Supplementary Material

1 Initialization Method

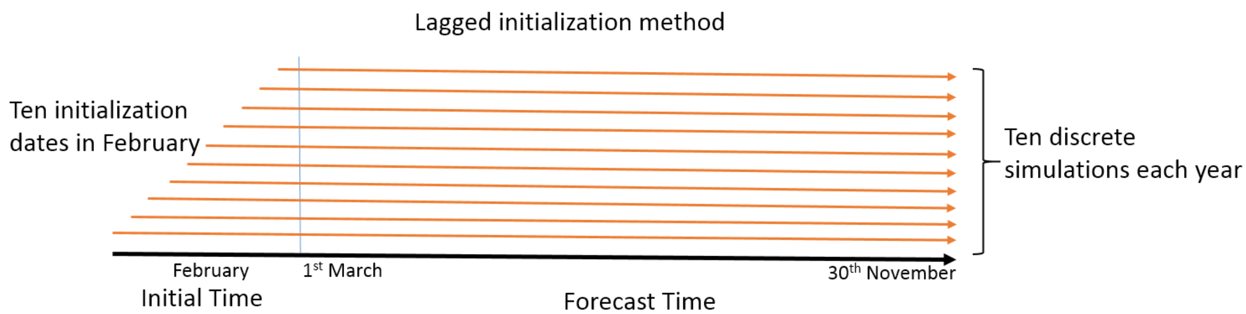
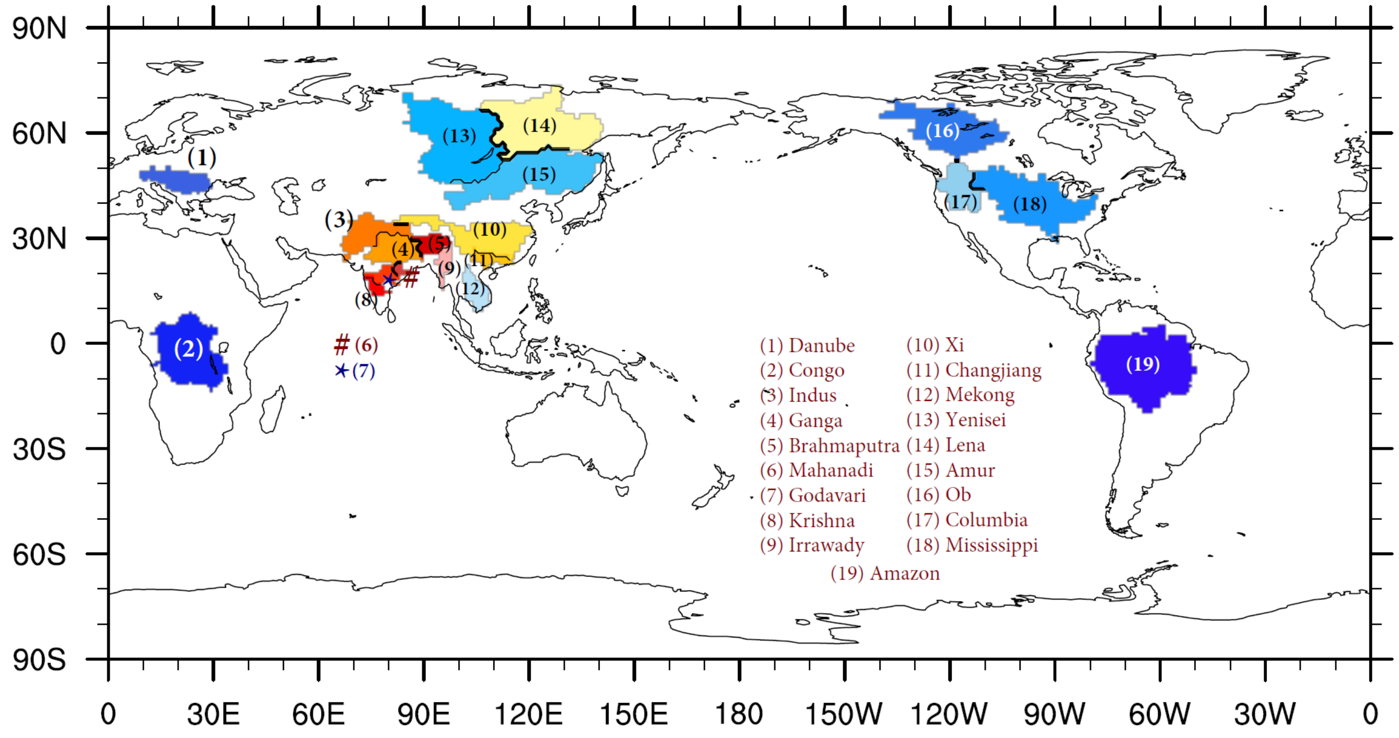


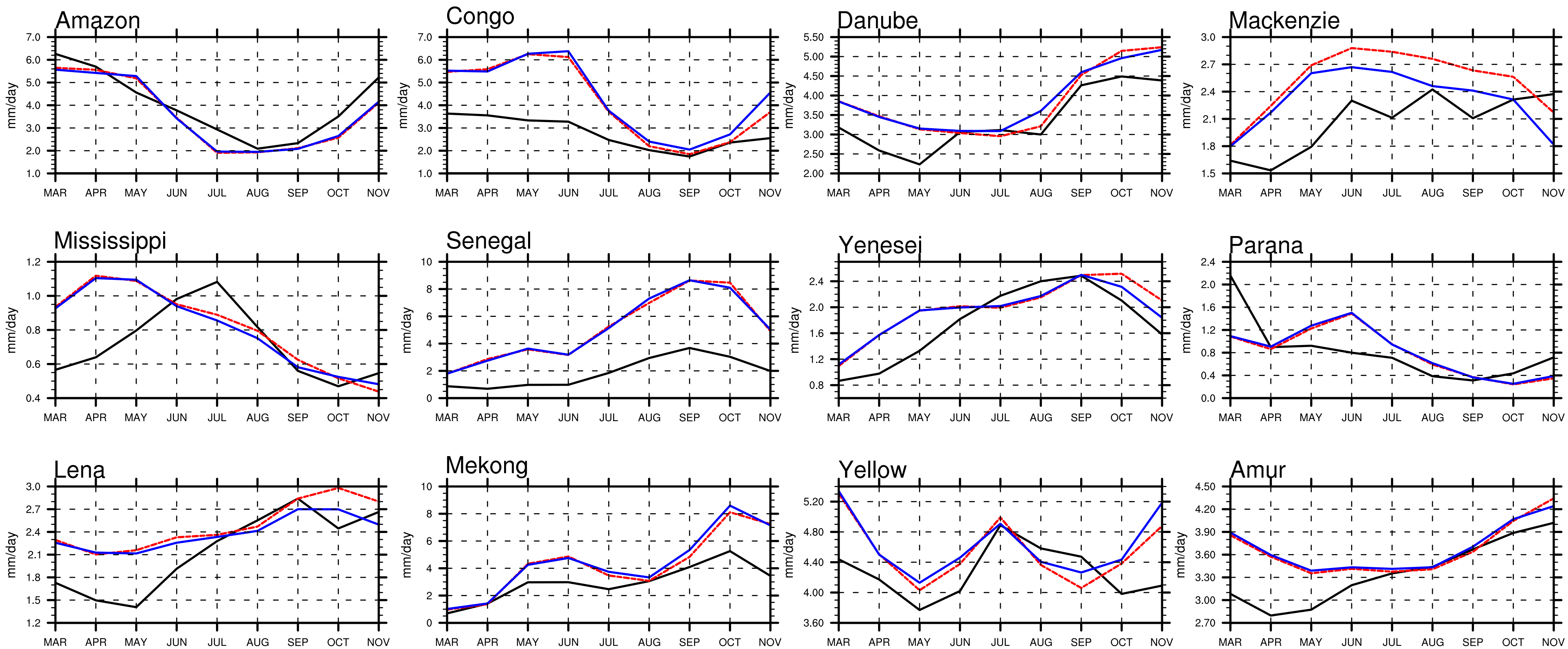
Figure Illustration of the lagged initialization method used for running CFSv2.

Weather forecasting is an initial value problem. For generating a seasonal forecast, the knowledge of the current state of weather is important. This current state of weather is what is called the initial condition. These initial conditions are fed to a GCM in order to generate forecasts. Despite the vast observational network, there are large data gaps. Further, the models used for data assimilation are not perfect. Therefore, it is not possible to accurately know the true state of weather. These small errors in the knowledge of initial conditions can quickly grow and occlude the forecasts. Therefore, in seasonal forecasting, the concept of ensemble forecast is important. Multiple forecasts with slightly different initial conditions are made to account for the uncertainty in observations. The mean of these ensembles is thought to account for the uncertainty in the initial conditions. Two methods of ensemble generation are widely used: the lagged ensemble method and the perturbed ensemble (PE) method. In the PE method, small perturbations to the initial condition are made in order to generate a set of initial conditions, which can be utilized to generate a set of forecasts. In the LE method, a set of initial conditions spread over different dates are used.

