

Detection of global runoff changes:
results from observations and CMIP5 experiments

Alkama R.*, Marchand L.*, Ribes A.*, Decharme B.*

* CNRM/GAME, Météo-France, CNRS, Toulouse, France

Abstract

This paper assesses the detectability of changes in global streamflow. First, a statistical detection method is applied to observed (no missing data which represent 42% of global discharge) and reconstructed (gaps are filled in order to cover a larger area and about 60% of global discharge) streamflow. Observations show no change over the 1958-1992 period. Further, extension to 2004 over the same catchment areas using reconstructed data does not provide evidence of a significant change. Conversely, a significant change is found in reconstructed streamflow when a larger area is considered. These results suggest that changes in global streamflow are still unclear. Next, changes in streamflow as simulated by models from Coupled Model Intercomparison Project 5 (CMIP5) using the historical and future RCP 8.5 scenario are investigated. Most CMIP5 models are found to simulate the climatological streamflow reasonably well, except over South America and Africa. Change becomes significant between 2016 and 2040 for all but three models.

I. Introduction

Human influence has now been documented in several parts of the water cycle: atmospheric water vapour [e.g. Willett et al., 2007; Santer et al., 2007], land precipitation [e.g. Zhang et al., 2007], or land evapotranspiration [e.g. Douville et al., 2012]. The case of runoff or river discharge is more contrasted. While some studies [thought](#) to have identified robust trends over some specific regions [e.g. Stahl et al., 2010 over Europe, Krakauer and Fung, 2008 over the US], other studies focused on the global scale have led to somewhat contradictory results. Based on 221 rivers, corresponding to 40% of global continental runoff, Labat et al. [2004] documented an increasing global runoff at the end of the 20th century compared to the beginning. In contrast, based on data from 925 rivers, corresponding to 80% of global runoff, Dai et al. [2009] show a slight decrease in global runoff over the second half of 20th century. This discrepancy can be explained by the differences either in the number of [gauging](#) stations used, in the period of investigation, or in the method used to fill the gaps. Indeed, both studies were using some reconstructions (meaning gap filling) in order to provide a more comprehensive spatio-temporal coverage.

Over the last few years, several studies have attempted to explain the supposed observed global runoff trend, sometimes based on Land Surface Models (LSMs). Labat et al. [2004] were the first to relate the supposed positive runoff trend to global warming. They pinpointed a positive feedback between warming, an increase in ocean evaporation and an increase in continental precipitation. This assumption was then contradicted by many other studies. For example, using MOSES LSM, Gedney et al. [2006] explained this positive trend by the decrease in transpiration as a result of the stomatal closure due to rising atmospheric CO₂.

Using ORCHIDEE LSM, Piao et al. [2007] concluded that the land use and climate change are primarily responsible for the observed positive runoff trend. In the same way but using the LPJmL model, Gerten et al. [2008] find that the impact of stomatal closure and land use changes are very small and that the main factor explaining runoff change is precipitation change. The relative importance of the fertilization and stomatal closure effects and land use is still very model-dependent [Alkama et al., 2010]. In a recent study, Alkama et al. [2011] hypothesize that the observed surface warming and the associated decline of permafrost and glaciers, not yet included in most LSMs, could have contributed to increased runoff at high latitude. They also emphasize that runoff trend is a regional scale issue, if not basin dependent. Finally, the majority of recent studies conclude that there was no significant global runoff trend in the late twentieth century [Milliman et al., 2008; Dai et al., 2009; Alkama et al., 2011].

This paper first aims to provide a novel assessment of the significance of recent observed changes. This assessment is based on the Temporal Optimal Detection (TOD) method [Ribes et al., 2010]. While most previous studies consider global mean runoff, the TOD method is able to provide a single global diagnostic based on continental-scale mean runoffs. The TOD method is applied to both observed data only (meaning no missing data) and reconstructed data (i.e. a substantial fraction of streamflow time-series is missing and reconstructed by Dai et al. 2009).

With regard to future projections, an intensification of the hydrological cycle over the 21st century is widely assumed [e.g., Liu et al., 2012]. However, regional patterns of human-induced changes in surface hydroclimate are complex and less certain than those in temperature. Indeed, both increases or decreases may be expected in future precipitation and

runoff, depending on the region [Milly et al., 2005; Alkama et al., 2010]. Our study investigates the large-scale runoff change over the late 20th and 21st century (with atmospheric greenhouse gas and aerosol concentrations from the [Representative Concentration Pathways](#) RCP 8.5 scenario), as simulated by 14 CMIP5 (<http://cmip-pcmdi.llnl.gov/cmip5/>) Atmosphere Ocean General Circulation Models (AOGCMs). First, we assess the extent to which these simulated runoffs are consistent with observations. To this end, a comparison with observed streamflow is performed for the past few decades. Second, the same experiments are used to investigate how the anthropogenic perturbation (green house gases) may lead to different responses, depending on the model. Third, the same detection technique is applied to climate change scenarios in order to determine the significance of the simulated changes. The date at which the changes become significant is of particular interest, and provides some information with respect to the consistency or inconsistency between observed and simulated changes.

This paper primarily addresses the following three major issues:

1. How does global observed and reconstructed stream flow change over time?
2. Are simulated streamflow reasonably consistent with observations?
3. How will streamflow change in the future?

II. Methodology

II.1. Data

To the best of our knowledge, the most complete downstream discharge dataset in existence was collected by Dai et al. [2009]. This data set represents historical monthly streamflow at

126 the farthest downstream stations for the world's 925 largest ocean-reaching rivers from 1900
127 to 2004. However, the length and reliability of the available time series vary greatly from one
128 river basin to another, and gaps are usually found. Observed streamflows are subject to some
129 uncertainties, and in particular measurement uncertainty (e.g. related to the estimation of
130 rating curves), potential homogeneity breaks, and missing values. Measurement errors are
131 very difficult to address and no homogenized datasets are currently available, so the results
132 provided in this study are conditional to this dataset, following previous work [e.g. Dai et al.,
133 2009] that also investigated the recent trends in global streamflows. Gaps may be filled by
134 using statistical techniques, numerical simulations using land surface models (LSMs), or a
135 combination of both. In the present application, gaps were filled by applying a statistical
136 linear correction to the river discharge simulated by a LSM with observed atmospheric
137 forcings [see Dai et al., 2009]. Such a reconstruction, however, is likely to introduce
138 additional uncertainty. Results may depend on potential inaccuracy of the LSM used,
139 homogeneity breaks in the atmospheric forcing, uncertainties coming from the observations
140 (sometimes only a few years) used to calibrate the statistical correction, and others. As a
141 consequence, this study carefully distinguishes between two different treatments. First, we
142 analyse observed streamflows only, by considering time-series with no missing data. In this
143 way, the number of the selected rivers is reduced to 161 over 1958-1992 period. This period
144 was chosen in order to find an optimal compromise between spatial and temporal covering.
145 Note that even under this restrictive treatment, the period investigated is similar to the one
146 considered in Gedney et al. [2006] or Alkama et al. [2010]. Second, in order to consider the
147 larger spatio-temporal coverage available, we apply the same analysis to the dataset including
148 reconstructed streamflows. As linear regression cannot be used if there is too much missing
149 data, Dai et al. [2009] succeeded in reconstructing only 687 gauging stations for the whole
150 1958-2004 period. We consider these 687 catchment areas over this period. Finally, a third,

[product](#) is used in order to extend the “observations” up to 2004. We then consider the rivers observed over the 1958-1992 period and allow missing / reconstructed values over the 1993-2004 period. This extension does include reconstructions, but the amount of reconstructed values is much reduced compared to the previous case (i.e. 687 rivers).

River discharge, in addition to [being](#) potentially influenced by anthropogenic climate change, may be affected by direct human intervention, due to water resource management (dams), water withdrawal (e.g. irrigation, industrial or domestic uses), changes in land use that impact evapotranspiration, and others. In terms of climate change detection attribution, these direct influences may be regarded as “confounding factors”, as they may cause a substantial trend without any climate change. Detailed estimation of such direct perturbations is very challenging and no global discharge data base of “naturalized streamflows” is currently available. However, several studies addressed the issue of quantifying these direct anthropogenic influences at the global scale, and suggested that they had minor impact on multi-year trends.

