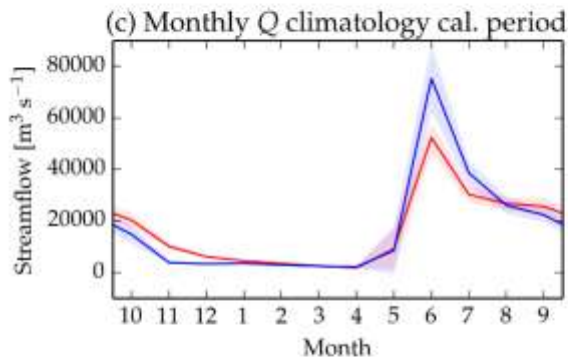
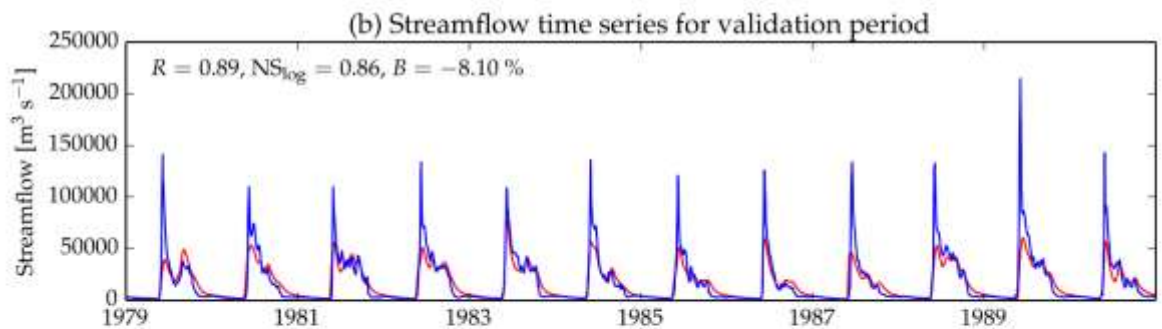
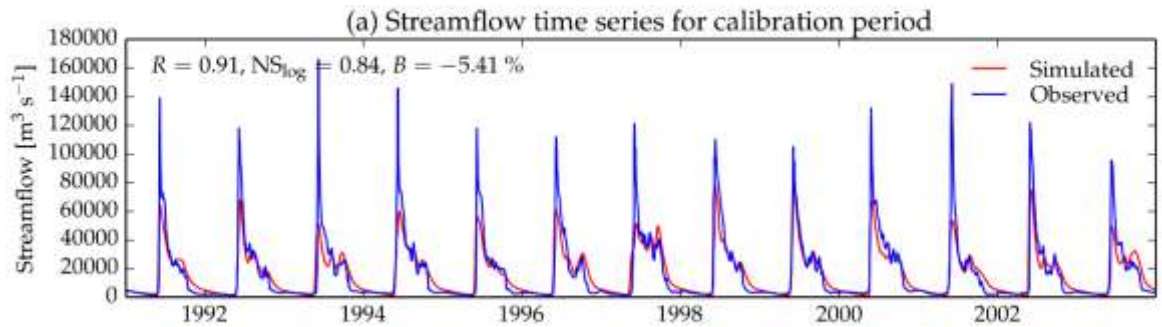
C810: Sava at S.Mitrovica



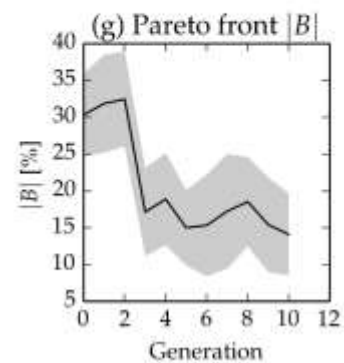
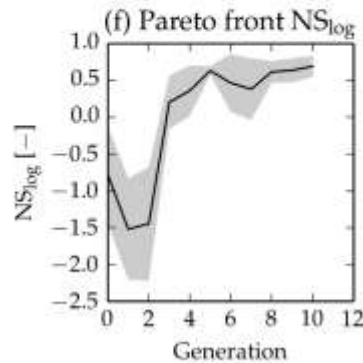
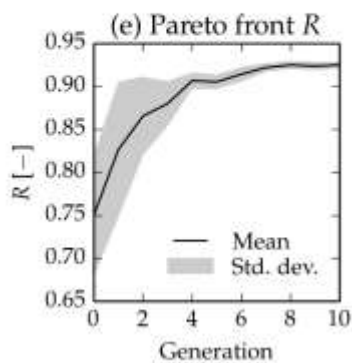
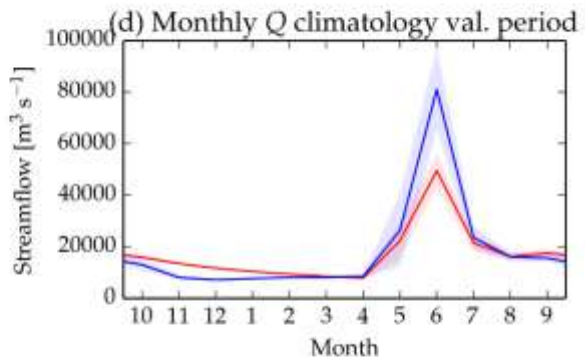
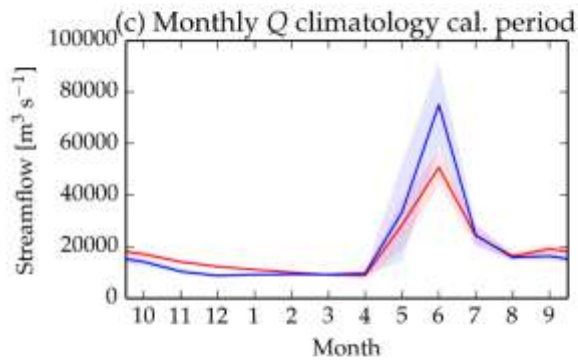
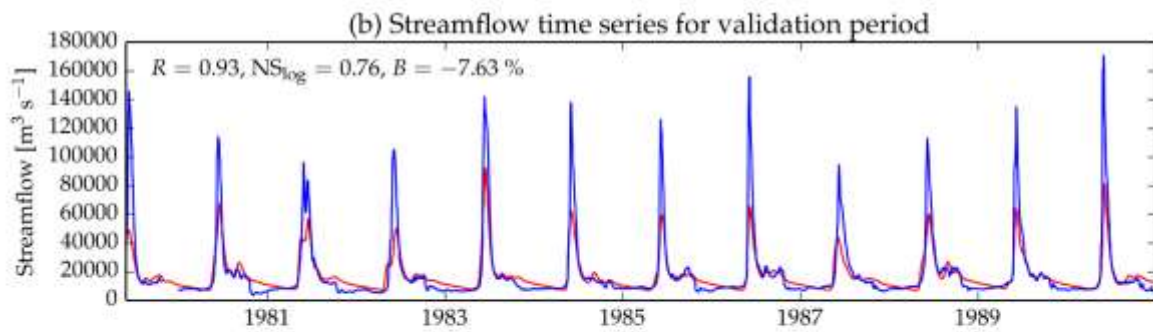
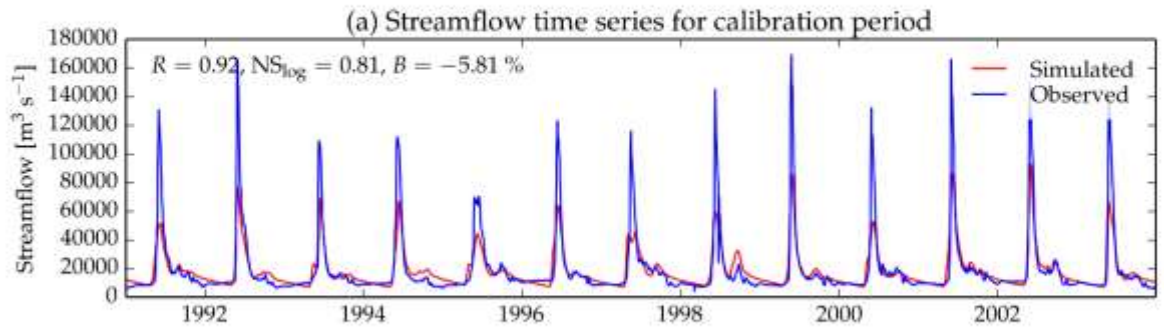
C817: Tisza at Senta / Szeged



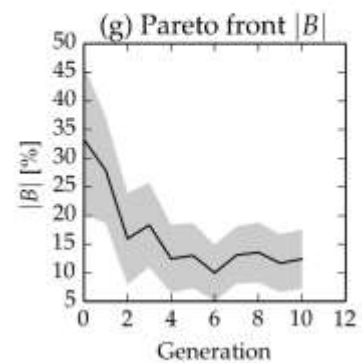
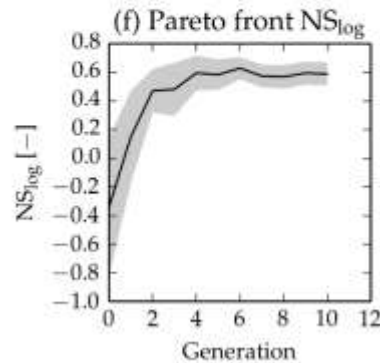
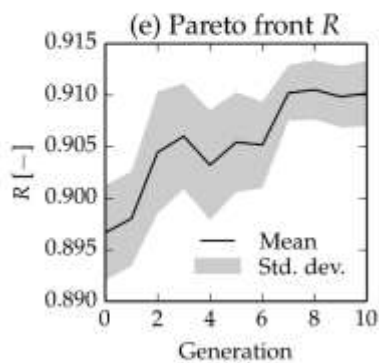
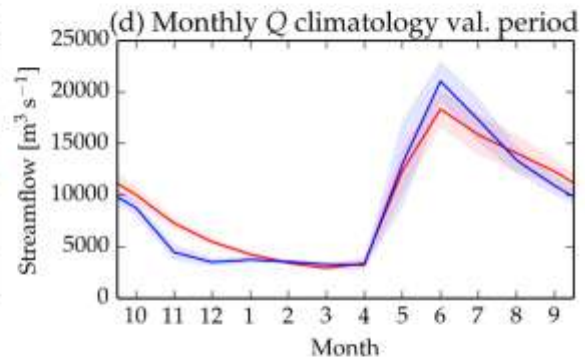
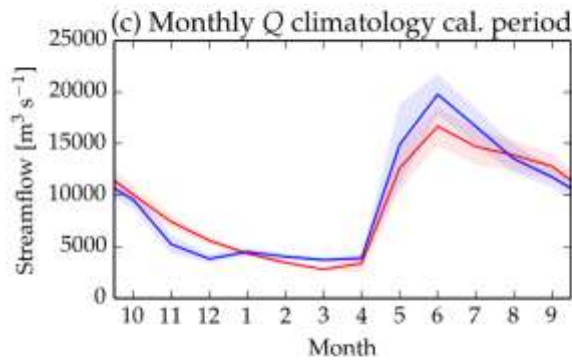
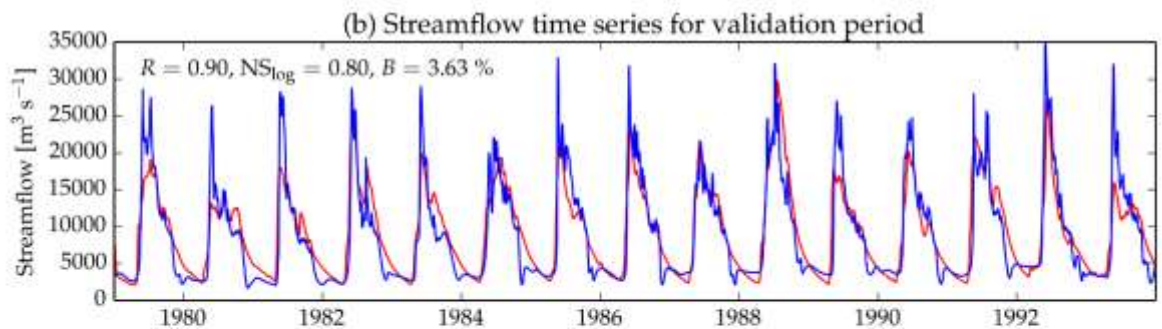
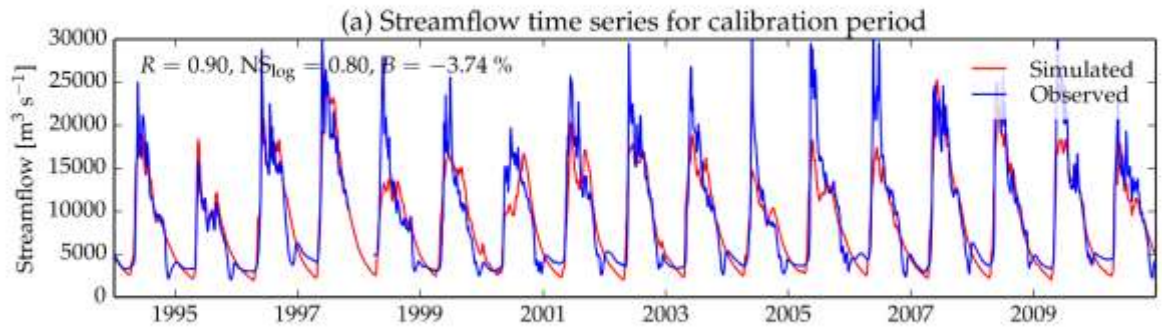
G0006: LENA at KYUSYUR (KUSUR)



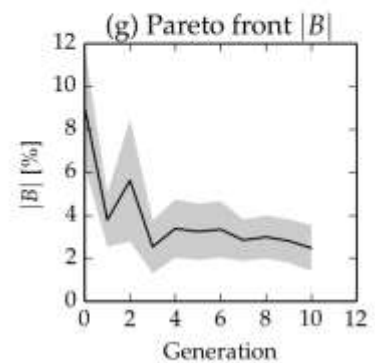
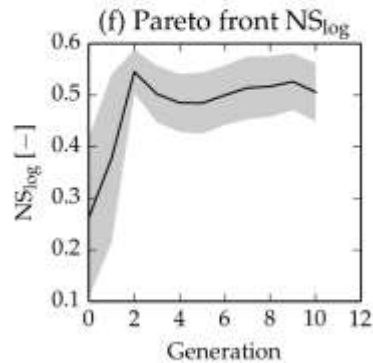
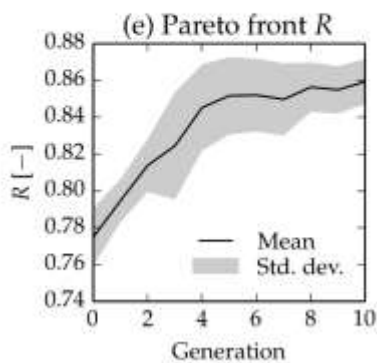
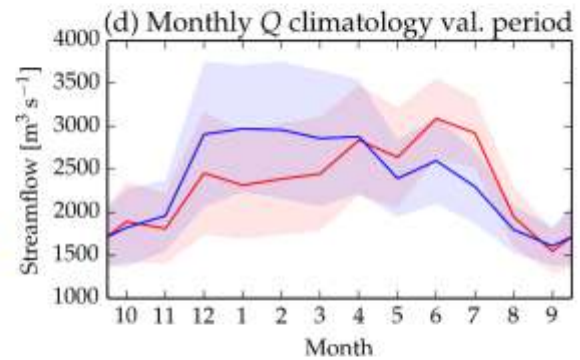
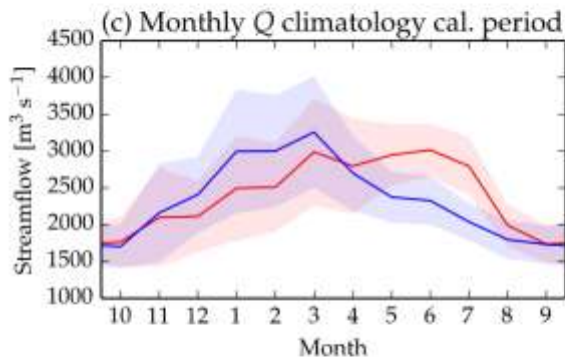
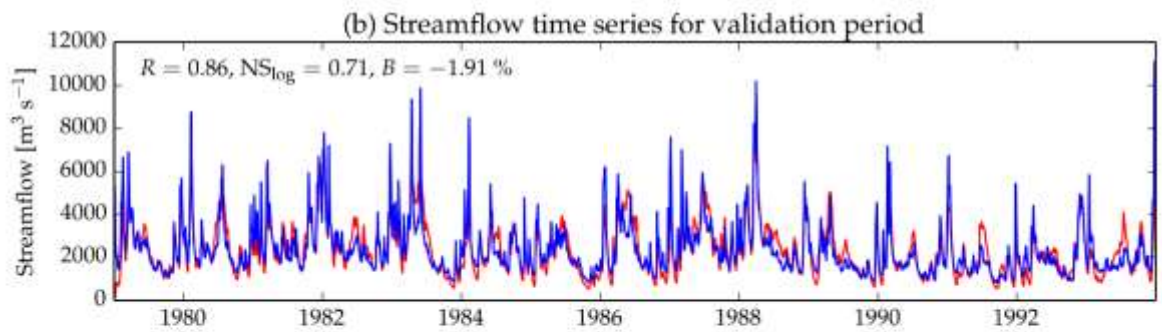
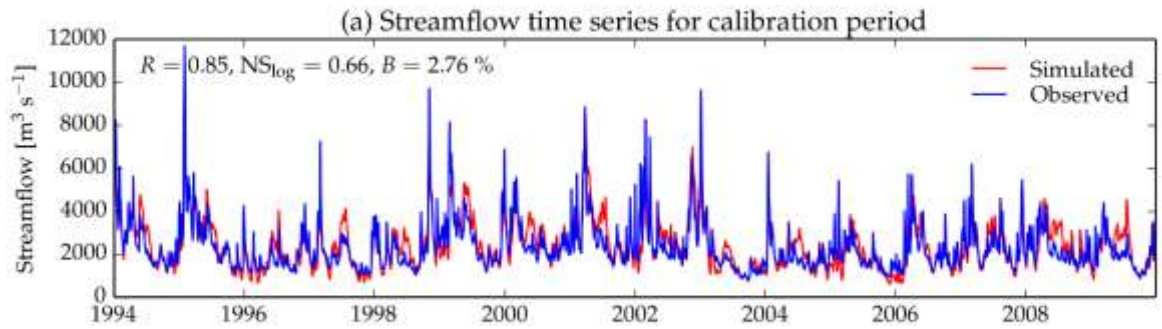
G0031: YENISEY at IGARKA



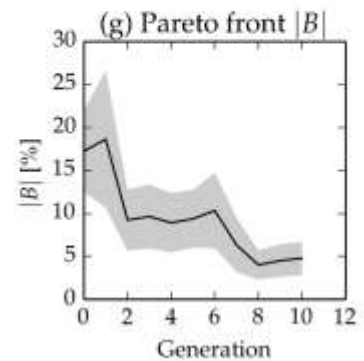
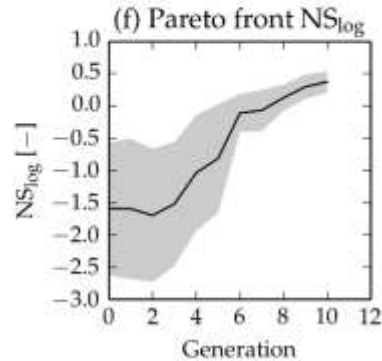
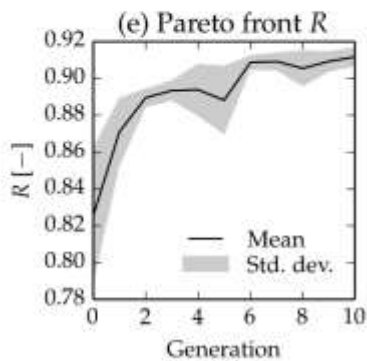
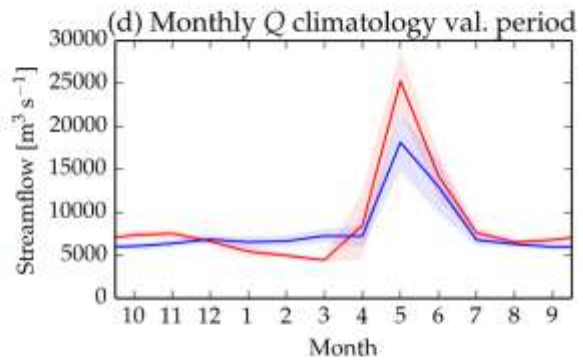
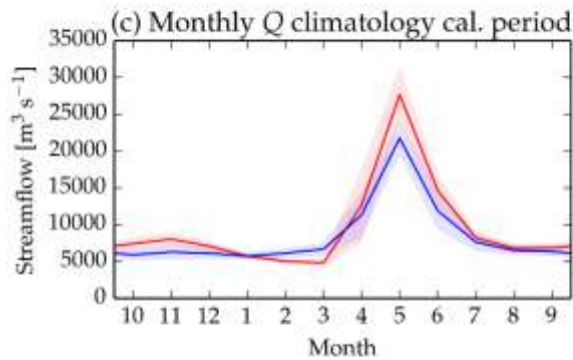
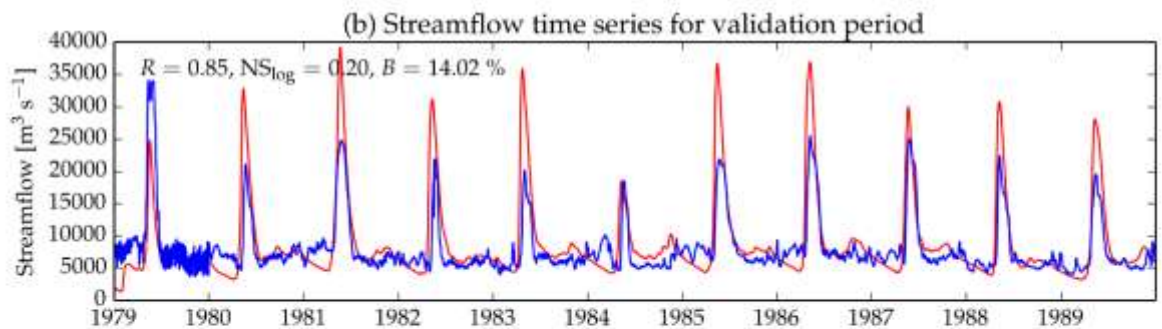
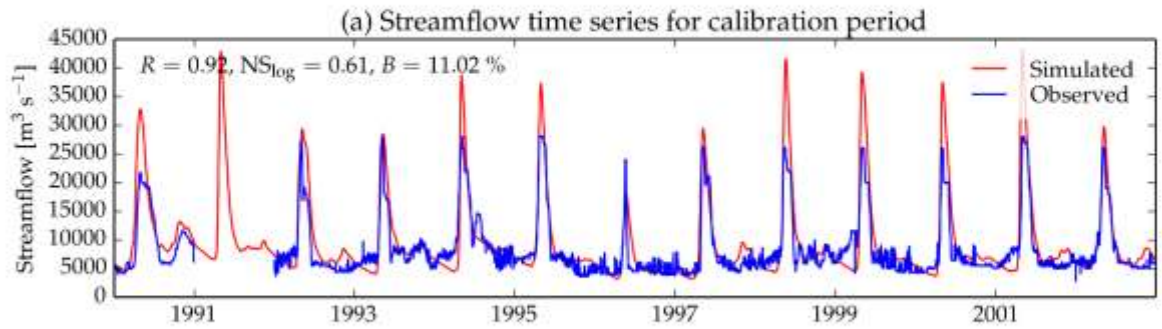
G0032: MACKENZIE RIVER at ARCTIC RED RIVER