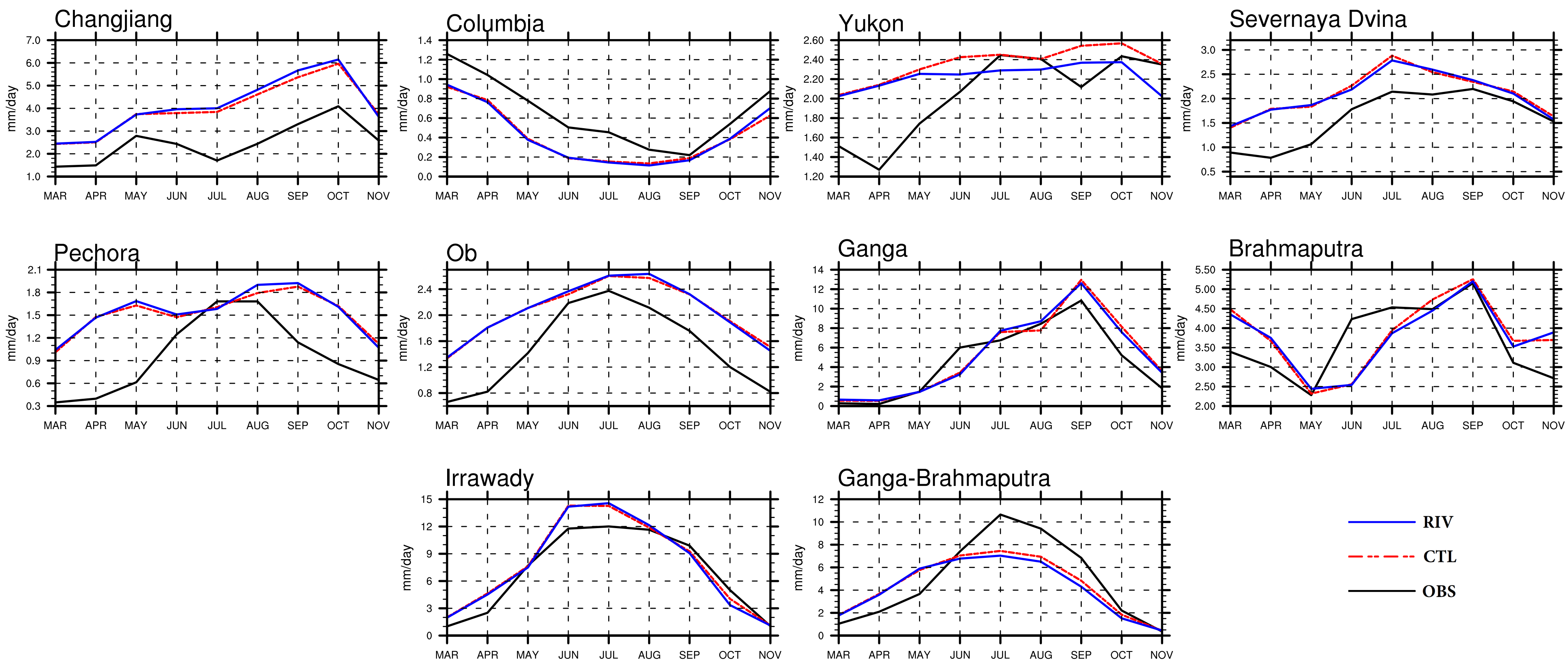
In this study, we use the LE method to initialize CFSv2. A set of ten dates are selected in February as shown in the figure above. The model is initialized on 3rd, 5th, 7th, 10th, 12th, 15th, 17th, 20th, 22nd, and 25th of February each year and integrated up to 30th November. This gives an ensemble of ten forecasts, whose common verification time is 1st March to 30th November. The mean of these ensembles for the common verification time (March-November) is used in this study. Once initialized with a particular date, model is allowed to run freely up to 30th November. Thus, we have 10 discrete simulations each year with common verification time of March to November. The hindcasts span for a period of 1981-2017.