Wisser et al. [2010] and [Sterling et al. \[2013\]](#), have quantified the impact of irrigation and reservoir operations over the 20th century. They concluded that “the [land use](#), expansion of irrigation and the construction of reservoirs has [considerably](#) and gradually impacted hydrological components in individual river basins. Variations in the volume of water entering the oceans annually, however, are governed primarily by variations in the climate signal alone with human activities playing a minor role”. The [later](#) is shown to hold at the continental scale (i.e. for individual oceanic basin, which corresponds to scales similar to the ones considered here, but with a different clustering). Other studies [e.g. McClland et al., 2004; Adam et al., 2007; Adam and Lettenmaier, 2008] confirmed that dams have altered the

seasonality of discharge, especially over [upstream](#) rivers, but are not responsible for changing annual values. Then, the impacts of land-use changes on land surface hydrology are still debated. On the one hand, when irrigation is neglected, land use can have an [important](#) influence on runoff via a decrease in surface evapotranspiration [Piao et al., 2007]. On the other hand, Lui et al. [2008] and Sun et al. [2008] indicate that deforestation over China, associated with irrigation, leads to increased evapotranspiration over the 20th century. Other studies, over individual river basins, suggested that the sign of land-use induced change was unclear (e.g. Twine et al. 2004 over the Mississippi river basin, VanShaar et al. 2002, over the Columbia River basin). Finally, some direct human influences via other activities have also been investigated and shown to have limited impact. For instance, McClland et al. [2004] demonstrated that increased forest fire frequency and severity may have contributed to changes in discharge, but can not be considered as a major driver.

We also used simulated runoff by different models from the Coupled Model Inter-comparison Project Phase 5 (CMIP5). Those runs supply three kinds of experiments: historical runs in which all external forcings come from observations, future runs which use greenhouse gas and aerosol emissions from the RCP 8.5 scenario, and finally piControl runs in which pre-industrial forcings are constant. The piControl runs are used to evaluate the internal climate variability. There are four RCPs types of possible future scenarios, and the RCP 8.5 involves the highest greenhouse gas concentrations at the end of the 21th century . It expects to [reach](#) the radiative forcing of 8,5 W/m² (~ 4 times more than the [current](#) value) at the end of 2100 which correspond approximately to 1370 ppmv of atmospheric CO₂ concentration. This scenario involves an intensive use of fossil fuels, with little mitigation stringency.

For this entire study, an ensemble of 8 zones where river basins are merged by continent and climate area was selected. The motivation for separating the northern cold climate from the tropics comes from Dai et al. [2009] and Alkama et al. [2011], which found significant runoff increase at high latitude that cannot be explained by the atmospheric forcing. While, the motivation of merging all of Africa's river basins in a single zone, even existing of large wide of climates, is coming from an ensemble of previous studies that shows that generally all largest African river basins had significantly decreased over the second half of 20th century [e.g. Alkama et al. 2011, Dai et al. 2009, Gedney et al. 2006; Labat et al. 2004]. The 8 selected zones are North America, Central America, South America, North Europe (including arctic basins), South Europe, North Asia (corresponding to Siberia), South Asia (including Oceania) and Africa (Figure 1).

Three steps are used before comparing the modelled runoff by different CMIP5 experiments and observations over the 8 regions: (1) Interpolate using bilinear method all of the CMIP5 modelled runoff into the same grid (0.5°x0.5°); (2) Compute the river basins at the 0.5°x0.5° grid. To be consistent with the data, river basins are defined at the known latitude and longitude of the observed gauged stations which are different to the river mouth; (3) Simulated runoff are then averaged over the computed river basins and merged into the 8 defined zones in the Figure 1.

II.2. Statistical method

II.2.a The Temporal Optimal Detection method

Detection is the process of demonstrating that an observed change is significantly different (in a statistical sense) that cannot be explained by natural internal variability. The statistical

method used for detection is the Temporal Optimal Detection method [Ribes et al., 2010]. We review the main concepts here but refer to Ribes et al. [2010] for full details about the method. The TOD method is based on a linear model:

$$Y(s,t) = a(s) + b(s) x(t) + \epsilon(s,t) \quad (1)$$

where $Y(s,t)$ denotes the observed streamflow at location s and time t , $a(s)$ is the climatological mean, $b(s)$ and $x(t)$ are respectively the spatial and temporal patterns of change, and $\epsilon(s,t)$ denotes the internal variability. TOD basically assumes that the temporal pattern of change is known while the spatial pattern is not. This is a substantial difference from other methods, such as optimal fingerprinting, in which the full spatio-temporal pattern of change is assumed to be known (up to a scaling factor, e.g. Hasselmann, [1993]). This assumption makes the TOD method particularly suitable here, because the spatial pattern of changes in global runoff is still under debate and somewhat model-dependent (see Section 3.2).

Regarding internal variability, the TOD method assumes that ϵ has a red noise structure (or autoregressive process of order 1, AR1, see e.g. Brockwell and Davis 1991). The red noise structure means that the random term $\epsilon(s,t)$ satisfies:

$$\epsilon(s,t) = \alpha \epsilon(s,t-1) + \eta(s,t), \quad (2)$$

where $\eta(s,t)$ is a white noise in time, i.e. $\eta(s,t)$ is independent from $\eta(s,t-1)$. This assumption also means that, e.g., the autocorrelation function decreases exponentially, with no long-range memory effect. An AR1 process is then described by a single parameter, α (see Eq.2), which is the one-year lag autocorrelation of ϵ . Note that in many statistical tests, residuals ϵ are assumed to be white noise (which corresponds here to $\alpha=0$), which makes the detection easier.

Given $Y(s,t)$, $x(t)$ and α , the inputs, the TOD method provides an estimate of the trend at each location $a(s)$. Based on this estimate, TOD performs a statistical test of the null-hypothesis

“ $a=0$ ”, and so returns a single P-value describing how significantly observations have changed.

II.2.b. Application to global runoff data

Here we discuss the choice of the parameters $x(t)$ and α , and the extent to which global discharges satisfy the assumptions behind the TOD method.

While assumed to be known, the temporal pattern of change $x(t)$ is commonly evaluated from simulations. In order to base our study on a very simple temporal pattern that is not model-dependent, we used only linear trends (i.e. $x(t)=t$). Note that the use of a linear trend instead of a potentially more complex smooth temporal pattern may be suboptimal. However, over short periods like the ones investigated here for observation (35 or 45 years) the non-linearity of the change is probably not the dominant feature (see also Fig 4). In addition to be very simple, this choice is consistent with several previous studies dealing with potential changes in the globally hydrology (e.g. Labat et al. 2004; Gedney et al. 2006; Dai et al. 2009, Alkama et al. 2010, 2011).

The choice of α , as well as the discussion on the accuracy of the red noise assumption, are here based on the analysis of pre-industrial control simulations. As observations are presumably influenced by external forcings, internal variability cannot be inferred directly from observations. Conversely, control simulations where external forcings are constant over time are expected to provide a physically-based description of internal variability, and using such simulations is quite common in detection and attribution studies (e.g. Hegerl and Zwiers 2011). Figure 2(a) shows the α value as estimated from the time-series of global runoff, for each CMIP5 control simulation. It is computed as the correlation between $y(t)$ and $y(t-1)$.

Although some discrepancies appear between different models, all values are between 0.04 and 0.3 with a medium value close to 0.2. These control simulations may also be used to check the accuracy of the red noise assumption. Figure 2(b) (c) and (d) illustrates the distribution of the P-value if the TOD test is applied to different 50-year segments from all control simulations using $\alpha=0$, 0.2 or 0.3 respectively. Such segments could be regarded as independent realizations under the null-distribution of the test. If internal variability is properly accounted for (in particular, if the red noise assumption is accurate), the P-values of the test applied to these segments should be distributed uniformly between 0 and 1. Figure 2(b) (c) and (d) suggests that the more suitable choice is $\alpha=0.2$ (distribution close to uniform), while $\alpha=0.3$ (resp. $\alpha=0$) is too conservative (permissive), leading to less (more) than expected values under the 5% threshold. In the following, we primarily discuss the results assuming $\alpha=0.2$ or $\alpha=0.3$ (which makes the detection more conservative and corresponds to the highest value found in individual models). The $\alpha>0.2$ may reflect an important ground water contribution to the stream flow. The results obtained with $\alpha=0$ (i.e. white noise) are also shown in some cases in order to provide a lower bound where no memory effect is accounted for. Note that the red noise assumption with $\alpha=0.2$ seems also consistent with the autocorrelation function of streamflow as simulated in control runs (not shown).

Finally, an important feature of the method with respect to our study is that it performs a multivariate diagnosis; i.e., it provide one single statistical diagnosis based on regional (continental-scale) streamflow. In particular, a change that generates increases or decreases in runoff depending on the region would be captured by this method. The TOD method as implemented here may then be regarded as a strategy for testing trend significance that allows the change to be spatially non-uniform and takes into account a non-white internal variability.

In particular, it differs from testing the significance of each regional trend individually, as a single test is performed for all regions simultaneously here.

III. Results

III.1. Statistical test on observed and reconstructed runoffs

The coverage of the 161 resp. 687 rivers worldwide is shown in Figure 1. Some regions (e.g. Western part of Asia, desert regions and Southern part of South America) suffer from lack of data. Indeed, only 31% (43%) of global land area excluding Antarctica are covered by the 161 (687) rivers basins which correspond to about 42% (60%) of global land discharge.