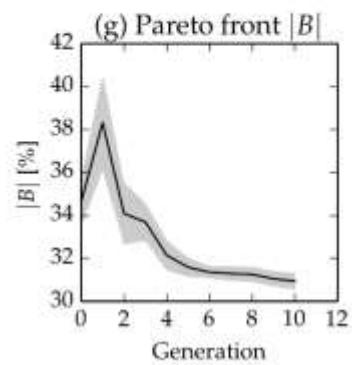
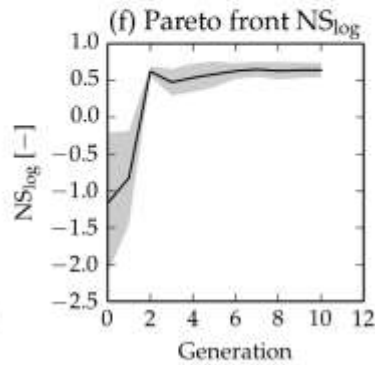
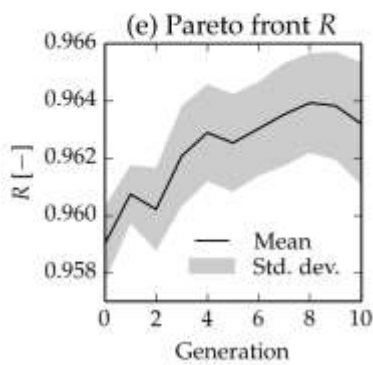
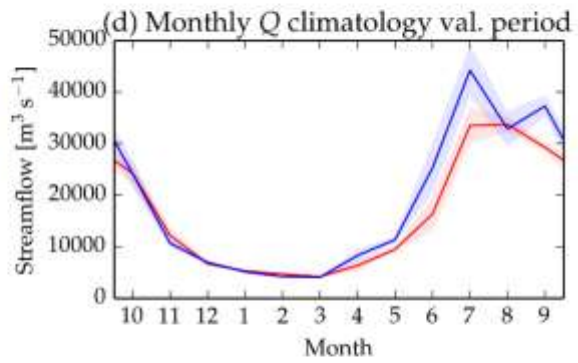
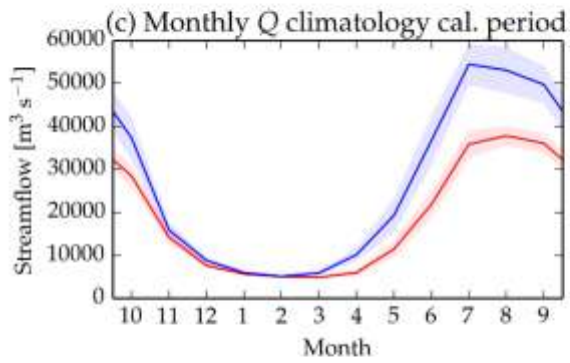
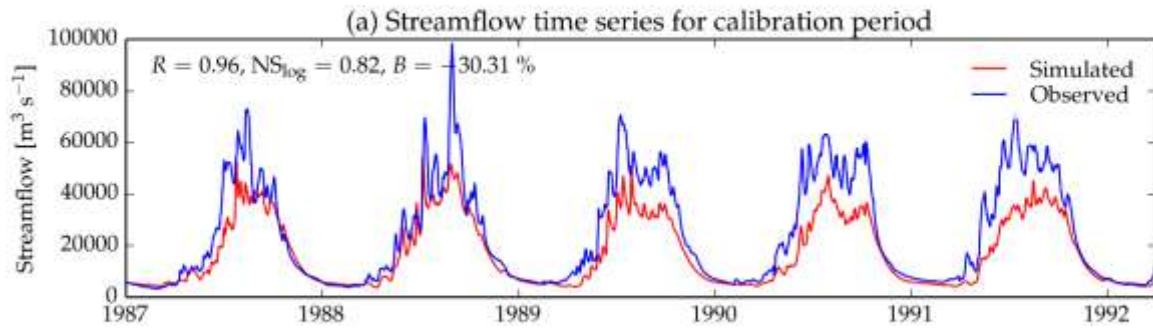
G0340: RHINE RIVER at REES



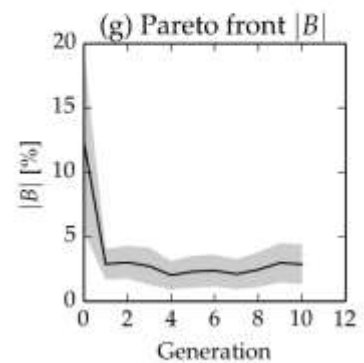
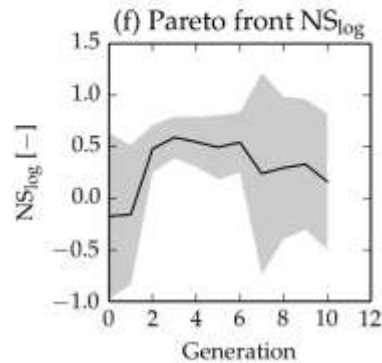
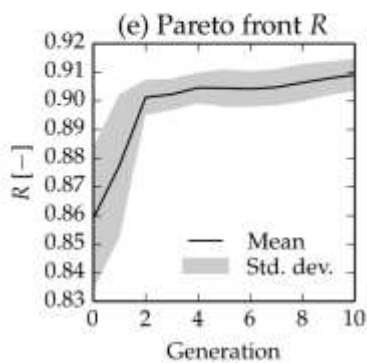
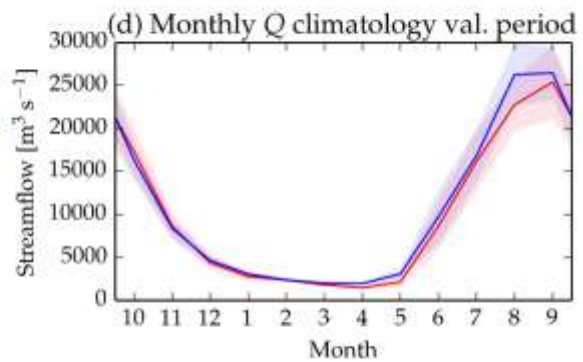
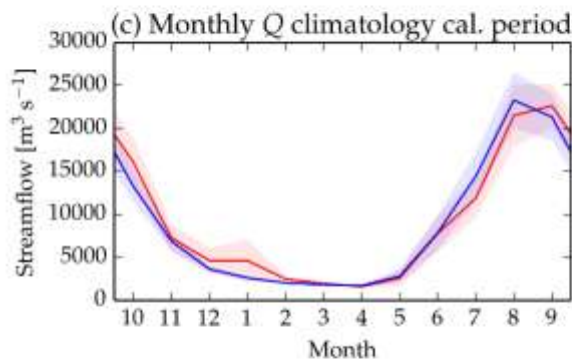
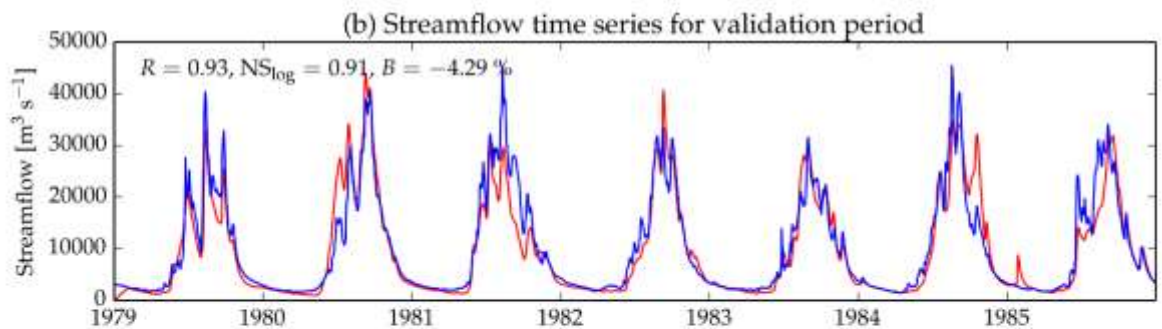
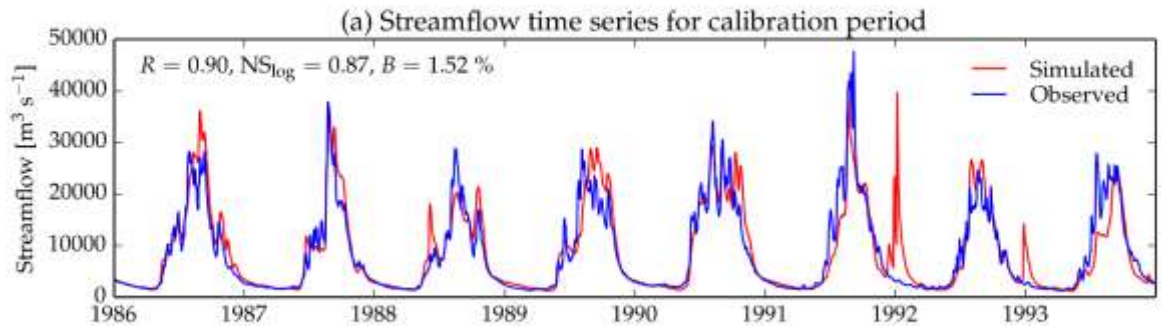
G0441: VOLGA at VOLGOGRAD POWER PLANT



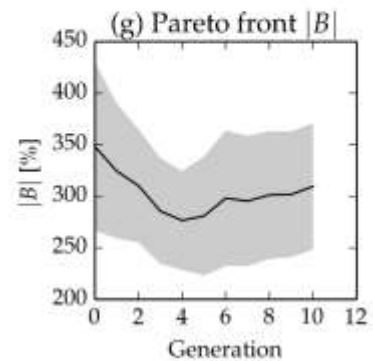
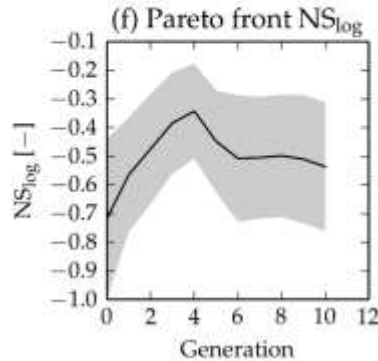
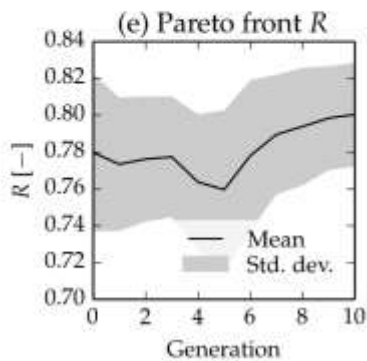
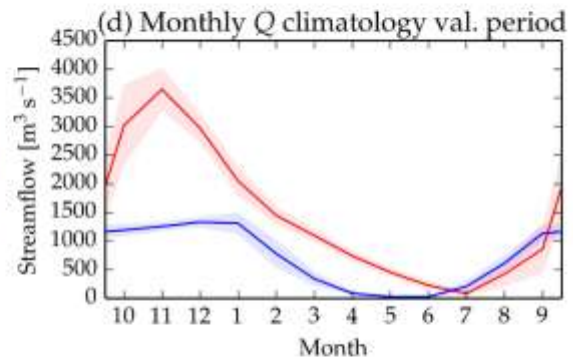
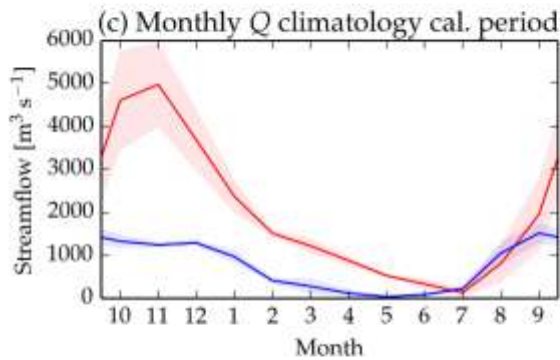
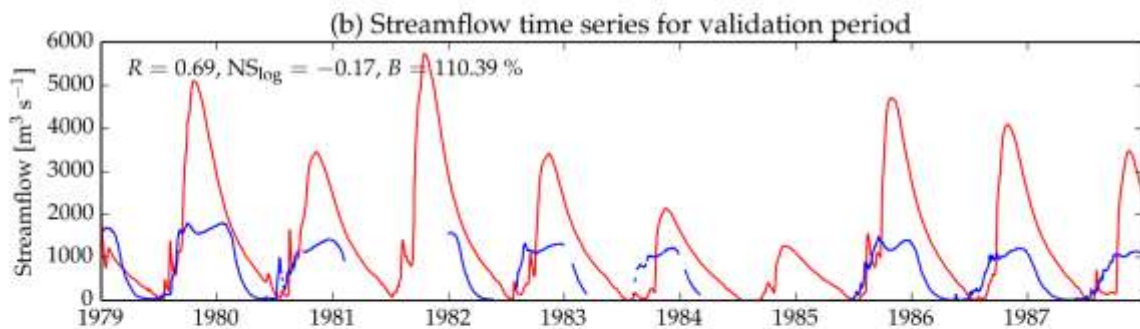
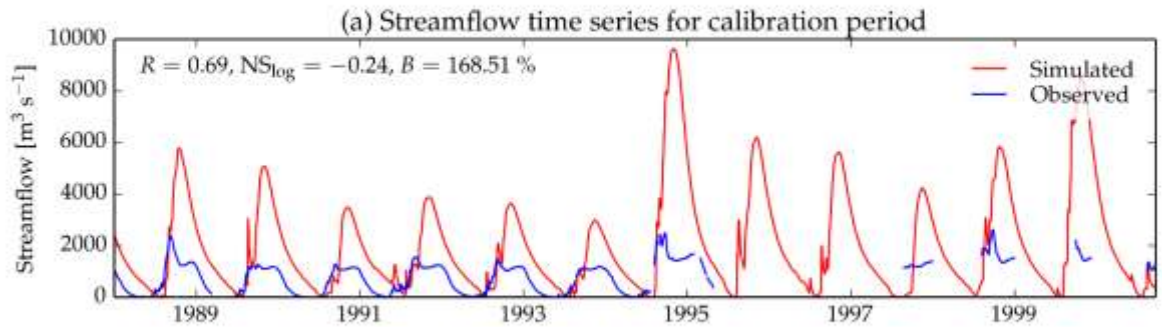
G0928: BRAHMAPUTRA at BAHADURABAD



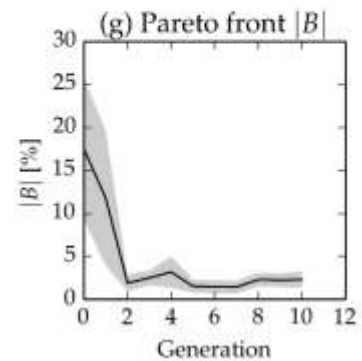
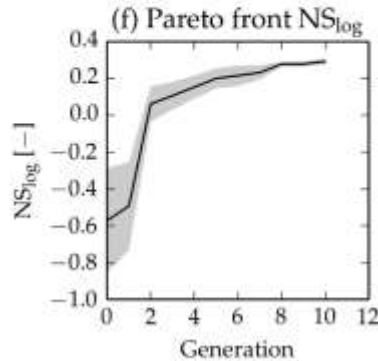
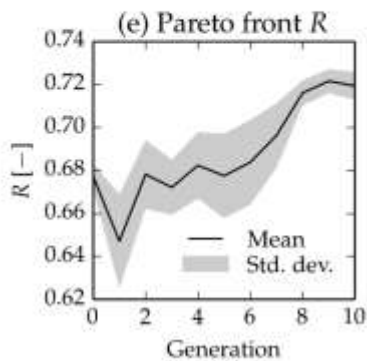
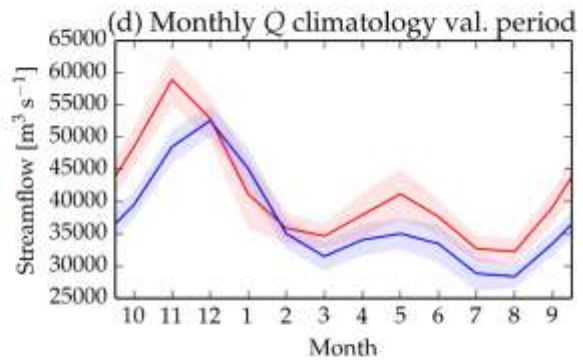
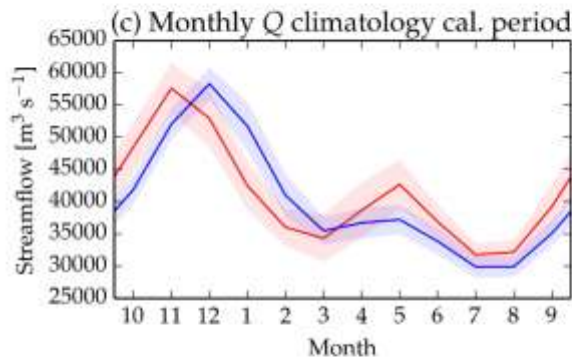
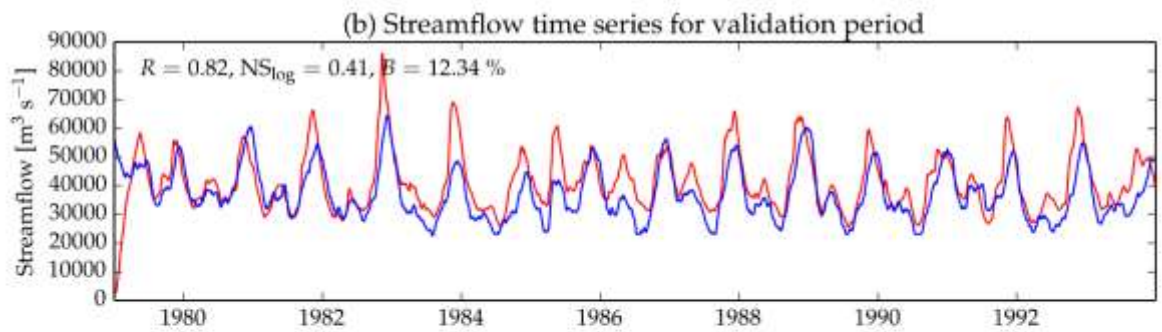
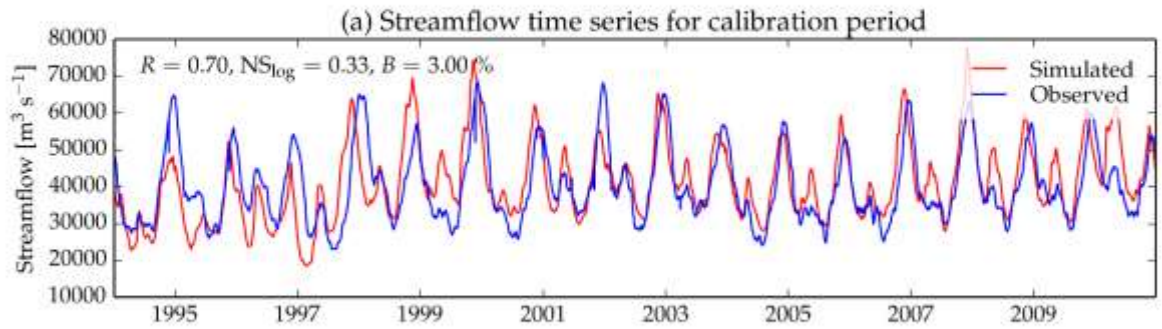
G0986: MEKONG at PAKSE



