ENTRO
NILE BASIN INITIATIVE
INITIATIVE DU BASSIN DU NIL

PCA
PRINCETON CLIMATE ANALYTICS

Advancing Drought Monitoring, Prediction, and Early Warning System in the Eastern Nile Basin

Final Report

Prepared by: Justin Sheffield, Hylke Beck, Ming Pan

Table of Contents

1. Introduction

This is the final report for the project: "Advancing Drought Monitoring, Prediction, and Early Warning System in the Eastern Nile Basin." It summarizes the work done on the project, including the objectives, approach, results and conclusions/recommendations.

The overall objective of the project has been to develop a demonstration seasonal hydrological forecast system for the Eastern Nile (EN) Basin. The system monitors climate and hydrology in near real-time, and produces climate and hydrological forecasts on a seasonal (out to 6 months) basis aimed at drought forecasting.The system is updated daily, and runs about 1–2 days behind real-time, which is based on the availability and latency of the contributing data products. The seasonal forecasts are updated once a month. This is supported byan assessment of historic drought hazard risk in the basin that integrates multiple sources of information from ground observations, satellite remote sensing and climate and hydrological models to provide the best estimate of drought conditions historically. The real-time monitoring and forecasting, coupled with the risk maps, form the basis for early warning and is integrated into a demonstration drought early warning system via an online interface and dashboard. This approach is based on our experience in developing, analyzing and applying hydrological forecasts at all scales (short-term to seasonal), including within the African Flood and Drought Monitor (AFDM; Sheffield et al., 2014), a continental-scale hydrological monitoring and forecasting system that has been running operationally since about 2010, and regional systems such as the Lake Chad Basin Flood and Drought Monitor (CHAD-FDM; Amani et al., 2019).

The EN system builds on this experience to develop a regional system with higher resolution, additional functionality for drought forecasting and data access. It also considersa new set of climate forecasts from the latest set of European climate models hosted by ECMWF. Climate model forecasts such as these are generally available at about 1.0° (~100km) resolution and aretherefore downscaled to the 5km resolution of the EN system, and bias-corrected to match the statistics of the observational climatology. A key part of developing a seasonal forecast system is to evaluate the skill of the seasonal climate forecasts that drive the hydrological forecasts and identify the best combination of climate models. The individual climate model forecasts are merged into a multi-model ensemble by weighting individual models based on skill evaluations. Skill weighted merging has been shown to outperform any individual model and simple model averaging for a variety of skill metrics (Wanders and Wood, 2016).

A historical database of gridded climate and hydrological variables has been developed for the EN system to provide historical context for the operational monitoring and forecasting system including the risk assessments. The historical hydrological data were generated by forcing a hydrological model with gridded precipitation and temperature data at daily and 0.05° (5km) resolution for 1979–2018, and have been evaluated in terms of comparisons with observed streamflow and soil moisture, and for drought risk. The climate forecasts drive the hydrological model to produce forecasts of hydrological variables and drought indices.

This final report provides an overview of this work including details of the datasets, models and methods used to develop and evaluate the system, and includes:

1. The approach to construction of the historic climate, hydrological and drought datasets.

2. Evaluation of this dataset against observational data, and analysis of derived drought risk and potential exposure.
3. A review of existing sources of climate and drought forecasts from different forecast centers and databases
4. An evaluation of the skill of these forecasts for a range of seasonal meteorological and drought metrics.
5. Development and evaluation of a multi-model weighted ensemble forecast.
6. Coupling of the climate forecasts with the hydrological model, to produce seasonal hydrological forecasts, and evaluation of the skill for a range of metrics.
7. An overview of the online dashboard that provides access to the drought early warning information and datasets.

2. Methodology

2.1. Historic Reconstruction of Climate and Hydrology

The historical reconstruction of hydrology for the EN Basin has been derived at a high 0.05° (□5 km) resolution based on the HydrologiskaByråns Vattenbalansavdelning (HBV) hydrological model (Seibert and Vis, 2012) and provides daily historic data for several key hydrological variables, including soil moisture, evaporation, and runoff. The meteorological forcing is derived by merging a wide range of station, satellite, and atmospheric model data to obtain the best possible predictions across all climate zones in the basin. The precipitation forcing is taken from the gauge-, satellite-, and model-based MSWEP product (Beck et al., 2017a, 2019a) for the historic period. The air temperature and wind speed forcings are taken from the gauge- and model-based MSMet product for the historic period. The runoff estimates from HBV are routed downstream using the highly computationally efficient RAPID streamflow routing scheme. To improve the runoff estimates of HBV, we implemented an innovative parameter regionalization approach that yields optimized model parameters that vary according to landscape and climate characteristics for the entire model domain, including ungauged regions.

2.1.1. MSWEP

Multi-Source Weighted-Ensemble Precipitation (MSWEP) is a highly innovative gridded precipitation dataset spanning 1979–2017 (Beck et al., 2017a, 2019a). MSWEP is unique in several aspects: (i) high spatial (0.1°) and temporal (3 hourly) resolution, increasing the local relevance of the precipitation estimates; (ii) optimal merging of precipitation estimates based on rain gauges (WorldClim, GHCN-D, GSOD, GPCC, and others), satellites (CMORPH, GridSat, GSMaP, and TMPA 3B42RT), and reanalyses (ERA-Interim and JRA-55), to obtain the best possible precipitation estimates at any location in the basin from the lowlands to the highlands; (iii) Cumulative Distribution Function (CDF) corrections, to improve the precipitation frequency; (iv) incorporation of daily (instead of monthly) gauge observations; and (v) a gauge correction scheme that accounts for reporting times of gauges, to avoid temporal mismatches between the uncorrected estimates and the gauge data. In addition, systematic biases frequently found in mountainous areas were ameliorated by first inferring the “true” precipitation from observed streamflow data using a Budyko model, and then deriving global gap-free precipitation bias correction maps using

random forest regression (Beck et al., 2019c). Figure 1 outlines the main steps involved in the production of MSWEP.

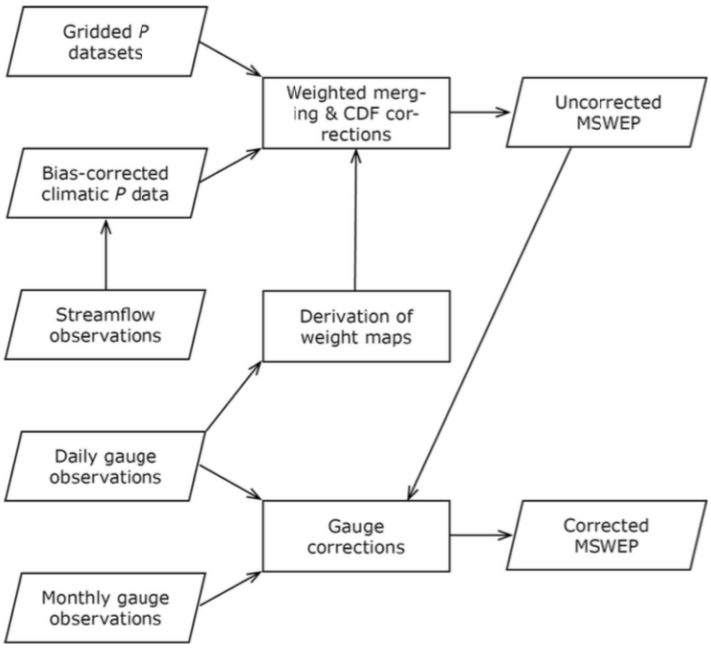
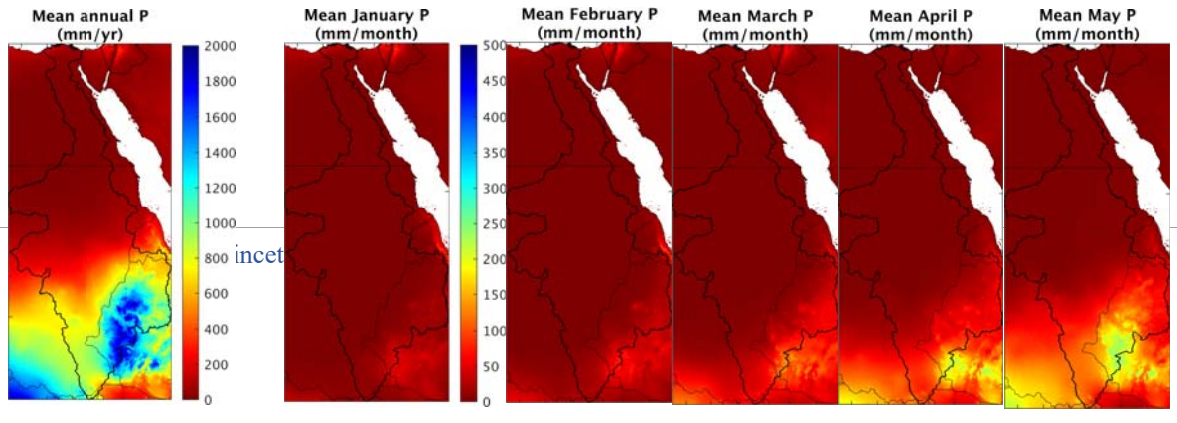


Figure 1. Flow chart of the main processing steps involved in the production of MSWEP.

MSWEP is the only precipitation dataset to take advantage of the complementary performance of (re)analysis and satellite precipitation datasets: (re)analysis datasets perform consistently better in frontal-dominated regions, whereas satellite datasets perform consistently better in convection-dominated regions. For every 0.1° grid-cell, including those in regions of complex orography, the data sources are merged depending on the performance of the respective data sources at nearby gauges. In addition, MSWEP is the only dataset to account for reporting times when applying the gauge corrections, the only fully global dataset with a 0.1° spatial resolution, and the only dataset to use streamflow observations for bias correction of precipitation in mountainous regions. Consequently, MSWEP exhibits substantially more realistic spatial patterns in mean, magnitude, and frequency compared to other state-of-the-art precipitation datasets. Comparison with other precipitation datasets suggest that those consistently underestimate precipitation amounts in mountain regions. For the EN system, the data were downscaled to 0.05° using bilinear interpolation. Figure 2 presents the MSWEP mean annual precipitation map for the EN Basin.



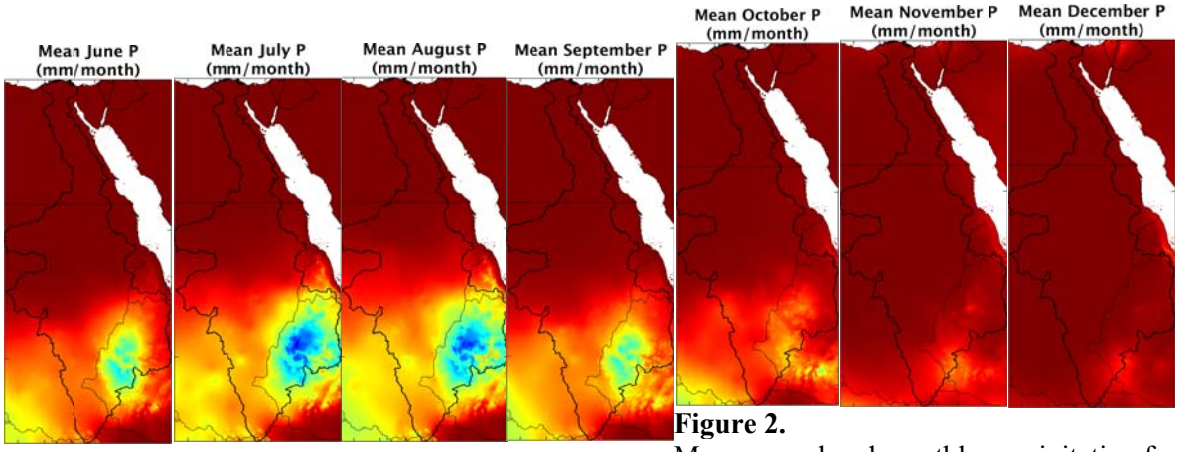


Figure 2. Mean annual and monthly precipitation for the Eastern Nile basin

2.1.2.MSMet

Multi-Source Meteorology (MSMet) is a gridded surface meteorological dataset spanning 1979–present with a daily temporal and 0.05° spatial resolution derived by merging three bias-corrected (re)analysis products. Among the included meteorological variables are the metrics daily mean, minimum, and maximum air temperature, and wind speed. Two reanalyses are used, ERA-Interim (0.75° resolution; 1959–present) and JRA-55 (0.56° resolution; 1979–present), and one analysis, GDAS-Anl (0.25° resolution; 2015–present). Each product is first downscaled to 0.05° and bias corrected using the station-based, topographically-adjusted WorldClim climatic dataset (1-km resolution; http://worldclim.org). The bias-corrected products are subsequently averaged to obtain temporally consistent, high-resolution meteorological data. The MSMet record is extended until the near present multiple times per day using the available data sources. Estimates less than one week old are upgraded once additional data sources become available to ensure the best possible accuracy at any moment in time. Figure 3 outlines the main steps involved in the production of MSMet. The MSMet daily minimum and maximum temperature data are used to compute the potential evaporation following Hargreaves (1994).

Figure 3 presents maps of the MSMet-based mean annual and monthly potential evaporation. Figure 4 presents the elevation for context.

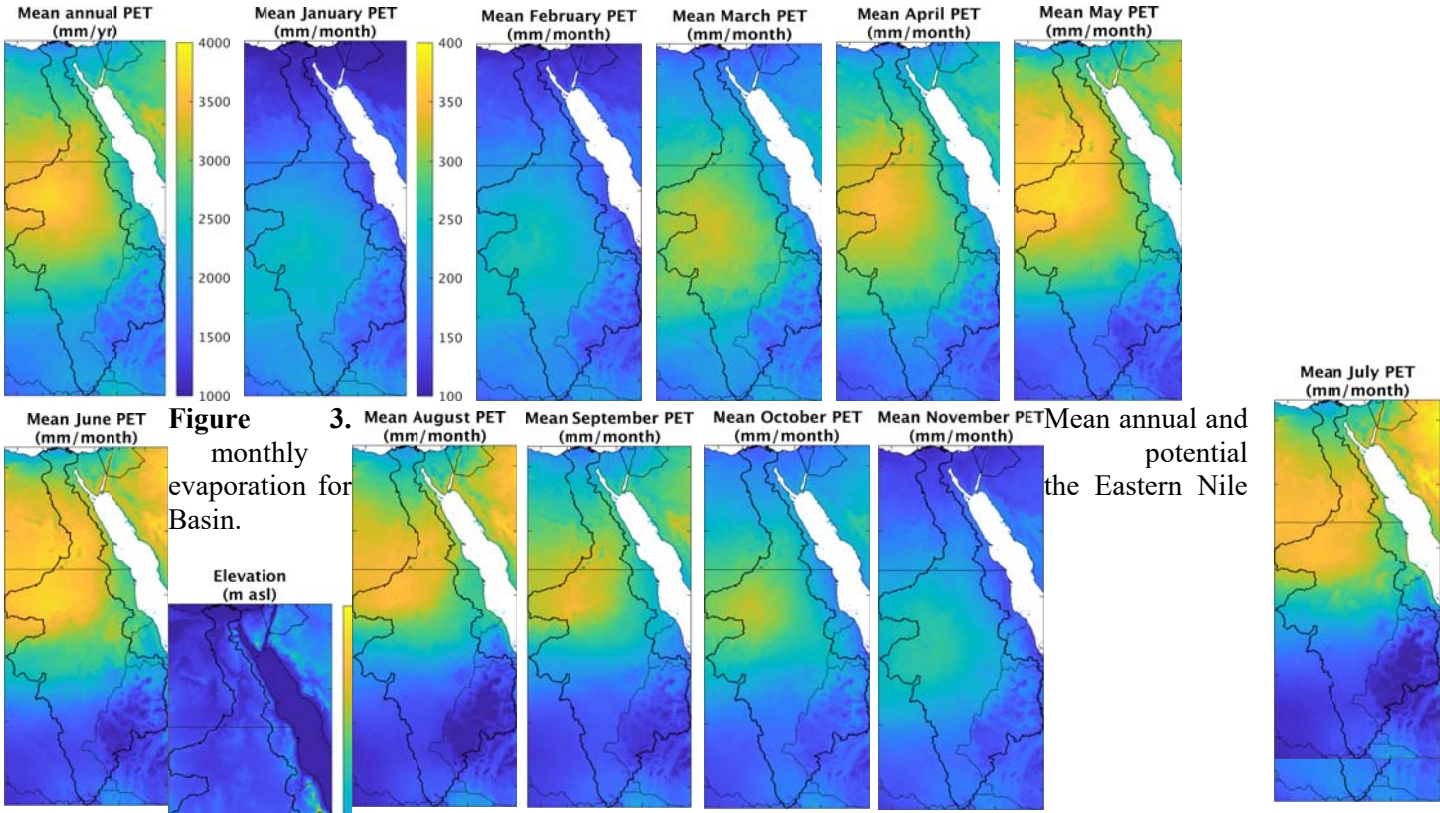


Figure 3. Mean annual and monthly potential evaporation for the Eastern Nile Basin.

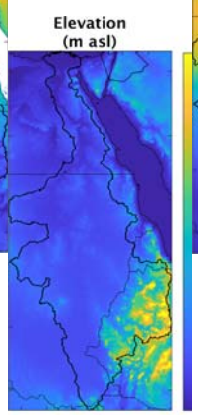


Figure 4. Surface elevation of the Eastern Nile basin.

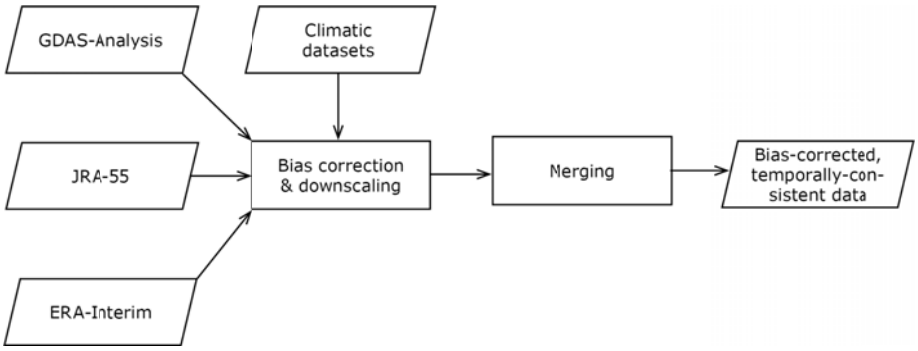


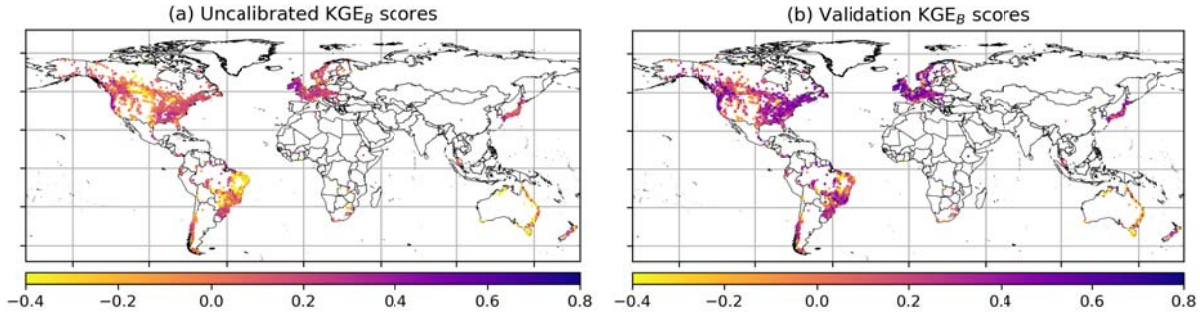
Figure 5. Flow chart of the main processing steps involved in the production of MSMet.

2.1.3. Hydrological model

The HBV hydrological model (Seibert and Vis, 2012) was used to obtain estimates of key hydrological variables including soil moisture, evaporation, and runoff. HBV was used because of its low complexity, high agility, and computational efficiency. Additionally, the model has been successfully used in numerous studies spanning a wide range of climate and physiographic conditions including several studies focusing on parts of the Eastern Nile basin (e.g., Rientjes et al., 2011; Gebrehiwot et al., 2013; Worqlul et al., 2017). The model runs at a daily time step, has one unsaturated-zone store, two groundwater stores, and 12 calibratable parameters.

All hydrologic models (both physical and conceptual) need to be calibrated in order to obtain accurate streamflow estimates. We implemented a novel parameter regionalization approach that involves the optimization of transfer equations linking model parameters to eight predictors related to climate, land cover, topography, and soils (Beck et al., in review). The optimization was performed in a fully spatially distributed fashion at high resolution (0.05°), instead of at lumped catchment scale, using an unprecedented database of daily observed streamflow from 4229 headwater catchments (<5000 km^2) worldwide. The optimized equations were subsequently applied to the Eastern Nile basin to obtain gap-free parameter maps necessary for the simulations. Running the model at high resolution accounts for the spatial heterogeneity in rainfall-runoff processes and the nonlinear relationship between model parameters and runoff estimates, and should therefore result in more robust estimates of the coefficients. Previous regionalization studies tended to ignore these factors by running the model at a lumped catchment scale (e.g., Yokoo et al., 2001; Hundecha and Bardossy, 2004; Bastola et al., 2008).