We first applied the TOD method to observed runoff over the 1958-1992 period, based on 8 zones in which observed streamflow at 161 downstream gauged stations are merged. Results are shown in Figure 3a in terms of P-value, for three values of the α coefficient. The P-value of year 1980, for instance, is obtained by applying the test to the data before 1980 i.e., the 1958-1980 period. This allows us to analyze the time evolution of the P-value. The P-value shown in 1992 provides the result of the statistical test over the full period of interest, 1958-1992. As may be expected, larger year-to-year variations are observed at the beginning of the period compared to the end, as one single year has stronger relative impact on the P-value (the size of the sample being smaller). The change in the annual runoff is detected at the year “t” only if the $P\text{-value}(t + i, \forall i \geq 0)$ is less than the significance threshold value and $P\text{-value}(t - 1)$ is greater.

Figure 3a (left) shows that the P-value remains higher than the significance threshold, 0.05, over the 1958-1992 period. This reveals no significant change in observed streamflow until 1992. This result is very robust here as it is obtained even under the white-noise (ie $\alpha=0$ which is very unlikely) assumption.

However, one can wonder what these results could have been over a more recent period. TOD is then applied over the same 161 river basins for the whole 1958-2004 period. As mentioned above, this extension requires to use a few reconstructed data over the 1993-2004 period, and cannot be regarded as “observations only”. After 1992, Figure 3a reveals that changes are still not detected for $\alpha=0.2$ or $\alpha=0.3$, as the P-value remains mainly higher than 0.1. It permits us to conclude that there is no significant change in observed global runoff on the observed 161 gauging stations from 1958 to 2004. Note that the P-value for $\alpha=0$ becomes lower than 0.05 but, as discussed before, this doesn’t allow us to reasonably claim that a change is detected. The relative anomaly (trend over the whole period compared to the runoff mean value) distribution in regional runoff shown in Figure 3a (right) reinforces this conclusion. Trends are rather small compared to the mean streamflow, except over Africa, where it reaches -30%. This result is confirmed by applying the TOD test over each individual region: a significant change in runoff is detected only over Africa since 1980 (Figure 4) which is consistent with the previously published results (e.g. Alkama et al. 2011) which pinpoint to a large decrease in precipitation and runoff over the 2nd half of 20th century.

In the same way, TOD is applied over the 687 river basins and over the same period (1958-2004). Figure 3b shows a significant change in reconstructed runoff since 2000 at the 95% significance level for $\alpha=0$ and, to a lesser extent, for $\alpha=0.2$. For $\alpha=0.2$, the P-value remains below but close to the significance level of 5%. For $\alpha=0.3$, no changes are found. Here, we

conclude that a change is detected, because detection does occur with a medium value of α . This result is not very robust, however, as detection no longer holds with a more conservative choice of α . Taking into account new rivers (687 rather than 161) and/or reconstructed rather than observed streamflow then seems to impact the results. However, the distribution and intensity of the relative discharge anomaly are not notably affected (Figure 3b). Africa is still the only region that exhibits a significant (negative) runoff trend.

III.2. Evolution of observed and simulated runoff

The evaluation of the CMIP5 simulated runoff was performed over the 8 zones and at the global scale corresponding to the 161 river basins. Figure 5 shows the temporal evolution of the yearly mean runoff (mm/d) from 1958 to 2100 for each of the 14 CMIP5 models. The temporal evolution of the observed runoff from 1958 to 1992 corresponding to the same zones is also shown. There are two outlying models, BCC and GISS, which show a large underestimation of both the global and the regional runoffs (Figure 6) as well as a low variability (Figure 7). At the global scale, all other CMIP5 models seem to simulate runoff in terms of mean state reasonably well (simulated runoff = observed runoff \pm 25%). They generally underestimate global runoff slightly, except for the MIROC model, which simulates a global runoff overestimated by about 15%. The runoff simulated over South America is underestimated by all models. The simulated runoff is also underestimated by the BCC, GISS MRI and INM models over Africa. In contrast, more than 50% overestimation is shown by all of NorESM, MIROC, IPSL, CSIRO, GFDL, CCSM, CanESM and FGOALS models. Over this continent, the runoff simulated by both CNRM-CERFACS and MPI is closer to the observations. Over all other regions, the error made by all models (except of course BCC and GISS) did not exceeds 50%, except FGOALS over South Europe and Central America, and

MRI over central America. The error in the standard deviations is also well simulated (Figure 7). Indeed, it did not exceed 50% except over Africa (North America) in where only CNRM, MPI, FGOALS, CanESM and GFDL (GFDM and INM) are reasonable. We can also note the large error of CSISRO (MPI) over South America (Centrale America).

Over the same period 1958-1992, the observations shows large 0.05 mm/yr^2 positive (negative) runoff trends over South America (Africa) but only MRI and MIROC (FGOALS and MIROC) shows reasonable trends (Figure 8). The rest of models exhibits slight or opposite trends. Over all regions, the models did not show any consensus. Some of them shows positive trends and others are negative. Compared to the observations, some of theme are reasonable and others are not (opposite trends). The previous exercises CMIP3 (Milly et al. 2005) and CMIP4 (Nohara et al. 2006) shows the same conclusion with the present study in terms of the comparison between simulated and observed runoff means, inter-annual variability and trends.

Despite the different bias existing in different model simulation, no bias correction methods are done in this study. It is known that bias correction method can alter the inter-annual variability of the discharge of small river basins and consequently could impact the significance of the trend signal. Whereas, the inter-annual variability of the large river basins and especially when regionally merged are relatively well simulated by GCMs (e.g. Milly et al. 2005; Nohara et al. 2006). Thus, the bias correction method had a small impact on the detection method used in this study.

Over the 21st century, all models show a positive global runoff trend except the INM model, in which the simulated global runoff decreases. At the regional scale, all models are in

agreement and show a positive trend over Northern Asia, Scandinavia, North America and South Asia. In contrast, the models simulate a negative trend over South Europe, except CNRM-CERFACS and IAP models, which are positive. For the other regions (South America and Africa), the models are in disagreement amongst themselves. For example, over South America, a negative trend is simulated by CCCMA and CSIRO, while MIROC, NCC, IPSL show a positive trend. Over Africa, the simulated runoff increases in CNRM-CERFACS, MIROC, MPMIP, NCAR, CCCMA, IPSL and NCC, decreases in CSIRO. Over Central America, models show no clear trend except GFDL (positive trend) and NCC and IAP (negative trend).

III.3. Statistical test on simulated runoff

In Figure 5, the TOD test is applied to each CMIP5 model over the 161 global river basins still merged into 8 zones, and over the whole 1958-2100 period. In order to highlight an **central** behavior, the median of these 14 P-values is computed each year, and its time evolution is illustrated. The P-values are also shown for the observed and the reconstructed data presented before. Note that the P-values of observed and simulated runoff are calculated over the 161 river basins whereas reconstructed data are computed over the 687 river basins. The P-values obtained from the CMIP5 runoff calculated over the 687 rivers are very similar to those computed over the 161 rivers. The results are only shown for $\alpha = 0.2$ (medium value). We define the date at which detection occurs as the first year for which the P-value remains lower than the 0.05 threshold up to 2100. The two models that simulate low runoff variability, BCC and GISS, are the first to detect a significant runoff change (in 2002 and 2005, respectively). The INM model, simulating no global significant trend, is the last to detect a significant change, in 2060. All other models detect a significant change between

2016 and 2040. This means that changes in runoff, as simulated by current climate models, are expected to become significant in the coming decades. As a consequence, the result previously obtained on observed runoff (161 river basins) appears to be very consistent with climate model projections.

The result obtained on reconstructed runoff (687 river basins) seems less consistent, as a change was found from 2000 onwards. In particular, Figure 9 suggests that the P-value computed from reconstructed runoff is on the border (if not outside) of the set of climate model projections. This feature, together with the substantial difference between the results obtained on observed and reconstructed data, may call into question the quality of reconstructions, and/or the accuracy of climate models projections. The difference found in detection date with 161 or 687 river basins could raise a lot of questions and the recent direct human influence on global trends could be one of the explanations. But this hypothesis is less supported because we did not detect change even using 161 basins over the same period with the 687 basins (i.e. 1958-2004). The 161 river basins represent 42% of global discharge and the 687 basins represents about 60% of global discharge. The main difference (60%-42%) of discharge are mainly coming from the northern Canadian and South Asian rivers. Without doubt, the human influence are far from being the main driver of the Northern Canadian rivers. Whereas, the large land use and irrigation changes over South Asia could be one possible explanation of this difference.

The P-value spread between the CMIP5 models is very large especially over 20th century, and reduces at the end of 21st century becoming significant (less than 0.05) over northern high latitude regions (Figure 4). While, over the other regions (South Asia, South and Centre

America), no change is detected by CMIP5 models till 2100. Figure 4 suggests that the northern high latitude regions represent the key of the future runoff change.