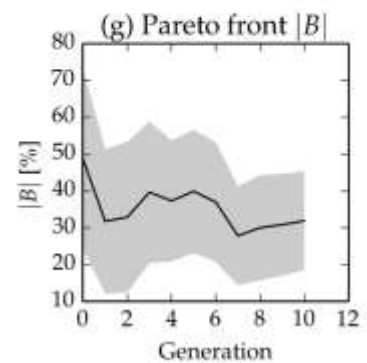
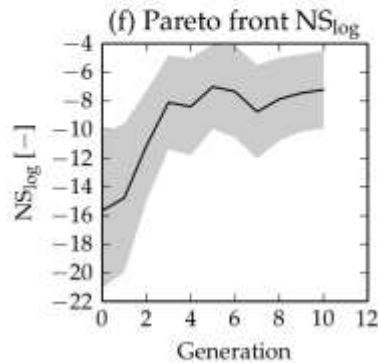
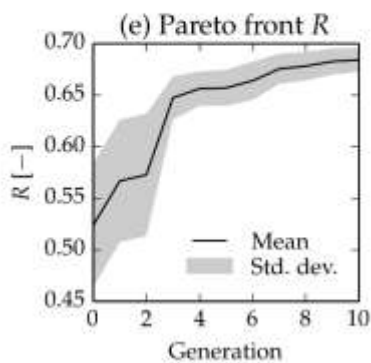
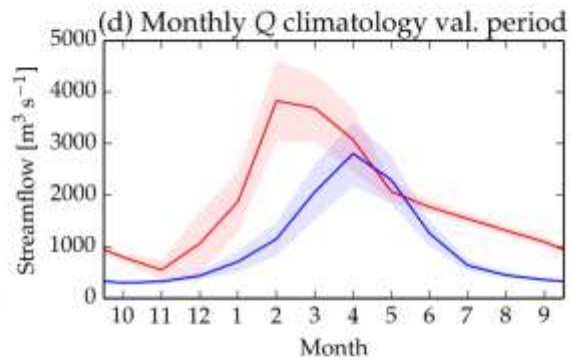
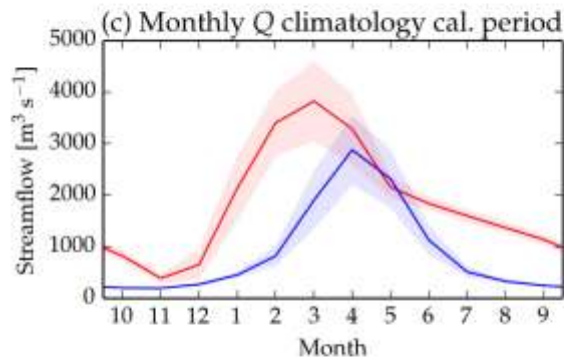
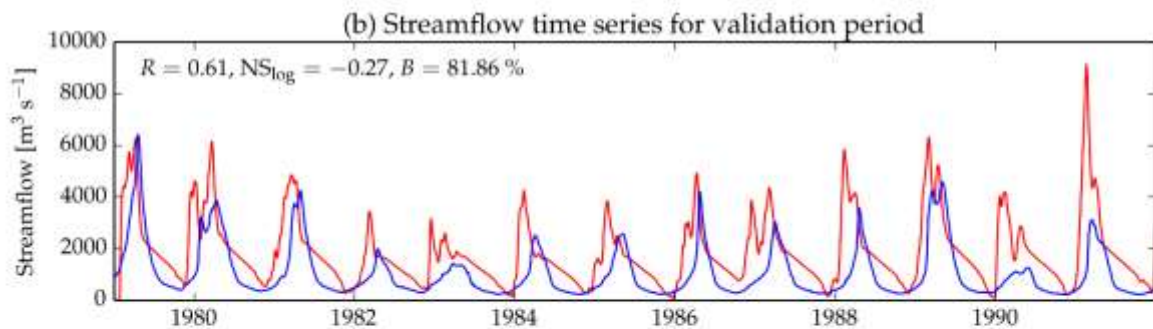
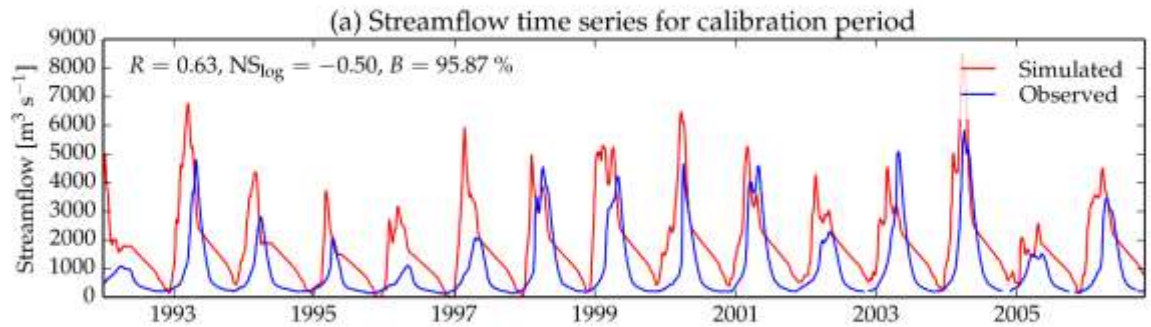
G1042: NIGER at MALANVILLE



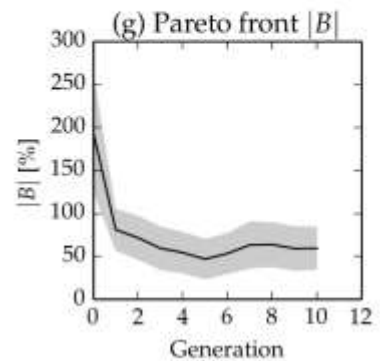
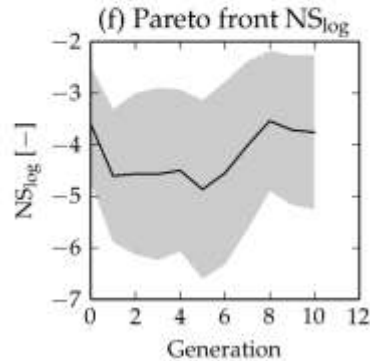
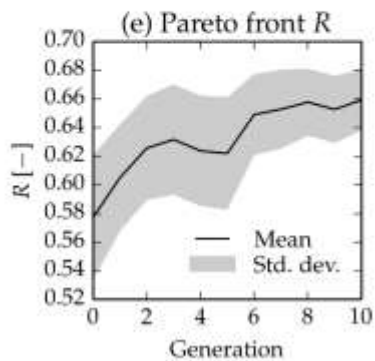
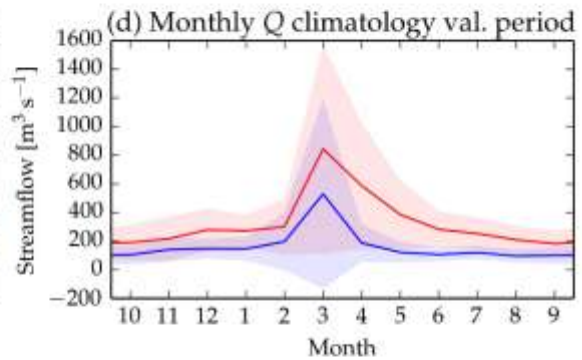
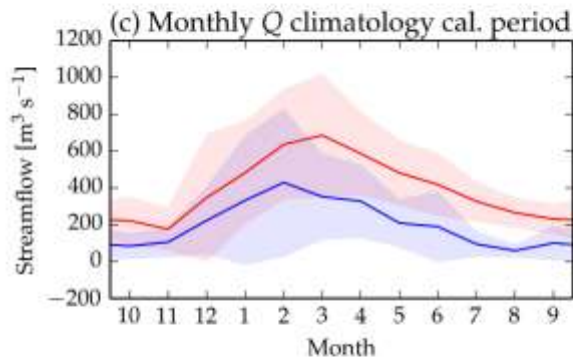
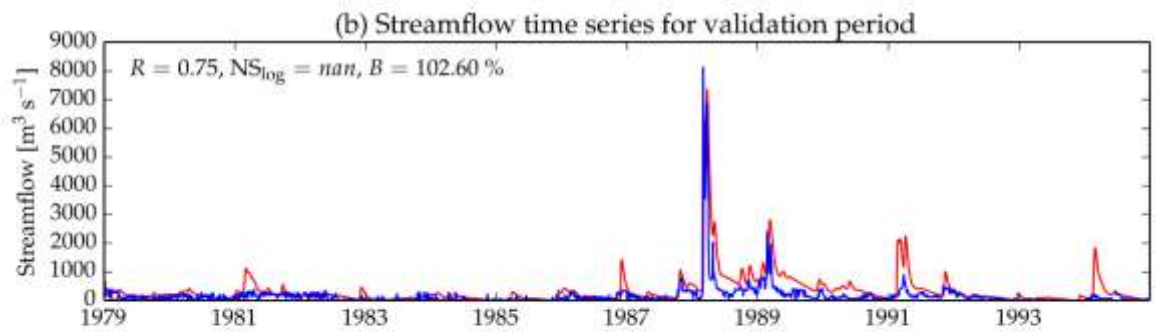
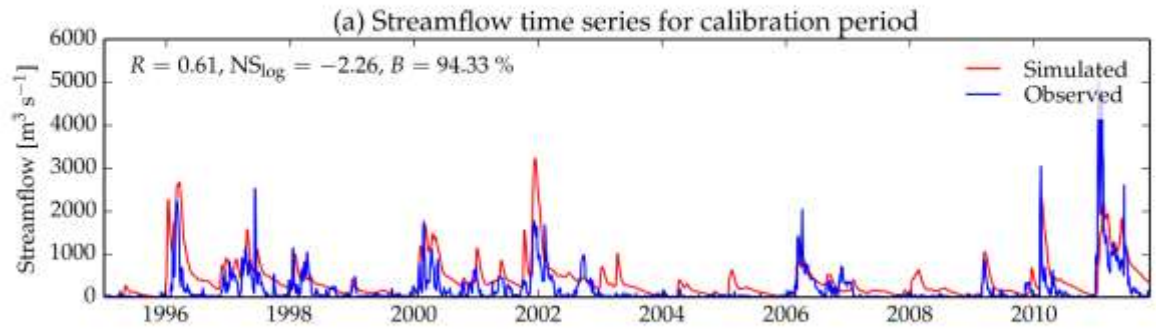
G1180: CONGO at KINSHASA



G1373: ZAMBEZI at KATIMA MULILO (64370001)



G1457: ORANGE at VIOOLSDRIF (27811003)



4. REFERENCES FOR PART A

- Alfieri, L., Thielen, J. and Pappenberger, F.: Ensemble hydro-meteorological simulation for flash flood early detection in southern Switzerland, *J Hydrol*, 424–425, 143–153, doi:10.1016/j.jhydrol.2011.12.038, 2012.
- Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J. and Pappenberger, F.: GloFAS – global ensemble streamflow forecasting and flood early warning, *Hydrol Earth Syst Sci*, 17(3), 1161–1175, doi:10.5194/hess-17-1161-2013, 2013.
- Alfieri, L., Salamon, P., Bianchi, A., Neal, J., Bates, P. and Feyen, L.: Advances in pan-European flood hazard mapping, *Hydrol. Process.*, 28(13), 4067–4077, doi:10.1002/hyp.9947, 2014.
- Alfieri, L., Burek, P., Feyen, L. and Forzieri, G.: Global warming increases the frequency of river floods in Europe, *Hydrol Earth Syst Sci Discuss*, 12(1), 1119–1152, doi:10.5194/hessd-12-1119-2015, 2015.
- Beck, H. E., de Roo, A. and van Dijk, A. I. J. M.: Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments, *J. Hydrometeorol.*, 16(4), 1478–1501, doi:10.1175/JHM-D-14-0155.1, 2015.
- Burek, P., Knijff van der, J. and Roo de, A.: LISFLOOD, distributed water balance and flood simulation model revised user manual 2013., Publications Office, Luxembourg. [online] Available from: <http://dx.publications.europa.eu/10.2788/24719> (Accessed 12 December 2014a), 2013.
- Burek, P., van der Knijff, J. and Ntegeka, V.: LISVAP, Evaporation Pre-Processor for the LISFLOOD Water Balance and Flood Simulation Model - Revised User Manual, Joint Research Centre –Institute for Environment and Sustainability. EUR 26167 EN. 36 pp., 2013b.
- De Roo, A., Odijk, M., Schmuck, G., Koster, E. and Lucieer, A.: Assessing the effects of land use changes on floods in the meuse and oder catchment, *Phys. Chem. Earth Part B Hydrol. Oceans Atmosphere*, 26(7-8), 593–599, 2001.
- Feyen, L., Vrugt, J. A., Nualláin, B., van der Knijff, J. and De Roo, A.: Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model, *J. Hydrol.*, 332(3-4), 276–289, doi:10.1016/j.jhydrol.2006.07.004, 2007.
- Forzieri, G., Feyen, L., Rojas, R., Flörke, M., Wimmer, F. and Bianchi, A.: Ensemble projections of future streamflow droughts in Europe, *Hydrol. Earth Syst. Sci.*, 18(1), 85–108, doi:10.5194/hess-18-85-2014, 2014.
- Harris, I., Jones, P.D., Osborn, T.J. and Lister, D.H. (2014). Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology* 34, 623-642
- Heim Jr. RR. 2002. A review of twentieth-century drought indices used in the United States. *Bulletin of the American Meteorological Society* 83(8): 1149-1165.
- Lei Y, Duan A. 2011. Prolonged dry episodes and drought over China. *International Journal of Climatology* 31(12): 1831-1840.

- Lloyd-Hughes B, Saunders MA. 2002. A drought climatology for Europe. *International Journal of Climatology* 22(13):1571-1592.
- McKee, T. B., Doesken, N. J., and Kleist, J. 1993: The Relationship of Drought Frequency and Duration to Time Scales, Proc. 8th Conf. on Appl. Clim., 17–22 January 1993, Anaheim, CA, 179–184, 1993.
- van der Knijff, J. M., Younis, J. and de Roo, A. P. J.: LISFLOOD: A GIS-based distributed model for river basin scale water balance and flood simulation, *Int J Geogr Inf Sci*, 24(2), 189–212, 2010.
- Revilla-Romero, B., Beck, H. E., Burek, P., Salamon, P., de Roo, A. and Thielen, J.: Filling the gaps: Calibrating a rainfall-runoff model using satellite-derived surface water extent, *Remote Sens. Environ.*, 171, 118–131, doi:10.1016/j.rse.2015.10.022, 2015.
- Rojas, R., Feyen, L., Bianchi, A. and Dosio, A.: Assessment of future flood hazard in Europe using a large ensemble of bias-corrected regional climate simulations, *J. Geophys. Res. Atmospheres*, 117(17), doi:10.1029/2012JD017461, 2012.
- Thiemig, V., Rojas, R., Zambrano-Bigiarini, M. and De Roo, A.: Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin, *J. Hydrol.*, 499, 324–338, doi:10.1016/j.jhydrol.2013.07.012, 2013.
- Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index, *J. Climate*, 23, 1696–1718, 2010
- Wanders, N., Wada, Y., & Van Lanen, H. A. J. (2015). Global hydrological droughts in the 21st century under a changing hydrological regime. *Earth System Dynamics*, 6(1), 1.
- Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O. and Schewe, J.: The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework, *Proc. Natl. Acad. Sci.*, 111(9), 3228–3232, doi:10.1073/pnas.1312330110, 2014.
- Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J. and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, *Water Resour. Res.*, 50(9), 7505–7514, doi:10.1002/2014WR015638, 2014.
- Wesseling, C. G., Karssenbergh, D.-J., Burrough, P. A. and Van Deursen, W. P. A.: Integrating dynamic environmental models in GIS: The development of a dynamic modelling language, *Trans. GIS*, 1(1), 40–48, 1996.
- Wilks, DS. 2011. *Statistical Methods in the Atmospheric Sciences*, 3rd ed. International Geophysics, Volume 100, Pages 2-676, Elsevier Inc., Academic Press, Oxford (UK), ISBN: 978-0-12-385022-5.
- Younis, J., Anquetin, S. and Thielen, J.: The benefit of high-resolution operational weather forecasts for flash flood warning, *Hydrol. Earth Syst. Sci.*, 12(4), 1039–1051, 2008.
- Zampieri M, D'Andrea F, Vautard R, Ciais P, de Noblet-Ducoudré N, Yiou P. 2009. Hot European summers and the role of soil moisture in the propagation of Mediterranean drought. *Journal of Climate* 22(18): 4747-4758.

PART B. EVALUATING AND IMPROVING THE SIMULATION OF PEAK RIVER DISCHARGE BY GLOBAL HYDROLOGICAL MODELS

Within Task4.3 of HELIX we evaluated Global Hydrological Models (GHMs) contributing historical simulations forced by observed weather to the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP). We focused on whether an implementation of a state-of-the-art global river routing and flooding model CaMa-Flood would improve the river discharge simulations, especially the amplitude of peak discharge, which is essential for simulating flooding. The models were driven by the same observed climate data, and they provide simulated global gridded daily discharge from 1971 to 2010 based on their own routing algorithms. Additionally, they also provide simulated global gridded daily runoff, before routing is performed. These runoff data are used as forcing for CaMa-Flood. Then for each GHM, both its original discharge and its runoff-driven, CaMa-Flood generated discharge are evaluated against monthly and daily GRDC discharge observations. Nash–Sutcliffe efficiency (NSE), Pearson's correlation, percent bias of mean, standard deviation and maximum discharge are used as performance metrics, and the previously mentioned two types of discharge simulations are evaluated at each station. Subsequently, the differences in performance between the two types of discharge simulations are computed, which are then used to derive basin-average changes in the performance metrics for each of 219 major river basins in the world.

Models on average are able to simulate relatively good monthly discharge over southern part of Asia, western Europe, most of South America and middle Africa (basin-average correlation greater than 0.6), but performance is generally poor over arid regions. With the application of CaMa-Flood routing, the most significant effect of seem to reduce both peak and total river discharge, and delay the peak discharge by about two weeks, likely related to its floodplain storage. As a result, a substantially better simulation of maximum monthly discharge is achieved with regard to the multi-model ensemble mean, which shows considerably reduced bias for basins with about 2 billions of people (51% of population in the 219 basins), compared to increased bias for basins with 0.3 billion people (7%). Higher improvement (or less degradation) is found with daily data, or for human-influenced basins, because CaMa-Flood routing acts similar to reservoir on discharge regulation. Especially considerable improvement is found in three models (CLM, H08, LPJmL), suggesting the importance of routing in flood simulations. A detailed look with Amazon discharge illustrate that CaMa-Flood routing tends to shift peak flow to a later time and reduce the amplitude, which results in performance improvements in some models, but decrease in some other models if the amplitude is already smaller than observation. Overall, notable improvement on the existing ISIMIP multi-model simulation archive of monthly and daily river discharge is possible with CaMa-Flood and can provide value to flood simulation efforts.

This work is written in a publication led by PIK. Zhao et al. "*Effect of routing on the performance of discharge simulations*" to be submitted to ISIMIP special issue of ERL (2016).

PART C. EVALUATION OF MEAN AND LOW HYDROLOGICAL STATES IN JULES

1. INTRODUCTION

The objectives of this area of work were as follows.

- Evaluate the JULES impact model for its ability to represent mean and low hydrological states
- Assess the model's ability to simulate past drought events
- Investigate the model's sensitivity to climate drivers, with regards to the representation of the various hydrological states and drought events.

2. METHODOLOGY

2.1. SIMULATIONS AND FORCING DATA

For the task of historical validation the Joint UK Land Environment Simulator (JULES) was forced with three different observational datasets. The simulations were performed at the global scale, with a spatial resolution of 0.5 degrees. The forcing datasets were the following: i) the Princeton Dataset (PGFv2), the GSWP3 dataset and the combined dataset of WATCH and WFDEI. The meteorological input followed a daily time-step and the simulations were performed with a 1- hourly time step. The simulation period spans from 1970 to 2010, while the year 1969 was used as the model spin up period.

2.2. HISTORICAL VALIDATION OF MEAN AND LOW HYDROLOGICAL STATES

In the context of performing a historical global scale model validation model simulations are compared with a gridded dataset of composite monthly runoff (Fekete & Vorosmarty 2011), spanning from 1986 to 1995. This comparison is also done for runoff aggregated at the basin level, for a group of 16 major basins covering different geographic and climatic regions. Validation focuses on mean and low hydrological state, which are expressed as mean and 10th lower percentile runoff production respectively. *Figure 12* shows the location and extent of the study basins.



Figure 12. 16 delineated major hydrological basins ranked by decreasing basin size.