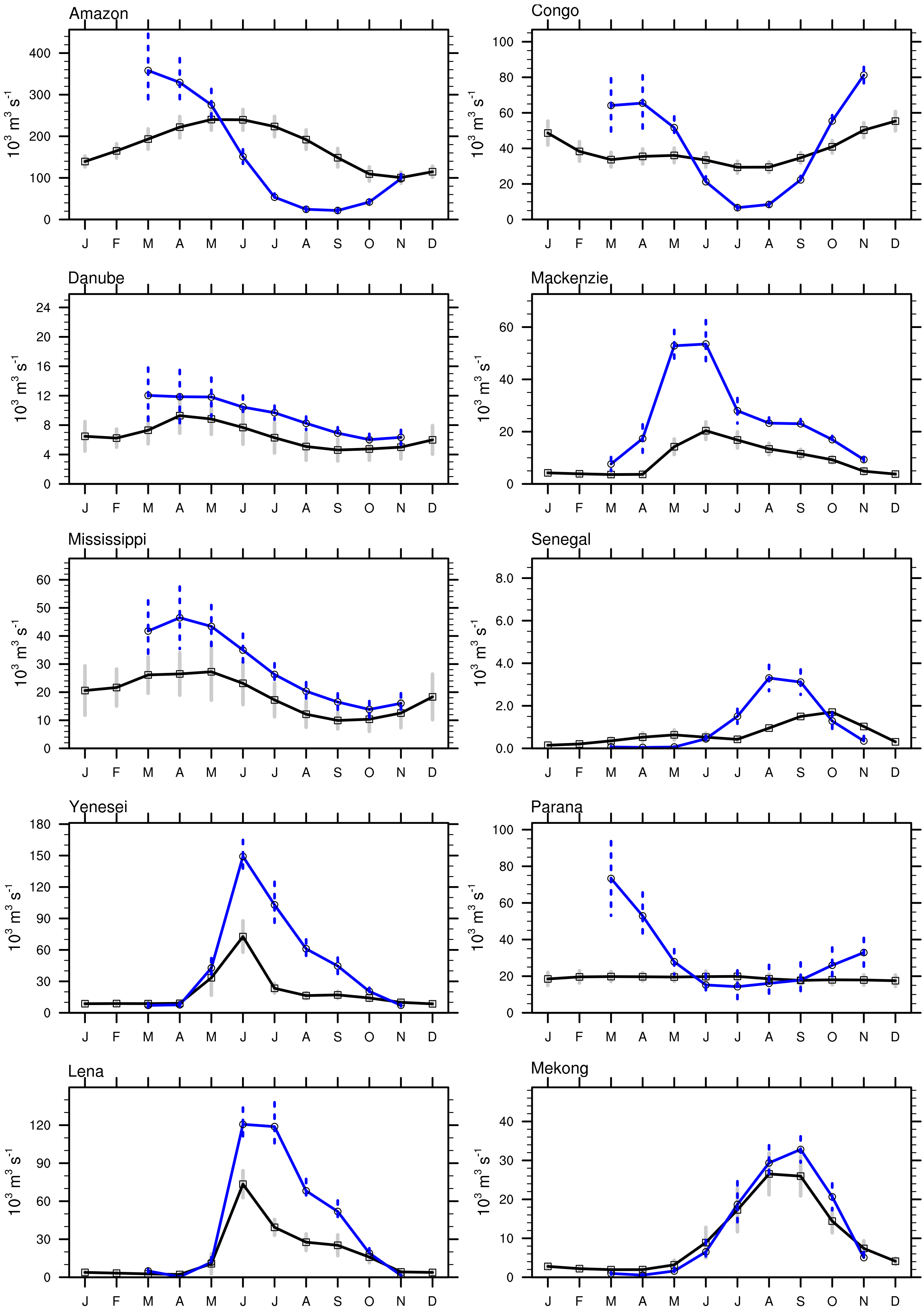
Supplementary Figure 1 The major global river basins dilineated in the routing model using the Digital Elevation Map (DEM).