VI. Summary and conclusions

In this work, the TOD statistical test [Ribes et al., 2010] is used to evaluate the possible changes on recent and future (RCP 8.5 conditions) runoff, based on fourteen CMIP5 experiments and streamflow data from Dai et al. [2009]. This evaluation is made over 8 zones, merging the world's 161 largest rivers. Our analysis suggests some answers to the three issues raised in the introduction.

1. How does global observed and reconstructed streamflow change over time?

No significant runoff change is found in the observations over the whole set of 161 rivers from 1958 to 1992. Extension to 2004, using reconstructed streamflows over the same catchment areas, does not lead to a different conclusion. This confirms previous results by Dai et al. [2009] and Alkama et al. [2011]. In contrast, reconstructed data over 687 rivers shows significant change at the 95% confidence level over the 1958-2004 period, at least with a medium assumption regarding the internal variability persistence. This change is not robust to a more conservative choice regarding internal variability. This result seems rather contradictory with the conclusions by Dai et al. [2009] who found no significant trend on global rivers discharge based on the same data. This discrepancy much likely comes from differences in the statistical method used. While Dai et al. [2009] were only looking at the global mean time-series, our diagnosis is based on continental scale discharge, and can be

explained by opposite changes over different continents, that tend to compensate themselves and result in little change on the global mean.

Taken as a whole, these results suggest that changes in global runoff are still unclear. Indeed, positive detection is only obtained when considering a dataset where a substantial amount of data comes from reconstruction. It is not robust to a narrowing of the spatial domain, nor to a little change (considering an important slow process such as groundwater on global streamflow) in the description of the internal variability (i.e. using $\alpha=0.3$ instead of $\alpha=0.2$). The use of reconstructions also arises additional questions with respect to the accuracy of the reconstruction, which depends on the quality of the atmospheric forcing used, the capabilities of the LSM, the relevance of the statistical correction applied, and others. We finally conclude that changes in global discharge cannot be robustly identified from observations over the recent decades.

2. Are simulated streamflows reasonably consistent with observations?

Except for BCC and GISS, which show large underestimations of global runoff, the other CMIP5 simulations perform reasonably well. However, regional biases are far from being negligible, as the model bias can exceed 50% of the mean observed runoff over some regions. These biases are comparable to those found in the last CMIP3 exercise [Nohara et al. 2006, Milly et al. 2005].

3. How will streamflow change in the future?

The majority of CMIP5 models under RCP 8.5 conditions simulate an increase in runoff over South Asia, Northern Europe, Northern Asia and North America, and a decrease over Southern Europe. However, no significant change appears over Central America, and no consensus can be found over South America and Africa. These features are similar to what Milly et al. [2005], Nohara et al. [2006] and IPCC [2007] have already shown. More globally, all models show an intensification of the global hydrological cycle over the 21st century. Indeed, the global continental precipitation, evaporation and runoff tend to increase. Change in global runoff becomes significant between 2016 and 2040 for all but three models. This suggests that our finding of no clear change from the observations is rather consistent with current projections for the next century.

References

Adam JC, Haddeland I, Su F, Lettenmaier DP. (2007), Simulation of reservoir influences on annual and seasonal streamflow changes for the Lena, Yenisei and Ob rivers. *Journal of Geophysical Research*. 112:D24114. doi: 10.1029/2007JD008525.

Adam JC, Lettenmaier DP. (2008), Application of new precipitation and reconstructed streamflow products to streamflow trend attribution in northern Eurasia. *Journal of Climate*. 21(8):1807–1828. doi: 10.1175/2007JCLI1535.1.

Alkama, R., B. Decharme, H. Douville, and A. Ribes. (2011), Trends in global and basin-scale runoff over the late twentieth century: Methodological issues and sources of uncertainty, *Journal of Climate*, doi/abs/10.1175/2010JCLI3921.1

519 Alkama, R., M. Kageyama and G. Ramstein. (2010), Relative contributions of climate
 520 change, stomatal closure, and leaf area index changes to 20th and 21st century runoff change:
 521 A modelling approach using the Organizing Carbon and Hydrology in Dynamic Ecosystems
 522 (ORCHIDEE) land surface model, *J. Geophys. Res.*, 115, D17112,
 523 doi:10.1029/2009JD013408.
 524

525 [Brockwell, P. J., and R. A. Davis \(1991\), Time Series: Theory and Methods, Springer, 584pp.](#)
 526

527 Dai, A., T. Qian, and K. E. Trenberth. (2009), Changes in continental freshwater discharge
 528 from 1948 to 2004, *Journal of Climate*, 22, 2773-2791.
 529

530 Douville, H., B. Decharme, A. Ribes, R. Alkama, J. Sheffield. (2013), Anthropogenic
 531 influence on multi-decadal changes in reconstructed global evapotranspiration. *Nature*
 532 *Climate Change* (3), 59–62. doi:10.1038/nclimate1632.
 533

534 Gedney, N., P.M Cox, R.A Betts, O. Boucher, C. Huntingford, and P.A Stott. (2006),
 535 Detection of a direct carbon dioxide effect in continental river runoff records, *Nature*, Vol
 536 439-16. doi:10.1038/nature04504.
 537

538 Gerten, D., S. Rost, W. von Bloh, and W. Lucht. (2008), Causes of change in 20th century
 539 global river discharge, *Geophysical Research Letters*, 35, L20405,
 540 doi:10.1029/2008GL035258.
 541

542 Hasselmann, K. (1993), Optimal fingerprints for the detection of time-dependent climate
 543 change, *Journal of Climate*, 6, 1957-1971.

544

545 [Hegerl, G. and F.W. Zwiers \(2011\), Use of models in detection and attribution of climate](#)
546 [change. WIREs Clim Change, 2: 570–591. doi: 10.1002/wcc.121.](#)

547

548 IPCC (2007) Climate change 2007: the physical science basis. Contribution of working group
549 I to the forth assessment report of the intergovernmental panel on climate change, 504-511 pp.

550

551 Krakauer, N.Y. and I. Fung (2008), Mapping and attribution of changes in streamflow in the
552 coterminous United States. *Hydrology and Earth System Sciences*, 12, 1111-1120.

553

554 Labat, D., Y. G. Aeris, J. L. Probst, and J. L. Guyot. (2004), Evidence for global runoff
555 increase related to climate warming, *Advances in Water Resources*, 27, 631-642.

556

557 Lambert, F. H., P. A. Stott, M. R. Allen, and M. A. Palmer. (2004), Detection and attribution
558 of changes in 20th century land precipitation, *Geophysical Research Letters*, vol 31, L10203,
559 doi:10.1029/2004GL019545.

560

561 Liu, C., R. P. Allan and G. J. Huffman. (2012), Co-variation of temperature and precipitation
562 in CMIP5 models and satellite observations, *Geophysical Research Letters*, vol 39, L13803,
563 8 PP., doi:10.1029/2012GL052093.

564

565 Liu ML. et al., (2008), Effects of the and land-cover change on the evapotranspiration and
566 water yield in China during 1900-2000, *J. of the American Water Resources Association*
567 44, 5, 1193-1207 .

568

McClelland, J. W., R. M. Holmes, B. J. Peterson, and M. Stieglitz. (2004), Increasing river discharge in the Eurasian Arctic: Consideration of dams, permafrost thaw, and fires as potential agents of change. *J. Geophys. Res.* 109.D18102, doi:10.1029/2004JD004583.

Milliman, J.D., K.L. Farnsworth, P.D. Jones, K.H. Xu and L.C. Smith. (2008), Climatic and anthropogenic factors affecting river discharge to the global ocean, 1951-2000. *Global and Planetary Change*, 62,187-194.

Milly, P. C. D., K. A. Dunne and A. V. Vecchia. (2005), Global pattern of trends in stream-flow and water availability in a changing climate. *Nature*, 438-17 doi:10.1038/nature04312.

Nohara, D., A. Kitoh, M. Hosaka and T. Oki. (2006), Impact of Climate Change on River Discharge Projected by Multimodel Ensemble. *J. Hydrometeor*, 7, 1076–1089.

Piao, S., P. Friedlingstein, P. Ciais, N. de Noblet Ducoudre, D. Labat, and S. Zaehle. (2007), Changes in climate and land use have a larger direct impact than rising CO₂ on global river runoff trends, *PNAS*, 104(39): 15242–15247.

Ribes, A., J. M. Azaïs, and S. Planton. (2010), A method for regional climate change detection using smooth temporal patterns, *Climate Dynamics*, doi:10.1007/s00382-009-0670-0.

Santer, B.D., et al. (2007), Identification of human-induced changes in atmospheric moisture content. *Proceedings of the National Academy of Sciences of the USA*, 104, 15248-15253.

Stahl, K., et al. (2010), Streamflow trends in Europe: evidence from a dataset of near-natural catchments, *Hydrology and Earth System Sciences*, 14, 2367-2382, doi:10.5194/hess-14-2367-2010.