Among the eight predictors, three are related to climate (aridity index, mean annual precipitation, and mean annual potential evaporation), two to land cover (mean normalized difference vegetation index and open water fraction), one to topography (mean surface slope), and two to soils (mean clay and sand content of the soil). To optimize the coefficients of the transfer equations, we used the (μ+λ) evolutionary algorithm implemented using the Distributed Evolutionary Algorithms in Python (DEAP) toolkit (Fortin et al., 2012). As objective function, we used a bounded version of the Kling-Gupta Efficiency (KGE_B). The KGE is an objective performance metric introduced by Gupta et al. (2009) and modified by Kling et al. (2012) that combines correlation (related to event dynamics), bias (related to runoff totals), and variability (related to the event distribution). The regionalization approach yielded improvements in streamflow simulation performance for 88 % of the independent validation catchments, confirming the effectiveness of the approach in improving streamflow simulation performance (Figure. 6).



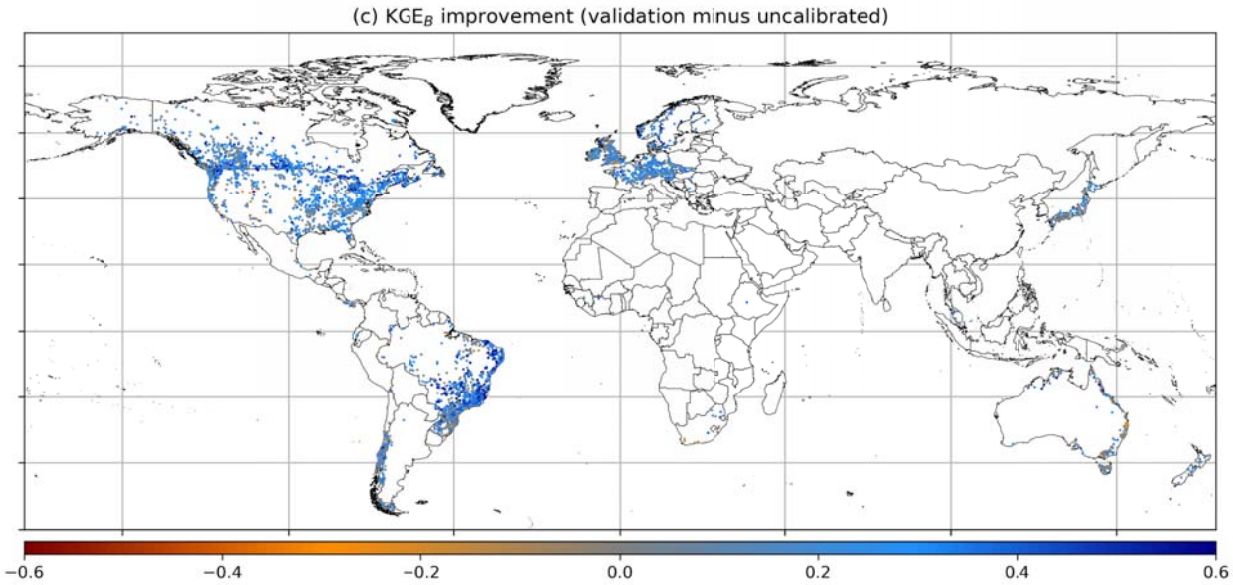


Figure 6. (a) Transformed Kling-Gupta Efficiency (KGE_B) scores for the first generation of the optimization process (i.e., using uncalibrated parameters). (b) KGE_B scores obtained for the independent validation catchments using regionalized parameters. (c) The improvement in KGE_B scores after regionalization calculated as validation minus uncalibrated KGE_B scores. Blue indicates improved performance, while orange and red indicate deteriorated performance. The scores were calculated using daily observed and simulated streamflow data. Each data point represents a catchment centroid (N=4229). Adapted from Beck et al. (in review).

2.1.4. RAPID river routing model

The runoff estimates from HBV were routed downstream using the Routing Application for Parallel computatIon of Discharge (RAPID) routing scheme developed by David et al. (2011). RAPID is a vector-based river network routing scheme which makes it substantially more computationally efficient than conventional grid-based routing schemes. RAPID uses a matrix-based version of the Muskingum method to calculate flow and volume of water in all reaches of a river network. We used river network data from the MERIT Hydro hydrography database (Yamazaki et al., 2019) which is based on the latest elevation data (MERIT DEM; Yamazaki et al., 2017) and water body datasets (G1WBM, GSWO, and OpenStreetMap), and contains in total ~3.2 million river reaches globally. The MERIT DEM (90-m resolution) combines the NASA SRTM3 DEM, JAXA AW3D DEM, and Viewfinder Panoramas' DEM, with NASA forest height data, and corrects for bias, stripe noise, speckle noise, and tree height bias. The river network for the Eastern Nile Basin is shown in Figure 7, along with the details in the eastern sub-basins. To account for the non-linear flow dynamics and responses to rainfall, a variable flow velocity method based on Manning's equation was adopted, where river bed slope and Manning's channel roughness were parameterized as a function of topography, sinuosity, bedrock material, etc. The hydraulic radius was estimated using a set of pre-defined channel trapezoidal shapes varying as a function of stream order and calculated streamflow at each model time step.

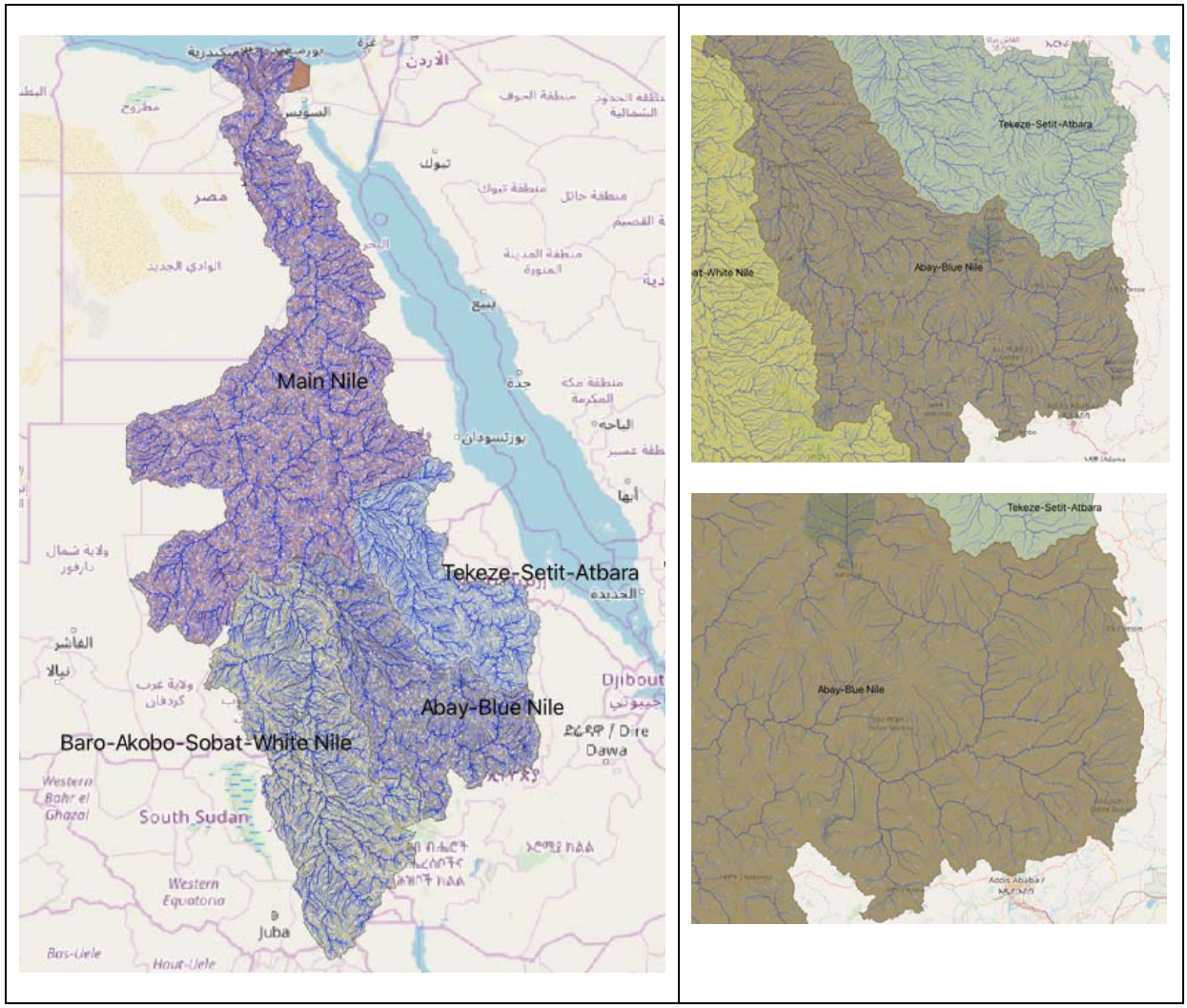


Figure 7. Maps of the RAPID vector-based streamflow routing network based on the MERIT/HydroSheds database for (left) the Eastern Nile basin, and (right) zoomed in on the Tekeze-Setit-Atbara and Abay-Blue Nile sub basins. River reach colors are dependent on stream order.

2.2. Development of Climate and Hydrological Forecasts

2.2.1. Existing sources of forecasts

Climate forecasts are taken from a range of climate models which provide spatially distributed forecasts of a range of physical variables at relevant time scales (sub-daily and daily resolution for at least 6 months). These criteria are based on the fact that the proposed EN system needs to provide forecasts across the basin, at different time scales, and that improved forecasts are likely to be obtained by merging multiple models rather than relying on a single model. This is advantageous compared to statistical forecasts, which are generally developed only for specific forecast regions or points, and are for aggregated (e.g. seasonal rainfall) or very specific (e.g. start of season) metrics, and so do not provide the flexibility required. Using climate model forecasts, as opposed to statistical forecasts, also provides the opportunity to diagnose the forecasts and their skill, as they provide related physical quantities, such as atmospheric variables.

Candidate climate models that have been considered are from a range of climate centres and forecast projects that have been developed and made available at a range of time scales (Table 1). In addition to the skill of the model forecasts, criteria for choosing a set of models is based on their resolution (space and time), availability of hindcasts (historic forecasts) to evaluate their skill, and availability of forecasts operationally. Ideally, a model will provide at least daily data for precipitation and temperature, at a spatial resolution of no coarser than 1.0° (100km), have at least 20 years of hindcast data and be available operationally.

We have worked with the models listed in Table 1 previously and have experience in their utility and skill characteristics (Yuan et al., 2015; Wanders et al., 2019) and merging them seamlessly across time (Yuan et al., 2014). We have used the North American Multi-Model Ensemble phase 1 (NMME-1) previously (e.g. Yuan and Wood, 2012; Yuan et al., 2013a,b,c; Kam et al., 2014; Yuan et al., 2014; Yuan et al., 2015; Thober et al., 2015) and this was a core part of the AFDM seasonal forecasts (Sheffield et al., 2014). The NMME-1 only provides monthly data, and so is less suitable for forecasts of sub-seasonal metrics (e.g. probability of dry spells, start of the rainy season, etc) and for hydrological forecasting as the hydrological model requires daily climate data. The latest phase 2 version (NMME-2), does provide sub-daily information that can resolve the key intra-seasonal events and be used to drive the hydrological model without any downscaling, but currently does not do so operationally and so is not suitable for the proposed demonstration of an operational EN system. A further set of climate models is available from the EU Copernicus Climate Service (C3S) implemented by the European Centre for Medium range Weather Forecasting (ECMWF), which consists of five European models that have been run in a consistent way and the data are provided on a consistent grid and format. Forecasts are provided for daily and sub-daily variables and the forecasts are available as hindcasts and operationally. We have previously used a subset of these models to develop a demonstration hydrological seasonal forecasting system for Europe (Samaniego et al., 2019) and have evaluated their skill across a range of metrics and scales across this continent (Wanders et

al., 2019). Given their availability operationally and daily resolution, we have selected the European set of models for the EN system. Note that all products are freely available. The Copernicus products are freely available for non-commercial use.

Table 1. The sets of forecast models considered.

Model set	Source	Spatial Resolution	Temporal Resolution	Lead time	Hindcast period
			Dynamical models		
NMME1, 2	NMME Project	1.0°	Monthly (NMME-1), sub-daily (NMME-2)	1–12 months	~1982–2010
Copernicus C3S	Copernicus Climate Data Store (CDS)	1.0°	Sub-daily	1–7 months	~1993–2016
			Statistical/climatological models		
Statistical	As available	Aggregated[1]	Generally monthly/seasonal	~1–3 months	Various[2]
ESP	MSWEP, MSMet	5 km	Daily	1–12 months	1979-2018

[1]Forecasts are for an aggregated spatial area such as a river basin.
[2]The hindcast period for statistical models depends on the record length of the contributing data.

A benchmark to evaluate the climate forecasts against is climatological forecasts (also known as Ensemble Streamflow Prediction, ESP) which is a standard approach to hydrological forecasting that is based on randomly sampling years from the historic observational record to form an ensemble of forecasts. This type of forecasting can also be conditioned on states of the climate that are known to be associated with wetter or drier conditions in the region (e.g. the El Niño Southern Oscillation (ENSO)). For example, if the current year is experiencing El Niño type conditions, then only historic El Niño years are sampled.

2.2.2. Evaluation of Climate Forecasts

Table 2 provides details of the five climate models in the ECMWF model ensemble. We evaluate the models for a common forecast length of 180 days (6 months) and hindcast period of 1993–2016. The forecasts are made at the model's native grid, but are provided by ECMWF on a slightly reduced 1.0° grid, globally. We evaluate the models against gridded observational data: precipitation is from the 0.1° resolution Multi-Source Weighted-Ensemble Precipitation (MSWEP) dataset (Beck et al., 2017a,b, 2019); temperature is taken from the 0.05° Multi-Source Meteorology (MSMet) dataset. The climate forecast data are interpolated to the observational resolution by simple bilinear interpolation. Note, that we bias-correct the climate model forecasts for the hydrological modeling at the 0.05° grid of the proposed EN forecast system (see section 2.4).

MSWEP has been evaluated globally against other similar gridded precipitation products and outperforms them in terms of 1) comparison with available precipitation gauges (for uncorrected gridded products where no gauge data are included directly), and 2) comparison with streamflow

gauges when used to force a hydrological model for the gauge-corrected version (Beck et al., 2017a,b, 2019). The uncorrected version performs well against precipitation gauges for the wetter upper part of the EN basin, but with lower performance over the lower part of the EN region in terms of the 3-day mean correlations, although with higher performance for monthly correlations(Beck et al., 2017a,b). The poorer performance in the lower part is typical of other gridded products (reanalysis and satellite) and is likely because the region is so dry and the little precipitation that does fall is highly variable in time and space (Beck et al., 2017a,b).The gauge corrected version of MSWEP that is used in the EN forecast system incorporates data from 15 available gauges across the EN basin that are determined to have sufficient quality (out of about 30 gauges available in total), and therefore removes much of the error in the uncorrected version.

Table 2. ECMWF European forecast models and their attributes.

Model	Forecast length (days)	Forecast initialization dates	Hindcast initialization dates	Forecast ensemble members	Hindcast ensemble members	Hindcast period	Spatial resolution
ECMWF	215	1st of month	1st of month	51	25	1981–2016	36–80 km
UKMO	215	Each day of month	1st, 9th, 17th, 25th	2	7	1993–2016	0.83° x 0.56°
Meteo-France	215	1st of month	1st of month	25	24	1993–2016	0.5°
DWD	180	1st of month	1st of month	50	30	1993–2017	~100 km
CMCC	180	1st of month	1st of month	50	40	1993–2016	~1°

Model skill is evaluated in multiple ways at the monthly scale and for a set of drought metrics. Evaluations at the monthly scale are done in three ways: 1) mean seasonal cycle (climatology) to assess the overall forecast bias; 2) correlation over the whole hindcast period to quantify how well the forecasts follow the observed variability; and 3) correlation of anomalies over the whole hindcast period to quantify how well the forecasts replicate the observed variability after removal of the mean seasonal cycle (which otherwise inflates the skill). For (2) and (3), results are shown for all forecast target months, and for selected target months in the wet season. Results are shown averaged over the entire EN basin, the three main sub-basins of the lower part of the EN basin (Abay-Blue Nile, Baro-Akobo-Sobat-White Nile, and Tekeze-Setit-Atbara) (see Figure 7), and for each grid cell as correlation maps. The skill of the models is considered for different lead times (how far in advance the target month or season is forecast) from 1 to 6 months. We focus on all lead times, from 1–6 months, and summarize skill in terms of correlation as a function of lead time and target month. All monthly evaluations are done using the ensemble mean for each model, which facilitates the evaluations and their interpretation, but we also evaluate and show the uncertainty around the mean from the range of individual ensemble members.

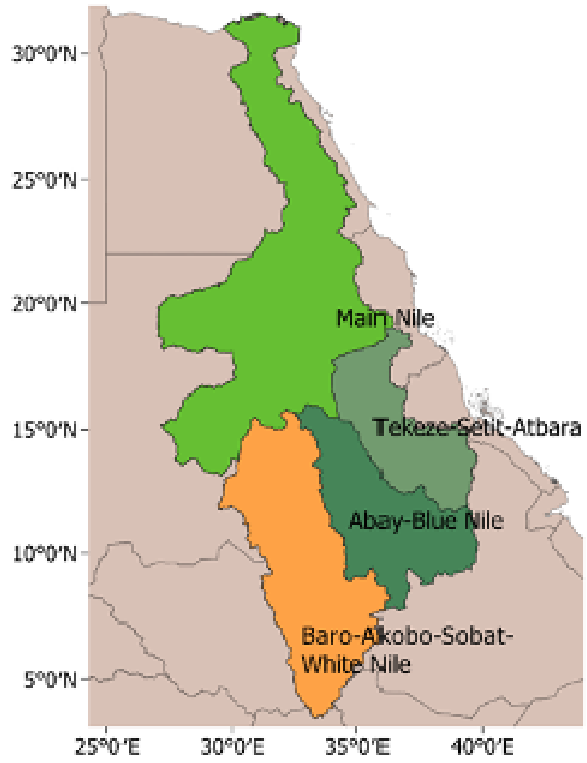


Figure 7. Map of the EN basin and sub-basins.

Model skill is also evaluated for meteorological drought using the Standardized Precipitation Index (SPI), and for specific historic drought events, based on each model's ensemble mean and full ensemble. The SPI is a well-used meteorological drought index that is the recommended index by the World Meteorological Organization (WMO; McKee et al., 1993). We compare the SPI from the model forecasts with the MSWEP observational dataset over a range of time scales (1, 3, 6, and 12 months), which are designated as SPI-1, SPI-3, SPI-6, and SPI-12. Standard forecast metrics are used, including *hit*, *miss* and *false alarm rates* for drought events that are based on a 2x2 contingency table. Hit rates are also called "probability of detection" and represent the percentage of events that are correctly forecast, compared to all observed events. Miss rates are the complement of this, i.e. the percentage of events that were not forecast. False alarm rates are the percentage of non-events that were forecast as droughts. These metrics are calculated as follows: define a as the number of events when drought occurs in both the forecast and observation, b as the number of drought occurrences in the forecast but not in the observation, and c for when drought occurs in the observation but not in the forecast. Then, the hit rate is $a/(a+b)$, and the false alarm ratio is $b/(a+b)$. These are calculated for each ensemble member and then aggregated across the ensemble.

Skill of the forecasts of drought events is also evaluated in probabilistic terms taking into account the uncertainty in the forecasts from the individual model ensemble members using the Brier Score (BS), which measures the average square error of a probability forecast. This is equivalent to the mean square error of a deterministic forecast, but the forecasts are given in probabilities (i.e the fraction of ensemble members that forecast the drought): $BS = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2$, where N is the number of months, o_i is the binary indicator for the observations (1 if the event occurs; 0 if the event does not occur), and f_i is the probability of occurrence for the forecast (fraction of ensemble members).

2.2.3. Development of a Multi-model Ensemble

Finally, we demonstrate the potential for improving the forecast skill by combining forecasts from individual models. As noted in the introduction, model averaging will generally lead to improved skill and we take this further by combining individual models based on their skill, such that a more skillful model will contribute more to the multi-model ensemble. These results are compared to simple model averaging. In the results section, we first show the improvement in skill from an equally weighted multi-model mean (MMM, unweighted mean), and then show how this can be improved further by taking into account the skill of individual models (MMWM, weighted mean). The weights applied to each model are calculated based on the covariability of individual model forecasts with the observational data, such that $w_{c,i} = \frac{\rho_{c,i}}{\sum_{c=1}^{N} \rho_{c,i}}$, where $w_{c,i}$ is the weight given to model c at grid cell i, and ρ is the correlation between the model and observation.