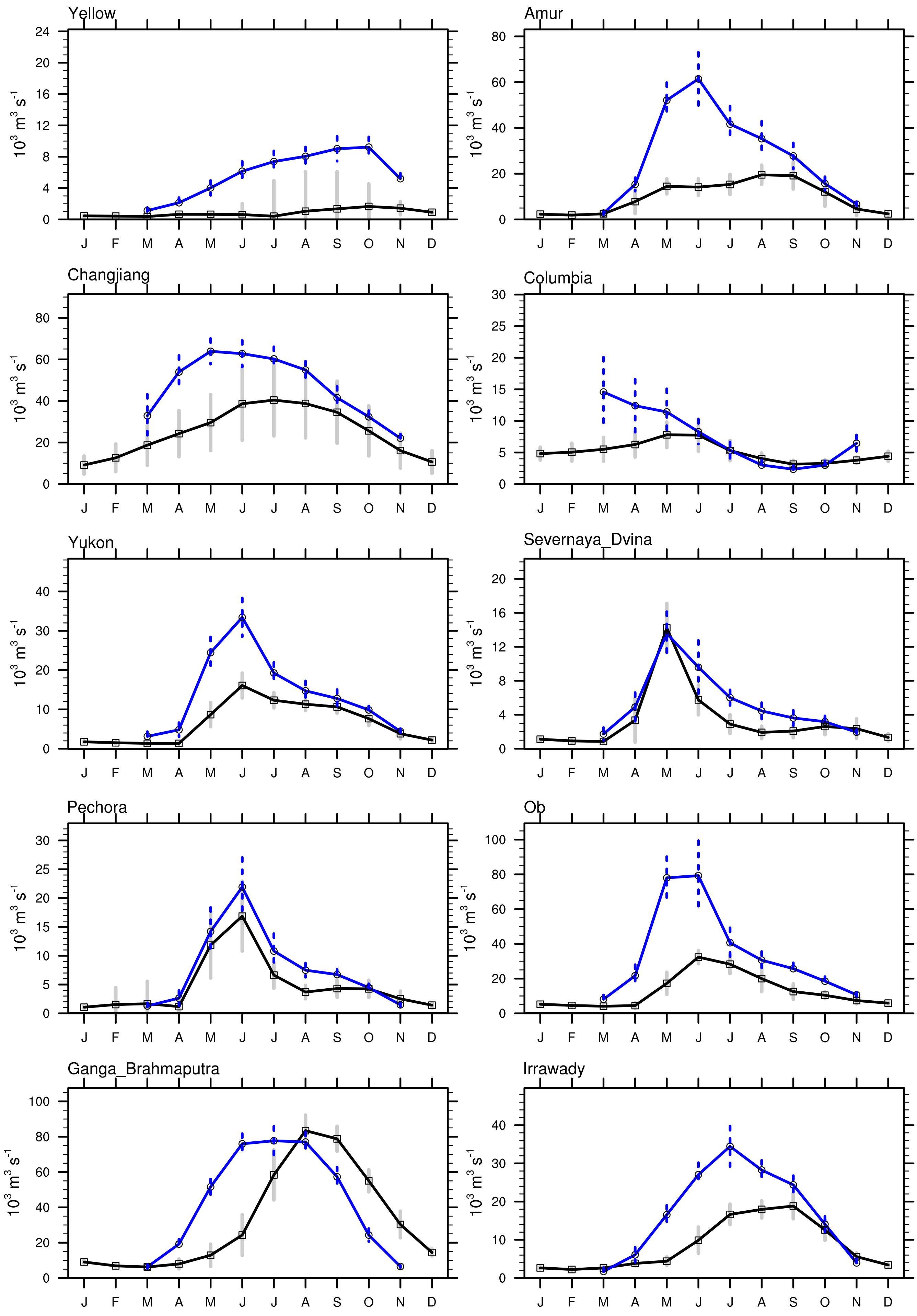
Supplementary Figure 2 The climatological annual cycle of precipitation over the major river basins in GPCP observations (black), CTL run (red) and RIV run (blue).