[Sterling S., A. Ducharne, J. Polcher \(2013\), The impact of global land-cover change on the terrestrial water cycle. *Nature climate change*. doi: 10.1038/nclimate1690.](#)

Sun G., et al., (2008), Watershed evapotranspiration increased due to changes in vegetation composition and structure under a subtropical climate. *J. of the American Water Resources Association* 44, 5, 1164-1175.

Twine TE., C. J. Kucharik, J. A. Foley (2004), Effects of land cover change on the energy and water balance of the Mississippi River basin, *J. of Hydrometeorology* 5, 4, 640-655.

VanShaar JR., I. Haddeland, DP. Lettenmaier (2002), Effects of land-cover changes on the hydrological response of interior Columbia River basin forested catchments. *Hydrological Processes*, 16, (13), 2499-2520.

Willett, K.M., N.P. Gillett, P.D. Jones, P.W. Thorne (2007), Attribution of observed surface humidity changes to human influence. *Nature*, 449, 710-712, doi:10.1038/nature06207.

Wisser, D., B. M. Fekete, C. J. Vörösmarty, A. H. Schumann (2010), Reconstructing 20th century global hydrography: a contribution to the Global Terrestrial Network- Hydrology (GTN-H), *Hydrology and Earth System Science*, 14, 1-24, doi:10.5194/hess-14-1-2010.

Zhang, X.B. et al (2007) Detection of human influence on twentieth-century precipitation trends. *Nature*, 448, 461-464, doi: 10.1038/nature06025.

Figure captions

Figure 1. Coverage of 161 (687) river basins up (down) over the 8 selected zones which are: 1 South America, 2 Africa, 3 South Asia including Oceania, 4 North Asia corresponding to Siberia, 5 South Europe, 6 North Europe including arctic basins, 7 Central America and 8 North America. The circles represent the in-situ gauged stations for each river accounted for in this study.

Figure 2. (a) Estimated alpha (α) based on the global runoff time-series from each CMIP5 model (piControl simulations); (b), (c) and (d) are the distribution of the P-value when TOD test is applied to different segments of 50year periods of all CMIP5 control runs using $\alpha=0$, 0.2 and 0.3 respectively.

Figure 3. (a) (left) Temporal evolution of observed (1968-1992) and reconstructed (1992-2004) runoff P-value over 161 river basins merged over 8 zones. The full horizontal black line represent the threshold level at 5%. (right) distribution of the runoff relative anomalies ($\Delta Q/Q$) in percentage over 161 river basins. (b) same as (a) but using reconstructed data over 687 river basins rather than observations over the whole 1968-2004 period.

Figure 4. Temporal P-value of observed (black) and simulated (light blue) runoff over the 8 regions merging 161 river basins using α at 0.2. The median of the 14 is in blue.

Figure 5. 1958 to 2100 global and regional time-series of the simulated (colors) and observed (black) annual runoff. The median of the fourteen models is given by the thick red line.

Figure 6. Percent error ($100(\text{Sim-Obs})/\text{Obs}$) of simulated runoff averaged over 1958-1992.

Figure 7. Percent error ($100(\text{Sim-Obs})/\text{Obs}$) of the standard deviation runoff simulated by CMIP5 models over 1958-1992.

Figure 8. Observed and simulated runoff trends over 1958-1992.

Figure 9. Temporal P-value of observed (black), reconstructed (red) and simulated (light blue) global runoff over 161 river basins using α at 0.2. The median of the 14 CMIP5 models is in blue.