We are focused on drought and potentially other extremes, which have timescales of monthly or less (e.g. rapid intensification or recovery), and averaging forecasts (unweighted or weighted) does not retain the inherent uncertainty in forecasts and smooths out the actual day to day variations that happen as part of weather variability. This makes it difficult to represent drought evolution and other extremes properly. Averaging the climate forecasts also makes it difficult to develop the daily forcings for the hydrological model. Therefore, the development of a multi-model ensemble mean will be used to demonstrate the improved skill, whilst in a practical sense, the multi-model ensemble will be implemented in an operational system as a set of individual forecasts that are sub-sampled from the individual forecasts of each model based on the skill weighting of each model. The multi-model ensemble will therefore contain more ensemble members from a skillful model compared to a less skillful model.

2.2.4. Development of Hydrological Forecasts

The hydrological forecasts are based on output from the HBV+RAPID hydrological modeling framework. This framework couples the HBV hydrological model with the RAPID river routing model to provide forecasts of hydrological variables (e.g. soil moisture, runoff, evapotranspiration, streamflow). The HBV model is run at daily and 0.05° resolution over the whole Nile basin. Runoff from the HBV model is routed down the river network by the RAPID

model, which simulates streamflow on a very high definition vector river network with predictions made at 69,744 river reaches. Modeling the whole of the Nile basin is necessary to predict flows from the White Nile into the EN basin. An historic simulation of HBV+RAPID has been run for 1979–2018, forced by the MSWEP precipitation and MSMet meteorology data, and this forms the observational historic benchmark for comparing the hydrological seasonal forecasts. This dataset is described in the report for Component 1 (Part 2).

The hydrological forecasts are forced by downscaled and bias-corrected climate model forecasts for the hindcast period (1993–2016), and these are evaluated against the observational simulation to determine the skill of the forecasts for different hydrological variables, lead times, and for the different climate models. Downscaling of the daily climate forecasts is carried out using simple bilinear interpolation from the 1.0° resolution of the climate model data to 0.05°. Bias-correction is then carried out to remove the biases in the precipitation and temperature forecasts at the daily, 0.05° scale. Bias-correction of temperature is done by simple removal of biases on a monthly basis, by developing correction factors for each 0.05° grid cell and each month of the year as $BC_{m,i} = T_{o,m,i} - T_{c,m,i}$, where BC is the bias correction factor for month of the year (m) and grid cell (i), and T is the temperature for the observation (o) and climate model (c). The bias correction factor is added to each daily temperature value in the forecast to remove the bias. For precipitation, empirical cumulative distribution functions (CDF) are developed for each grid cell and month of the year for the observation and for the forecast, and these are used to map the forecast quantile of each day to the equivalent quantile for the observation. Mathematically, this is equivalent to: $P_{d,i} = F^{-1}_{o,m,i}\left(F_{c,m,i}\left(P_{c,d,i}\right)\right)$, where P is the precipitation for day (d) and F is the CDF for the month and grid cell. This approach is preferable over a simple monthly scaling approach (as done for temperature) as it reduces the overestimation of the number of low rainfall days in the climate models.

Each forecast is initialized on the first of the month using the hydrological states (soil moisture, canopy storage, channel storage) of the previous day taken from the historic observational simulation. The simulation is run for 6 months using the downscaled and bias-corrected climate forcing data from each of the 5 climate models and their ensemble members, for each year and month in the hindcast period (1993–2016). This is equivalent to 36,288 hindcast simulations, however, as there are missing data in the CDS archive, the actual number of the simulations is about ⅔ of that. The forecasts are evaluated in the next section against the observational historic simulation. They are also evaluated against a benchmark simulation based on ESP.

3. Results

3.1. Historic Hydrology Evaluations

The historic reconstructed hydrology was evaluated against available observations, namely streamflow data from within the basin, and soil moisture observations globally (as there are no such observations within the Eastern Nile basin as far as we can tell).

3.1.1. Streamflow

Readily available observed streamflow data is limited for the Eastern Nile Basin (for example from the Global Runoff Data Centre), and so evaluations are necessarily focused on records for a handful of stations available from ENTRO and national agencies. Evaluations are carried out at daily and monthly scale. Figure 8 shows the location of nine stations with available data. Results at daily and monthly scale are shown in Table 3. The results show reasonable performance with correlations ranging from 0.4 to 0.83 at the daily scale and 0.46 to 0.92 at the monthly scale. Note that no direct calibration of the model was carried out in the basin because of the lack of long-term observed data across the basin, and the simulations are reliant on the suitability of the regionalized parameter values described above. Future work will potentially focus on improving model performance via further detailed evaluation and calibration against additional observed data if available. This will certainly reduce biases in the model, and may contribute to improvements in the other statistics.

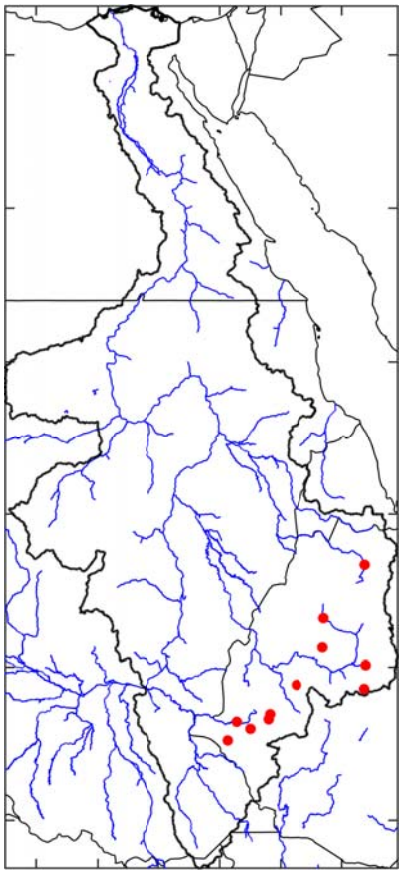


Figure 8. Location of streamflow gauging stations within the Eastern Nile Basin that are used to evaluate the HBV+RAPID hydrological modeling system.

Table 3. Streamflow stations and model performance statistics. KGE is Kling-Gupta Efficiency (Kling et al., 2012), CC is correlation coefficient, α is variability ratio, and β is bias ratio.

	Location			Daily Model Statistics				Monthly Model Statistics			
Station	Lat	Lon	Period	KGE	CC	α	β	KGE	CC	α	β
Baro nr Gambella	8.23°	34.57°	1980–2009	0.59	0.83	1.17	1.34	0.72	0.92	1.17	1.2
Gilo nr Pignudo	7.62°	34.27°	1980–1990	-0.5	0.73	2.46	1.21	-0.49	0.87	2.48	1.17

Sor nr Metu	8.32°	35.60°	1980–2006	0.71	0.76	0.87	1.09	0.84	0.92	0.87	0.97
Gebba nr Suppe	8.48°	35.65°	1980–2005	0.44	0.7	1.47	1.03	0.49	0.83	1.48	0.92
Abbay nr Bahidar	11.60°	37.41°	1980–2002	-0.12	0.4	1.48	1.82	0.02	0.46	1.48	1.67
Wechit nr AlemKetema	10.08°	38.78°	1997–2010	-49.46	0.59	51.44	2.18	-49.11	0.68	51.1	1.97
Anger nr Nekemet	9.43°	36.52°	1994–2004	0.33	0.66	1.43	1.4	0.47	0.76	1.42	1.21
Birr nr Jiga	10.65°	37.38°	1980–2002	0.67	0.69	0.91	0.95	0.85	0.88	0.91	1
Muger nr Chancho	9.30°	38.73°	1980–2002	0.52	0.75	0.6	0.91	0.59	0.92	0.6	0.92
Tekezze nr Yechi	13.35°	38.75°	1994–2003	0.61	0.7	0.87	0.79	0.75	0.86	0.88	0.83
Upper Baro nr Masha	8.00°	35.00°	1988–2007	0.37	0.54	0.58	0.92	0.41	0.66	0.57	0.7

3.1.2. Soil moisture

We assessed the temporal dynamics of the HBV model forced by the MSWEP precipitation against 14 state-of-the-art global surface soil moisture products, including five based on satellite retrievals, four based on "open-loop" models, and five based on models that assimilate satellite observations. As reference, we used in-situ soil moisture measurements at approximately 5-cm depth from 949 probes globally, due to a lack of measurements in the Eastern Nile Basin. Figure 9 presents the results of our assessment. Among the three single-sensor satellite-based products (AMSR2, SMAPL3, and SMOS), the L-band-based SMAPL3 performed best by a significant margin. Among the two multi-sensor satellite-based products (TC and ESACCI), the triple collocation-based product TC achieved superior performance and outperformed SMAPL3 as well. The performance ranking of the four open-loop models (GLDAS, HBV-ERA5, HBV-IMERG, and HBV-MSWEP) is consistent with previous precipitation dataset evaluations (e.g., Beck et al., 2017b, 2019c). HBV forced with MSWEP precipitation achieved the best performance not just among the open-loop models, but among all 14 products, justifying our choice of hydrological model and precipitation dataset for the present application. The five models that assimilate satellite observations (GLEAM, SMAPL4, HBV-ERA5+SMAPL3, HBV-IMERG+SMAPL3, and HBV-MSWEP+SMAPL3) generally outperformed the open-loop models and exhibited a smaller spread in performance.

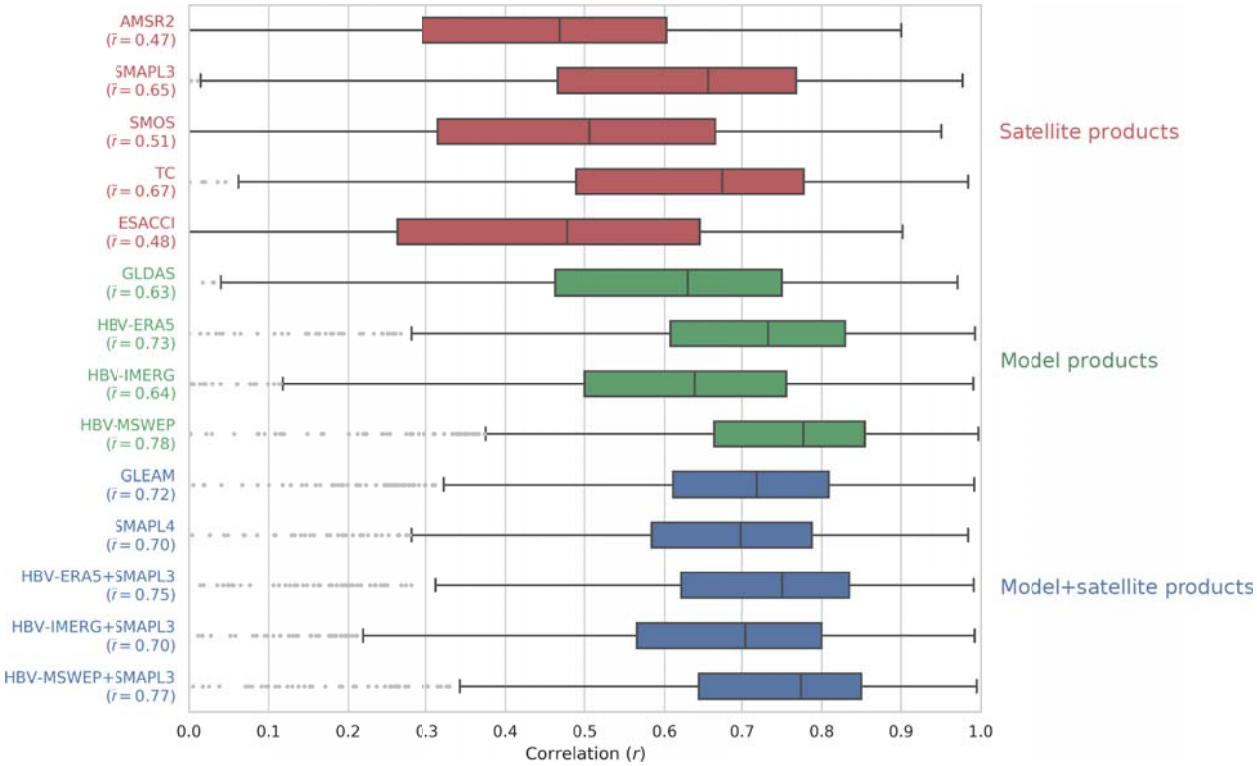


Figure 9. Performance of surface soil moisture products in terms of daily temporal correlation for the period May 25, 2015 to September 29, 2019. The left and right edge of the box represents the 25th and 75th percentile values, respectively, the line in the box represents the median value, the whiskers show the rest of the distribution (except outliers), and the grey points depict the outliers.

3.2. Drought Analysis

Drought is defined in general as a deficit in water relative to some threshold, and can be estimated for meteorological, agricultural and hydrological conditions. Typically a drought index is used to quantify the characteristics of drought, usually in terms of its severity relative to a drought threshold. There are hundreds of different indices that have been proposed for different types of drought, regions and applications. A discussion of these is given in various reviews (e.g. Mishra and Singh, 2010; Sheffield and Wood, 2011; Zargar et al., 2011). We focus on standard indices to represent each of these drought types: the Standardized Precipitation Index (SPI) at each grid cell for meteorological drought, soil moisture percentiles (SMp) at each grid cell for agricultural drought, and standardized flow volumes (StdVol) at each river reach for hydrological drought. These indices are used frequently for research and applications, have a long history of evaluation, and are relatively straightforward to calculate and interpret. The SPI is the recommended index for meteorological drought by the World Meteorological Organization (WMO). The SMp has been used by many authors to understand drought variability and the impacts of climate change (e.g. Sheffield and Wood, 2008; Berg and Sheffield, 2019). Standardized flow volumes are a recommended index for hydrological drought (van Lanen et al., 2004). As no index can by itself represent all aspects of drought, especially in the diverse and variable climates of the Eastern Nile basin, the use of these three indices provides good coverage of the variability of drought and its propagation through the hydrological system, from precipitation, to soil moisture, to streamflow. Results are presented here on agricultural and hydrological drought, as these have the most direct impacts on agriculture and water resources, respectively.

3.2.1. Drought variability

We first show average drought conditions as a monthly time series over the whole Eastern Nile Basin. Figure 10 presents the time series of soil moisture for the historic period, 1980–2018. Figure 10 also shows the area in drought based on the SMp, with a threshold of the 20th percentile. Key events are identified including droughts in 1983/84, 1991/92, the early 2000s, 2008/09 through to 2013, and most recently in 2015 and 2017 for the whole EN basin. The most extensive drought events reach a peak coverage of the basin of about 30% at the monthly time scale. Figures 11a-c present the same analysis but for the (a) Abay-Blue Nile, (b) Baro-Akobo-Sobat-White Nile, and (c) Tekeze-Setit-Atbara sub-basins. Again, events in 1983/84, 1991/92 are seen in the time series, but with weaker events during the mid 2000s. The last 10 years indicate generally drier than normal conditions (except for around 2014 and 2018) across these sub-basins.

Bayissa et al. (2018) and Kebede et al. (2019) analyzed drought over the upper Blue Nile using a range of indices and datasets including those based on soil moisture. Bayissa et al. (2018) looked at the period 1970-2010 for a set of six physical drought indices, and compared with impact data from the Emergency Events Database (EM-DAT), the international disaster database (http://www.emdat.be/database). EM-DAT identifies several events including 1983-1984, 1994-95, 2003-2004 and 2009. Kebede et al. (2018) further identified 1981 and 2015 as major drought events, based on summertime values of precipitation and soil moisture. 1983-84 is also noted here as one of the severest events on record over the EN basin, and most other sub-basins. The 1994-95 drought was likely part of a longer-term 1989-1995 drought, which is replicated here for SMp in most regions, including the Abay-Blue Nile sub-basin (Figure 11a). 2003-2004 is depicted here as a severe event over some sub-basins (Figures 11a-c) and 2009 is a large

event in most regions. 2015 is also picked up by SMp in some regions but 2014/2015 is a neutral/wet year in other regions. 2016-17 is quite a strong drought period in the Baro-Akobo-Sobat-White Nile sub-basin (Figure 11c).

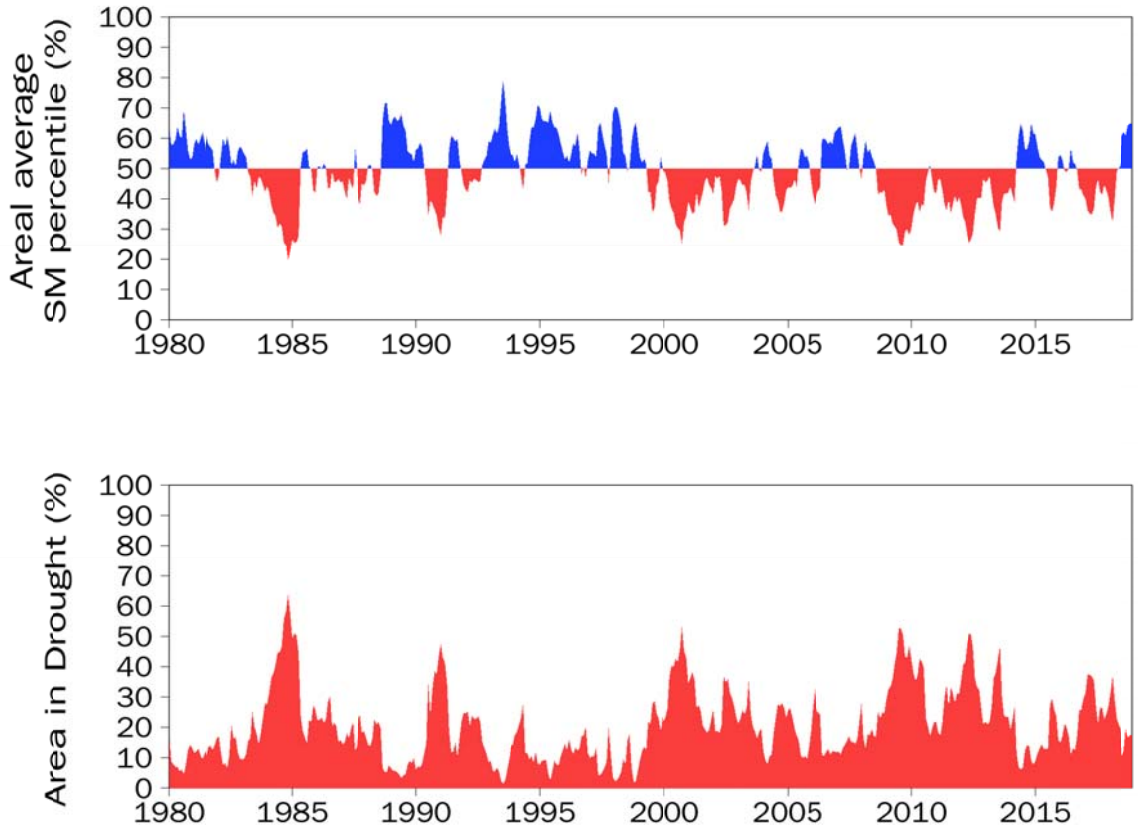


Figure 10. Time series of (top) areal average soil moisture percentile, and (bottom) area in drought for a threshold of the 20th percentile for the Eastern Nile Basin.

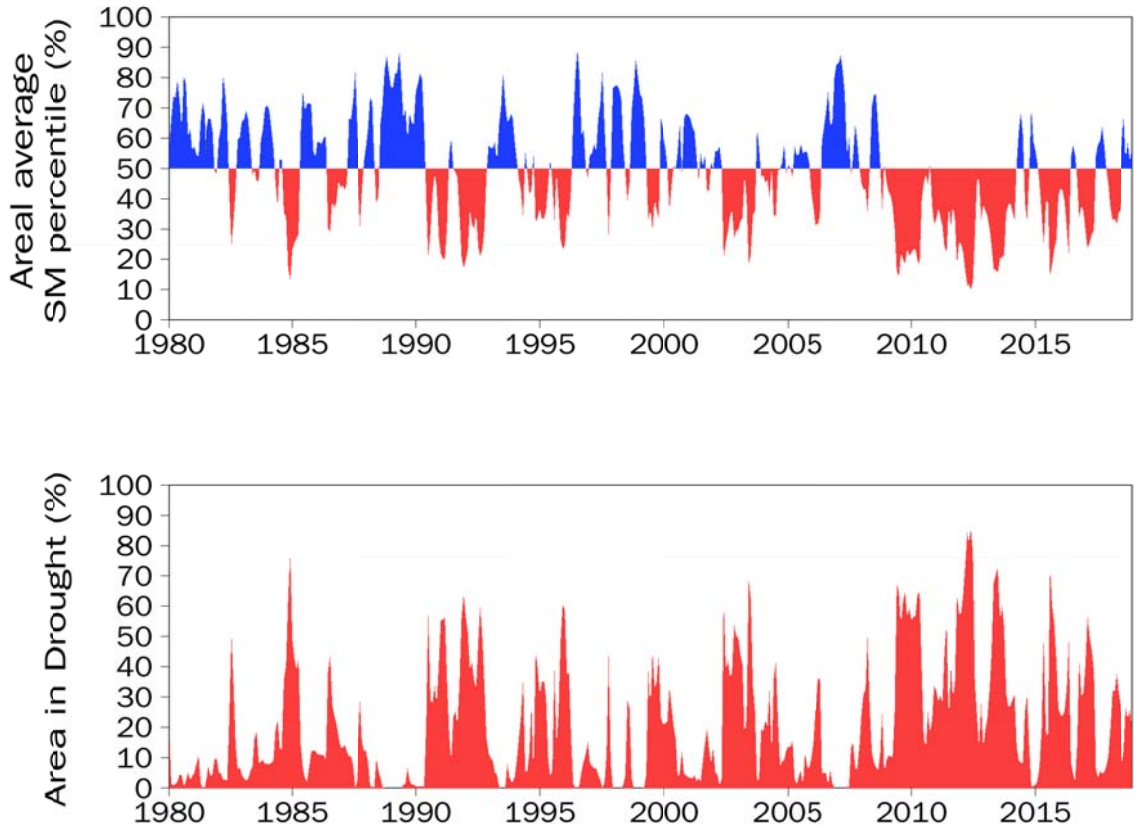


Figure 11a. Time series of (top) areal average soil moisture percentile, and (bottom) area in drought for a threshold of the 20th percentile for the Abay-Blue Nile sub-basin.