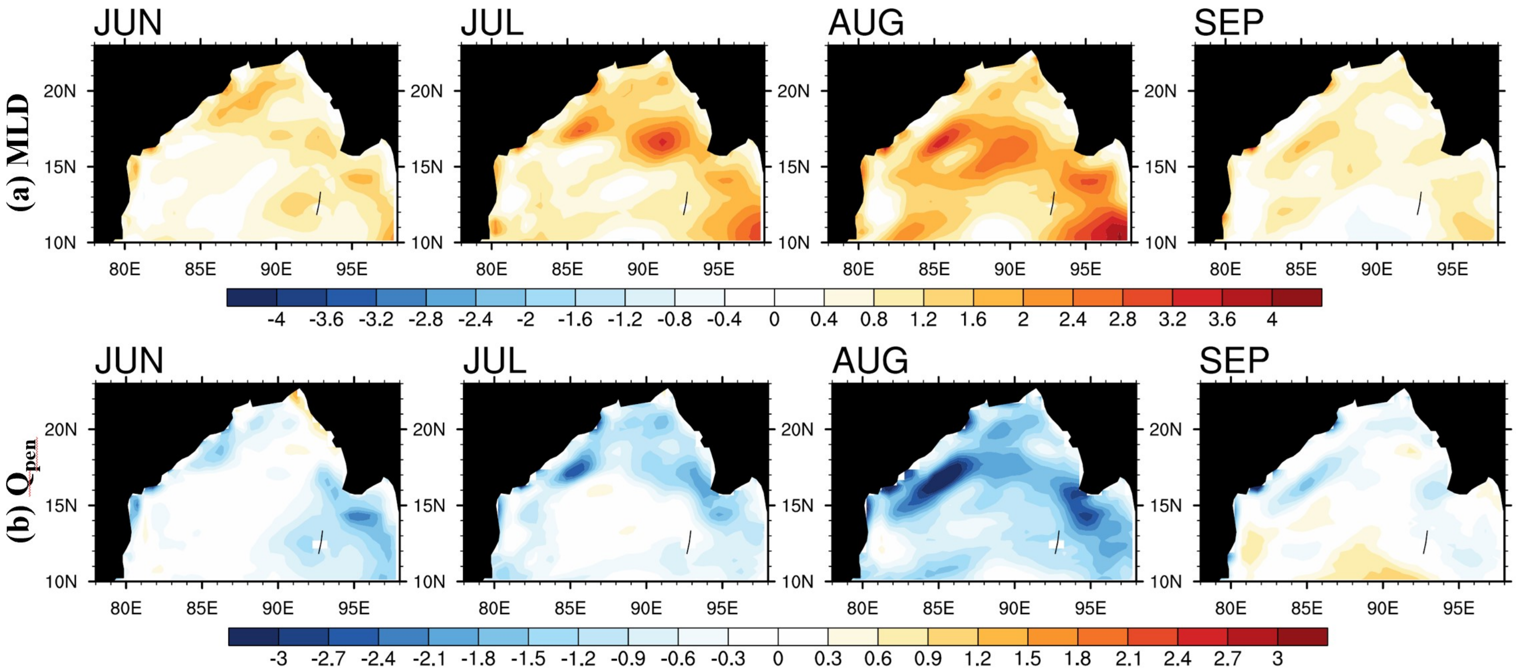
Supplementary Figure 2 Continued.



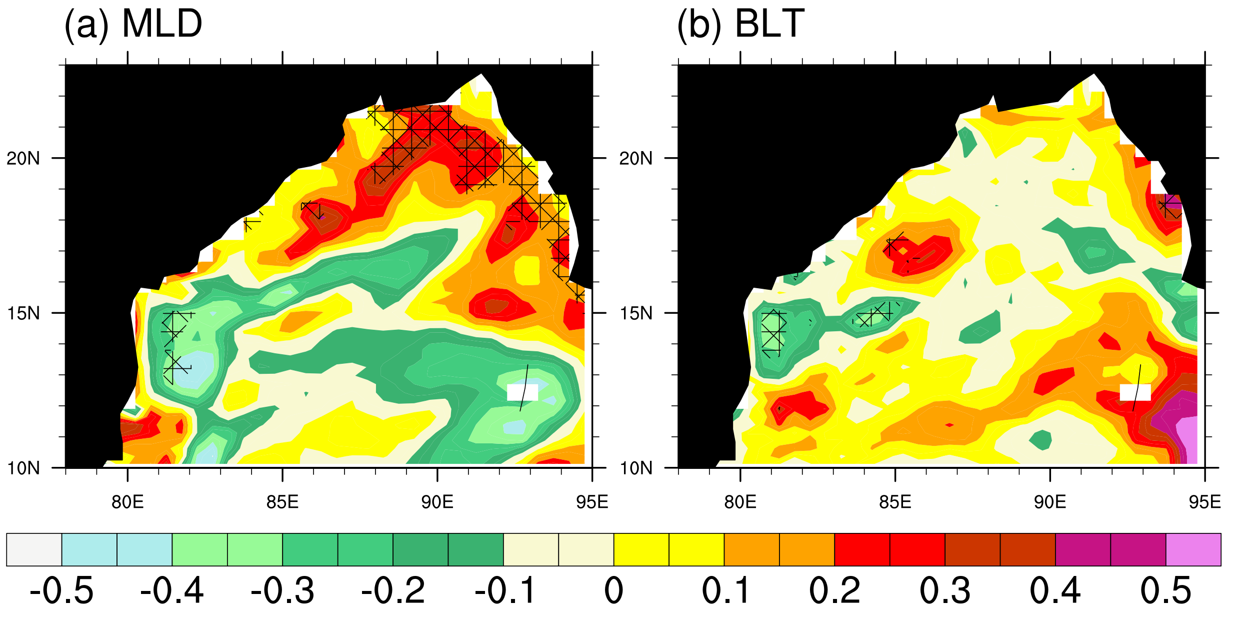
Supplementary Figure 3 The climatological monthly hydrographs from Dai and Trenberth (2021) database (black) and the RIV simulation (blue). Also shown as vertical error bars are the inter-annual standard deviations of the monthly hydrograph.