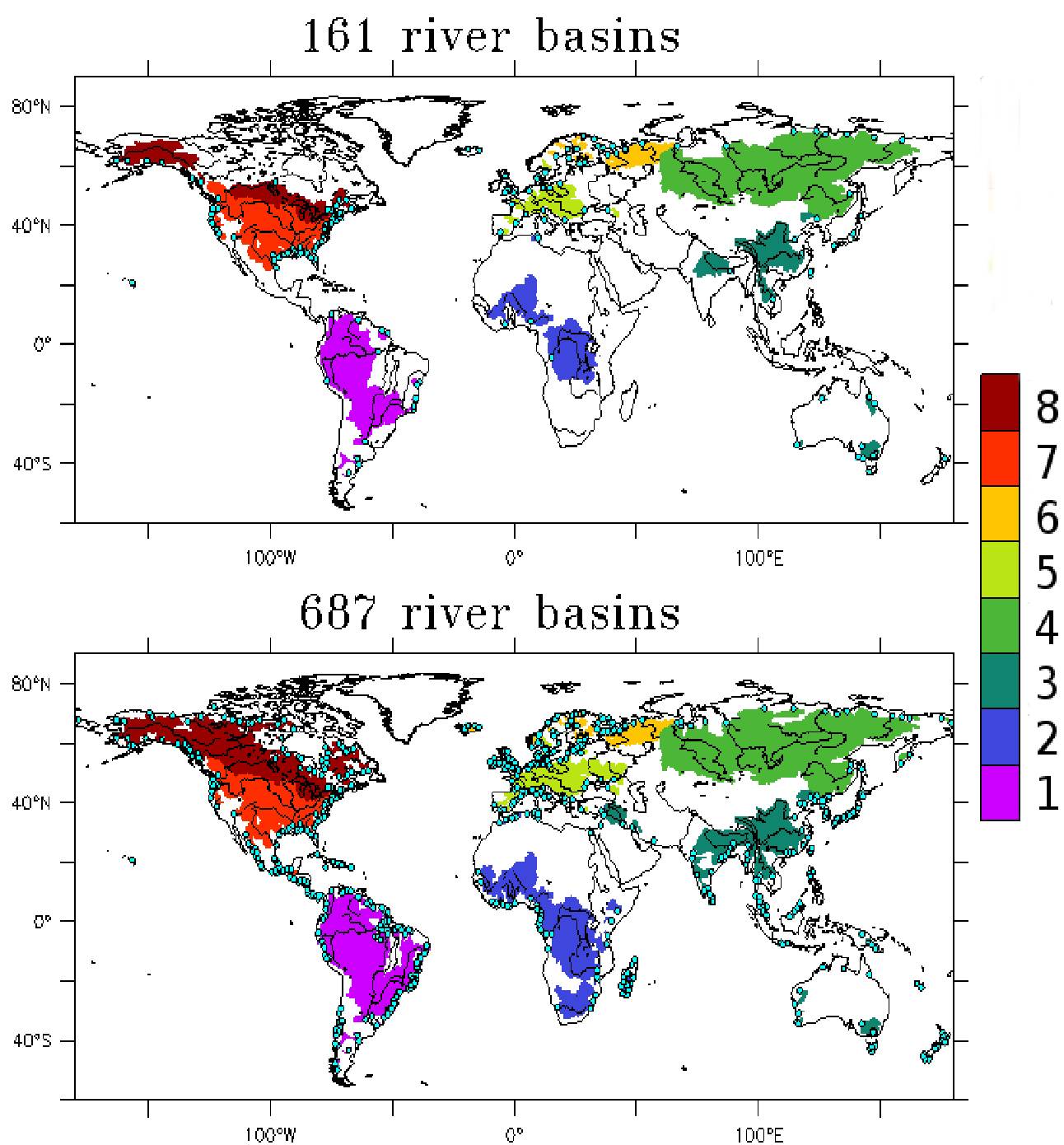


Fig. 1

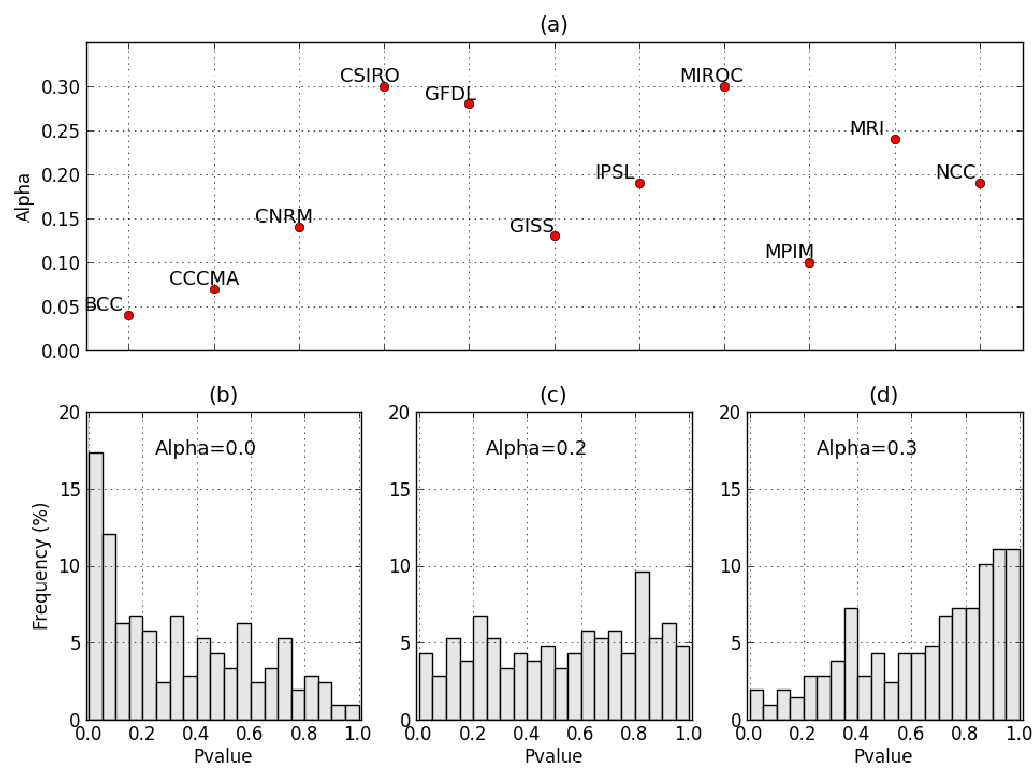
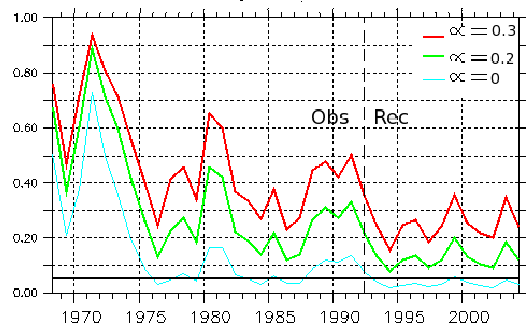
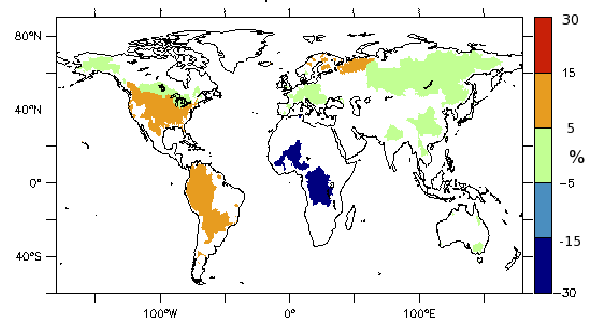