2.3. VALIDATION AGAINST DROUGHT EVENTS

2.3.1. Description of drought events

Four past drought events were selected in order to validate the JULES' model against them. The events selected are significant drought events that happened into the time-span of the simulations and affected different regions of the globe. A bibliographic description of the selected drought events follows.

- *Drought in the United States in 1988*

During the spring and summer of 1988 an extensive and severe drought event struck the North American region. The region that was mostly affected is the mid-western part of the United States, including the agricultural region of the Great Plains (Sud et al. 2003). It is estimated that the drought event costed 30 billion dollars in terms of agricultural production failure and numerous heat stress induced losses in human lives. The meteorological processes that led to the drought formation manifested in March 1988 and the drought related anomalies (mainly in atmospheric circulation and precipitation) continued in April, May and June 1988. Although precipitation returned to normal levels in July, the evolution of the hydrological drought continued, due to the high evapotranspiration rates of the summer period. The drought, expressed as soil moisture deficit conditions and heat waves lasted until August 1988 (Namias 1991; Trenberth & Branstator 1992). Other manifestations of this extensive drought event include severe precipitation deficits, heat waves and forest fires as well as river discharge depletion, with Mississippi hitting a record low streamflow in mid-June 1988 (Sud et al. 2003).

- *Drought in Europe in 1989-1990*

One of the most significant drought events of the last 30 years that affected the European region occurred in 1989-1990. The total area affected by drought exceeded 40% of the continent (Tallaksen et al. 2011). From a climatological perspective the onset of the 1989-1990 drought was triggered by a decline in precipitation in the Mediterranean region during the winter and spring of 1989. In the summer of 1989 short wet periods appeared sporadically in the biggest part of the European region, causing an interim withdrawal of drought conditions in the continent, apart from Turkey where drought did not recede. Precipitation deficiencies reoccurred the following winter (1990) in south-eastern Europe, causing conditions of severe hydrological drought in the late spring and summer of 1990. This drought event, during its peak in May 1990, affected most of southern and central Europe. The eastern Mediterranean was seriously affected by the precipitation deficits of two consecutive years and the resulting drought conditions. Some of the impacts documented are declines in agricultural production as well as surface and groundwater scarcity (European Drought Centre 2013).

- *Heat waves and droughts in Europe in summer 2003*

Another exceptional event for Europe was the 2003 drought, combining significant precipitation deficits with record-setting heat extremes. At its peak, nearly all of Europe was in drought, except for the Iberian Peninsula and the far eastern Mediterranean. As a consequence, large losses in crop yield and extremely low discharge levels of rivers were reported in large parts of Europe. The overall

death toll of the heat waves and of the drought was estimated at about 70,000 people (Robine et al. 2008).

The drought event began with a precipitation deficit in Scandinavia in early spring 2003, which further expanded and in June the center was located in Poland. Climatological drought remained moderate through June until July, when the extent and severity of the precipitation deficit rapidly expanded to cover most of Europe. This rapid expansion of drought was caused by a persistent blocking high-pressure pattern that lingered over Western Europe. The most severe precipitation deficits, which occurred in July and August of 2003, were accompanied by the warmest temperatures ever recorded in Europe until then. This greatly increased evapotranspiration, reducing available water. The meteorological drought began to quickly recede in late September and October, with drought conditions only remaining in northern Italy and southern France.

In terms of impacts the European drought of 2003 affected an area spreading from Portugal to Romania and Bulgaria (Demuth 2009; European Environment Agency 2010). It was characterized by diverse and far reaching effects resulting from an exceptional rainfall deficit combined with extended heat wave conditions.

Agriculture was particularly affected in Southern and Central Europe: French, Italian, German, Austrian, Swiss, Slovakian, Spanish and Portuguese agriculture but also Eastern countries have been among the most affected by the drought and the heat wave in 2003 (Swiss 2010). In many countries of the South-Eastern European region, like Hungary, Slovenia, Croatia, Serbia and Montenegro, 2003 was among the major agricultural droughts in recent years. The International Disaster Database EM-DAT registered drought disasters in 2003 in Bosnia-Herzegovina, Croatia and Hungary. Significant decreases in yields compared to previous years were common, however with big variations per region. European livestock farmers were worst hit due to the big impact on green fodder supply. Agricultural losses were estimated to amount to more than 10 billion Euros (Swiss 2010). Governmental measures to mitigate the effects for the farmers were taken in several countries and also by the European Commission.

Local limitations and serious shortage problems in public water supply were reported for some rural mountainous areas in Italy, Austria, Switzerland, France and Germany. A few communities and single farms, which had depended exclusively on the use of local spring waters (or traditional private wells), had to be supplied by emergency actions. However, immense demands during the hot summer period, restrictions on water use and abstraction and a strong overall depletion of resources, i.e. dried up springs and boreholes, extremely low groundwater levels and reservoir stocks, were common across Europe.

In several European rivers, extremely low flows and high water temperatures were recorded. The energy sector was challenged by a reduced potential of hydropower production, widespread problems with cooling of nuclear and thermal power plants as well as unusually high demands. Thermal and nuclear power plants throughout Europe had to operate at reduced capacities or even shut down due to the high river water temperatures. In August emergency exemptions from environmental legislation were granted for several power plants in France, the Netherlands and Germany in order to avoid disruptions and ensure continuity in the supply. Conditions of power supply in Italy (Cassardo et al. 2007) and France (e.g., Poumadere et al. 2005) were among the most critical ones.

- *Drought in the Amazon in 2005*

In 2005 a severe drought event hit the Amazon. Its manifestation was particularly intense due to the combination of the 2002-03 El Niño with a dry spell in 2005 (Zeng et al. 2008). The drought affected mostly the western and southern regions of the Amazon basin (Zeng et al. 2008), while the central and eastern parts remained unaffected (Marengo et al. 2008). The drought conditions resulted in historical lows being reported for the water levels of the Amazon River, suspension of the navigation in the river and many wildfires. Drought consequences were also pronounced for agriculture and hydroelectric energy production (Marengo et al. 2008). The drought peaked in September of 2005, and although precipitation returned to normal levels in October, conditions of hydrological drought persisted until the beginning of 2006 (Marengo et al. 2008; Chen et al. 2009). This underlines the significant sensitivity of the Amazon's hydrology to temporally extended droughts in contrast to the shorter and seasonal drought events against which the ecosystem has developed coping mechanisms (Zeng et al. 2008).

2.3.2. Drought event modelling

For the drought events that happened during the period covered by the gridded runoff dataset (US 1988 and Europe 1990) the spatial variability and extent of the drought as captured by the observed dataset and the model results was compared. For each grid-box, drought thresholds are calculated taking into account all the values of the 10-year runoff record. To examine the variability in the intensity of drought conditions, three thresholds are considered: the lower 10th, 20th and 30th percentiles. The results are calculated and presented for a six-month period around the drought peak, which is defined by literature sources.

Moreover, a basin scale analysis of the model's ability to represent the extreme events discussed in the above section is performed. Five hydrological basins that were affected by the drought condition are examined:

- Amazon - affected by the 2005 Amazon drought event
- Mississippi - affected by the US drought in 1988
- Danube, Rhine and Elbe – affected by the European 1990 and 2003 droughts.

For the basin scale analysis, the drought events were identified employing two variations of the threshold level method: the constant threshold method (CTM) and the daily varying threshold method (DVTM), with the former being more appropriate for identifying streamflow droughts and the latter streamflow deficiencies (Stahl 2001). A streamflow deficiency corresponds to a severe drought in case it happens during the low flow season.

For the CTM for drought event definition, the Q70 (flow equalled or exceeded for 70% of the time) of the 1976 to 2010 period is used as threshold. Time-series of streamflow are smoothed using a 10-day moving average, in order to avoid minor and mutually dependent events (Stahl 2001). For the DVTM the threshold is established as follows: for each Julian day k the threshold is defined as the Q90 (or 10th percentile) of a 31-day window discharge centering at day k , applied to all years of the study period (1976-2010) (Stahl 2001; Prudhomme et al. 2011). These two procedures are followed for observed and modelled streamflow time-series separately. Then observed and modelled streamflow of the drought year are compared with their respective historical threshold and the number of days that flow fell below that threshold are counted and presented as a cumulative density function (cdf) of days under drought threshold in one year.

3. RESULTS

3.1. HYDROLOGICAL STATES

Figure 13 shows the differences in mean monthly runoff production between the simulations and the observed runoff dataset. The biggest differences appear in the west and central equatorial African region and in south America, around the Amazon. Concerning the rest of the land surface, runoff underestimations of up to 25 mm/month are found at the northern regions while overestimations of the same magnitude are found in regions of central north America, Europe, east Asia and Oceania. According to *Figure 14*, which illustrates the differences in the lower 10th percentile of runoff between simulations and observations, low runoff is underestimated by the model in west and south equatorial Africa, east south America and Indonesia. Regions of low runoff overestimation are northern Europe and west north America, together with west south America.

For a closer insight on the differences between the simulations and the observational dataset as well as on the effect of the forcing dataset on the simulations, the gridded data were spatially aggregated and compared at the basin scale. The basin aggregated mean runoff values of simulations and observations are shown in *Figure 15*, along with a scatterplot of simulations versus observations. In general terms the deviations between observations and simulations are small and their correlation is strong (R-squared is 0.91, 0.93 and 0.85 for the WATCH+WFDEI, PGFv2 and GSWP3 datasets respectively). The larger runoff underestimation is reported for Congo and overestimation for Amazon and Volga. However, this concerns the absolute difference between observed and modelled values. Expressing these deviations as relative (percent) differences (see Appendix) it is found that Amazon is one of the basins with the smaller deviations from observed runoff (5.7% to 17.0%). Other basins with small relative differences between simulations and observations are Mississippi, Danube and Rhine. The largest runoff underestimation in relative terms is found for Guadiana (-71.6% to -80.2%) followed by Niger and Congo. The forcing dataset used has a considerable impact on mean runoff simulations. The range of the three simulations expressed as percent of the average varies between 10.1% (for Amazon) to 83.5% (for Kemijoki). For some basins with large range of simulations (Lena, Oder and Kemijoki) it is observed that a single forcing dataset leads to very small values of percent difference from runoff observations while the other datasets result in large deviations from the observed values.

The comparison and correlation between observed and simulated values of 10th percentile runoff are shown in *Figure 16*. Low runoff simulations are well correlated with observations, although their correlation is slightly weakened compared to the one documented for mean runoff (R-squared values are 0.86, 0.87 and 0.77 respectively for the WATCH+WFDEI, PGFv2 and GSWP3 datasets). Concerning the differences of simulations from observations, they appear much smaller in absolute terms but are significantly more pronounced in relative terms. The range of the simulations is also augmented for low runoff (21.9% to 287.4 %).

3.2. DROUGHT EVENTS

The spatial extent of the 1988 drought over the north American region for a 6-month period is shown in *Figure 17*. Regarding the drought extent, the simulations result in a higher percentage of the land being under drought conditions for all the examined months. The temporal evolution of simulated drought patterns however, matches that of the observed drought. Regarding the representation of the 1990 drought event that affected the European region (*Figure 18*), observations and simulations result in a similar pattern and evolution of the event, although the simulations show a greater area affected by the drought conditions in the first time-step of the analysis (March 1990). *Table 3* and *Table 4* show the spatial average of correlations between observed and simulated values that fall under the 10th percentile drought threshold, for the 1988 and 1990 drought events respectively.

The results of the basin scale drought event analysis are illustrated in *Figure 19*. It can be observed that for most events and basins the DVTM gives a greater agreement between observed and simulated drought event evolution. The best agreements of cdfs of observed and simulated days under drought threshold is documented for Mississippi (1988 drought) and Danube (2003 drought). For the Danube 1990 event only one of the three forcing datasets (GSWP3) results in a good representation of the observed cdf while for the Elbe 1990 event none of the forcing datasets results in simulations that can depict the observed evolution of the event.

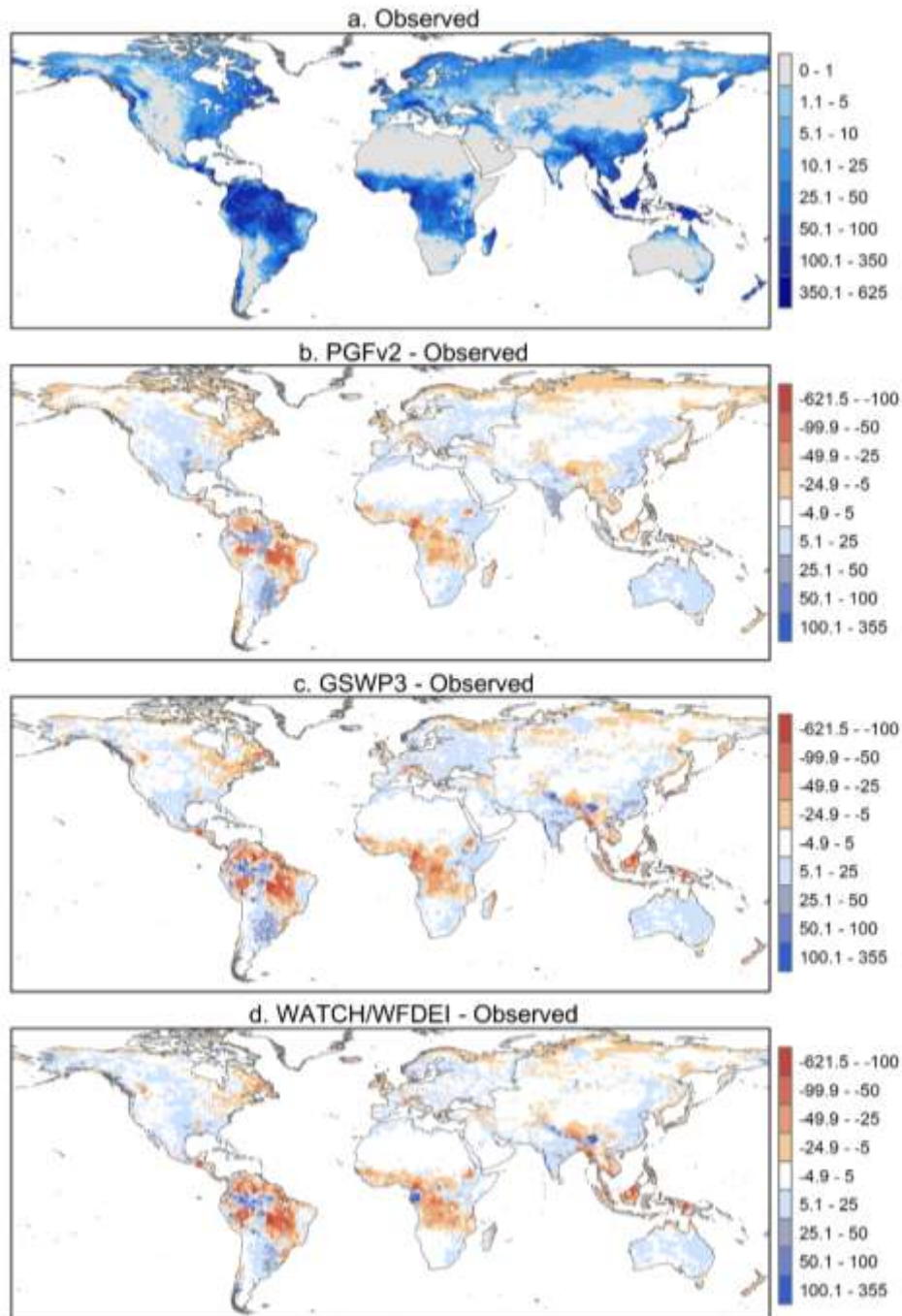


Figure 13. Mean runoff [mm/month] of the period 1986-1995, b-d. Difference between mean modelled and observed runoff, for the three forcing observational datasets.

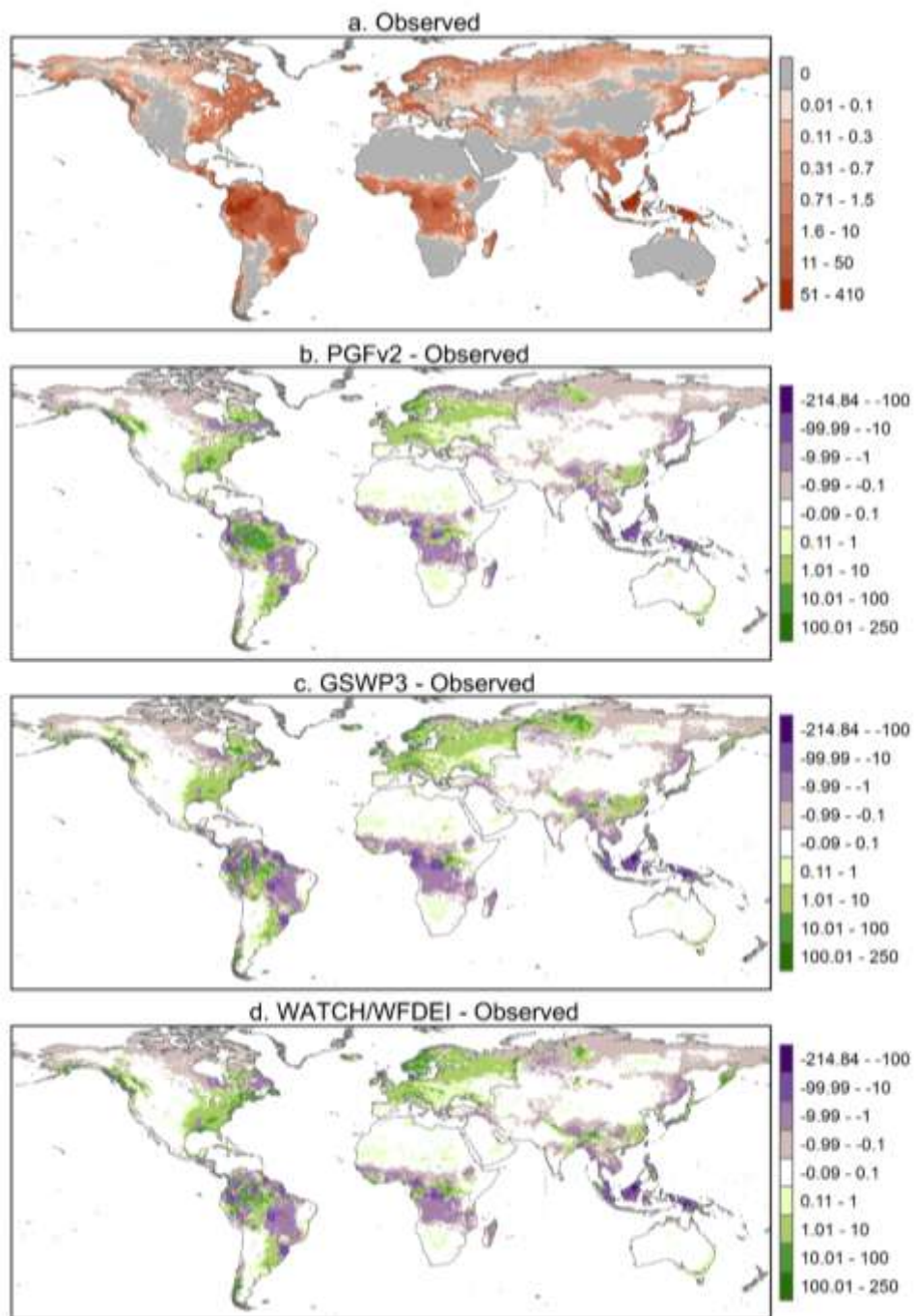


Figure 14. As Figure 2, but for the lower 10th percentile runoff.

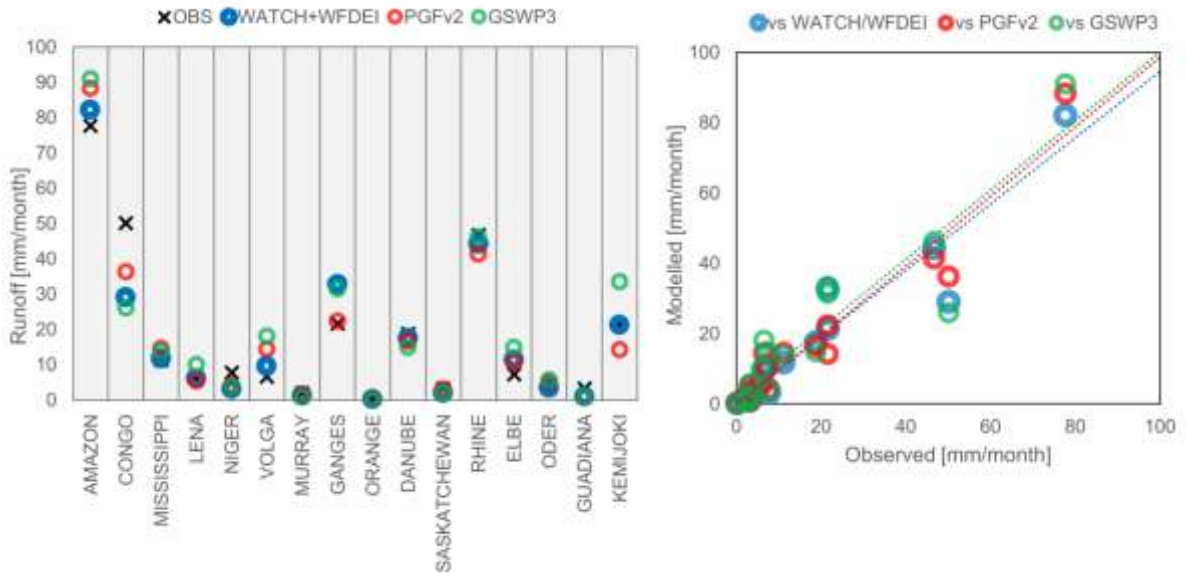


Figure 15. Basin aggregated mean runoff production. Comparison between 3 observational datasets.