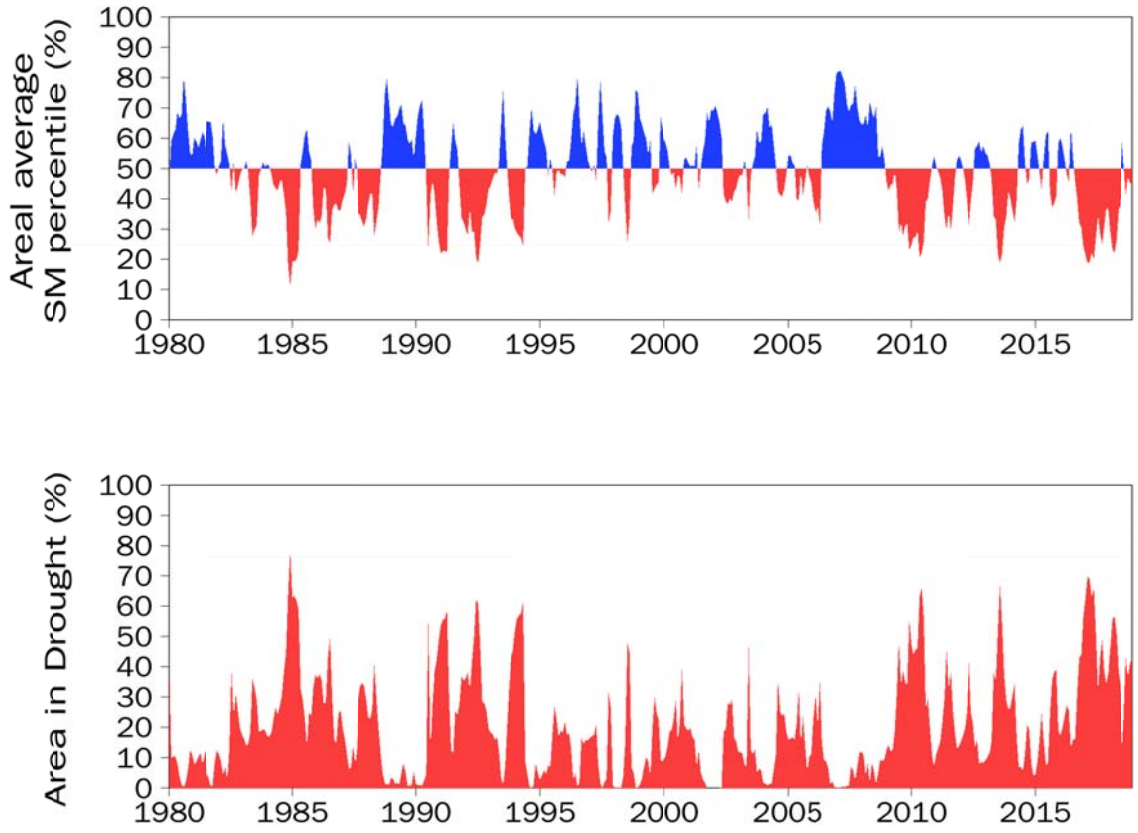


Figure 11b. Time series of (top) areal average soil moisture percentile, and (bottom) area in drought for a threshold of the 20th percentile for the Baro-Akobo-Sobat-White Nile sub-basins.

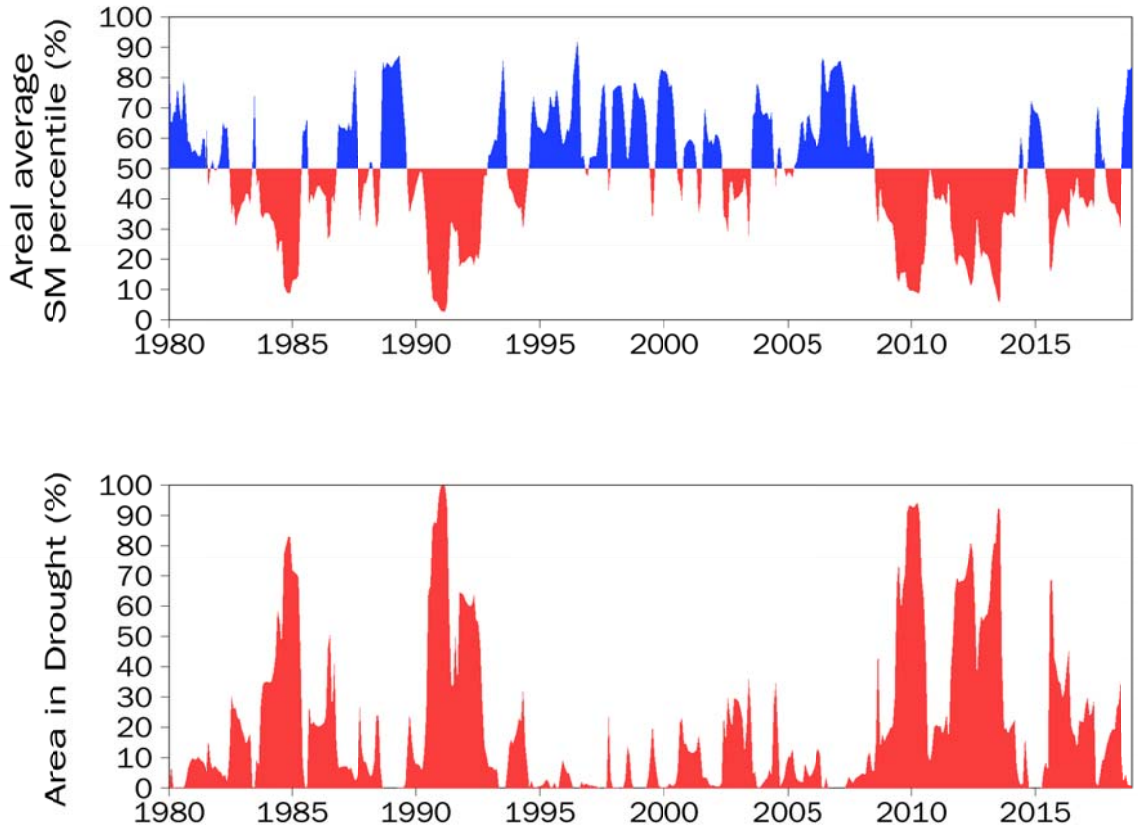


Figure 11c. Time series of (top) areal average soil moisture percentile, and (bottom) area in drought for a threshold of the 20th percentile for the Tekeze-Setit-Atbara sub-basin.

3.2.2. Drought risk

Drought risk is quantified based on the frequency of event occurrence. Figures 12-15 show maps of frequency and return period based on SMp for different drought durations (1–3, 4–6, 7–12 and >12 months duration), and for different drought thresholds (10 and 20th percentiles). Return periods are estimated based on the drought frequencies over the 39 year time period (1980-2018), and are quite uncertain for the low risk events which are infrequent or do not occur in the time period. Each of these types of drought will likely have different impacts due to their duration, severity and timing. Shorter duration droughts (1–3 months) are more frequently occurring (low return period and therefore high risk), and most prevalent in the upper basin where hydrological variability is higher in the relatively wetter climate. In this region, for the 10th percentile threshold (Figures 12-13), the number of droughts exceeds 10 events over the ~39 year time period (1980–2018), which is equivalent to about one event every four years. The frequency reduces with increasing drought duration (e.g. 4–6 months, and 7–12 months) with a maximum of about 3–5 events in total (or an approximate return period of 7–13 years). The area of maximum frequency shifts northwards with increasing drought duration as driven by the wet to dry gradient and associated stronger climate persistence. Droughts of 12 or more months in duration are restricted to a few locations in the northern, very dry part of the basin. Droughts defined by the less severe 20th percentile threshold (Figures 14-15) occur more often by definition, with 1–3 month duration droughts occuring up to around every two years in the southern part of the basin (Figure 15).

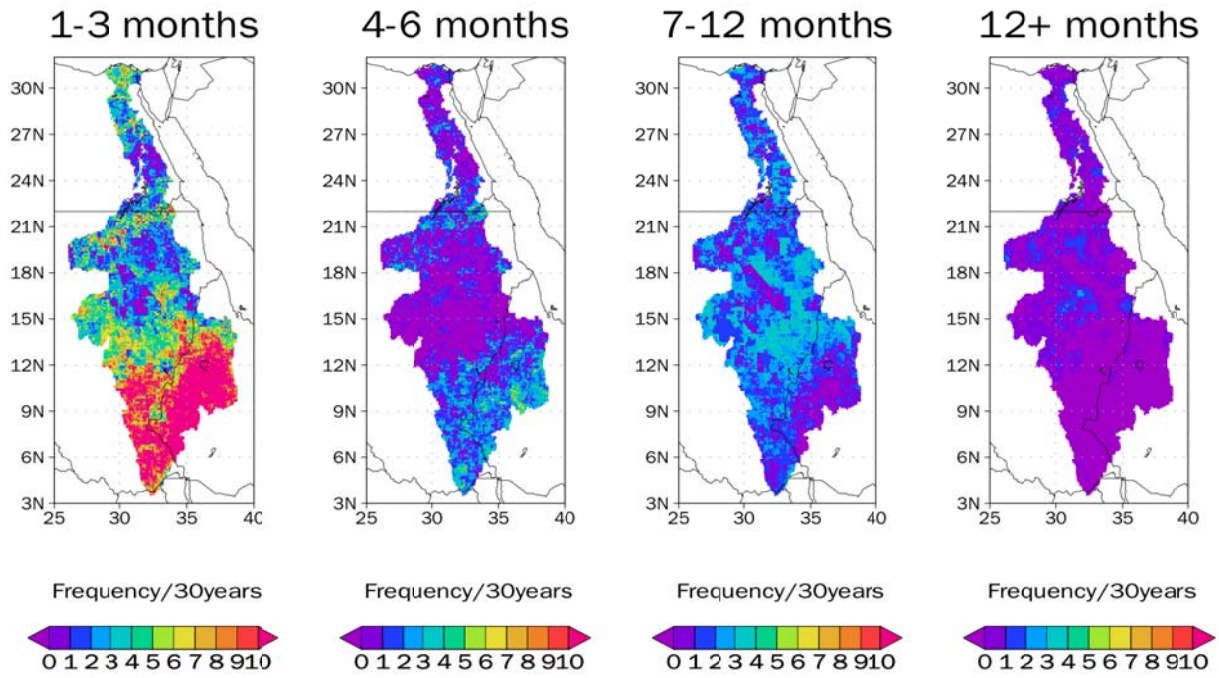


Figure 12. Frequency of different duration drought events, based on soil moisture percentiles for a 10th percentile threshold over the historic period of 1980-2018. Drought events are identified at 1–3, 4–6, 7–12 and >12 months duration.

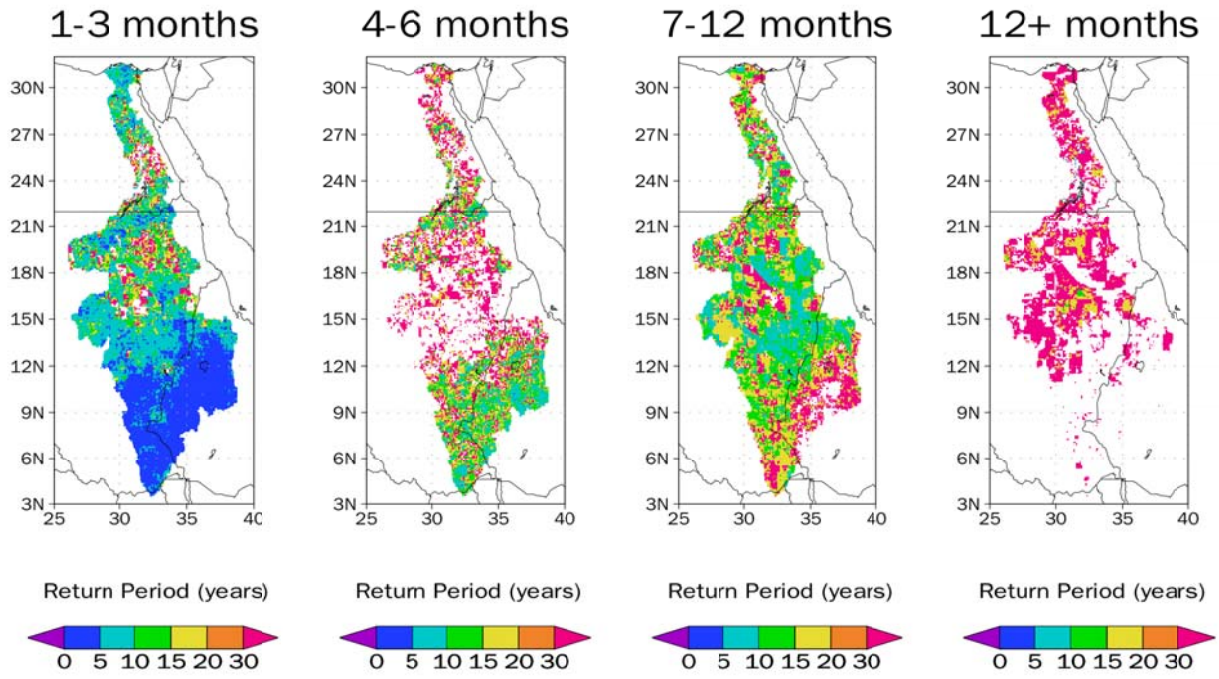


Figure 13. Same as Figure 12 but for the return period.

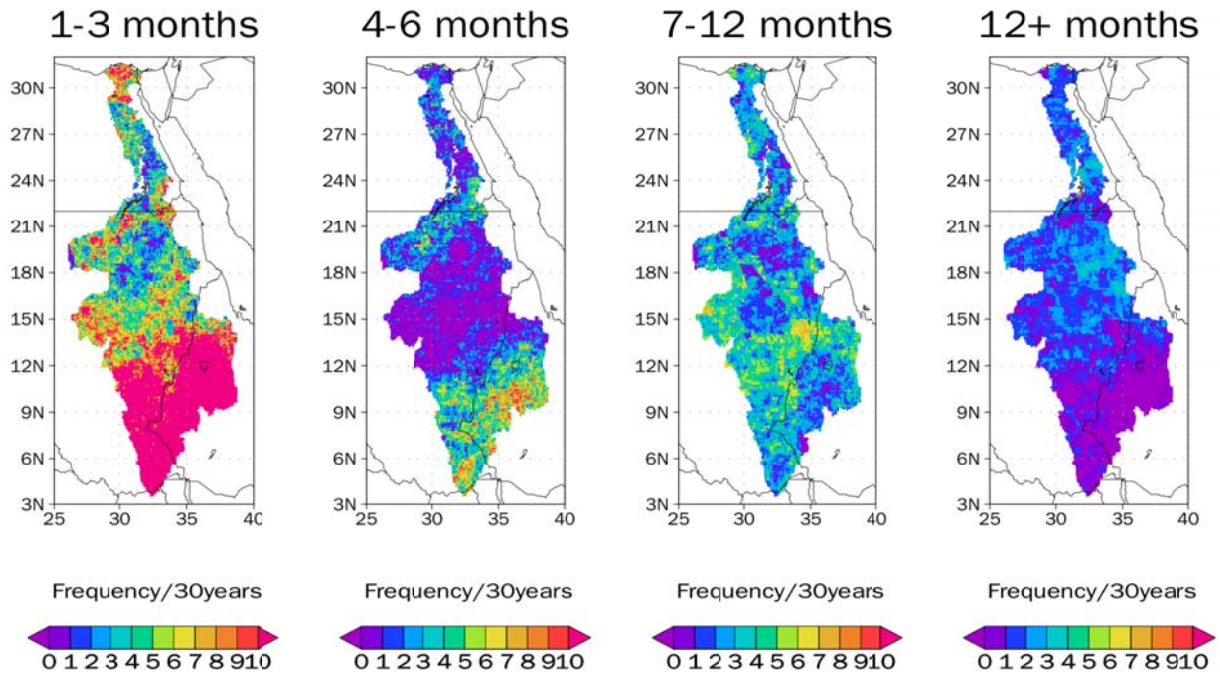


Figure 14. Same as Figure 12 but for the 20th percentile threshold.

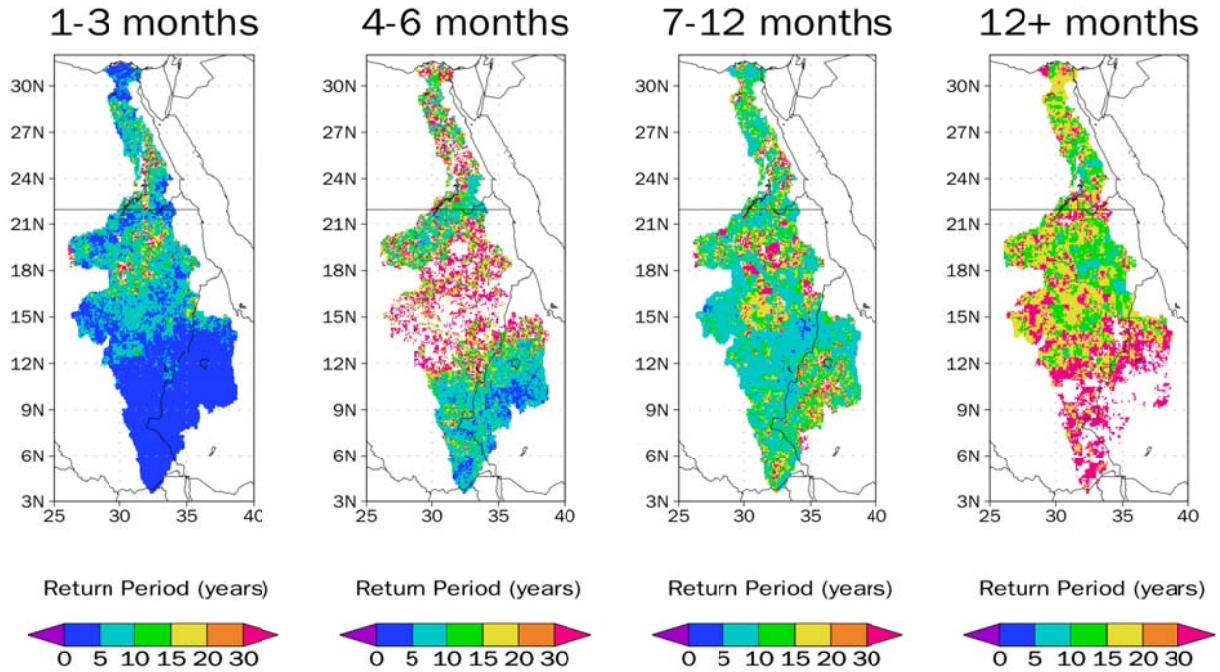


Figure 15. Same as Figure 13 but for the 20th percentile threshold.

Hydrological drought is presented and analyzed slightly differently to focus on the daily time scale, and for the number of days that are exceeded for 80% of the time (Figure 16) and the cumulative deficit below this exceedance value (Figure 17). The results are also shown for three decadal periods to highlight any changes over time. The drought duration analysis indicates a maximum in the middle part of the basin (> 85 days), where there is the largest convergence of flows and the climate tends to be less variable. There is a slight increase in the area of highest durations (> 85 days) over time, which is consistent with the evolution of conditions for the Tekeze-Atbara and Blue Nile regions seen previously. Hydrological droughts in the southernmost and eastern part of the basin are less than about 50 days, and as low as less than 30 days in the Victoria basin. The cumulative deficits (Figure 17) show a clear north-south gradient, with higher deficits in the headwaters of the basin and the highest deficits along the main stem in absolute terms.