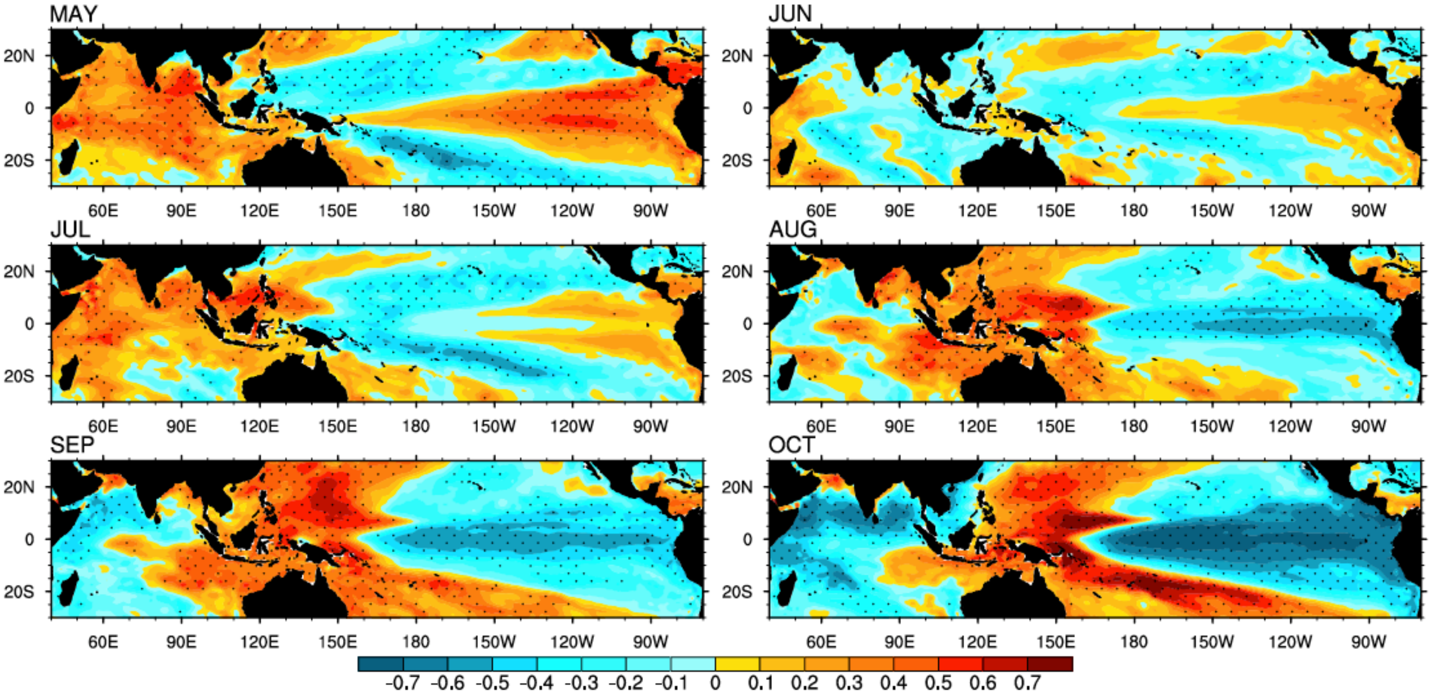
Supplementary Figure 3 Continued.