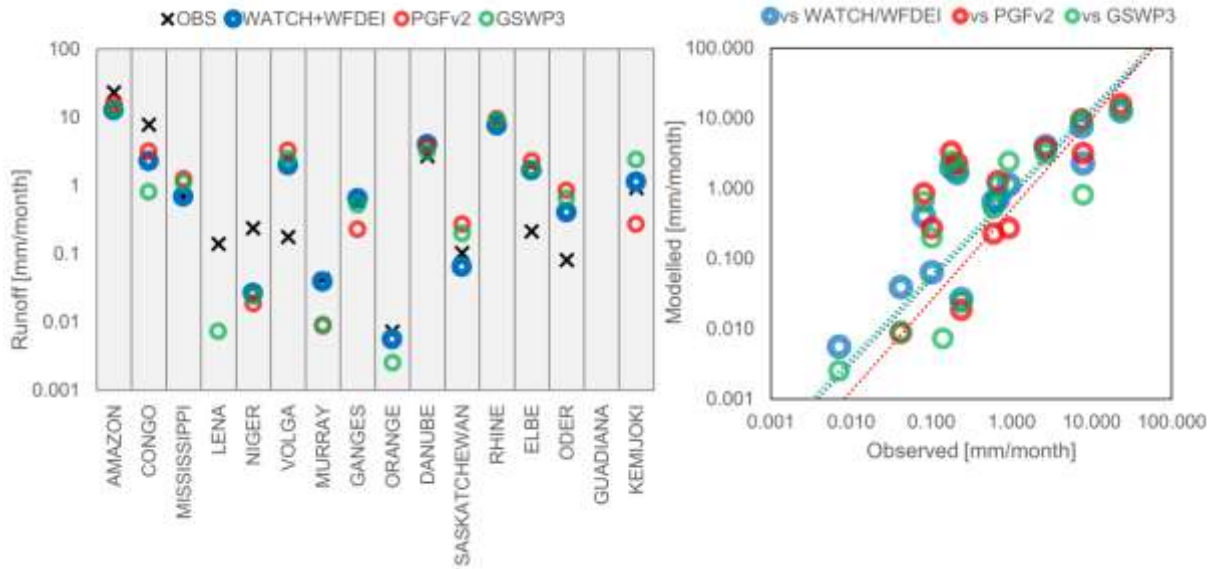


Figure 16. As Figure 6, but for 10th percentile runoff.

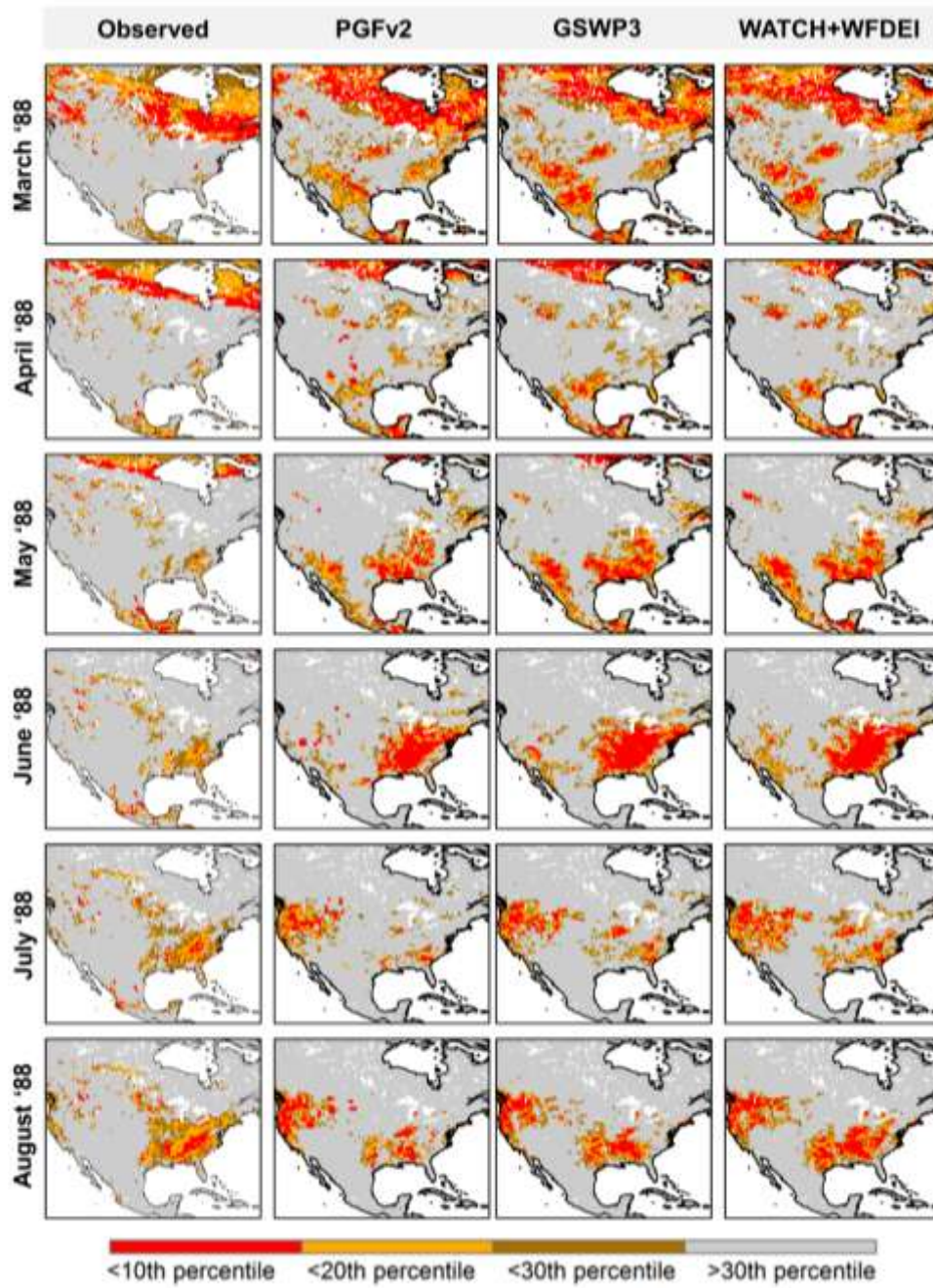


Figure 17. Spatio-temporal evolution of the 1988 drought event affecting North America.

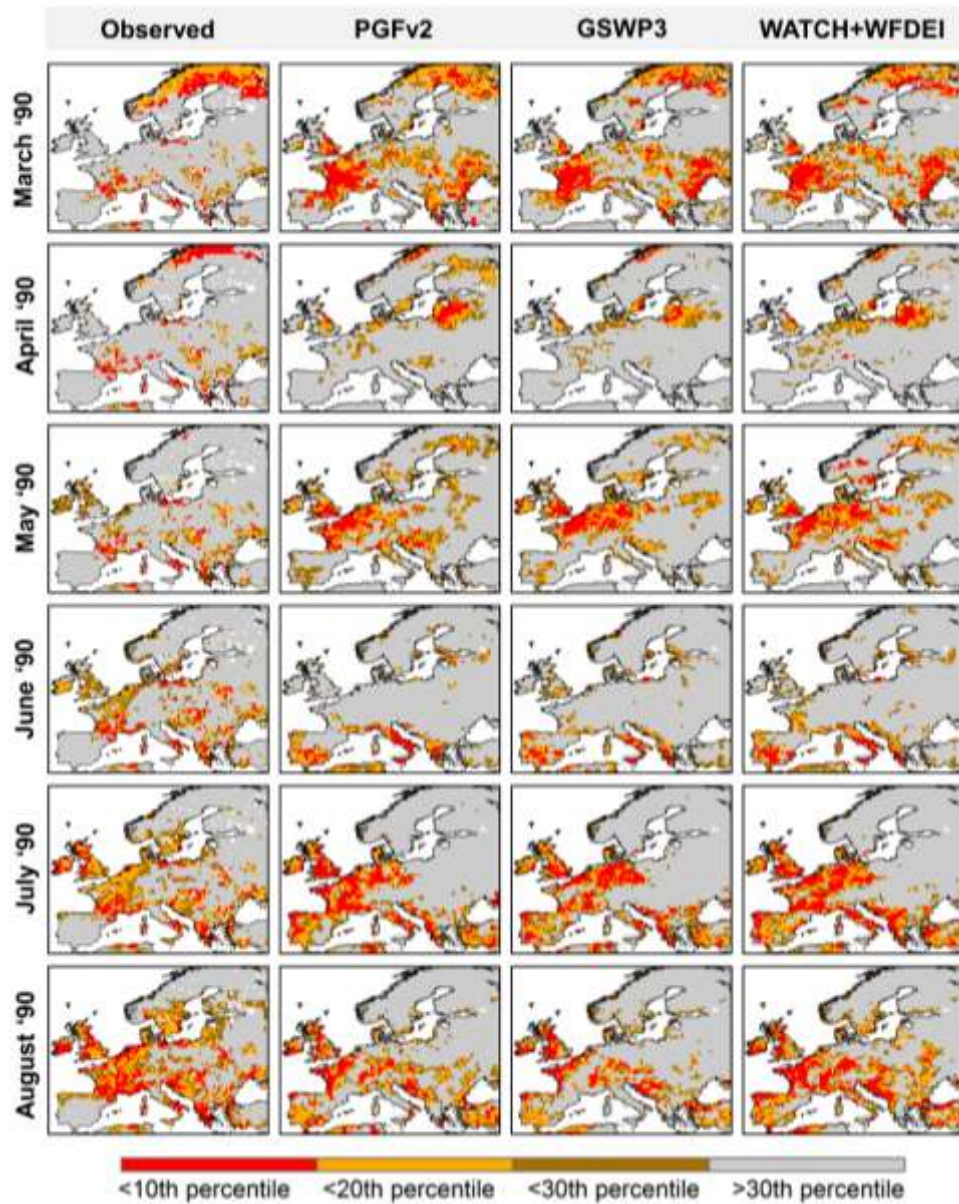


Figure 18. Spatio-temporal evolution of the 1990 drought event affecting Europe.

Table 3. Spatial average of correlations between observed and simulated values corresponding to the <10th percentile drought threshold, for the 1988 drought event affecting North America.

	WATCH+WFDEI	PGFv2	GSWP3
Mar-88	0.49	0.50	0.44
Apr-88	0.33	0.35	0.49
May-88	0.02	0.03	0.11

Jun-88	0.18	0.18	0.21
Jul-88	0.21	0.27	0.20
Aug-88	0.37	0.42	0.36
Average	0.27	0.29	0.30

Table 4. Spatial average of correlations between observed and simulated values corresponding to the 10^{th} percentile drought threshold, for the 1990 drought event affecting Europe.

	WATCH+WFDEI	PGFv2	GSWP3
Mar-90	0.41	0.45	0.45
Apr-90	0.27	0.27	0.42
May-90	0.03	0.06	0.08
Jun-90	0.18	0.19	0.18
Jul-90	0.25	0.25	0.22
Aug-90	0.38	0.40	0.37
Average	0.25	0.27	0.29

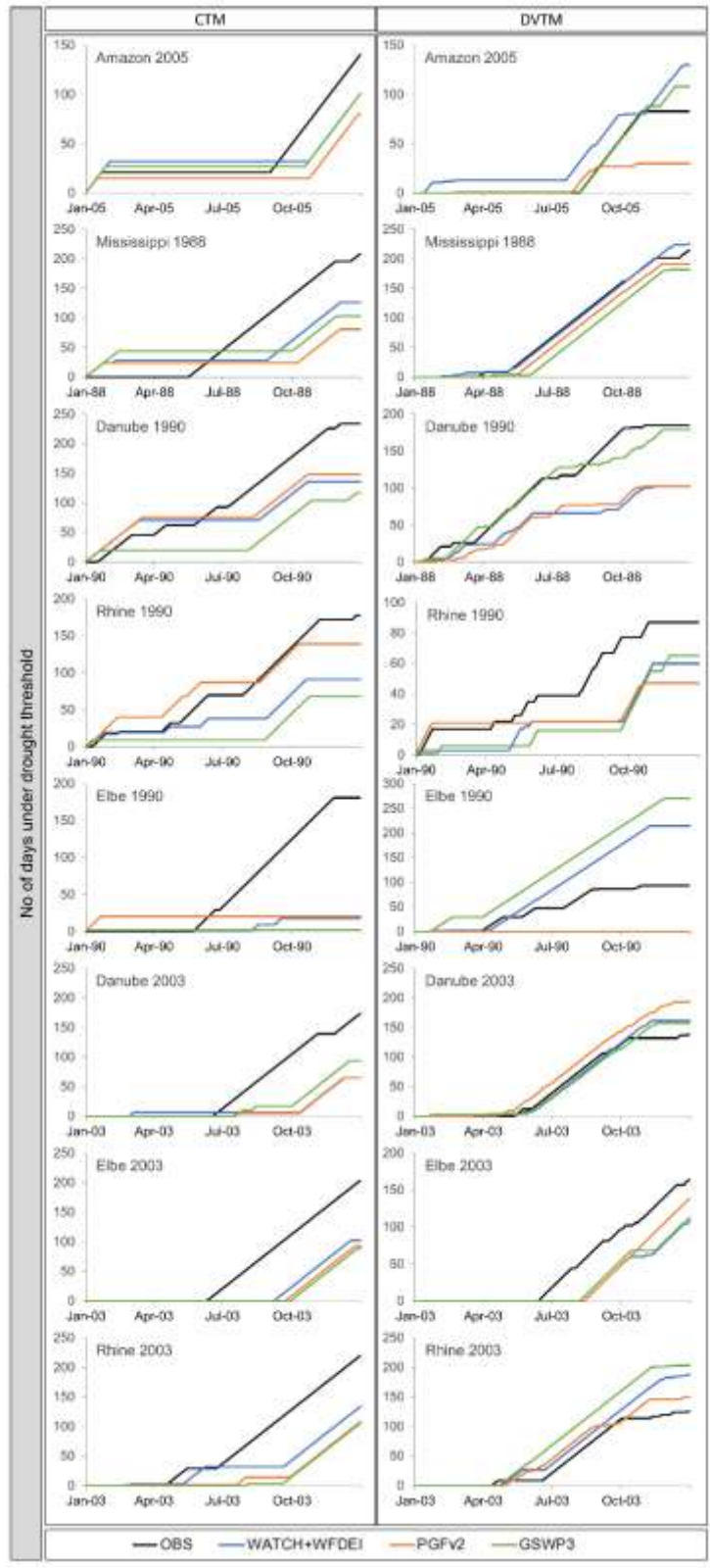


Figure 19. Temporal evolution of drought events at the basin level. Two drought identification methods (CTM-left column and DVTM-right column) are applied.

4. CONCLUSIONS OF PART C

This work focused on assessing the sensitivities of the JULES model to climate drivers and on the evaluation of the model's ability to represent mean and low hydrological states and drought events and led to the following conclusions:

- The JULES model can substantially represent simulate mean and low hydrological states. However, model performance varies significantly between the examined hydrological basins and for the different forcing datasets.
- Representation of past drought events is more challenging than the modelling of mean and low hydrological states. Model performance on this matter is also dependent on the adopted drought definition.
- The use of different forcing datasets was found to have a considerable effect on the hydrological outputs. Thus JULES exhibits high sensitivity to the climate drivers.

REFERENCES FOR PART C

- Cassardo, C., Mercalli, L. & Berro, D., 2007. Characteristics of the summer 2003 heat wave in Piedmont, Italy, and its effects on water resources. *Asia-Pacific Journal of Atmospheric ...* Available at: <http://www.dbpia.co.kr/Journal/ArticleDetail/773247> [Accessed December 1, 2015].
- Chen, J.L. et al., 2009. 2005 drought event in the Amazon River basin as measured by GRACE and estimated by climate models. *Journal of Geophysical Research: Solid Earth*, 114(5), pp.1–9.
- Demuth, S., 2009. Learning to live with drought in Europe. *World Sci (Nat Sci Q Newslett)*. Available at: https://scholar.google.gr/scholar?hl=en&q=+Learning+to+live+with+drought+in+Europe&btnG=&as_sdt=1%2C5&as_sdt=#0 [Accessed December 1, 2015].
- European Drought Centre, 2013. Drought of 1989/1990. Available at: http://www.geo.uio.no/edc/droughtdb/edr/DroughtEvents/1989_Event.php [Accessed December 1, 2015].
- European Environment Agency, 2010. Mapping the Impacts of Natural Hazards and Technological Accidents in Europe: An Overview of the Last Decade. Available at: https://scholar.google.gr/scholar?q=Mapping+the+impacts+of+natural+hazards+and+technological+accidents+in+europe+%E2%80%94+An+overview+of+the+last+decade&btnG=&hl=en&as_sdt=0%2C5#0 [Accessed December 1, 2015].
- Fekete, B. & Vorosmarty, C.J., 2011. ISLSCP II UNH/GRDC Composite Monthly Runoff. *ISLSCP Initiative II Collection*, edited by: Hall, FG, Collatz, G., Meeson, B., Los, S., Brown de Colstoun, E., and Landis, D., Data set, available at: <http://daac.ornl.gov/>, from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, 10.
- Marengo, J.A. et al., 2008. The Drought of Amazonia in 2005. *Journal of Climate*, 21(3), pp.495–516. Available at: <http://journals.ametsoc.org/doi/abs/10.1175/2007JCLI1600.1>.
- Namias, J., 1991. Spring and Summer 1988 Drought over the Contiguous United-States - Causes and Prediction. *Journal of Climate*, 4(1), pp.54–65.
- Poumadere, M. et al., 2005. The 2003 heat wave in France: dangerous climate change here and now. *Risk analysis*. Available at: <http://onlinelibrary.wiley.com/doi/10.1111/j.1539-6924.2005.00694.x/full> [Accessed December 1, 2015].
- Prudhomme, C. et al., 2011. How Well Do Large-Scale Models Reproduce Regional Hydrological Extremes in Europe? *Journal of Hydrometeorology*, 12(6), pp.1181–1204. Available at: <http://journals.ametsoc.org/doi/abs/10.1175/2011JHM1387.1> [Accessed January 23, 2014].
- Robine, J.-M. et al., 2008. Death toll exceeded 70,000 in Europe during the summer of 2003. *Comptes rendus biologiques*, 331(2), pp.171–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18241810> [Accessed December 1, 2015].

- Stahl, K., 2001. Hydrological drought-a study across Europe. Available at: <http://www.freidok.uni-freiburg.de/volltexte/2001/202/\npapers2://publication/uuid/961E50D8-7935-4643-91A8-E05C680A1754>.
- Sud, Y.C. et al., 2003. Simulating the Midwestern US drought of 1988 with a GCM. *Journal of Climate*, 16(23), pp.3946–3965. Available at: <Go to ISI>://WOS:000186684500008.
- Swiss, R., 2010. Natural Catastrophes and man-made disasters in 2009. *Sigma. Zurich: Swiss Re*. Available at: https://scholar.google.gr/scholar?q=Natural+catastrophes+and+man-made+disasters+in+2003&btnG=&hl=en&as_sdt=0%2C5#1#1 [Accessed December 1, 2015].
- Tallaksen, L.M., Stahl, K. & Wong, G., 2011. SPACE-TIME CHARACTERISTICS OF LARGE-SCALE DROUGHTS IN EUROPE DERIVED FROM STREAMFLOW OBSERVATIONS AND WATCH MULTI-MODEL SIMULATIONS. , (48).
- Trenberth, K.E. & Branstator, G.W., 1992. Issues in establishing causes of the 1988 drought over North America. *Am. Meteorol. Soc*.
- Zeng, N. et al., 2008. Causes and impacts of the 2005 Amazon drought. *Environmental Research Letters*, 3(1), p.014002. Available at: <http://stacks.iop.org/1748-9326/3/i=1/a=014002?key=crossref.c44c784f922d8737bbc15f6081e9d5c0>.