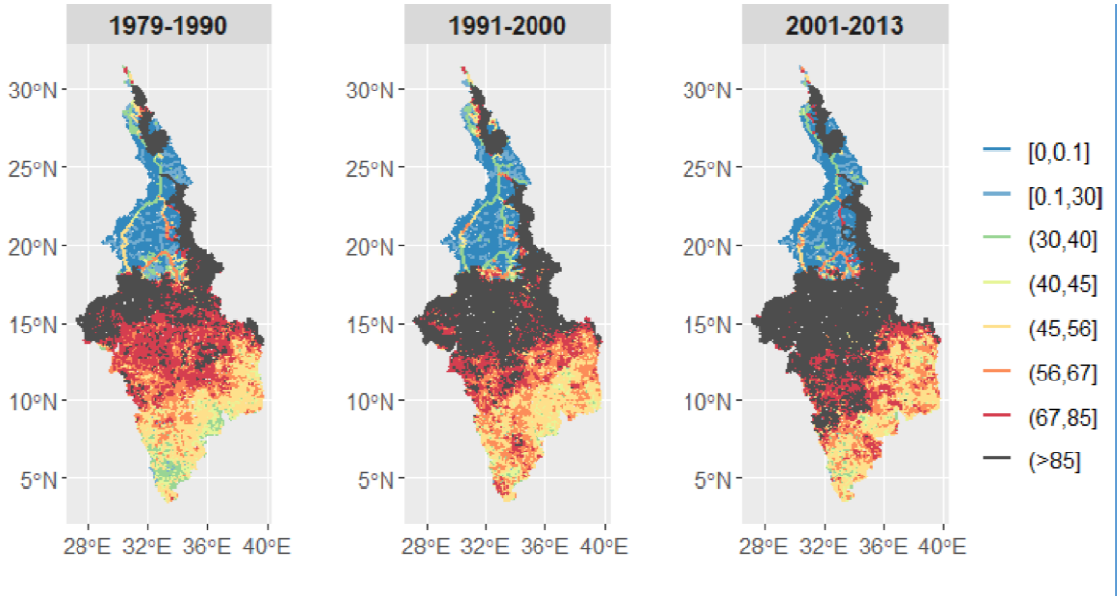


Figure 16. Hydrological drought duration (days) for flows exceeded 80% of the time, shown for three near-decadal time periods between 1979 and 2013.

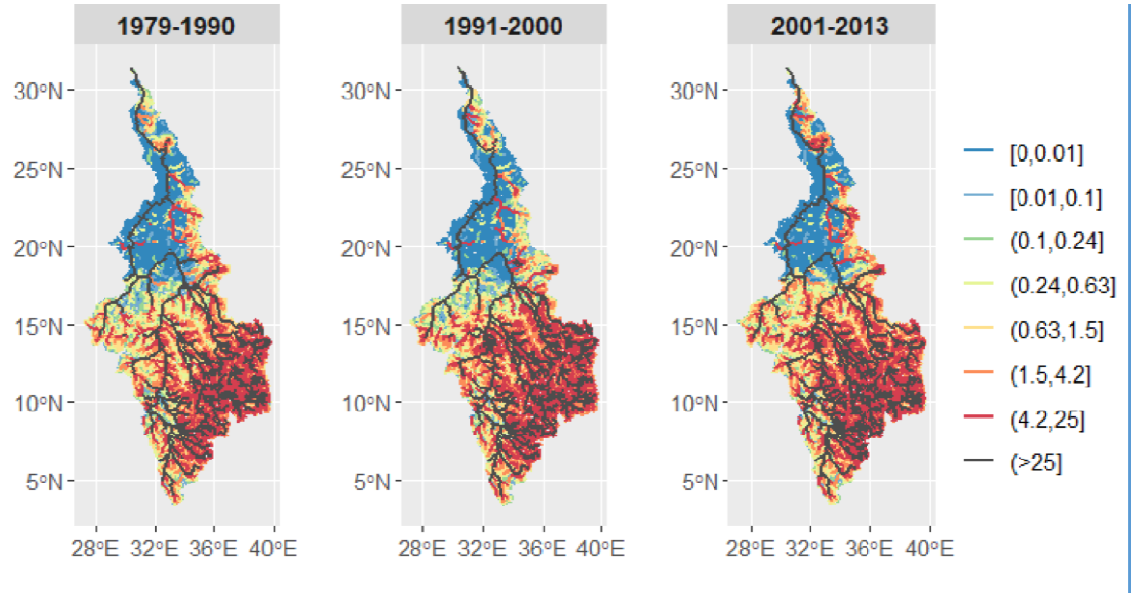


Figure 17. Cumulative daily streamflow deficit (m) computed from the flow duration curve as flow exceeded 80% of the time, and shown for three near-decadal time periods between 1979 and 2013.

3.2.3. ENSO impacts

The impact of ENSO on soil moisture is shown in Figure 18. ENSO is known to impact across the basin, in particular over the Ethiopian Highlands in all seasons (e.g. Camberlin and Philippon 2002; Seleshi and Zanke, 2004; Diro et al. 2008; Abtew et al., 2009; Segele et al. 2009) and more generally across the Nile basin (Awange et al., 2014), although the sign and strength of the relationship varies with region and season (Gleixner et al., 2017). Climate model experiments show that up to 50% of the summertime (Kiremt) rainfall anomalies over the Ethiopian Highlands is related to variability of SSTs in the equatorial Pacific.

Figures 18a-c show composite maps of SMp for neutral, strong El Niño and strong La Niña years, which are shown in Table 4. The SMp composite is calculated for 6-month periods (Oct-Mar for the ENSO year and the following year, Jan-Jun for the following year and Apr-Sep for the following year). For example, for the strong El Niño of 1982-83, SMp for Oct-Mar is averaged over Oct-Dec 1982 and Jan-Mar 1983; the Jan-Jun value is calculated over Jan-Jun 1983 and so on. The values for 1982-83 are then averaged with the values for all other strong El Niño years to form the composite.

ENSO and neutral years are defined using the NOAA CPC Oceanic Niño Index (ONI) (https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php), which is the 3 month running mean of ERSST.v5 SST (Huang et al., 2017) anomalies in the Niño 3.4 region (5°N-5°S, 120°-170°W). Years are classified based on identification of events defined as 5 consecutive overlapping 3-month periods at or above a SST anomaly threshold of +-0.5°C, with strong events defined with a threshold of +-1.5°C.

Table 4. Years classified as neutral, strong El Niño and strong La Niña, as based on the NOAA CPC ONI.

Neutral	1980 1981 1985 1989 1990 1992 1993 1996 2001 2003 2012 2013 2019
Strong El Niño	1957 1965 1972 1982 1987 1991 1997 2015
Strong La Niña	1973 1975 1988 1998 1999 2007 2010

The impact of ENSO is reasonably clear across the basin, with El Niño years related to generally lower SMp in the middle and southwestern parts of the basin, and La Niña years associated with wetter conditions, mostly in the middle part of the basin. There is a strong dry signal in El Niño years in the lower parts of the three sub-basins (Abay-Blue Nile, Baro-Akobo-Sobat-White Nile, and Tekeze-Setit-Atbara sub-basins), and a less well defined wetter signal during La Niña years. The signal in the Ethiopian highlands is less clear, but there is a tendency for the composites to be wetter in La Niña years, which is consistent with previous studies. The El Niño signal is less obvious. The strength of the ENSO signal across the basin diminishes from ONJDFM through to JFMAMJ and AMJJAS, which reflects the tendency of ENSO and its teleconnections to be strongest in the boreal winter. There is a wet-dry signal with El-Niño/La-Niña in the Nile Delta, although it should be noted that this is based on modeled SM and not the influence of the Nile River on the delta wetlands.

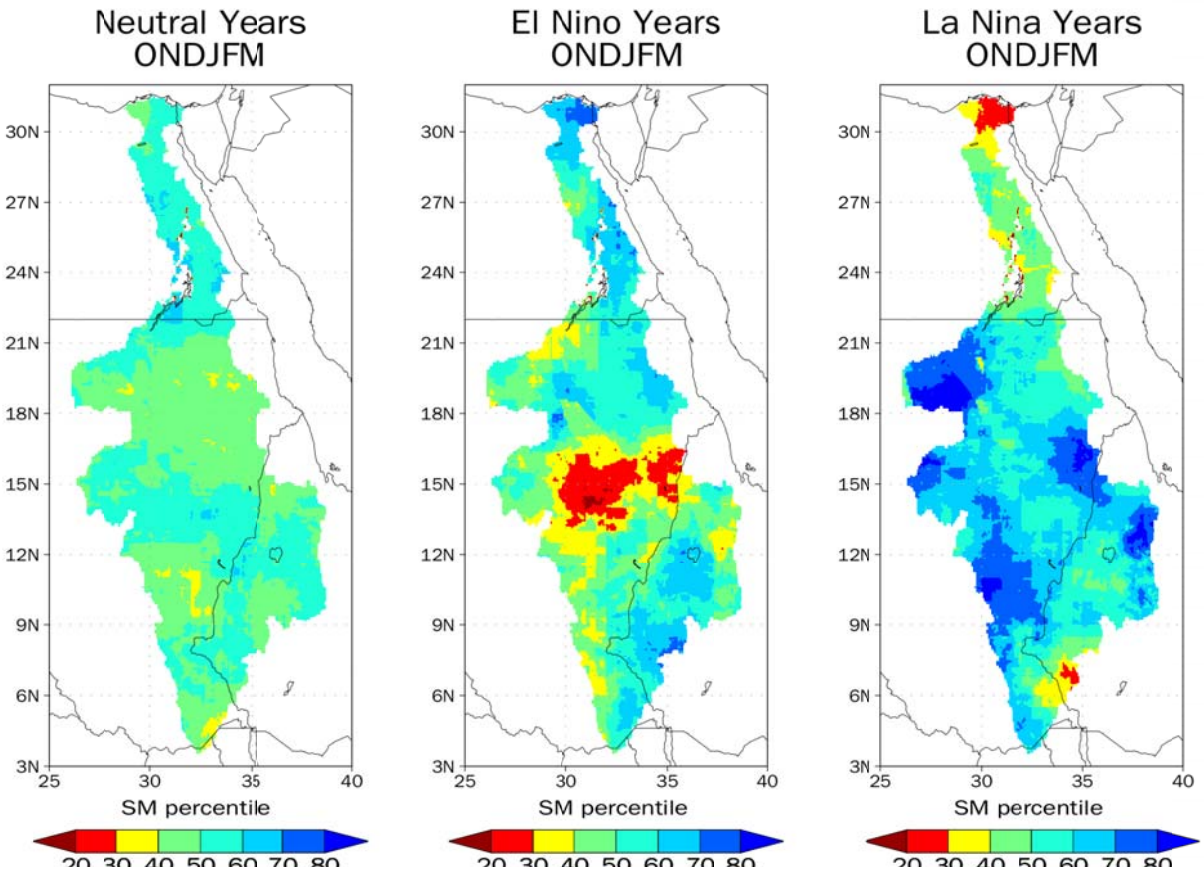


Figure 18a. ONDJFM composite SMp over 1979-2018 for (left) neutral ENSO years, (middle) strong El Niño years and (right) strong La Niña years.

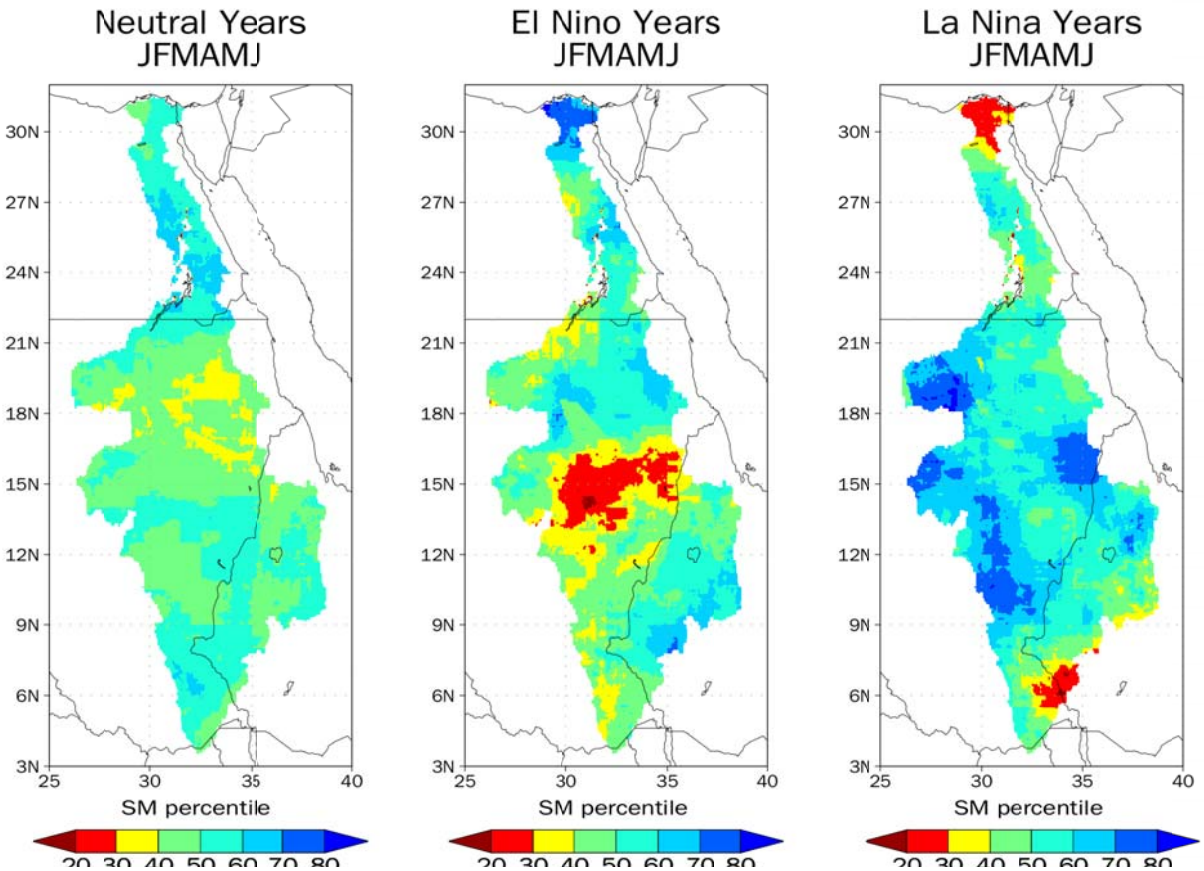


Figure 18b. JFMAMJ composite SMp over 1979-2018 for (left) neutral ENSO years, (middle) strong El Niño years and (right) strong La Niña years.

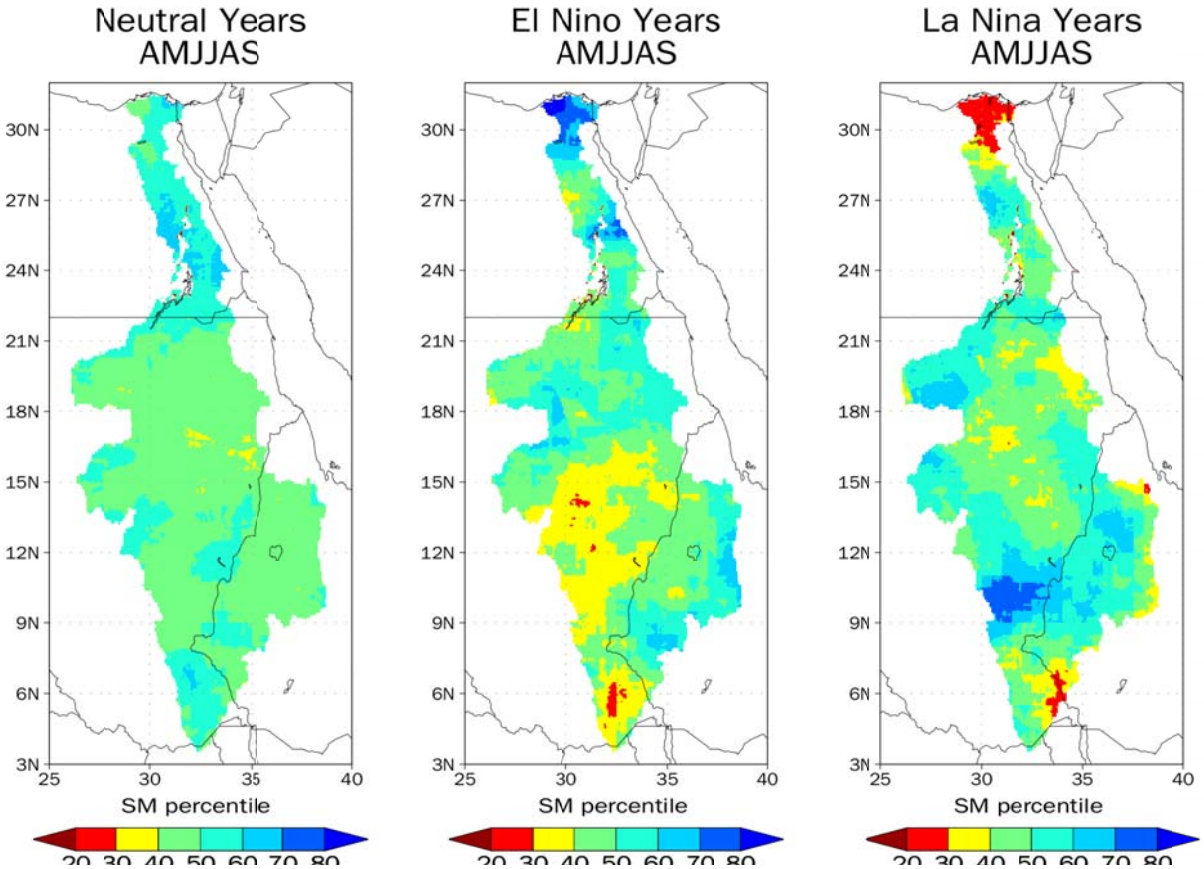


Figure 18c. AMJJAS composite SMp over 1979-2018 for (left) neutral ENSO years, (middle) strong El Niño years and (right) strong La Niña years.

3.2.4. Potential drought exposure

This final section shows some preliminary analysis of exposure of populations and agricultural lands to drought hazard. Exposure is one component of the risk of impacts, which also includes the vulnerability (or conversely the resilience) of sectors and populations, but does provide a first order estimate of potential risks. Exposure here is based on 1) the location and number of people exposed to a certain duration drought frequency, and 2) the location and area of agricultural land (crops) exposed.

Figure 19 shows the the population of Eastern Nile Basin, the intensity of cropland (crop area percentage), and the estimated return period frequency of 1-3 month droughts (with a 20th percentile threshold). The blue areas of the drought map in Ethiopia indicate an approximate return period of 2 years

(~20 events over the 39 year period), which could be considered high risk. There is high overlap between areas of dense population and agriculture, and high short-term drought risk, especially in the districts south of Lake Tana (West/East Gojjam and AgewAwi). The intense region of cropping in southern Sudan and South Sudan are exposed to slightly lower risk of short-term droughts but have higher exposure to medium to long-term (7-12) month droughts (see Figure 14). They are also exposed to high impacts during strong El Niño events (see Figure 18) although this is most prominent in the dry season of the boreal winter (ONDJFM) when impacts will generally be lower than in the rainy season of AMJJAS.

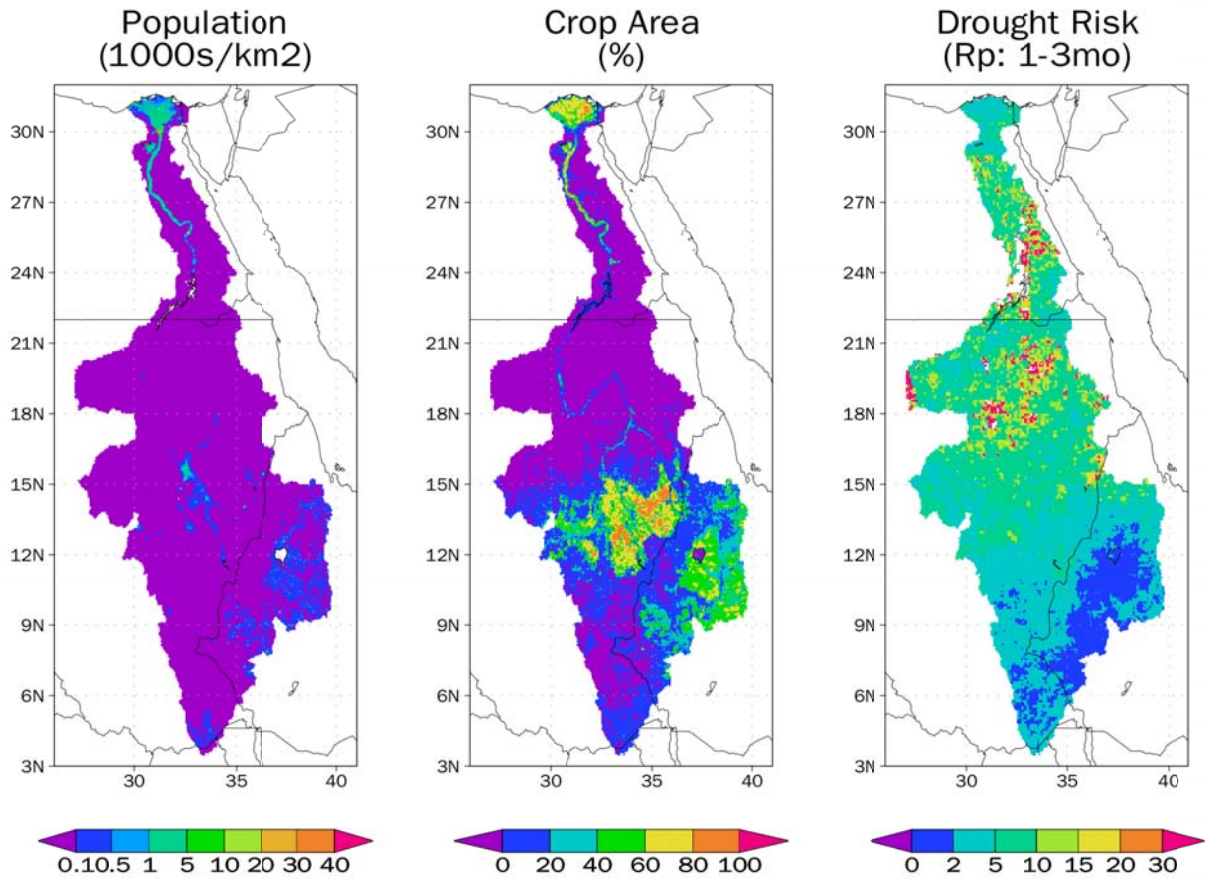


Figure 19. (left) population of the Eastern Nile Basin estimated for 2019 from the Worldpop (2018) dataset, (middle) percent crop cover based on the Copernicus Land Cover Service product (Buchhorn et al., 2019), and (right) return period of 1–3 months duration droughts based on soil moisture percentile index (SMp) and a 20th percentile drought index.