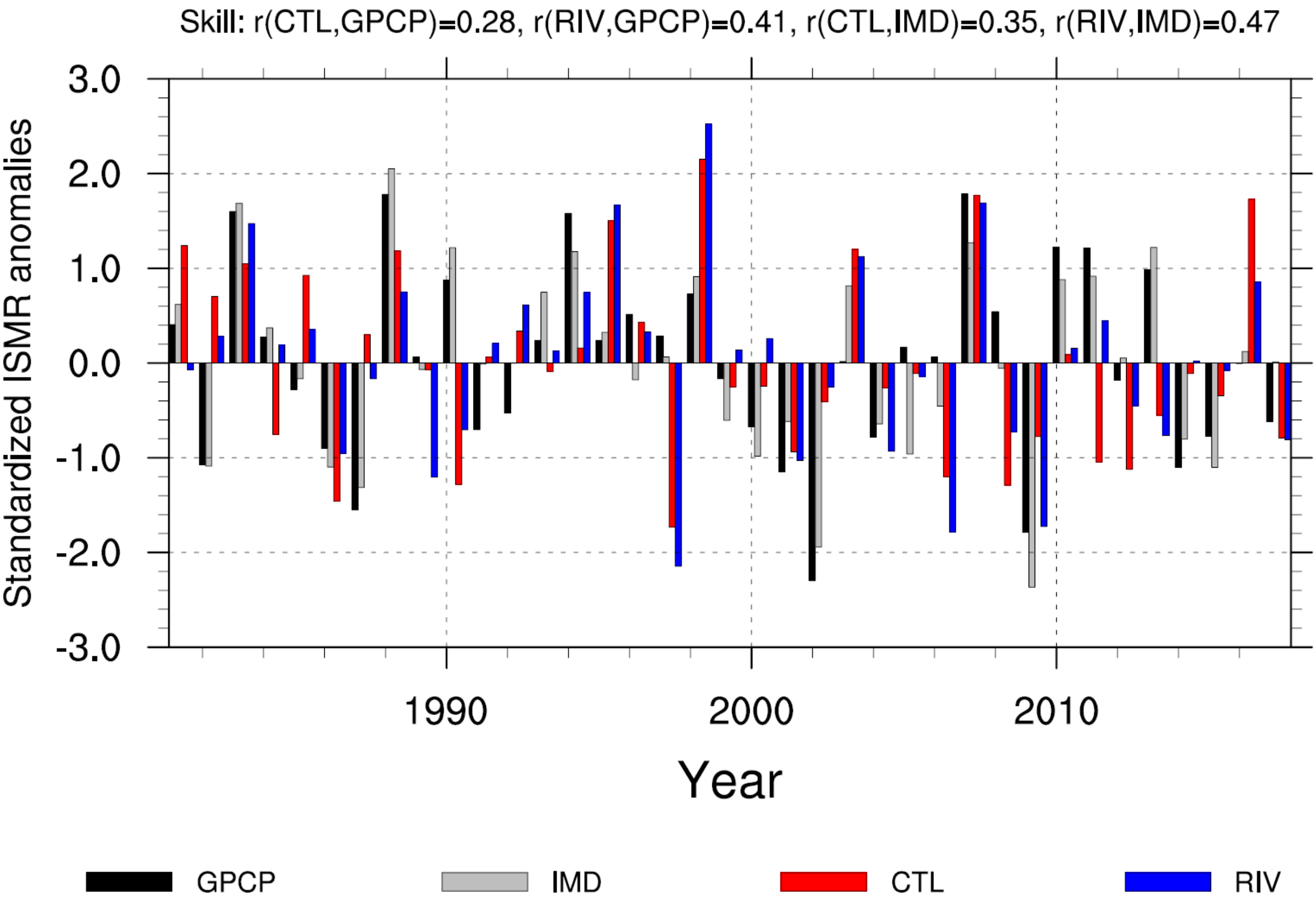
Supplementary Figure 4 The difference (RIV-CTL) in (a) the mixed layer depth (MLD), and (b) the penetrative shortwave radiation (Q_{pen}).



Supplementary Figure 5 The difference (RIV-CTL) in the inter-annual standard deviation (meters) of (a) the mixed layer depth (MLD), and (b) the barrier layer thickness (BLT). The hatched regions denote the area where the variances are significantly different at 90% confidence level using an F-test.



Supplementary Figure 6 The correlation between the total runoff discharged in the Bay of Bengal and the monthly mean sea surface temperatures in the RIV run. The correlations significant at 90% confidence level are stippled.



Supplementary Figure 7 The standardized time series of all India averaged rainfall anomalies from GPCP, IMD, CTL run and RIV run. The skill, defined as the correlation coefficient between observations and model, are also indicated.