APPENDIX TO PART C

Mean runoff production [mm/month]						
Basin	OBS	WATCH+ WFDEI	PGFv2	GSWP3	Range of simulations	Range of simulations as percent of average value
AMAZON	77.73	82.15	88.28	90.96	8.81	10.11%
CONGO	50.17	29.09	36.29	26.00	10.29	33.79%
MISSISSIPPI	11.23	11.80	14.63	13.79	2.83	21.12%
LENA	5.90	6.11	5.35	10.04	4.68	65.34%
NIGER	7.73	3.13	4.18	3.91	1.05	28.06%
VOLGA	6.56	9.50	14.44	18.10	8.59	61.33%
MURRAY	2.02	1.27	1.42	0.82	0.61	52.06%
GANGES	21.65	32.83	22.37	31.50	10.46	36.18%
ORANGE	0.30	0.18	0.30	0.15	0.15	69.04%
DANUBE	18.82	17.53	16.52	14.93	2.60	15.95%
SASKATCHEWAN	3.33	1.93	3.14	2.10	1.21	50.50%
RHINE	46.73	44.22	41.34	46.23	4.89	11.12%
ELBE	7.28	11.41	10.50	14.93	4.43	36.06%
ODER	3.29	3.43	5.18	5.70	2.27	47.58%
GUADIANA	3.28	0.93	0.71	0.65	0.28	36.78%

KEMIJOKI	21.66	21.26	14.25	33.44	19.19	83.50%
----------	-------	-------	-------	-------	-------	--------

Mean Runoff	Difference from observations [Simulation – Observation] [mm/month]			Percent difference from observations		
Basin	WATCH+ WFDEI	PGFv2	GSWP3	WATCH+ WFDEI	PGFv2	GSWP3
AMAZON	4.42	10.55	13.23	5.69%	13.57%	17.02%
CONGO	-21.08	-13.88	-24.17	-42.02%	-27.67%	-48.18%
MISSISSIPPI	0.56	3.40	2.56	5.02%	30.23%	22.79%
LENA	0.21	-0.54	4.14	3.61%	-9.20%	70.23%
NIGER	-4.60	-3.55	-3.83	-59.52%	-45.96%	-49.48%
VOLGA	2.94	7.88	11.53	44.77%	119.99%	175.71%
MURRAY	-0.76	-0.60	-1.21	-37.40%	-29.62%	-59.69%
GANGES	11.17	0.72	9.85	51.61%	3.32%	45.48%
ORANGE	-0.12	-0.01	-0.15	-39.27%	-1.70%	-49.84%
DANUBE	-1.29	-2.30	-3.89	-6.86%	-12.24%	-20.69%
SASKATCHEWAN	-1.40	-0.19	-1.22	-41.98%	-5.70%	-36.78%
RHINE	-2.50	-5.39	-0.50	-5.35%	-11.53%	-1.07%
ELBE	4.13	3.22	7.65	56.76%	44.27%	105.11%
ODER	0.14	1.89	2.40	4.14%	57.50%	73.09%
GUADIANA	-2.35	-2.57	-2.63	-71.61%	-78.40%	-80.17%
KEMIJOKI	-0.40	-7.41	11.79	-1.84%	-34.20%	54.43%

Low runoff production [mm/month]						
Basin	OBS	WATCH+ WFDEI	PGFv2	GSWP3	Range of simulations	Range of simulations as percent of average value
AMAZON	23.09	12.55	15.71	13.07	3.16	22.94%
CONGO	7.82	2.27	3.18	0.81	2.37	113.72%
MISSISSIPPI	0.68	0.69	1.25	1.10	0.56	55.32%
LENA	0.14	0.00	0.00	0.01	0.01	287.41%
NIGER	0.24	0.03	0.02	0.02	0.01	33.78%
VOLGA	0.18	1.99	3.29	2.46	1.30	50.43%
MURRAY	0.04	0.04	0.01	0.01	0.03	159.08%
GANGES	0.60	0.64	0.23	0.52	0.41	89.58%
ORANGE	0.01	0.01	0.00	0.00	0.00	171.11%
DANUBE	2.69	3.98	3.72	3.12	0.87	24.07%
SASKATCHEWAN	0.10	0.06	0.27	0.20	0.21	116.29%
RHINE	7.50	7.56	9.47	9.22	1.91	21.86%
ELBE	0.21	1.65	2.27	1.76	0.62	32.55%
ODER	0.08	0.40	0.84	0.65	0.44	69.80%
GUADIANA	0.00	0.00	0.00	0.00	0.00	53.07%
KEMIJOKI	0.92	1.13	0.27	2.40	2.13	168.06%

Low Runoff	Difference from observations [Simulation – Observation] [mm/month]			Percent difference from observations		
Basin	WATCH+ WFDEI	PGFv2	GSWP3	WATCH+ WFDEI	PGFv2	GSWP3
AMAZON	-10.54	-7.38	-10.02	-45.65%	-31.97%	-43.41%
CONGO	-5.55	-4.64	-7.01	-70.97%	-59.38%	-89.69%
MISSISSIPPI	0.01	0.57	0.42	1.31%	83.88%	62.54%
LENA	-0.14	-0.14	-0.13	-99.77%	-100.00%	-94.80%
NIGER	-0.21	-0.22	-0.21	-88.89%	-92.15%	-90.00%
VOLGA	1.81	3.11	2.28	1021.69%	1755.15%	1286.41%
MURRAY	0.00	-0.03	-0.03	-8.35%	-79.36%	-78.37%
GANGES	0.04	-0.38	-0.08	6.17%	-62.44%	-13.97%
ORANGE	0.00	-0.01	0.00	-23.40%	-91.85%	-64.74%
DANUBE	1.29	1.03	0.43	48.09%	38.21%	15.83%
SASKATCHEWAN	-0.04	0.17	0.10	-37.57%	164.43%	94.25%
RHINE	0.06	1.97	1.72	0.83%	26.33%	22.87%
ELBE	1.44	2.06	1.55	678.28%	968.73%	729.81%
ODER	0.32	0.76	0.57	392.87%	933.54%	697.52%
GUADIANA	0.00	0.00	0.00	-97.86%	-96.48%	-97.86%
KEMIJOKI	0.20	-0.65	1.48	22.14%	-70.59%	159.91%

PART D. ABILITY OF BIOME MODELS TO REPRODUCE THE MAGNITUDE OF CARBON FLUX CHANGES FOR EXTREME CLIMATE ANOMALIES PROVIDED BY LA NINA AND EL NINO OCCURRENCES OVER THE LAST 40 YEARS

In task 4.3 the JULES and ORCHIDEE models further developed in HELIX, were evaluated together with the eight ISI-MIP2a biome models against observed long-term net carbon fluxes (i.e., Net Biome Productivity, NBP) over terrestrial ecosystems during the recent five decades (1971-2010). We evaluated modeled global NBP against 1) the updated global residual land sink (RLS) plus land use emissions (E_{LUC}) from the Global Carbon Project (GCP), presented as R+L in this study from (Le Quéré et al., 2015), and 2) the land CO₂ fluxes from two inversion systems: Jena CarboScope (Rödenbeck et al., 2003; 2005) and MACC 2015 (Chevallier et al., 2005, 2010; Chevallier, 2013). The ensemble mean modeled NBP (include seven models with land-use change) is higher than but within the uncertainty of R+L, while the simulated NBP trend over the last 30 years is lower than that from R+L and from the two inversion systems (Table 1). ISI-MIP2a biome models well capture the interannual variation of global net terrestrial ecosystem carbon fluxes, compared to R+L from GCP and land CO₂ fluxes from the two inversion systems (Table 1). Tropical NBP represents $31 \pm 16\%$ of global total NBP during the past decades, and the year-to-year variation of tropical NBP contributes most of the variation of global NBP. Spatially, models agree on a carbon sink in the same grid-cell (i.e., positive NBP) with a positive trend over $47 \times 10^6 \text{ km}^2$ and $21 \times 10^6 \text{ km}^2$ respectively (i.e., 42% and 18% of vegetated land respectively; Fig. 1a and 1b). They agree on a carbon source with a negative trend over less areas (over $1.4 \times 10^6 \text{ km}^2$ and $8.1 \times 10^6 \text{ km}^2$ respectively; Fig. 1a and 1b). The simulated global carbon sink and its positive trend reflect dominantly a larger global NPP increase than heterotrophic respiration (R_h) in the past decades, while carbon source and/or negative NBP trends simulated in some regions could be due to the climate change and human-induced land-use change.

Table 1. Mean NBP and NBP trend estimation from ISI-MIP2a models, R+L, inversions and TRENDY models, and the correlation of NBP interannual variability among them. Values for different periods are shown depending on the time span of datasets used to compare. The ensemble mean NBP and its trend from ISI-MIP2a models account for 7 models with land-use change (i.e., excluding CARAIB), while the correlation on interannual variability account for all 8 ISI-MIP2a models.

Period	ISI-MIP2a		R+L	Inversions	
	PGFv2	GSWP3	(Le Quéré et al., 2015)	Jena CarboScope	MACC2015
Mean NBP (Pg C yr ⁻¹)	1.32 ± 0.33	0.76 ± 0.36	1.44 ± 1.08	0.79 ± 1.16	
1981-2010	1.34 ± 0.35	0.78 ± 0.38	1.41 ± 1.08	0.94 ± 1.21	0.90 ± 1.05
1990-2009	1.51 ± 0.34	0.75 ± 0.41	1.49 ± 1.03	1.24 ± 1.14	1.80 ± 1.24
Trend (Pg C yr ⁻²)					
1971-2010	0.019 ± 0.010	0.010 ± 0.010	0.036		
1981-2010	0.044 ± 0.020	0.038 ± 0.013	0.052*	0.065*	0.085*
1990-2009	0.044 ± 0.022	0.060 ± 0.024	0.026	0.052	0.057

Correlation 1971-2010

on	vs. R+L	$0.61 \pm 0.08^*$	$0.65 \pm 0.07^*$
Interannual	1981-2010		
variability	vs. R+L	$0.59 \pm 0.10^*$	$0.64 \pm 0.08^*$
(unitless)	vs. Jena		
	CarboScope	$0.64 \pm 0.12^*$	$0.65 \pm 0.09^*$
	vs. MACC2015	$0.58 \pm 0.15^*$	$0.61 \pm 0.10^*$

* indicates the NBP trend is significant with $p < 0.05$.

Var. in ensemble mean NBP from ISI-MIP2a models indicates the ensemble mean year-to-year variation \pm the standard deviation among models.

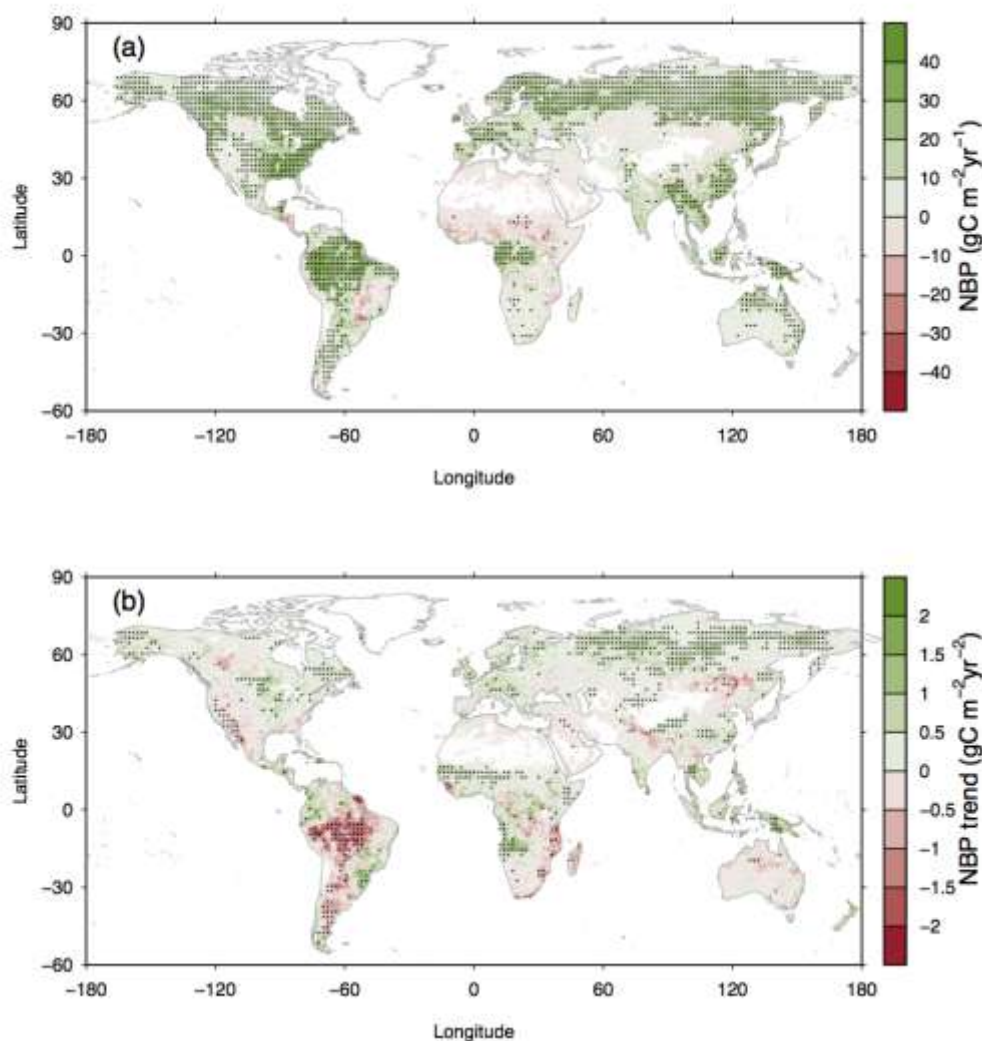


Figure 1. The spatial pattern of (a) ensemble mean NBP and (b) ensemble mean trends of NBP during the period 1971-2000. The stippled grids indicate the significant positive/negative average NBP, and over 85% (i.e., 12 out of 14) of model-forcing combinations with land-use change (excluding CARAIB) agree on the sign of the mean NBP. To present the stippled grids more clearly, each stipple is drawn at a resolution of $2^\circ \times 2^\circ$, where more than 8 out of 16 original grids ($0.5^\circ \times 0.5^\circ$ resolution) area stippled. The values over vegetated land (with annual mean NDVI larger than 0.1 over the period 1982-2010) are shown in the figure.

Significant global NBP anomalies from the long term mean between the two extreme phases of the El Niño Southern Oscillation (ENSO) is simulated by all models ($p < 0.05$), and also found in R+L ($p = 0.06$), which could be mainly attributed to NPP anomalies, rather than to Rh (Fig. 2). The global NPP and NBP anomaly during ENSO events is dominated by anomalies in tropical regions impacted by tropical climate variability (Fig. 2). Multiple regressions between R+L interannual variations and tropical climate variation reveal a significant negative response of R+L to tropical mean annual temperature variation, and a non-significant positive response of R+L to tropical annual total precipitation. However, the different responses derived from modeled NBP suggest some models may be under-sensitive to tropical temperature variation for the extreme range covered by strong El Niño to strong La Niña, but over-sensitive to tropical precipitation variation in their simulated NBP and NPP anomalies forced by ENSO.

This work is part of a publication by Chang et al. "Benchmarking of the ISI-MIP2 biome models" to be submitted to ISIMIP special issue of ERL (2016).

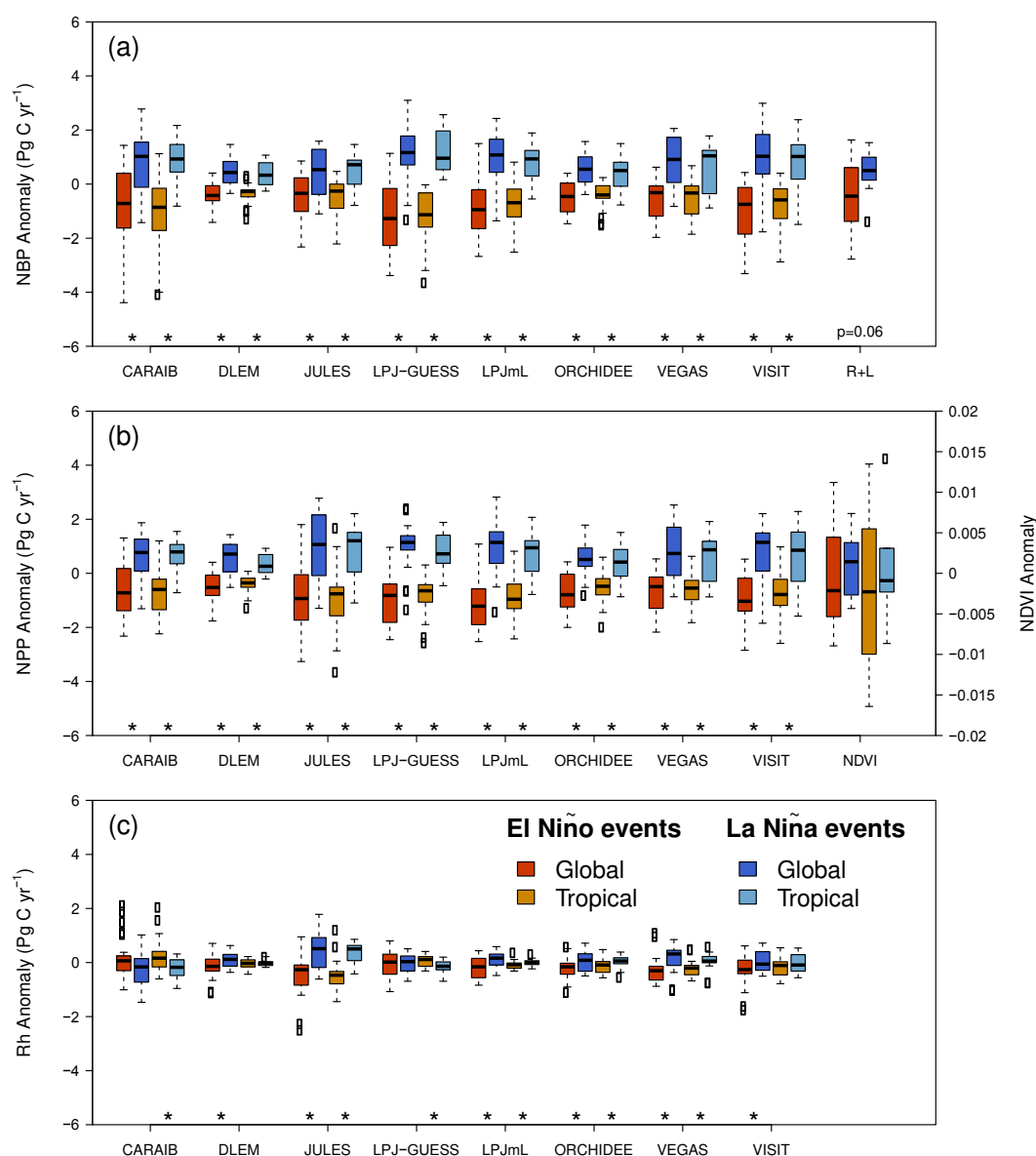


Figure 2. Comparison between detrended anomalies of (a) modeled NBP and the global Residual Land Sink (RLS) plus land use change emissions (E_{LUC}) (R+L; Le Quéré et al., 2015), (b) modeled NPP and the global NDVI, and (c) modeled Rh in ENSO events (i.e., El Niño years (Orange) and La Niña years (blue)). ENSO events are defined with MEI bimonthly ranks (Wolter and Timlin, 2011). The asterisks indicate the mean value of NBP/R+L, NPP/NDVI, or Rh between the El Niño events and La Niña events are significantly different (i.e., $p < 0.05$ with Student's T-statistic), while the p value is shown if the differences are not significant (i.e., $p \geq 0.05$).

REFERENCES FOR PART D

- Chevallier, F., Ciais, P., Conway, T.J., Aalto, T., Anderson, B.E., Bousquet, P., Brunke, E.G., Ciattaglia, L., Esaki, Y., Fröhlich, M. and Gomez, A. (2010) CO₂ surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, *J. Geophys. Res.*, 115, D21307, doi:10.1029/2010JD013887.
- Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F.-M., Chédin, A., and Ciais, P. (2005) Inferring CO₂ sources and sinks from satellite observations: Method and application to TOVS data, *J. Geophys. Res.*, D24309, doi:10.1029/2005JD006390.
- Chevallier, F. (2013) On the parallelization of atmospheric inversions of CO₂ surface fluxes within a variational framework, *Geosci. Model Dev.*, 6, 783-790, doi:10.5194/gmd-6-783-2013.
- Le Quéré, C., Moriarty, R., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Friedlingstein, P., Peters, G. P., Andres, R. J., Boden, T. A., Houghton, R. A., House, J. I., Keeling, R. F., Tans, P., Arneeth, A., Bakker, D. C. E., Barbero, L., Bopp, L., Chang, J., Chevallier, F., Chini, L. P., Ciais, P., Fader, M., Feely, R. A., Gkritzalis, T., Harris, I., Hauck, J., Ilyina, T., Jain, A. K., Kato, E., Kitidis, V., Klein Goldewijk, K., Koven, C., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lima, I. D., Metzl, N., Millero, F., Munro, D. R., Murata, A., Nabel, J. E. M. S., Nakaoka, S., Nojiri, Y., O'Brien, K., Olsen, A., Ono, T., Pérez, F. F., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Rödenbeck, C., Saito, S., Schuster, U., Schwinger, J., Séférian, R., Steinhoff, T., Stocker, B. D., Sutton, A. J., Takahashi, T., Tilbrook, B., van der Laan-Luijkx, I. T., van der Werf, G. R., van Heuven, S., Vandemark, D., Viovy, N., Wiltshire, A., Zaehle, S., and Zeng, N. (2015) Global Carbon Budget 2015, *Earth Syst. Sci. Data*, 7, 349-396, doi:10.5194/essd-7-349-2015.
- Rödenbeck, C., S. Houweling, M. Gloor, and Heimann, M. (2003) CO₂ flux history 1982-2001 inferred from atmospheric data using a global inversion of atmospheric transport, *Atmos. Chem. Phys.* 3, 1919-1964.
- Rödenbeck, C., T.J. Conway, and R.L. Langenfelds (2006) The effect of systematic measurement errors on atmospheric CO₂ inversions: A quantitative assessment, *Atmos. Chem. Phys.* 6, 149-161.
- Wolter, K., and Timlin, M. S. (2011) El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext). *Intl. J. Climatology*, 31, 14pp., 1074-1087.