The breakdown of exposure by drought event duration and risk level is shown in Figure 20. Here, risk categories are defined in terms of the estimated return period with low: Rp> 10 years; medium: 3 <Rp< 10 years; high: Rp< 3 years. The spatial variation in risk for short-term drought is shown in Figure 19 and in Figure 15 for all drought duration types. The exposure under different drought risk categories depends on the duration, with generally higher exposure to low risk and longer duration droughts. Conversely,

there is lower exposure to medium risk droughts (orange bars) for medium (4-6 months) to long-term (7-12 months) droughts, but higher exposure for short term (1-3 months) droughts. There are no areas under high risk of medium (4-6 months) to very long droughts (> 12 months). The high-risk droughts are limited to short-term (1-3 months) droughts, but the population exposed is about 36M, and the cropped area is about 0.08M (80,000) km^2.

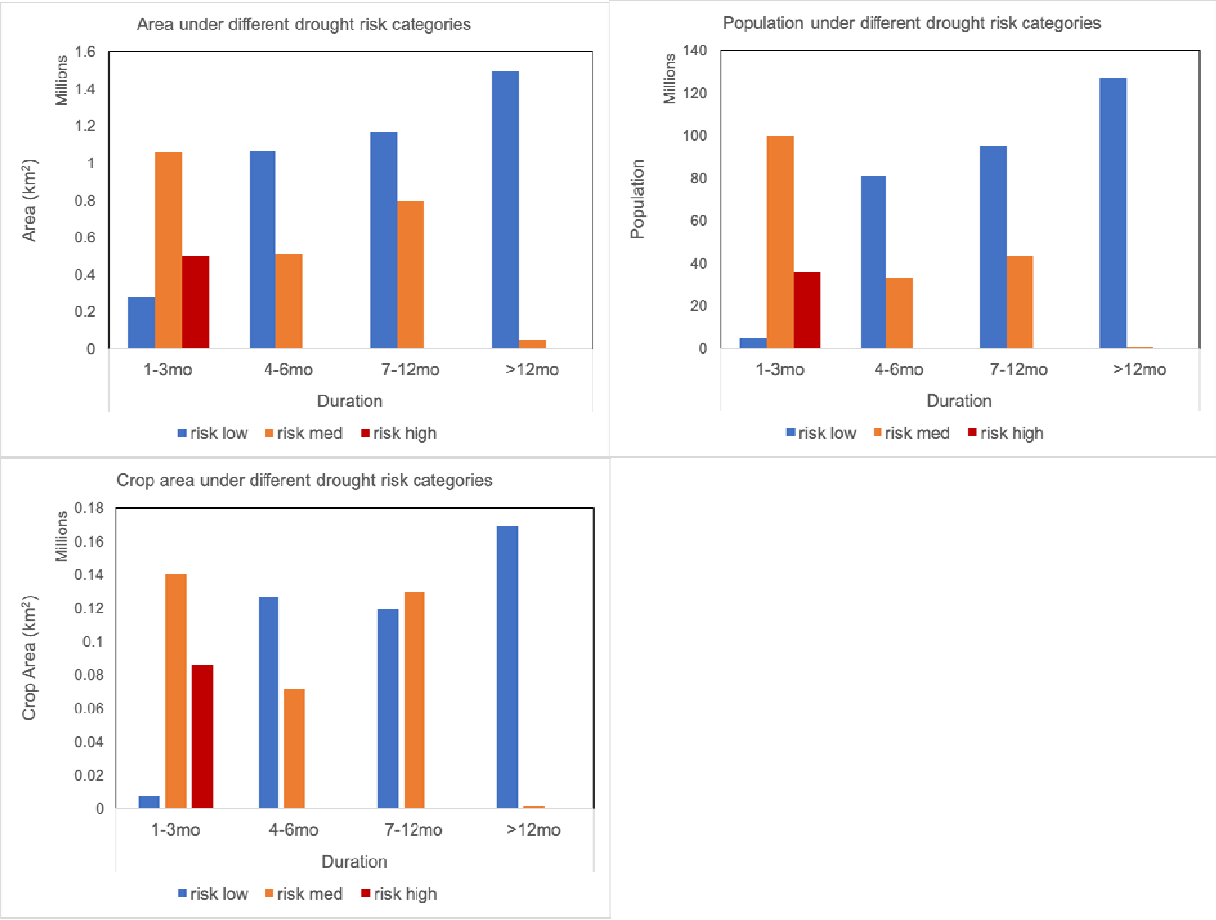


Figure 20. Area, population and cropped area under different drought risk categories for different durations with a 20th percentile threshold. Risk categories are defined in terms of the estimated return period (shown spatially in Figure 16) with low: Rp> 10 years; medium: 3 <Rp< 10 years; high: Rp< 3 years.

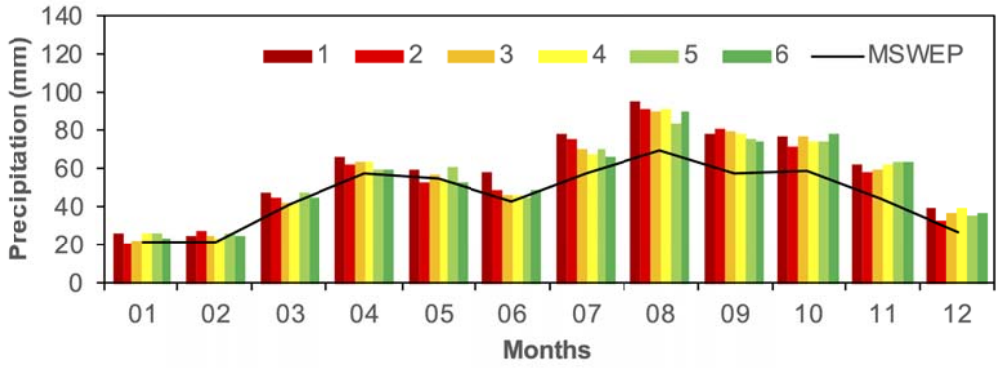
3.3. Evaluation of Climate Forecasts

3.3.1. Seasonal Climatology

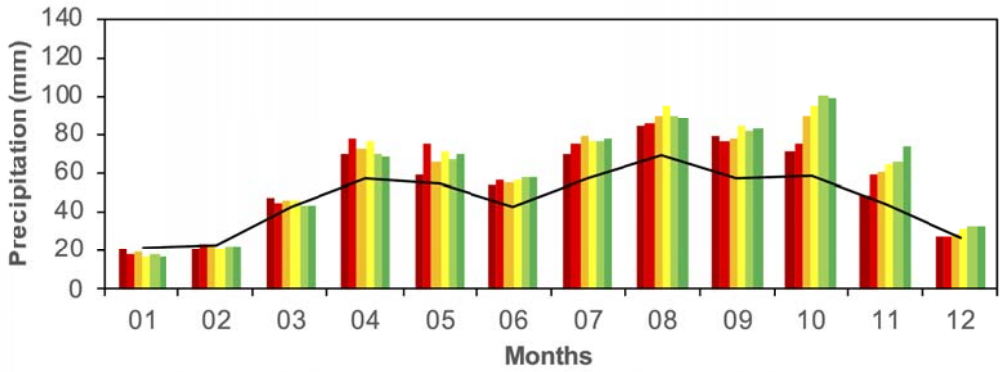
Figure 21 compares the EN basin mean seasonal cycle of observed precipitation from MSWEP with that from the five climate models. Similar figures for the sub-basins are shown in an appendix. Overall, the models reproduce the seasonal cycle of precipitation at all lead times.

Most models overestimate the observed seasonal cycle, and more so in the wetter summer months, except for the DWD model, which tends to underestimate in nearly all months and lead times. The overall bias is least in the DWD model and highest in the CMCC model (see Table 5). Although the biases tend to change with increasing lead time, they do not necessarily become larger. For example, monthly precipitation amounts tend to decrease with increasing lead time in the ECMWF model, but this can increase or decrease the bias depending on the month. This indicates that the applied bias correction should be specific to each month. See section 2.4 for details of the bias correction approach which is used to remove these biases.

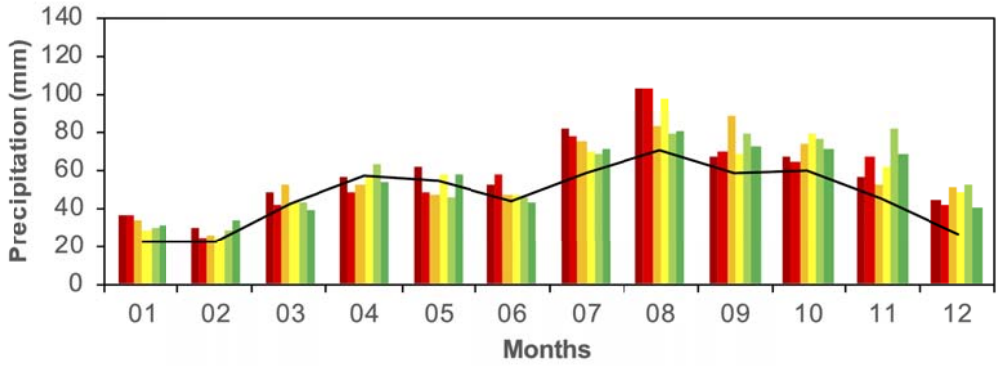
ECMWF



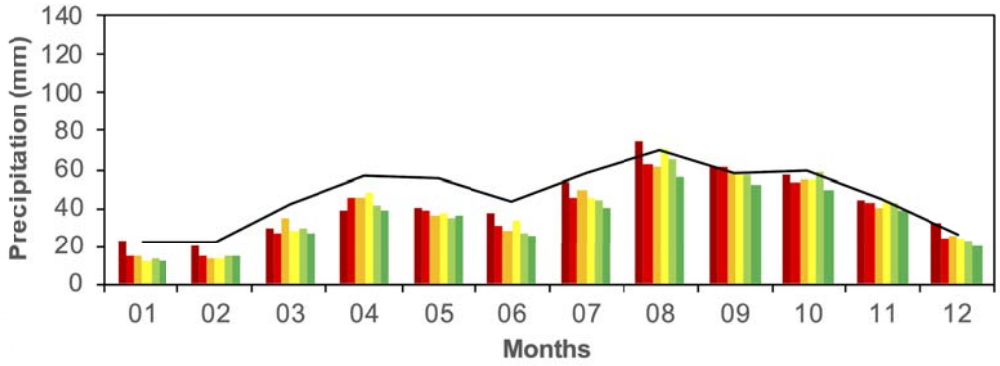
UKMO



Meteo-France



DWD



CMCC

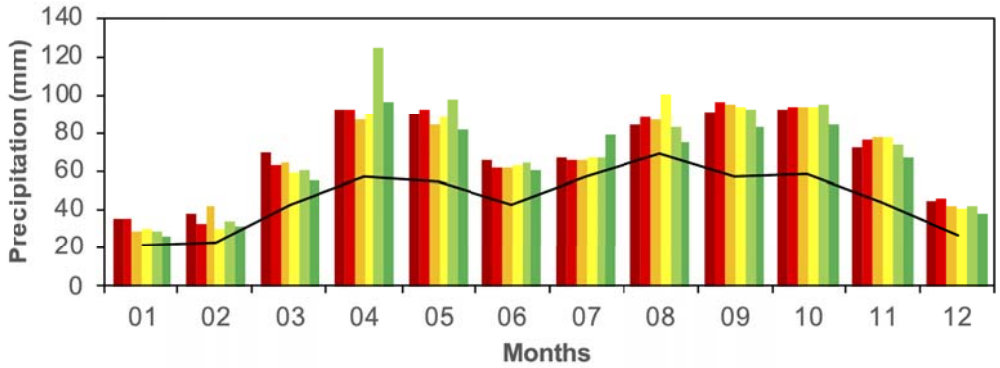
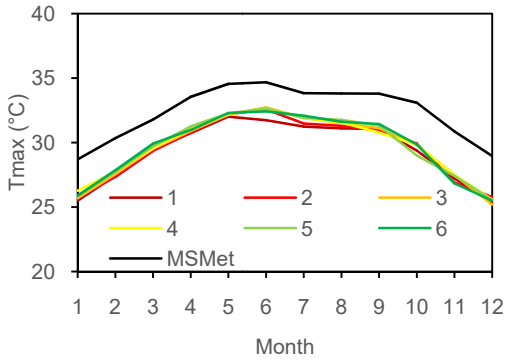


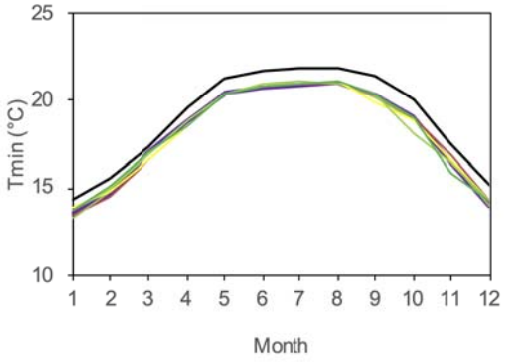
Figure 21. Precipitation monthly average climatological mean (1993–2016) averaged over the EN basin for MSWEP and the five climate models, for leads 1–6.

Figure 22 shows the mean seasonal cycle of monthly average Tmax and Tmin averaged over the EN basin for the observations (based on MSMet) and the five climate models. Similar figures for the sub-basins are shown in the appendix. All models underestimate Tmax for all months and the biases tend to be slightly larger in the dry season. Two models (ECMWF, UKMO) underestimate Tmin, and three models (Meteo-France, DWD, CMCC) overestimate Tmin. Model CMCC has the least bias annually and model ECMWF has the largest bias for Tmax. Models ECMWF, Meteo-France and DWD have the least bias annually for Tmin. Biases for the EN basin at the mean annual scale are summarized in Table 5.

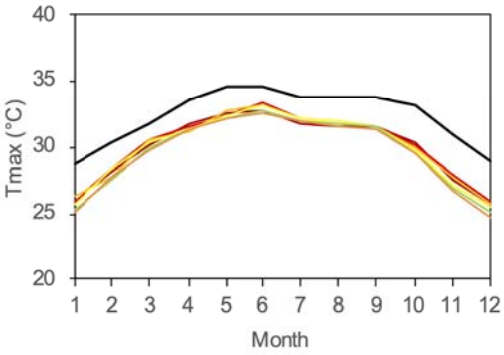
Tmax Tmin

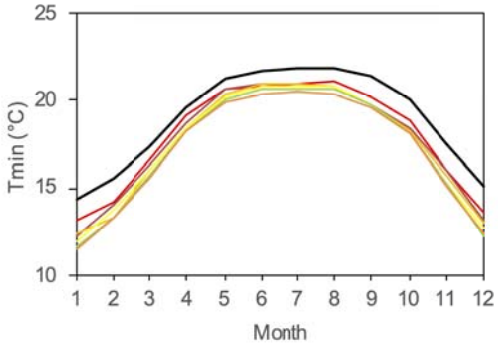
ECMWF



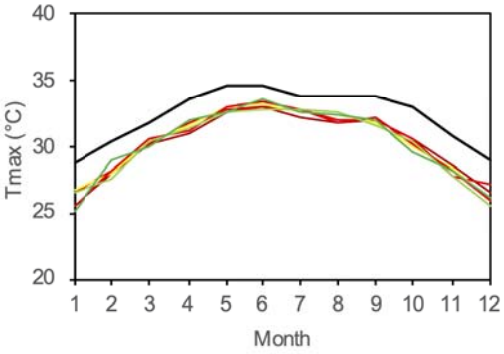


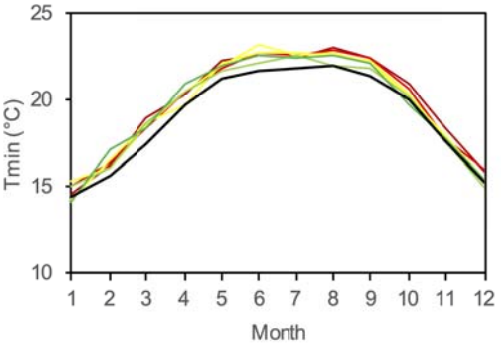
UKMO



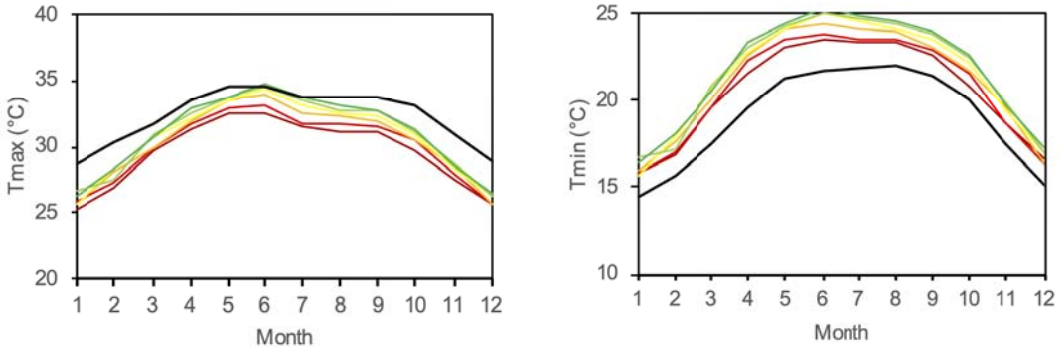


Meteo-France





DWD



CMCC

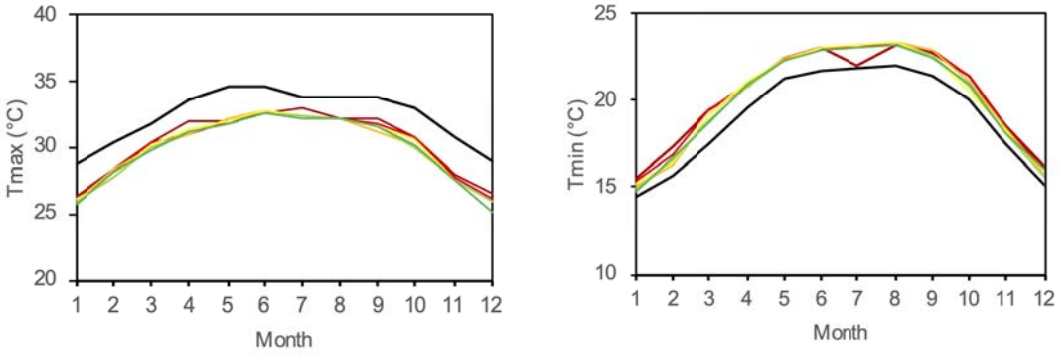


Figure 22. Tmax and Tmin monthly average climatological mean (1993–2016) averaged over the EN basin for MSMet and the five climate models, for leads 1–6 months.

Table 5. Mean annual bias of the climate model lead-1 forecasts for precipitation, Tmax and Tmin averaged over the EN basin.