Fig. 2

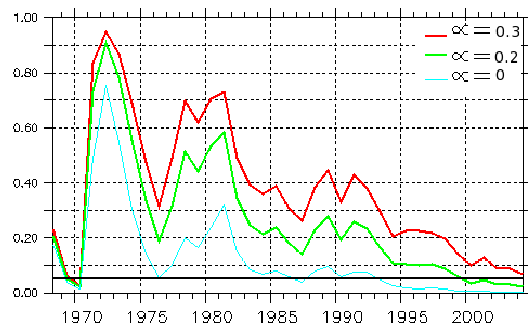
(a) P-value over 168 river basins



Trend ($\Delta Q/Q$) over 1958-2004



(b) P-value over 687 river basins



Trend ($\Delta Q/Q$) over 1958-2004

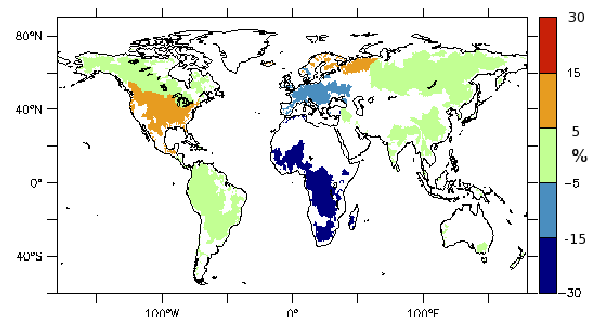


Fig. 3

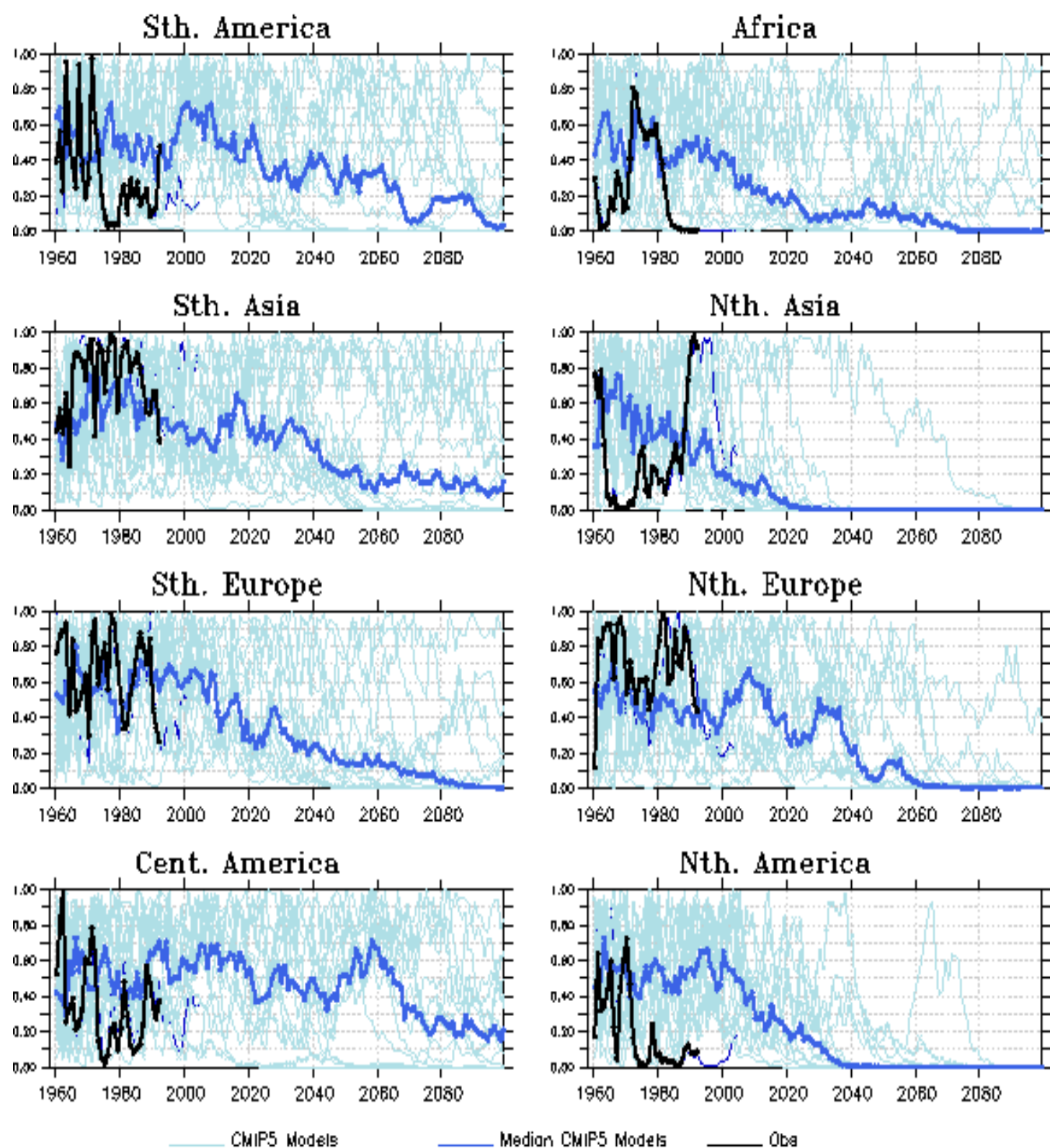