PART E. THE EXPONENTIAL RATE OF CHANGE IN GLOBAL BIOMASS STOCKS: PLANT MORTALITY, EL NINO ANOMALIESh

Introduction

The purpose of this work is twofold: firstly, to assess the ability of the ISI-MIP2 biome models to reproduce historical trends of biomass in terrestrial ecosystems, and their sensitivity to climate anomalies. Secondly, to contribute to the task of identifying areas for potential improvement in the model representation of those biophysical processes exerting a greater influence in the dynamics of global biomass.

As a test set for comparison, we consider recent satellite passive microwave reconstructions of above-ground biomass (AGB) (Liu Y. et al., 2015). According to these, AGB stocks exhibit a net decrease of 0.07 PgC yr^{-1} , corresponding to the period 1993-2012. Such a decline is attributed to losses of tropical forest, which are partially compensated by biomass increases in mixed forests, in boreal regions. In order to isolate the biophysical component of historical changes in biomass, the contribution of areas under intense land-use management is excluded from the analysis.

In line with the objectives of this work, as starting point we represent a value of biomass B_t at a time t as

$$B_t = B_0 \Phi(t_0, t) \quad (1)$$

The information about the drivers of change is contained in Φ , which denotes a one-to-one mapping between a value B_0 , at $t_0 \leq t$ and the actual value B_t . As imposed by conservation of biomass in stocks, we require Φ to satisfy the balance equation $\frac{d\Phi_t}{dt} = \frac{NPP_t}{B_0} - \frac{\Phi_t}{\tau_t}$, where NPP_t denotes the net primary productivity and τ_t the residence time of carbon in vegetation. This condition leads to

$$\Phi = e^{Kt} \quad (2)$$

with

$$K_t = \int_{t_0}^t k_{t'} dt' \quad (3)$$

the integral of the exponential rate of change of biomass (ERB)

$$k_{t'} = \frac{1}{\tau_b(t)} - \frac{1}{\tau_t} \quad (4)$$

Here τ_b denotes the residence time value for which the balance condition $\frac{dB}{dt} = 0$ would be satisfied, in virtual absence of external disturbances, i.e. at $\tau_t = B_t / NPP_t$ (see derivation in Datasets and Methods Section).

The inverse of τ in (4) embodies information on the overall rate of loss of carbon in vegetation and on tree mortality in particular. As recently pointed out (Friend A. et al., 2014), the analysis of τ deserves special attention since this quantity constitutes a major source of uncertainty in multi-model projections of global biomass. According to formulation (1)-(4), changes in biomass can be explained in terms of the statistical properties of the ERB trajectories. In turn, the ERB values can be interpreted as fast-scale changes in the rate of carbon losses, resulting in a symmetry breaking between the input and output fluxes of carbon in vegetation. We inquire on the sensitivity of ERB to anomalies in key climatic variables, such as temperature and precipitation.

In addition, it is important to stress the implications of the sign of integral (3) for the uncertainty of biomass dynamics. According to Eqs. (1)-(3), a difference of values between model and observations entails an exponential deviation between the corresponding biomass trajectories.

In terms of AGB the total biomass can be expressed as $B = AGB[1 + f(AGB)]$, with f denoting the ratio of below- to above-ground biomass, $f \propto AGB^{\alpha-1}$ (Niklas K., 2005). The allometric exponent α is, in general, plant type specific. Yet, empirical evidence suggests that when aggregating above-/below- ground biomass data, corresponding to different plant communities, the resulting overall α exponent is close to unity (Chang D., 2007). Under such conditions, the ERB values derived from both AGB and total biomass will become comparable (see derivation of ERB in Dataset and Methods). Due to this fact, it is meaningful to test model-based values of global K_t against their AGB counterpart.

From a dynamics viewpoint, the similarity between model and observations is assessed by applying a dynamic time warping (DTW) metric (Salvador, S., & Chan, P., 2007) to the sequences of signs of the detrended global time-series of ERB. Finally, we address the relation of ERB to climate-induced disturbances by attending to the sensitivity of ERB to variations in El Niño South-Oscillation (ENSO) index, similarly as in (J. Chang et al.).

Datasets and Methods

Datasets

- Above ground biomass:** the AGB dataset used for this analysis (VOD-AGB, 2015) has been derived from satellite passive wave radiation time-series, for the period 1993-2012, as a basis to generate maps of vegetation optical depth (VOD). The latter estimates the water content in woody and leaf vegetation tissues and thus constitutes an indicator for changes in AGB. Advantages attributed to this indicator include: low saturation effects in dense canopies; sensitivity to both photosynthetic and non-photosynthetic biomass; robustness of results vis-a-vis the prevailing atmospheric conditions during observation (Liu Y. et al., 2012). The VOD to AGB conversion factors were obtained (Liu Y. et al., 2015) using collections of benchmark AGB maps (averaged between 1998 and 2002). The resulting VOD-AGB calibration was extrapolated to produce a set of AGB time-series, each one corresponding to a grid cell at the ($0.25^\circ \times 0.25^\circ$) spatial resolution.
- ISI-MIP2a biome models:** this data comprises the simulations (at a $0.5^\circ \times 0.5^\circ$ resolution) generated by the biome models CARAIB, JULES, LPJ-GUESS, LPJmL, DLEM, JULES, VISIT and ORCHIDEE, following the ISI-MIP2 protocol. These models were forced using 4 climate reconstruction dataset, namely, WFDEI, WATCH, PGFv2 and GSWP3. Simulations using WATCH were excluded as these are limited to the period 1971-2000, whereas in the remaining simulations this range extends up to 2010. From these simulations we selected the Cveg variable, corresponding to total biomass. In most of models, Cveg is provided at annual scale. For Cveg values simulated at monthly resolution, a year average was applied. For all the time-series, the year 1993 was considered as initial time for the analysis. In the case of CARAIB, the Cveg data is provided for each plant functional type (pft). Thus for this model the gridcell values of total biomass values were estimated by the sum of biomass per pft, weighted by their corresponding cell fractions.
- Multivariate ENSO Index (MEI):** MEI time-series (<http://www.esrl.noaa.gov/psd/enso/mei/>) have been derived on the basis of six observed climate variables (Wolter and Timlin, 2011). The MEI time-series was restricted to the 1993-2010 period.

Land-mask

A land mask was generated in order to exclude grid-cells with a land fraction of natural vegetation cover below 75%, as well as those grids where the reliability of VOD estimations is compromised by the presence of large open water bodies (Liu Y. et al., 2015). As source for the land-cover fractions we used the HYDE3.1/MIRCA maps. From these, we selected the map corresponding to year 2005 (the last available year). The combined mask was applied to the AGB and ISI-MIP2 datasets.

Derivation of global ERB time-series

The discrete version of the biomass balance equation (with time steps $\Delta t=1\text{yr}$) can be expressed as

$$\Delta B_t = k_t B_t \Delta t \quad (5)$$

with $\Delta B_t = B_t - B_{t-1}$ and $k_t = \frac{NPP_t}{B_t} - \frac{1}{\tau_t}$. The solution to Eq.(5), with initial condition B_0 , reads

$B_{n\Delta t} = \frac{B_0}{\prod_{i=1}^n (1 - k_i \Delta t)}$. In the continuous limit ($\Delta t \rightarrow 0$) the product in this expression converges to a time-ordered exponential (Coddington, N. & Levinson A., 1955), or in our case Eqs. (2)-(3).

As a preliminary step, global time-series of AGB and modeled biomass were obtained by taking the integral over the un-masked grids of the maps. The ERB values were then derived by computing the $\frac{1}{B_t} \frac{\Delta B_t}{\Delta t}$ ratios for each of the global biomass time-series. Here, it is important to notice that if the allometric exponent α equals one, the scaling factors between above and below ground biomass cancel out in Eq. (5) and thus the ERB values obtained from either AGB or total biomass are equivalent.

Evaluation of similarities between time-series

Dynamic Time Warping: DTW is a standard machine-learning technique, and is considered a benchmark in a wide range of pattern similarity recognition and signal classification problems (Salvador, S., & Chan, P., 2007). Essentially, the DTW similarity detection method consists in estimating the number steps involved in order to achieve an optimal match between two given sequences. The advantage of DTW over Euclidian distance metrics lies on the robustness of the method to identify similarities between the underlying characteristic signatures of sequences, regardless the presence of time-shifts and sequence size miss-matches. In the present work a Python fast DTW implementation was applied (<https://pypi.python.org/pypi/fastdtw/0.3.0>).

Results

1. The ERB trends and similarities.

In the AGB observations we find a biomass decrease characterized by an exponent $K_t = -0.01$ (Figure 1a). In contrast, the ISI-MIP2 ensemble describes an overall increase with $K_t = 0.05$ (model ensemble mean). Among the models, VEGAS shows the smallest value $K_t = 0.01$. No ERB trend is observed for AGB. Significant ERB trends (in the sense of a p-value $p < .05$) are found for DLEM (WFDEI) $17 \times 10^{-5} \text{yr}^{-1}$ and ORCHIDEE (WFDEI, PGFv2) $13 \times 10^{-5} \text{yr}^{-1}$. Associated to these values, Figure 1b illustrates the corresponding divergence between model and observations at the level of biomass trajectories.

By applying DTW to the sign sequences of detrended ERB, the highest similarity between model and AGB (Figure 2a) is found for VEGAS, followed by the model ensemble mean. For most of the models the DTW-distances to AGB are sensitive to the forcings.

According to the DTW-distances between pairs of signals, the sign sequences of detrended ERB form three large clusters (Figure 2b): the first one mostly comprises the sequences of VISIT and ORCHIDEE; the second of VEGAS, JULES and DLEM and the third one of LPJmL and LPJ-GUESS. The highest similarities do not necessarily correspond to the sequences generated by a same model (when using different forcings).

2. The relation of ERB to ENSO events

The regression of ERB on MEI for AGB shows a value $R^2=.69$ (Figure 3a), whereas for the models $R^2 \leq 0.5$, where the highest value corresponds to the WFDEI and GSWP3 simulations of ORCHIDEE.

For AGB we find an ERB-MEI sensitivity (Figure 3b) of -0.48(%). For the models, significant sensitivities (with $p < .05$, Figure 3c) are found for LPJmL (WFDEI, GSWP3) (-.13(%)), VEGAS (WFDEI) (-.11%), ORCHIDEE (GSWP3 (-0.14%), WFDEI (-0.09%)) and CARAIB (WFDEI) (-0.17%). In addition, for $p < 0.06$, CARAIB GSWP3 (-0.2%). For AGB the intercept value is practically zero, while positive for all the models (Figure 3d). In general, the regression coefficients appear forcing dependent.

Discussion

The sign of the integral K_t (3) differs between AGB and models. It is a remarkable fact that, in addition of being positive, the K_t values of models are in general forcing dependent (Figure 1a) –as this entails that, in the long-run, a same model can give rise to divergent descriptions at the level of biomass trajectories (Figures 1b) if different climate data are used as forcing. Also noticeable is the fact that the greatest similarities among the sign sequences of detrended ERB do not necessarily correspond to those generated by the same model (Figure 2b).

The analysis of the regression of ERB on MEI suggests a great potential for model improvement by addressing the annual-scale processes underlying the loss of carbon in vegetation, and their relation to climate variability.

In contrast with the AGB case, for most of the models the ERB does not show a significant sensitivity to MEI. Moreover, as comparison of Figure 1a and Figure 3d shows, differences in the dynamics of biomass can be mostly attributed to differences in intercept values. For AGB the intercept value is practically zero, entailing that in absence of climate anomalies the global stock of AGB is near balance. In contrast, the positive intercept values in the models point to a systematic overestimation of the residence time –in the sense of $\tau_{t'} > \tau_B$. In addition, since for most of the models the sensitivity of ERB to MEI is weak, one may conjecture that the modeled residence time represents a quasi-constant fraction of the B/NPP ratio. This could be further investigated by attending to possible commonalities in the carbon loss representation among models belonging to a same similarity cluster (Figure 2b).

Conclusions

Satellite passive microwave reconstruction of global above-ground biomass shows an overall decrease during the last decades. This work has shown that the biophysical component of such a decline can be mainly attributed to the influence El Nino anomalies.

Important differences in biomass trends have been identified between model and observations. These differences have been addressed in terms of the exponential rate of change of biomass. According to this analysis, the model-observation differences in biomass can be attributed to an underestimation of the influence of climate in the model representation of carbon losses in vegetation, along with an overestimation of background values of residence time.

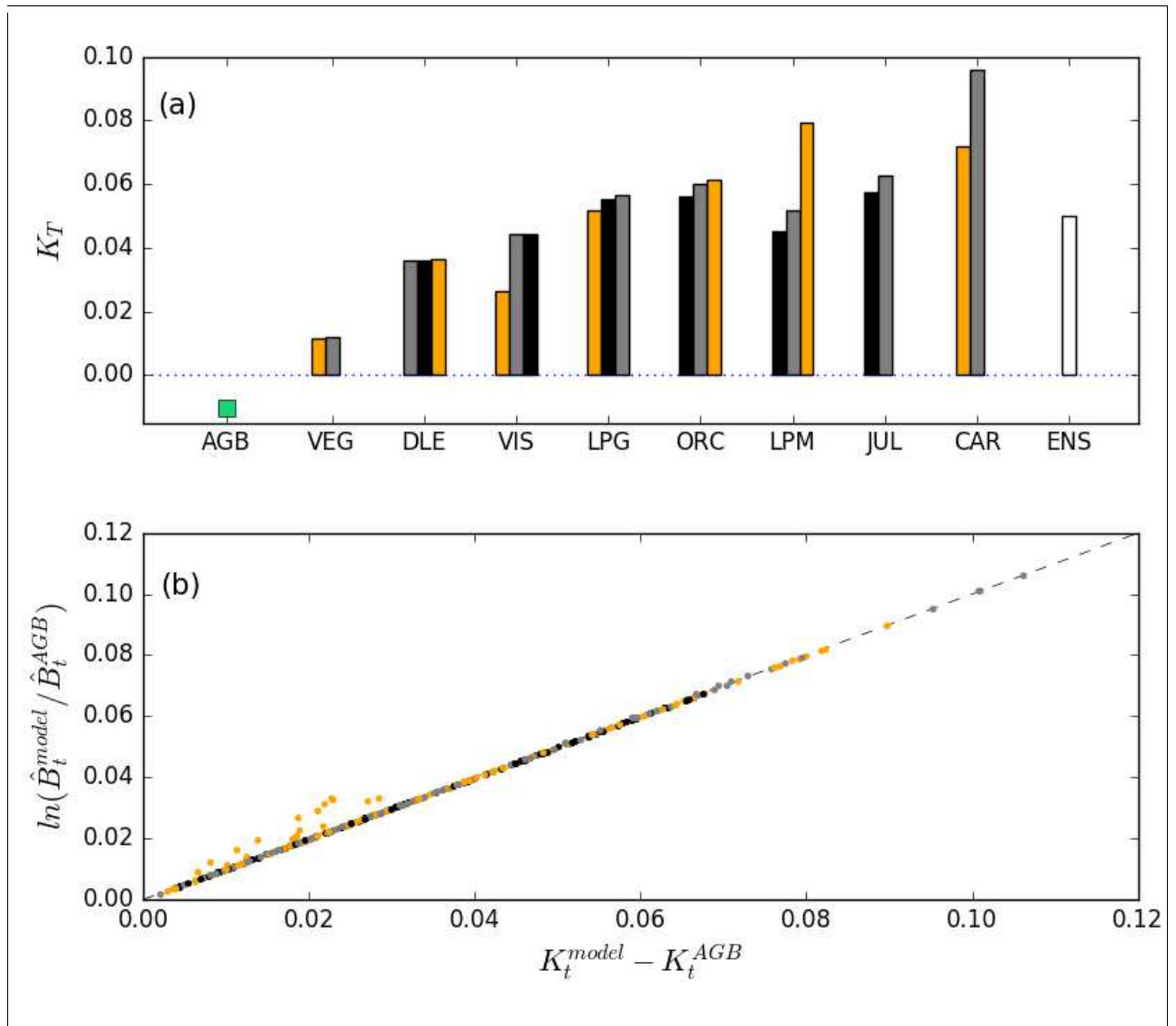


Figure 1. The integral of ERB and the uncertainties of global biomass change. For AGB (green), models (see color description below) and ensemble mean (white): a) the time integral, K_t , of ERB values; b) the divergence of model biomass and AGB trajectories (normalized to initial values) as a function of the differences of K_t values derived from AGB and models (dash line representing the identity). Models: VEGAS (VEG), DLEM (DLE), VISIT (VIS), LPJ-GUESS (LPG), ORCHIDEE (ORC), LPJmL (LPM), JULES (JUL), CARAIB (CAR). Climate forcings: PGFv2 (black), GSWP3 (gray) and WFDEI (orange).

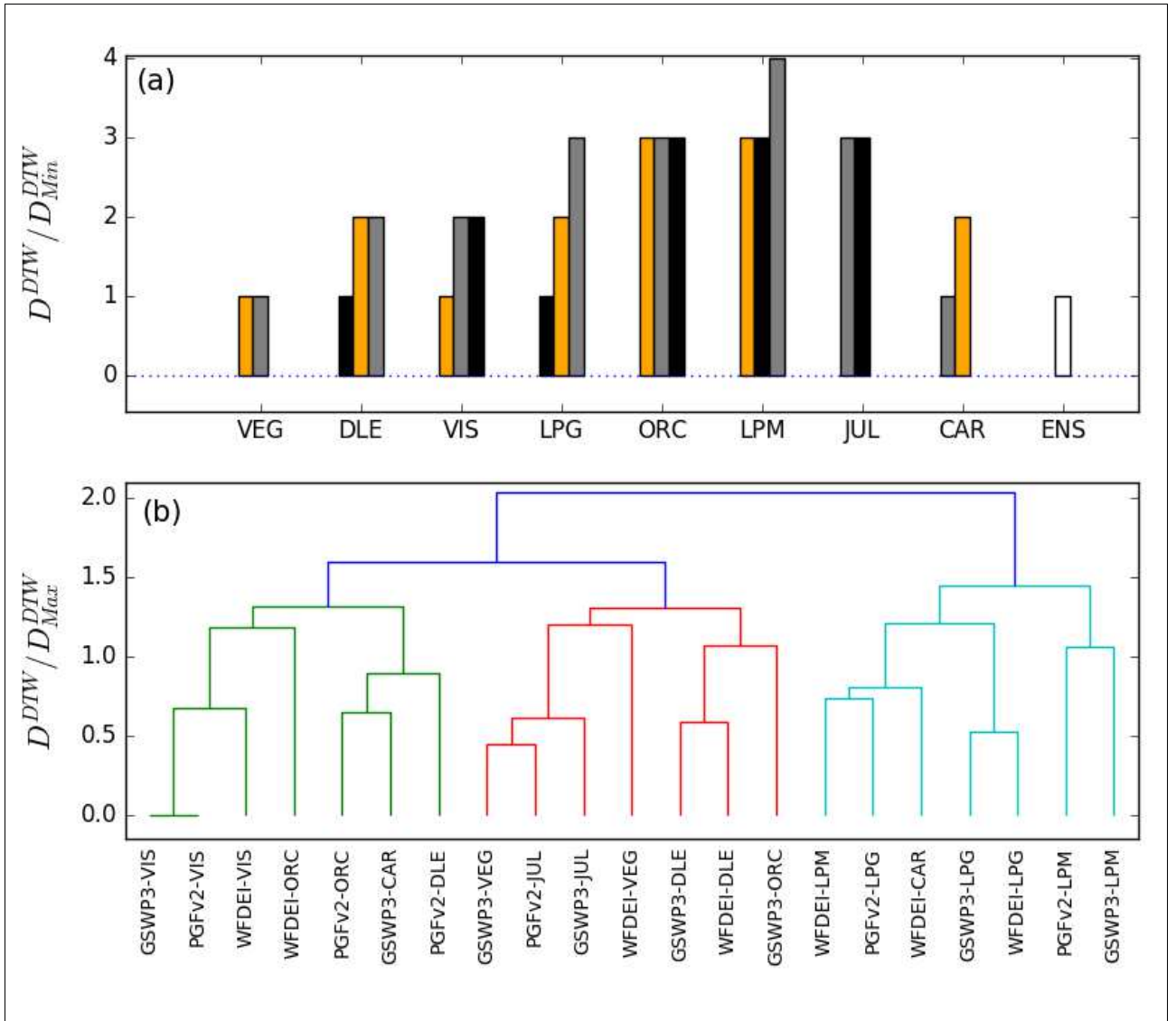


Figure 2. Similarities in ERB between models and observations. After applying DTW to the sequence of signs of the detrended ERB signals: (a) the model-AGB distances; (b) classification of the model sequences according to their mutual similarities.