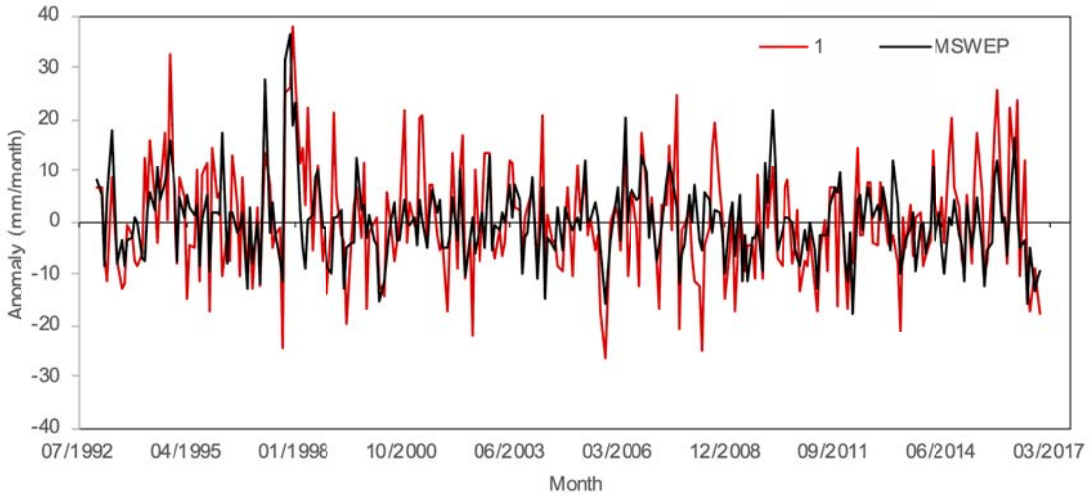
	Precipitation (mm)	Tmax (°C)	Tmin (°C)
ECMWF	158	−2.96	−0.95
UKMO	97.5	−2.41	−1.20
Meteo-France	145	−2.13	0.92
DWD	−45	2.70	−0.95
CMCC	291	−1.95	1.28

3.3.2. Forecast Skill for All Months

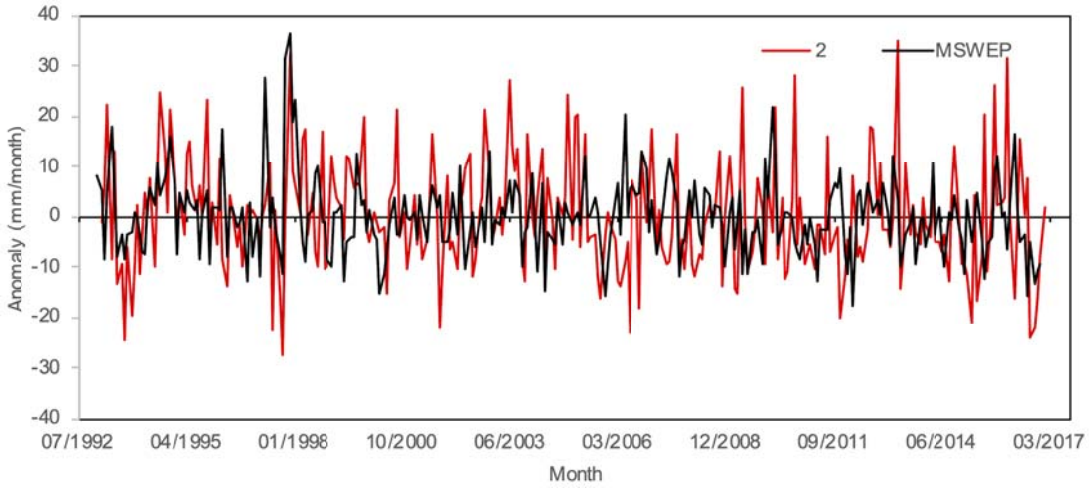
Time series of forecast anomalies of precipitation from the ECMWF model for the EN basin compared to anomalies of the observational data are shown in Figure 23 for 1993–2016 for a variety of lead times. Lead-1 forecasts represent the observed variability reasonably well,

for example for the wet period around 1997/98 and dry period of 2005/06. The match between the forecast and the observed data decreases with increasing lead time, although some extreme years are still captured. The correlations between the forecast and observed time series at lead times from 1–6 months are summarized for all models for monthly precipitation anomalies averaged over the EN basin in Figure 24. The ECMWF model has the highest correlation for lead-1 at 0.60, followed by Meteo-France, CMCC, UKMO and DWD. Correlations decrease with lead time to within the range 0.0 to 0.3 for leads 2–5. The ECMWF model tends to have higher correlations for all lead times.

(a)



(b)



(c)

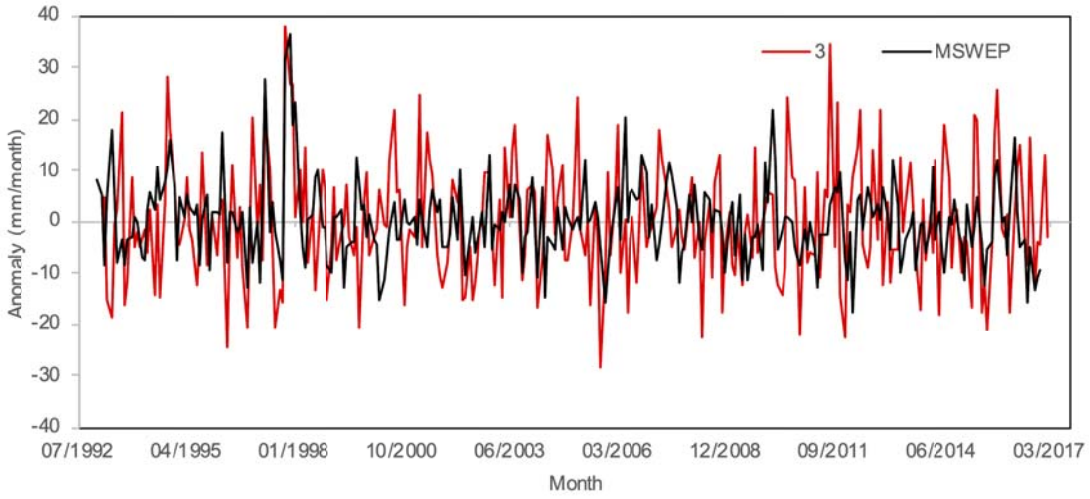
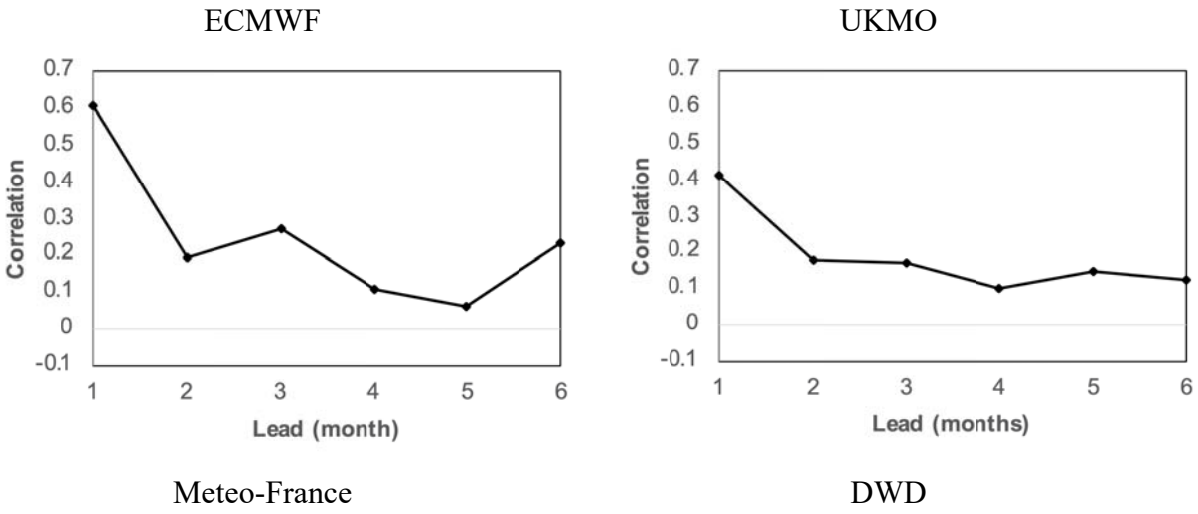


Figure 23. Time series of EN basin averaged precipitation monthly anomalies from MSWEP (black line) and ECMWF forecasts (colored lines) for 1993–2016 for all months and (a) lead-1, (b) lead-2 and (c) lead-3.



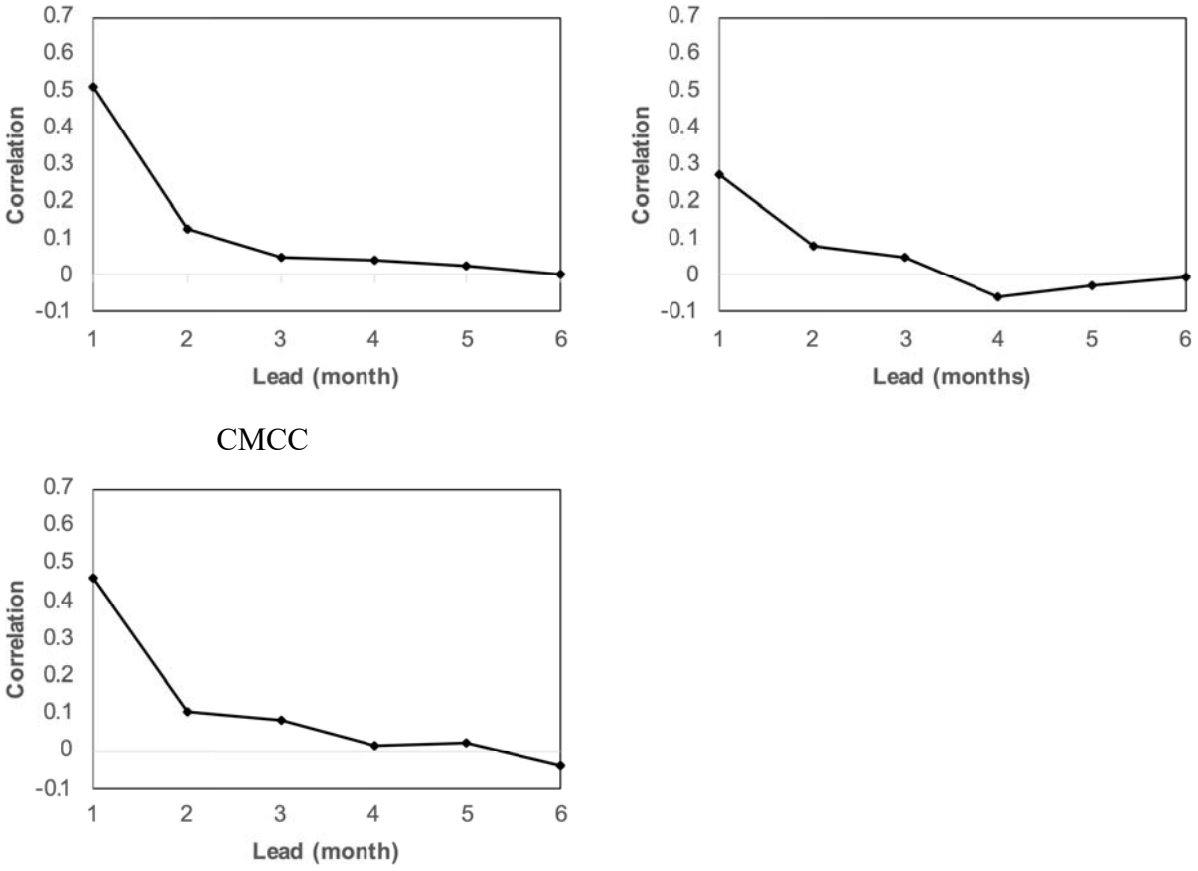
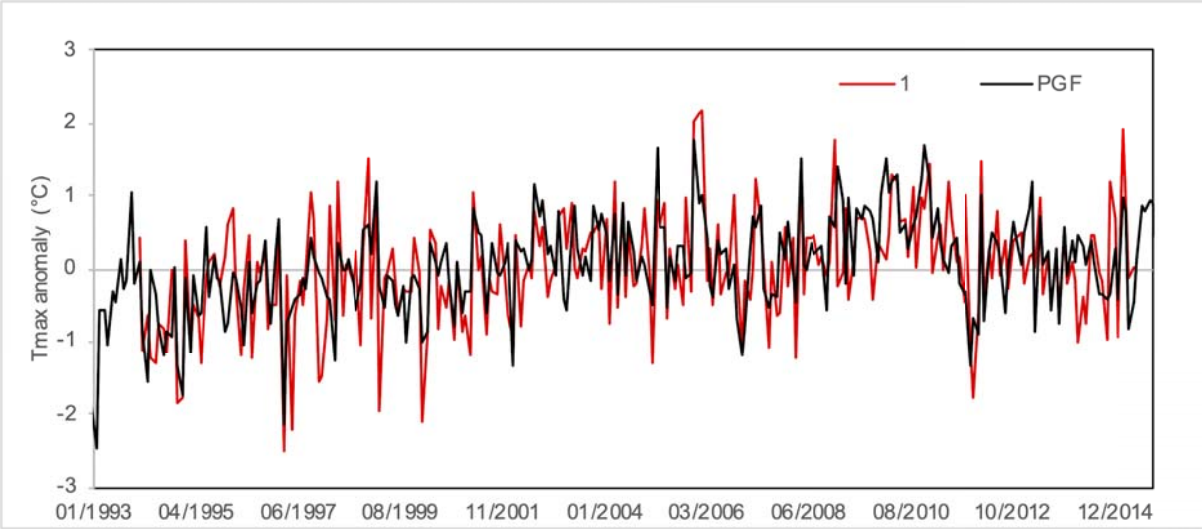


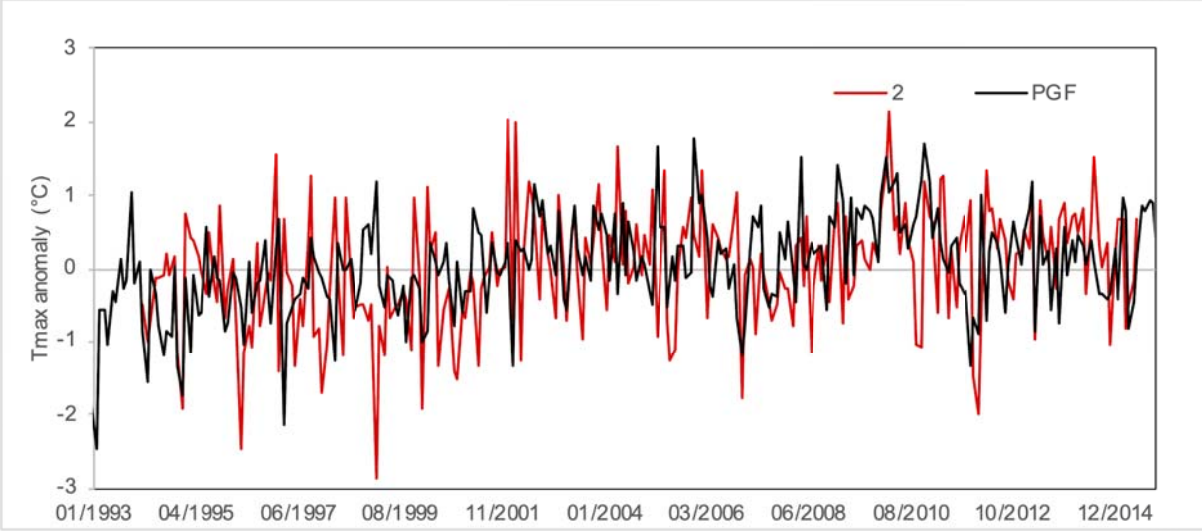
Figure 24. Correlation between anomalies of forecasts and observed monthly precipitation averaged over the EN basin, for 1993–2016 and lead 1–6 months for the five climate models.

Similar to Figure 23, time series of forecast anomalies of Tmax for the EN basin are compared to anomalies of the observational data in Figure 25. This is for the UKMO model, which is one of the better performing models for temperature. All lead forecasts represent the overall warming signal including the decadal variations, and most of the monthly and inter-annual variability in the anomalies. Lead-1 forecasts (Figure 25a) represent the observed variability well, and as for precipitation, identify some extreme events, such as the cool event in 1997 and the warm event in 2006. Similar to the precipitation forecasts, the correlation drops with increasing lead time, and the forecasts fail to match the observed variability at lead 3 (Figure 25c).

(a)


1
PGF
Tmax anomaly (°C)
3
2
1
0
-1
-2
-3
01/1993
04/1995
06/1997
08/1999
11/2001
01/2004
03/2006
06/2008
08/2010
10/2012
12/2014

(b)


2
PGF
Tmax anomaly (°C)
3
2
1
0
-1
-2
-3
01/1993
04/1995
06/1997
08/1999
11/2001
01/2004
03/2006
06/2008
08/2010
10/2012
12/2014

(c)

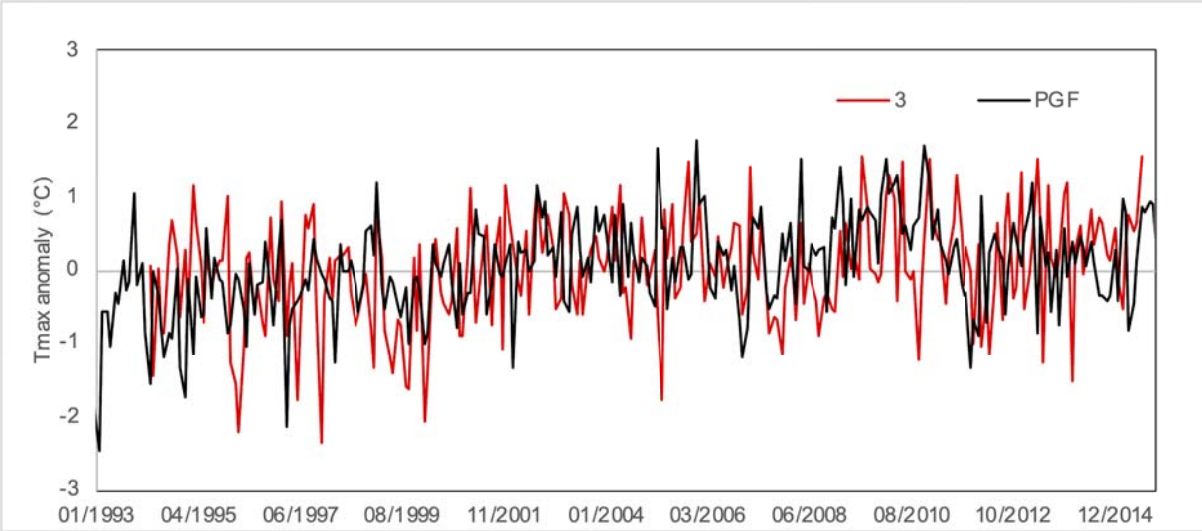
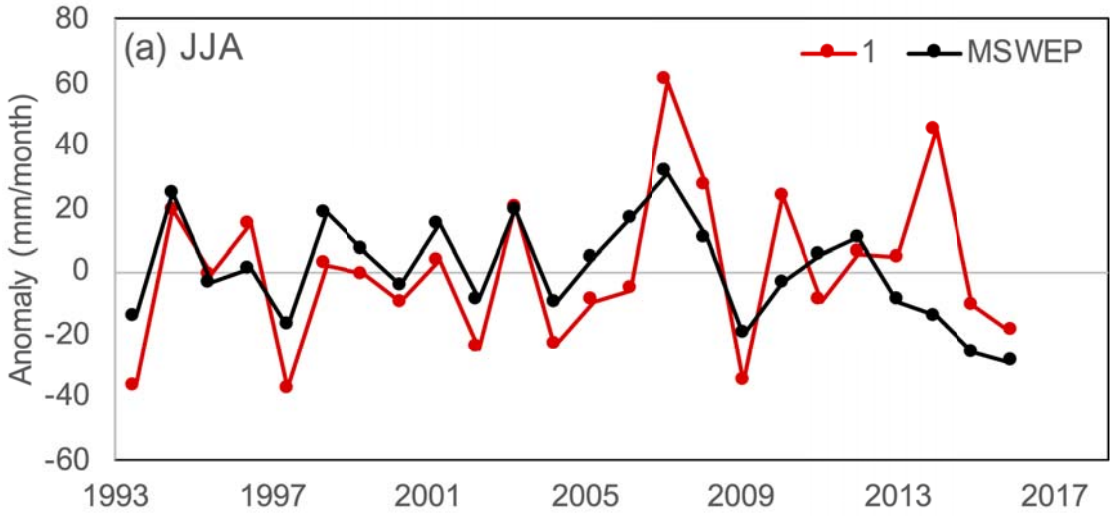
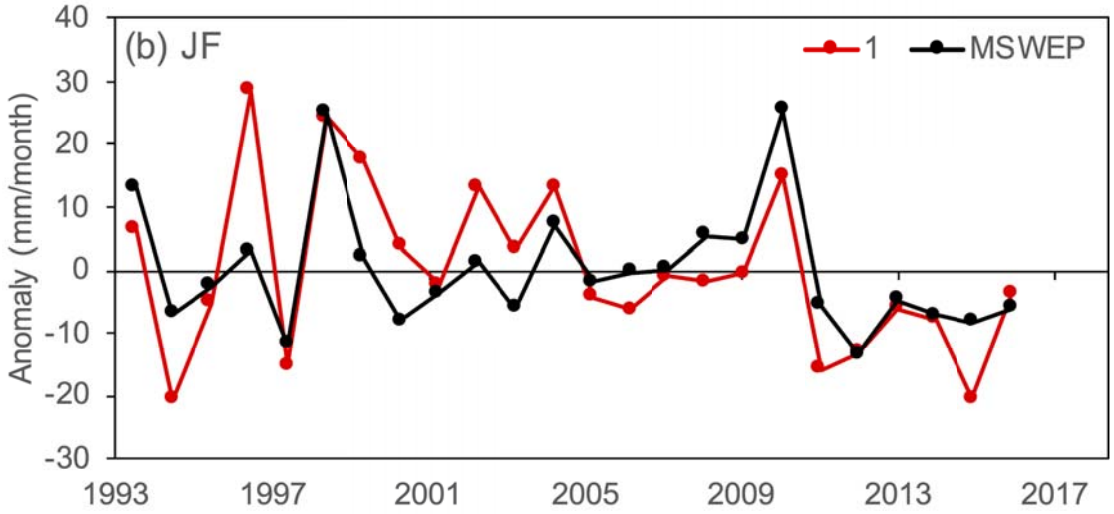


Figure 25. Time series of EN basin averaged Tmax anomalies from MSMet (black line) and UKMO (colored lines) forecasts for 1993–2016 for all months and for (a) lead 1 only, (b) lead-2 and (c) lead-3.