Fig. 4

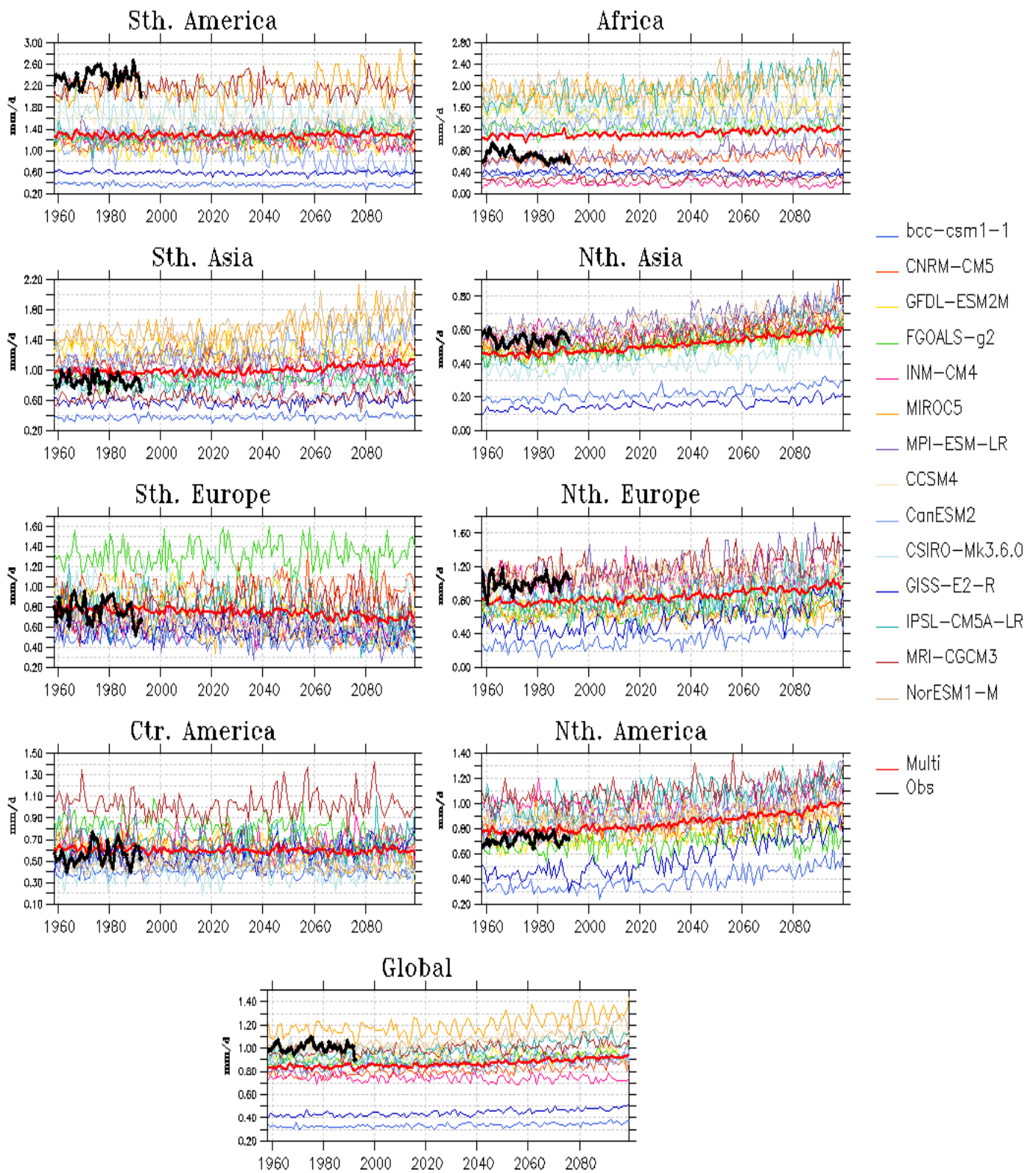


Fig. 5

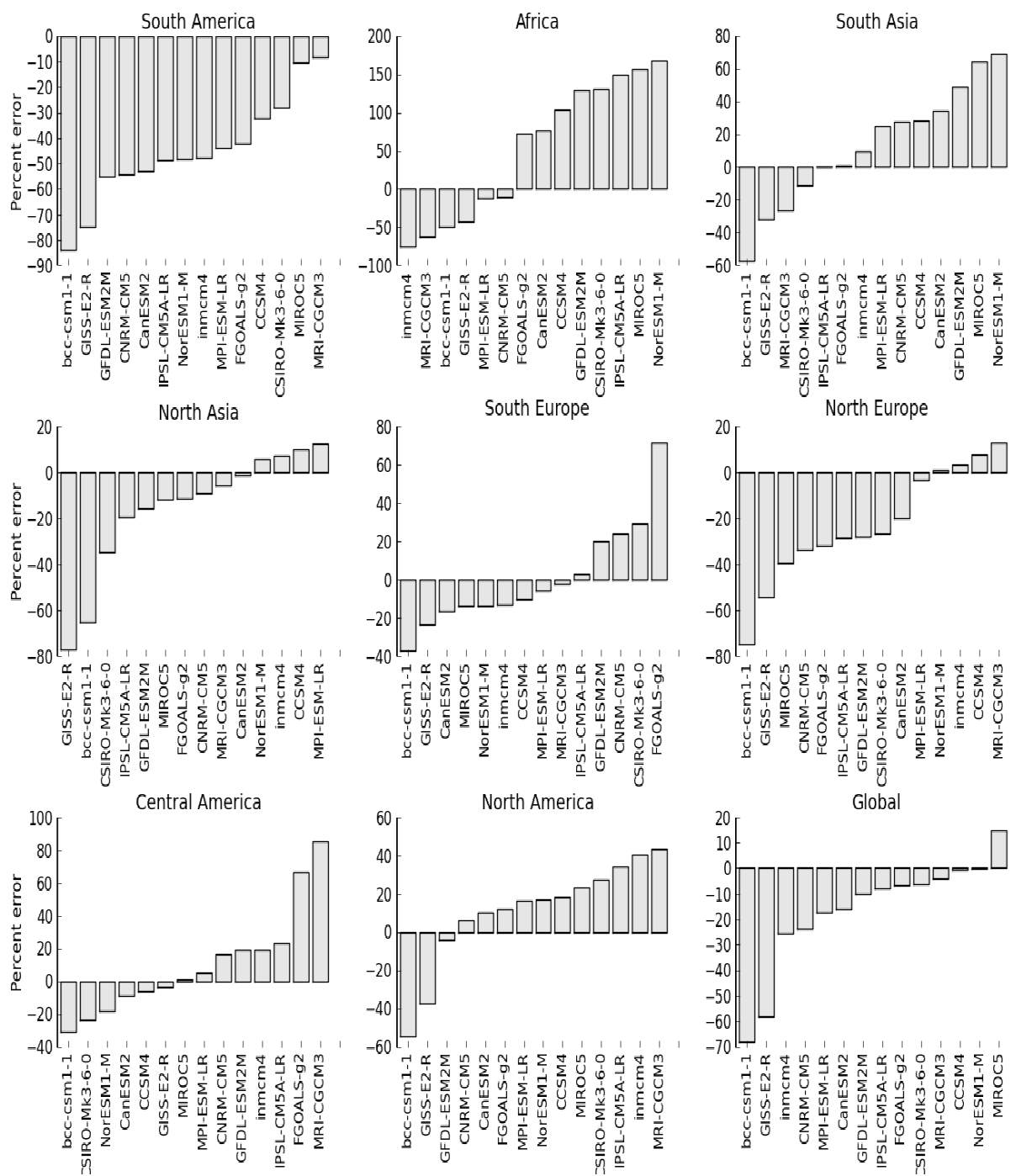


Fig. 6

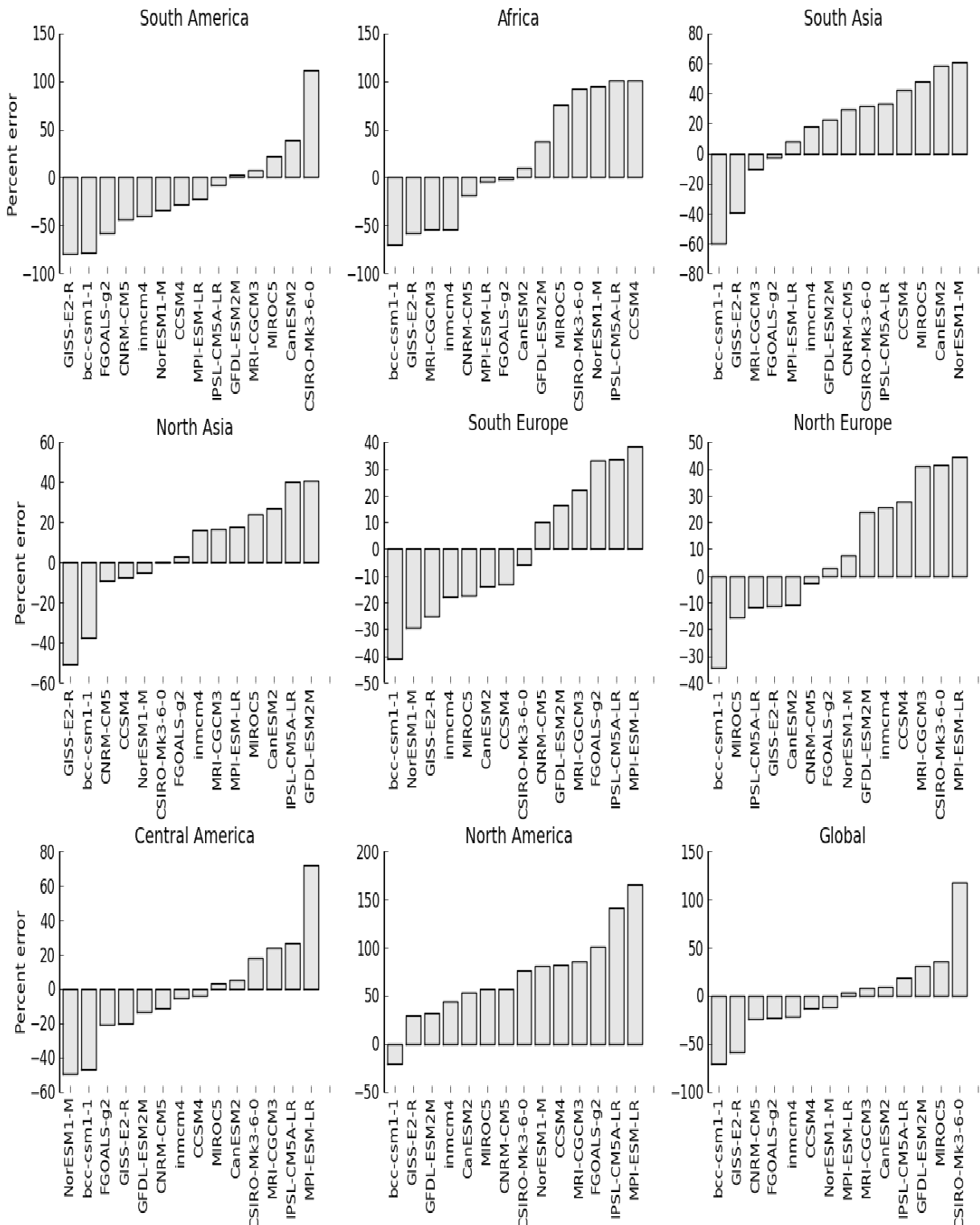


Fig. 7

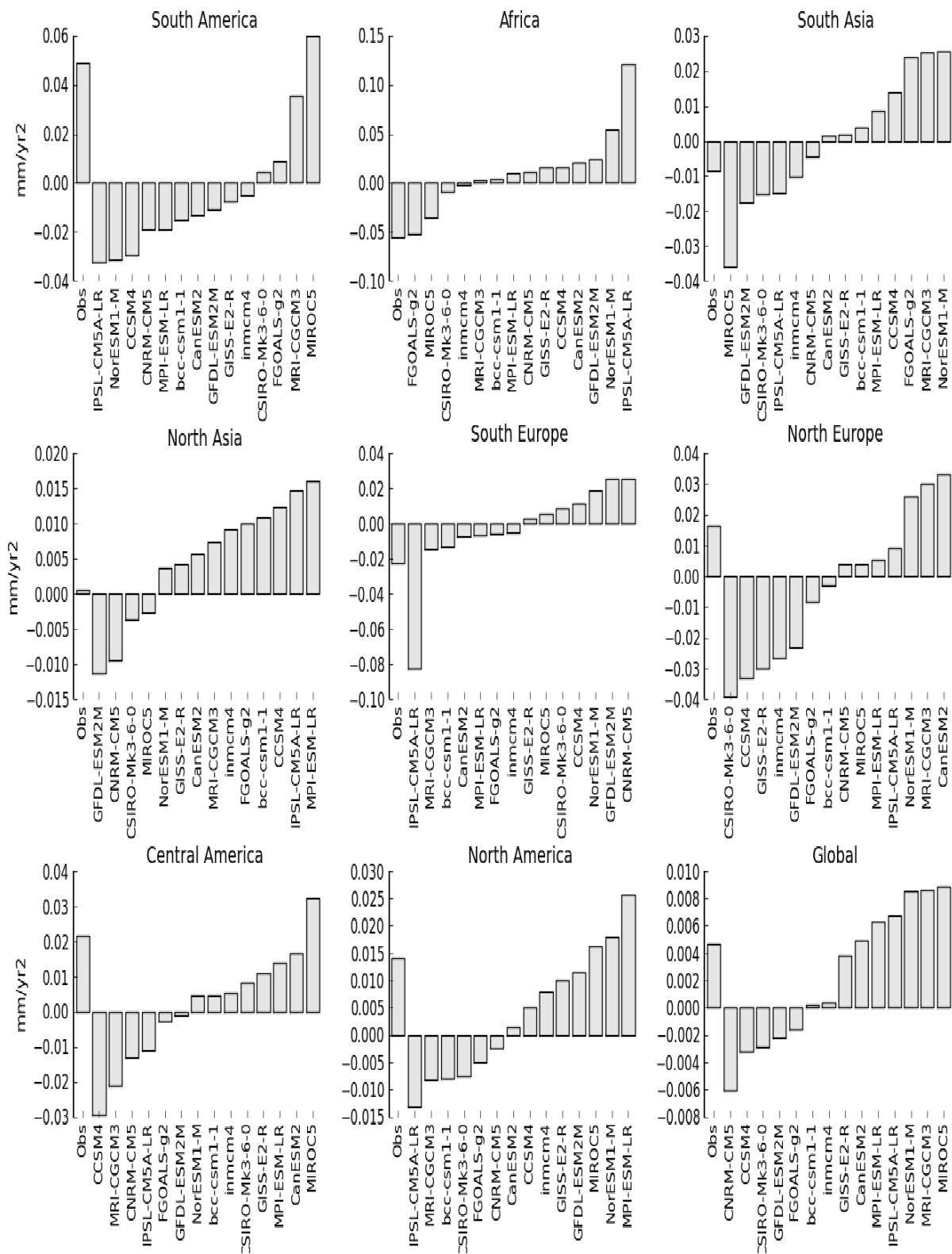


Fig. 8

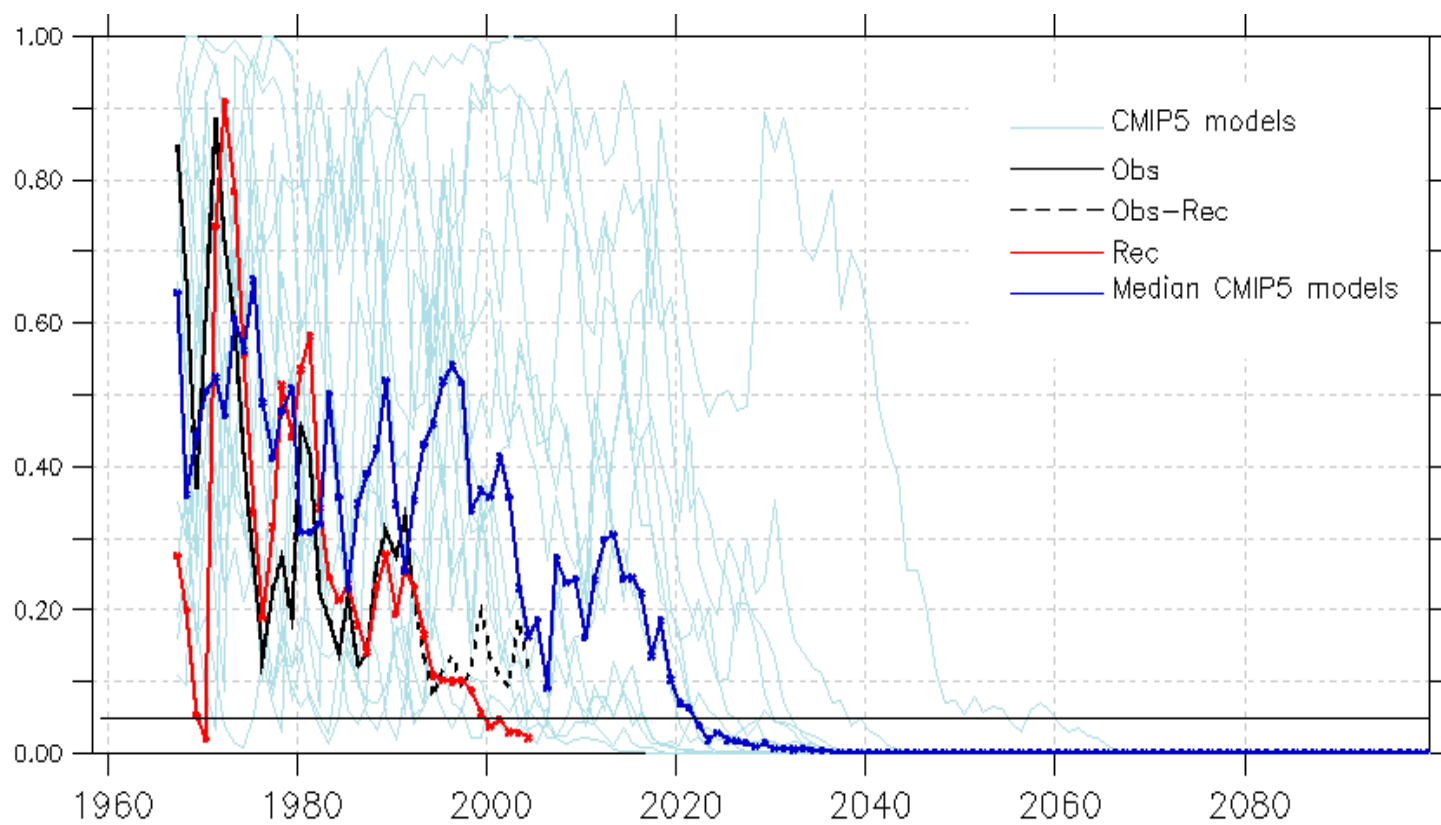


Fig. 9