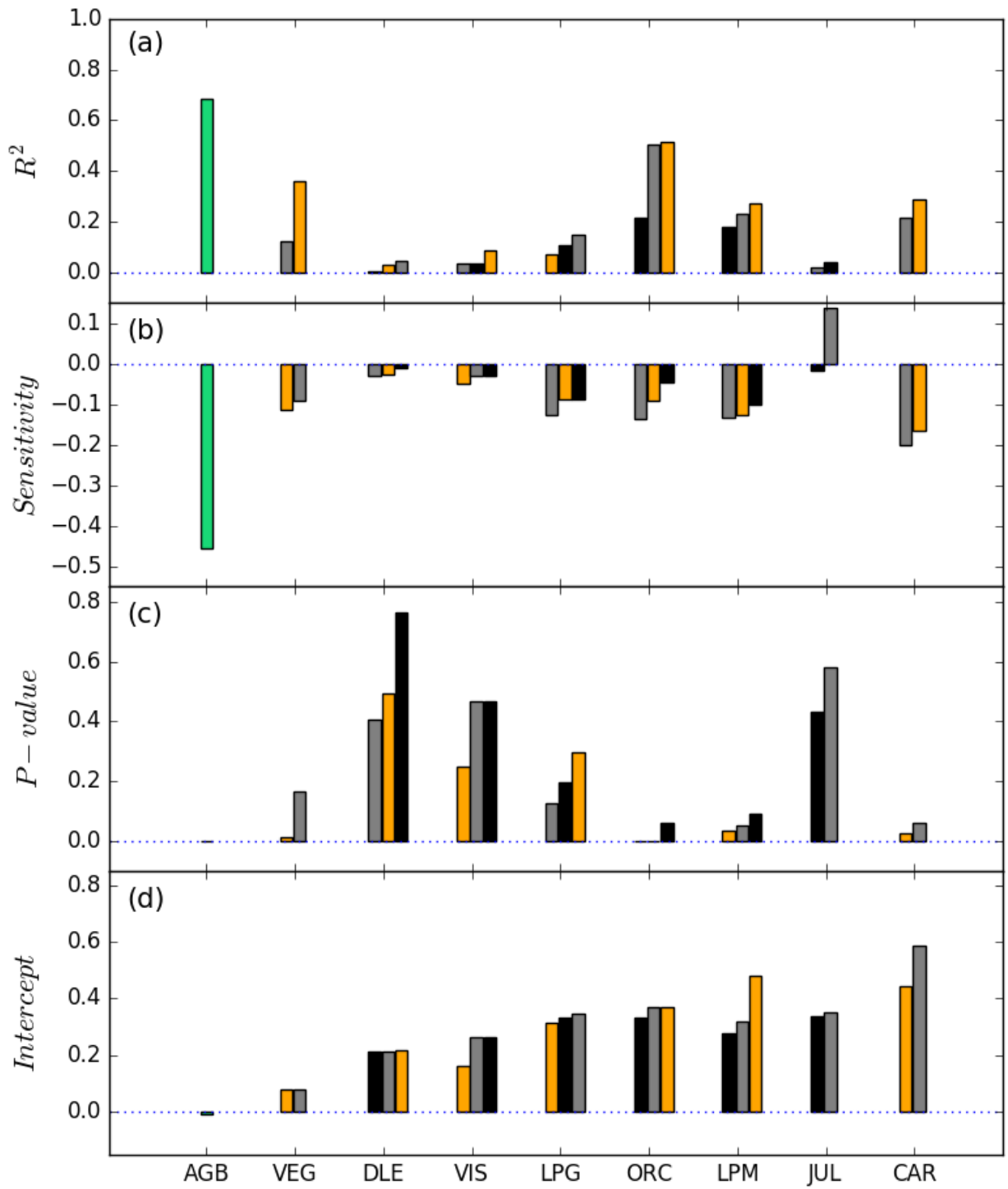


Figure 3. The relation of ERB to ENSO events. From (a)-(d), the coefficients of the linear regression of ERB on MEI (a). Color coding as in Figure 1.

References for Part E

- Liu, Y. Y., Van Dijk, A. I., De Jeu, R. A., Canadell, J. G., McCabe, M. F., Evans, J. P., & Wang, G. (2015). Recent reversal in loss of global terrestrial biomass. *Nature Climate Change*, 5(5), 470-474.
- Friend, A. D., Lucht, W., Rademacher, T. T., Keribin, R., Betts, R., Cadule, P., ... & Ito, A. (2014). Carbon residence time dominates uncertainty in terrestrial vegetation responses to future climate and atmospheric CO₂. *Proceedings of the National Academy of Sciences*, 111(9), 3280-3285.
- Niklas, K. J. (2005). Modelling below-and above-ground biomass for non-woody and woody plants. *Annals of Botany*, 95(2), 315-321.
- Cheng, D. L., & Niklas, K. J. (2007). Above-and below-ground biomass relationships across 1534 forested communities. *Annals of Botany*, 99(1), 95-102.
- Wolter, K., and M. S. Timlin, (2011). El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext). *Intl. J. Climatology*, 31, 14pp., 1074-1087.
- Coddington, N. & Levinson A., (1955). *Theory of Ordinary Differential Equations*. McGraw-Hill, New York.
- Salvador, S., & Chan, P. (2007). Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5), 561-580.

PART F. UNDERSTANDING THE WEATHER-SIGNAL IN NATIONAL CROP-YIELD VARIABILITY

Within Task4.3 of HELIX we evaluated Global Gridded Crop Models (GGCMs) contributing 1) historical simulations forced by observed weather and 2) future projections based on 5 Global Circulation Models (GCMs) to the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP). In a first step we assessed to what degree the historical simulations can reproduce reported national crop yield fluctuations from the Food and Agriculture Organization of the United Nations (FAO, 2013). As year-to-year variations in crop yields can have major impacts on the livelihoods of subsistence farmers and may trigger significant global price fluctuations with particularly severe consequences for people in developing countries it is important to identify the main drivers of these fluctuations which may help to introduce measures to dampen them. While real world yield fluctuations are not only forced by weather fluctuations but also by management changes, weeds, diseases, and pests, the considered process-based GGCMs primarily account for weather fluctuations. While the occurrence of weeds, diseases and pests is not represented at all fertiliser application rates were held constant (PEGASUS, pDSSAT, pAPSIM) or adjusted flexibly according to N stress (EPIC-IIASA, EPIC-BOKU, GEPIC) through time. Planting dates decisions were implemented based on daily weather conditions within a fixed (pDSSAT and LPJmL) or dynamic planting window (EPIC-BOKU, LPJ-GUESS, GEPIC, and PEGASUS). Harvesting dates were specified by the choice of cultivars (in terms of heat unit requirement to reach maturity) which was allowed to change through time according to changes in temperatures in PEGASUS, LPJ-GUESS, and to some extent in GEPIC (regarding the selection of winter and spring wheat). Given this more stylized representation of management the simulated time series are considered as proxies for weather fluctuations. Hence, we calculated the amount of the variance of the reported time series that can be explained by the simulations meaning estimating the fraction of the reported yield fluctuations that can be explained by weather given our process-understanding such as implemented in the GGCMs. Although an explicit quantification of weather-induced variability of crop yields and a deeper understanding of the underlying processes is essential for adaptation strategies it has so far been left to empirical models. Using state-of-the-art process-based crop model simulations and observational weather input we find that weather variations driving crop models can explain more than 50% of the variability in wheat yields in Australia, Canada, Spain, Hungary, Romania and Morocco. For maize we also find weather sensitivities exceeding 50% in six countries including the USA. For rice the explained variance exceeds 50% in Japan and South Korea and for soy in Argentina.

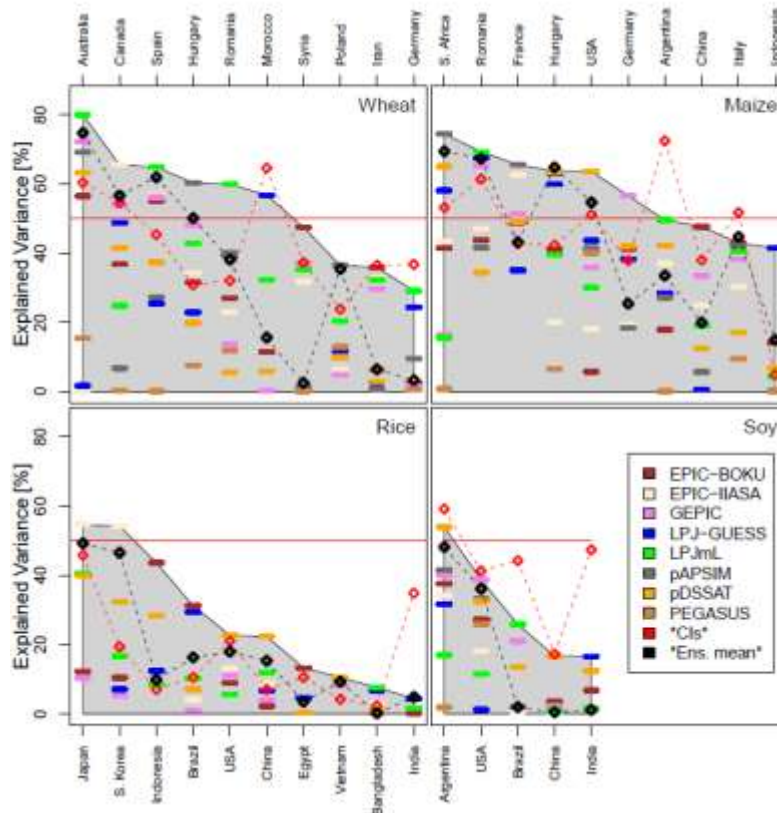


Figure 1: Quantification of lower bounds of weather-induced yield variability. Colored symbols in the lower part of each panel: Fraction of the variance of the reported year-to-year variability that is explained by individual crop models. Red diamonds: Highest variance explained by simple climate indicators (= CIs; see Fig S3 for the individual results of the three applied regression approaches). Black diamonds: Variance of reported yield fluctuations explained by the multi-model mean of simulated yields. Countries are ordered according to the highest variance explained by individual crop models. Grey polygons show the range from zero to the highest explained variance provided by the process-based crop models for each country.

For comparison to process-based crop yield simulations, we also provide results from four empirical approaches with one particularly addressing the issue of non-linear crop responses to temperature fluctuations by integrating temperatures below and above a certain threshold over time and allowing for individual responses (Lobell et al., 2011; Schlenker and Roberts, 2009). The empirical approaches are restricted to the detection of highly simplified relationships between yields and spatially and temporally aggregated weather indicators but could indirectly account for the effects of flood, storms, hail or weather-related occurrence of weeds, diseases or pests, in so far as they are correlated to weather indicators. The highest fraction of weather-induced variability derived from the four empirical models (Fig. 3, red diamonds) is generally lower than the highest explained variance provided by the process-based models except for soy. This is remarkable and underlines the good performance of the process-based models given the fact that their parameters have not been adjusted to represent reported annual yield fluctuations but rather represent a priori knowledge about the relevant physiological processes.

GGCMs offer alternative ways to identify the main drivers of the reported crop yield variability than simplified statistical models (Lobell et al., 2013). In particular, they allow for an artificial reduction of water stress by assuming optimal irrigation at each grid point. Here we calculate the explained variances assuming full irrigation for the model providing the highest variance in Fig. 4. The reduction in water stress reduces the explained variances to

20% or below for all countries where the assumed present day irrigation fraction is below 40% except for maize in the US and Indonesia (see Fig. 4).

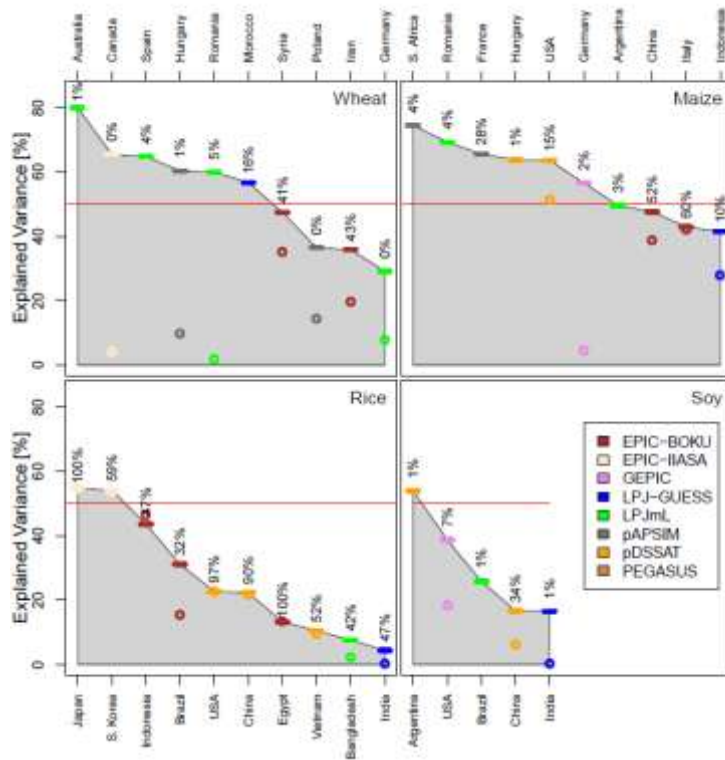


Figure 2: Similar to Fig. 4, but only showing the highest explained variances per country (colored bars identical to Fig 1) and the explained variances of the same model but assuming full irrigation (colored circles). Numbers indicate the “present day” irrigation fractions (MIRCA2000) underlying the “default” simulations represented by the colored bars. Explained variances with full irrigation are not shown if the underlying correlation is negative.

For wheat in Australia and Spain and maize in South Africa, Romania, France and Hungary there is no positive correlation any more. This supports the hypothesis that water stress is a main driver of the observed yield variations or does at least mediate the effects of temperature fluctuations in a way that can be strongly reduced by irrigation (Lobell et al., 2013).

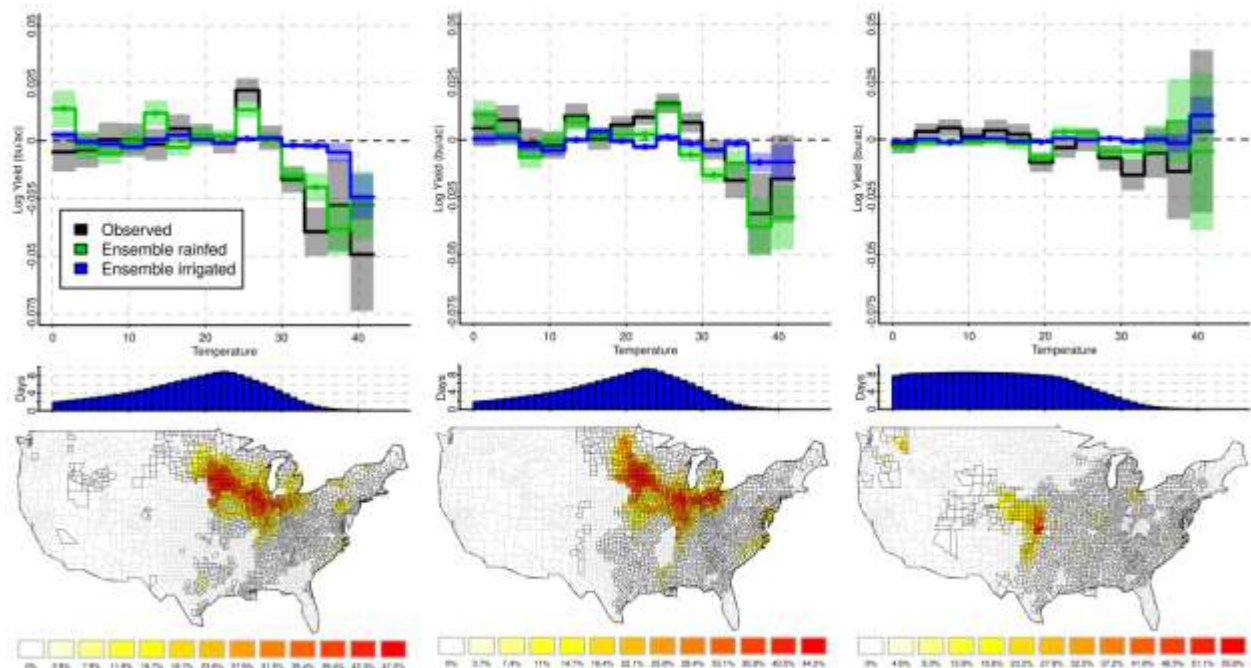
PART G. REPRESENTATION OF HEAT EFFECTS ON CROP YIELDS IN GLOBAL GRIDDED CROP MODELS

To evaluate the response of crop yields to weather fluctuations as implemented in the process-based GGCMs we compared it to the observed response derived from sub-national yield observation in the US. In a panel regression approach the response of crop yields to exposure to different temperatures was estimated for maize, soybean and wheat. The method is based on (Schlenker and Roberts, 2009) and analogously applied to historical crop yield simulations generated within ISIMIP2a. It is based on the following statistical model

$$\log Y_{it} = \alpha_0 + \sum_{h=0,3,6,\dots}^{39} \gamma_h [\theta_{it}(h+3) - \theta_{it}(h)] + \mathbf{z}_{it}\boldsymbol{\delta} + c_i + \varepsilon_{it}$$

where Y is yield, \log the natural logarithm, i the county and t the year. $\theta_{it}(h)$ is the cumulative distribution function of days during the growing season spent at temperature h , and the γ_h represent the estimated scaling coefficients

We show that the ensemble of nine crop models reproduces the observed average temperature responses. Each day above 30°C diminishes maize and soybean yields by up to 6% under rainfed conditions. Declines observed in irrigated areas, or simulated assuming full irrigation, are weak. This supports the hypothesis that water stress induced by high temperatures causes the decline. For wheat a negative response to high temperature is neither observed nor simulated under historical conditions, since critical temperatures are rarely exceeded during the growing season.



(a) Maize

(b) Soybean

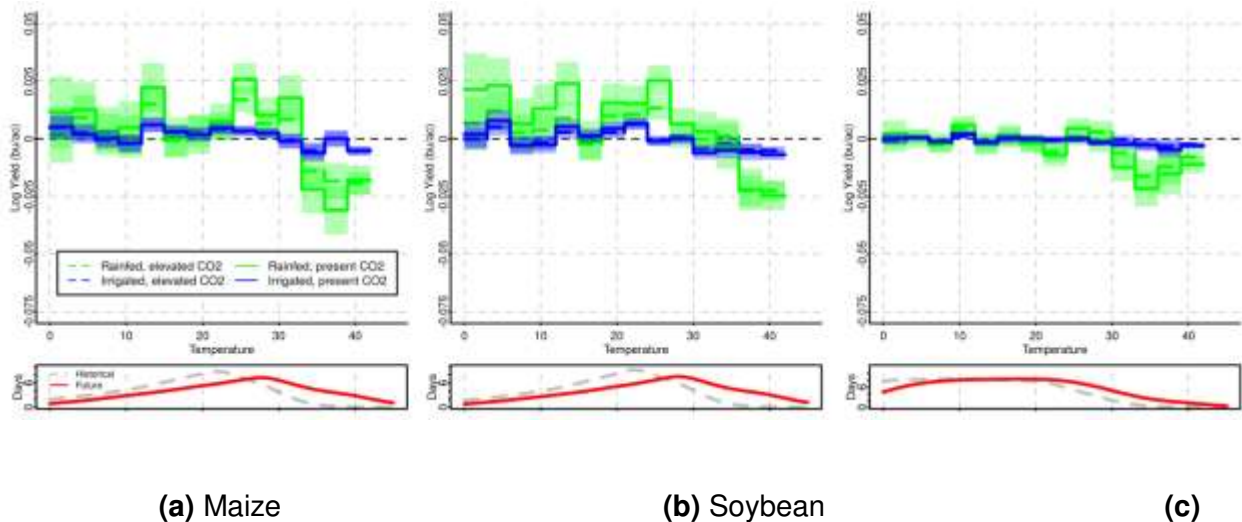
(c)

Wheat

Figure 6: Comparison of statistically estimated effects of temperatures on observed and simulated US yields. Panel (a) shows maize, panel (b) soybean and panel (c) wheat. The middle part of each panel shows the histogram of times spent in individual temperature bins as the sum of times derived

for each grid point across the growing seasons. The maps in the bottom show rainfed counties (black outlines) with their percent land-use share (colors) of the respective crop (for wheat only counties with predominantly winter wheat). Black lines in the top parts: Coefficients γ_h derived from log-transformed observed yields (eq. (1)). Green/blue lines: Coefficients of the ensemble median rainfed/irrigated simulated yields. Estimates are derived by a panel regression of US county data where the considered crop is grown under predominantly (> 90%) rainfed conditions. Shaded areas represent the 95% confidence intervals. Simulated coefficients are marked by colored dots if they are significantly different from the observed coefficients (confidence intervals do not overlap). The analysis is based on the assumption of a fixed growing season.

As rising levels of atmospheric CO₂ are expected to increase the water use efficiency of crops we additionally applied the above regression to the future yield projections generated in the ISIMIP Fast Track. We show that the implemented effect of CO₂ fertilization does not strongly reduce the temperature sensitivity of the simulated crop yields. In the future, yields are modelled to decline for all three crops at temperatures above 30°C. Elevated CO₂ can only weakly reduce these yield losses, in contrast to irrigation.



Wheat

Figure 7: Simulated yield responses to temperature under future climate change (2071-2099, RCP8.5) in rainfed counties (map in Figure 1). Panels are (a) maize, (b) soybean, (c) wheat. Green: Ensemble yield responses to temperature under rainfed conditions. Blue: Ensemble yield responses under irrigation. Solid lines: fixed present day [CO₂], dashed lines: elevated [CO₂] according to RCP8.5. Shaded areas are 95% confidence intervals. The bottom part of each plot displays the temperature histograms for the historic (dashed grey) and future (solid red) periods. SI Figures S25-27 contain analogous plots of the individual model simulations and model-specific growing seasons.