3.3.3. Forecast Skill for the Wet and Dry Seasons

The next series of figures summarizes the forecast skill for key target months, such as during the main wet season and compares this to results for all months and the dry season. Figure 26 shows time series of lead-1 ensemble mean forecasts of precipitation anomalies from the ECMWF model compared to the MSWEP observational data for June-July-August (JJA) and January-February (JF). JJA and especially July-August is the peak of the wet season across most of the EN basin (see Figure 21). JF is the peak of the dry season. The forecasts show modest skill for JJA anomalies with correlation of 0.45, and higher skill for the dry period of JF, with correlation of 0.8, indicating that forecasting the main wet season anomalies is challenging. Figure 27 summarizes the skill of the ECMWF model for all months, and different seasons for all lead times, showing that skill is highest for the dry season, moderate for all months combined, and lowest for the wet season months.


(a) JJA
1
MSWEP
Anomaly (mm/month)
80
60
40
20
0
-20
-40
-60
1993
1997
2001
2005
2009
2013
2017


(b) JF
1
MSWEP
Anomaly (mm/month)
40
30
20
10
0
-10
-20
-30
1993
1997
2001
2005
2009
2013
2017

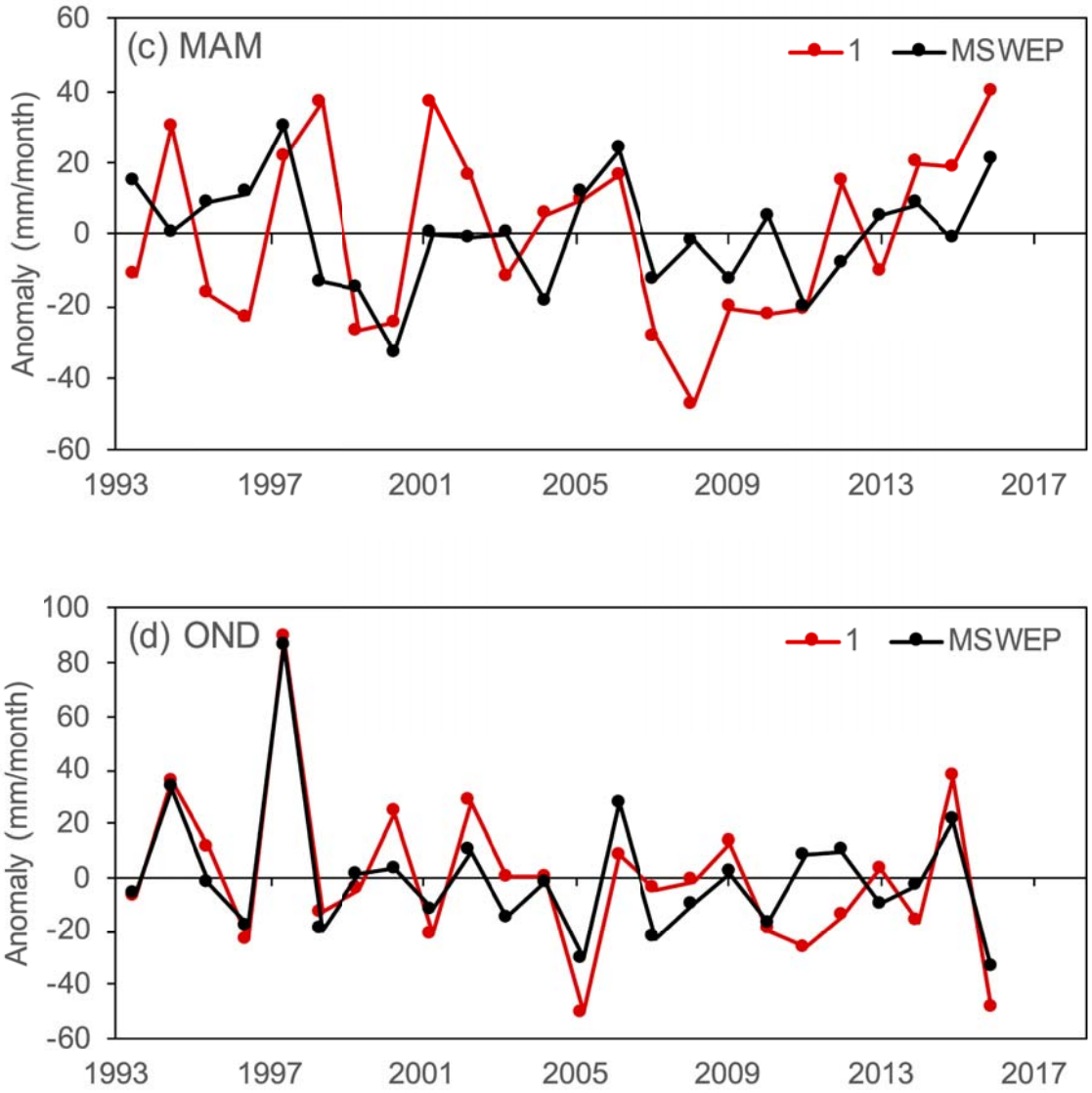


Figure 26. Time series of EN basin average seasonal precipitation anomalies of MSWEP and ECMWF lead-1 ensemble mean forecast for (a) June-July-August (JJA), (b) January–February (JF), (c) March–May (MAM), and (d) October–December (OND).

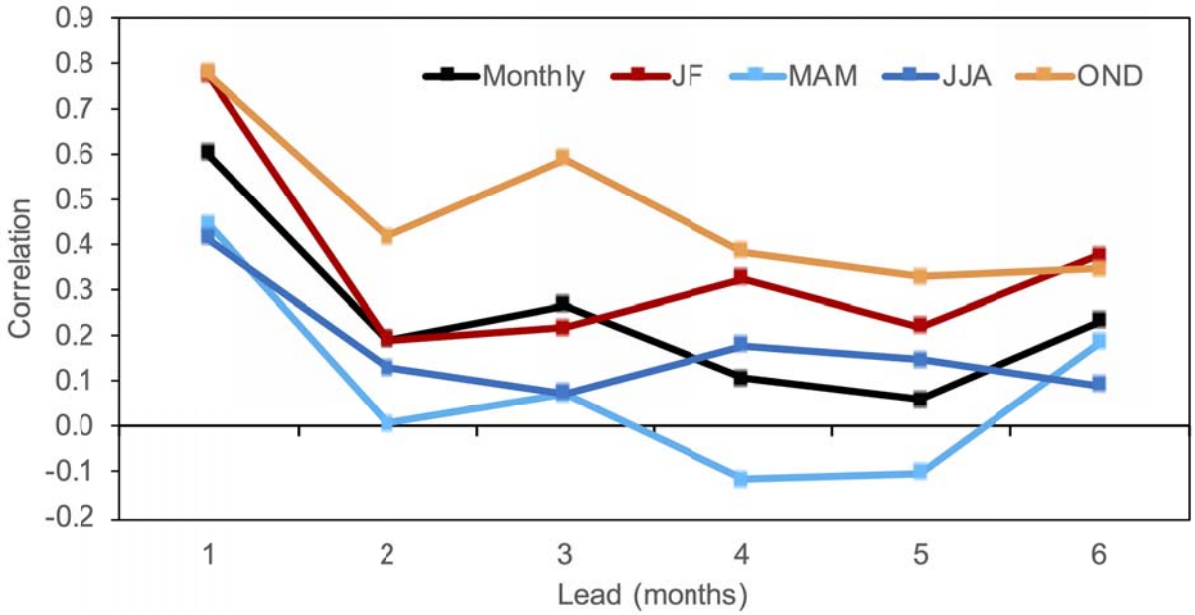
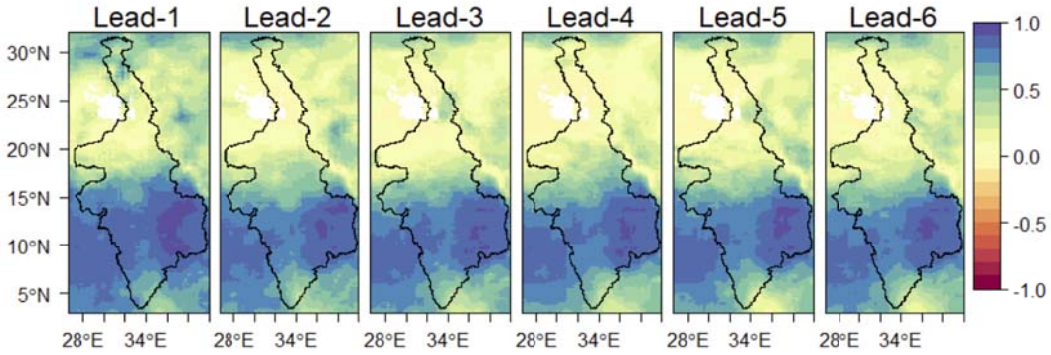


Figure 27. Summary of correlations of EN basin average precipitation anomalies of MSWEP versus ECMWF forecasts for lead 1–6 for all month (Monthly), JJA, MAM, JF and OND.

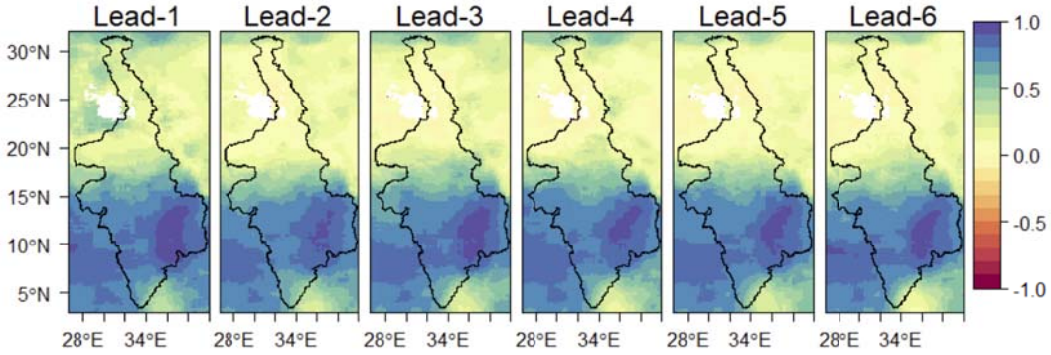
3.3.4. Spatial Distribution of Forecast Skill

Figures 28 and 29 show correlation maps for monthly precipitation (Figure 28) and monthly precipitation anomalies (Figure 29) for all months between the ensemble mean for each model and the MSWEP observational data. The correlations for monthly precipitation are around 0.9 for much of the lower part of the EN basin where the seasonal cycle of precipitation is well defined. In the drier parts of the basin in the north, the correlations are closer to zero, reflecting the general lack of a well-defined seasonal cycle in these dry regions. Correlation values tend to decrease with lead time, but remain relatively high in the lower part of the basin. The ECMWF model has higher correlations overall, with Meteo-France tending to have the lowest values, although these are still above 0.5–0.6 in the lower part of the basin.

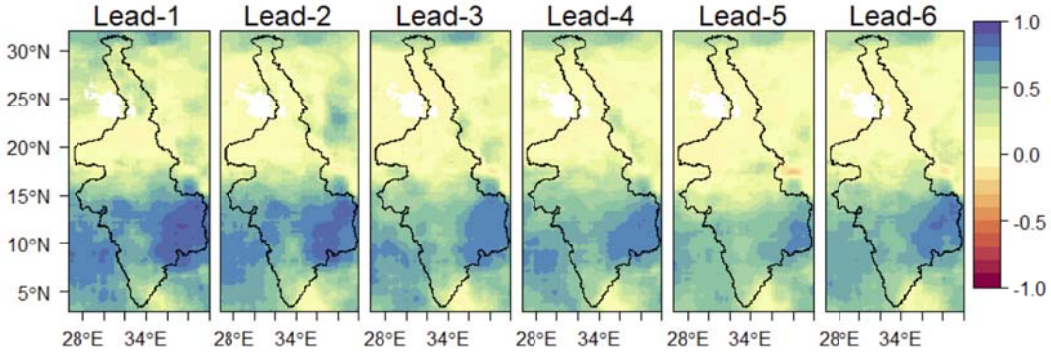
ECMWF



UKMO



Meteo-France



DWD

CMCC

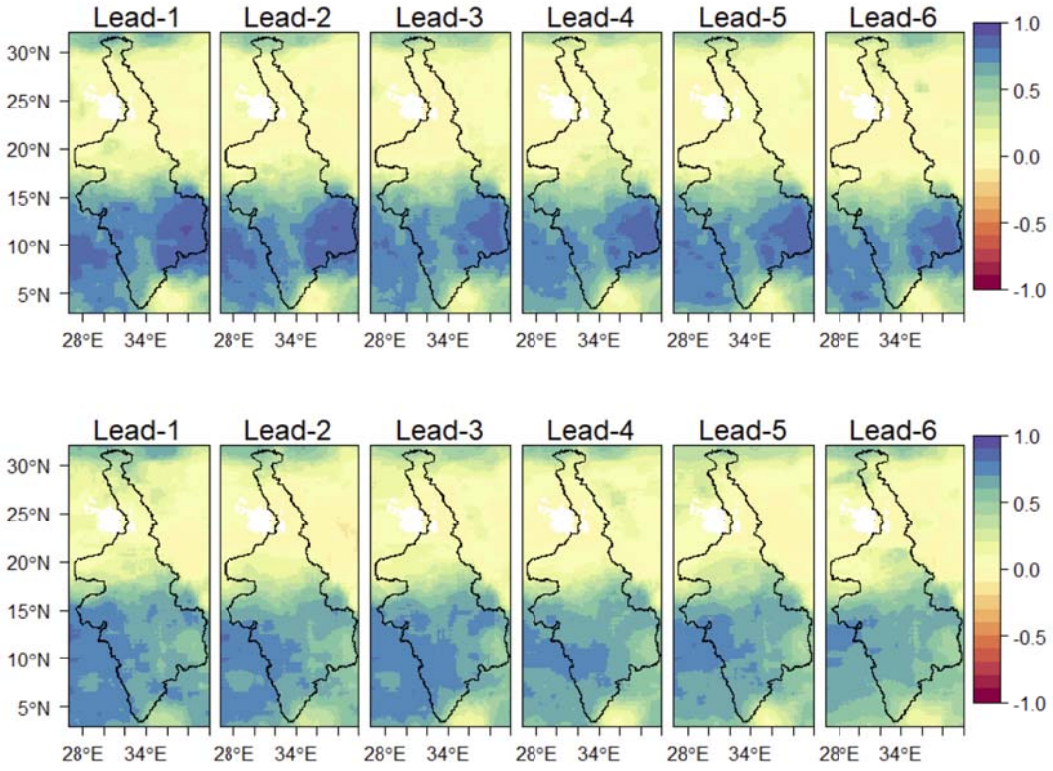
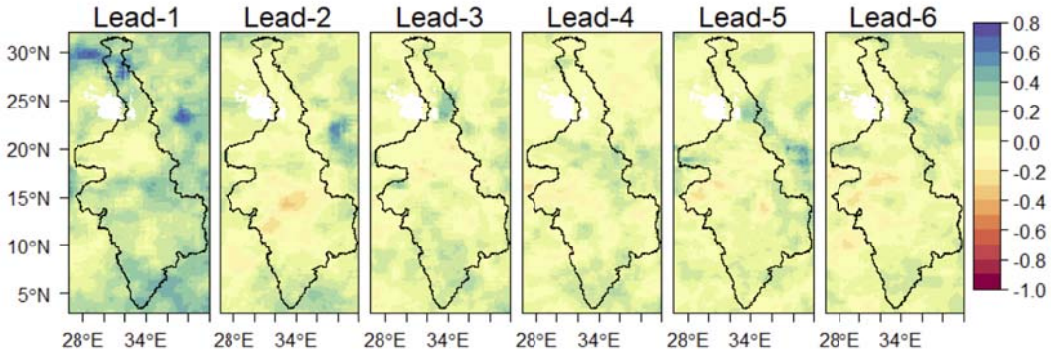


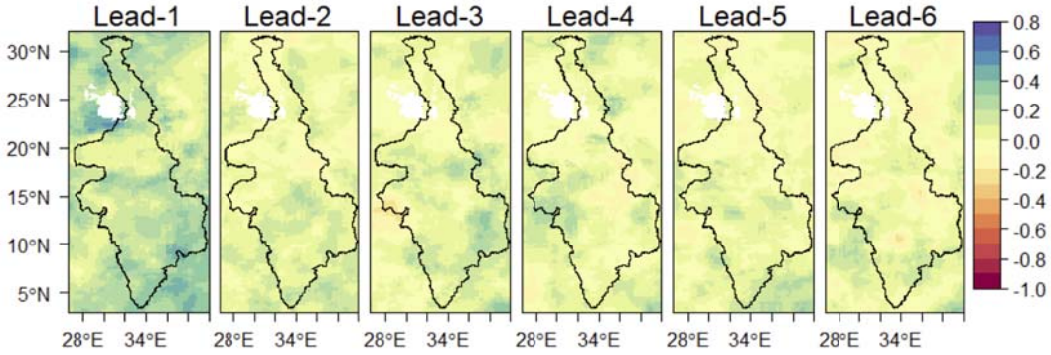
Figure 28. Correlation of the lead 1–6 ensemble mean forecasts of monthly precipitation for each model with MSWEP.

Correlations are much lower for the monthly precipitation anomalies (Figure 29), and decrease quite rapidly beyond lead-1. Again, the Meteo-France model shows the lowest correlation values at lead-1. The UKMO and DWD models show the highest correlation values at lead-1 in general. Although there is a sharp decline in correlation values beyond lead-1 for most models, some models retain some skill at lead-2 and -3, for example the DWD model.

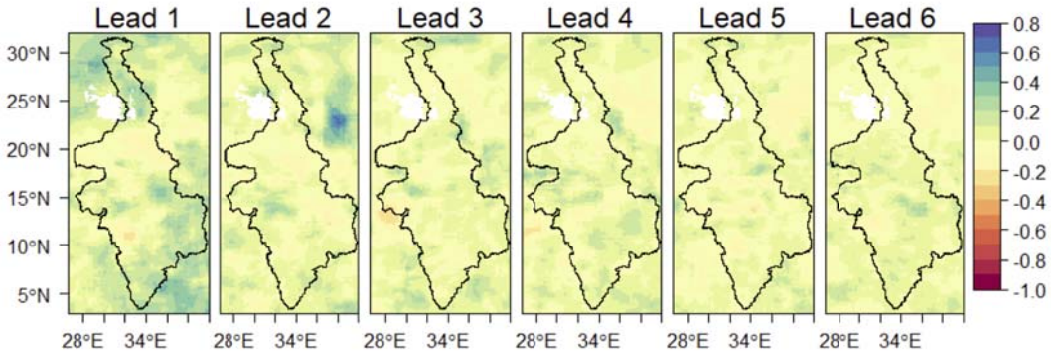
ECMWF



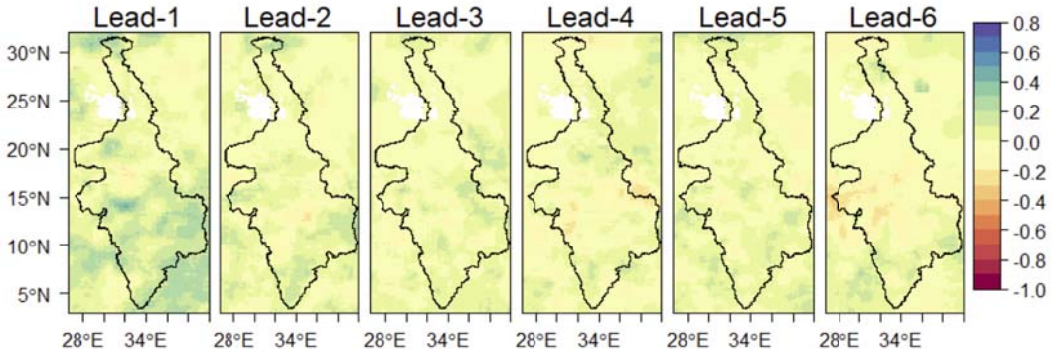
UKMO



Meteo-France



DWD



CMCC

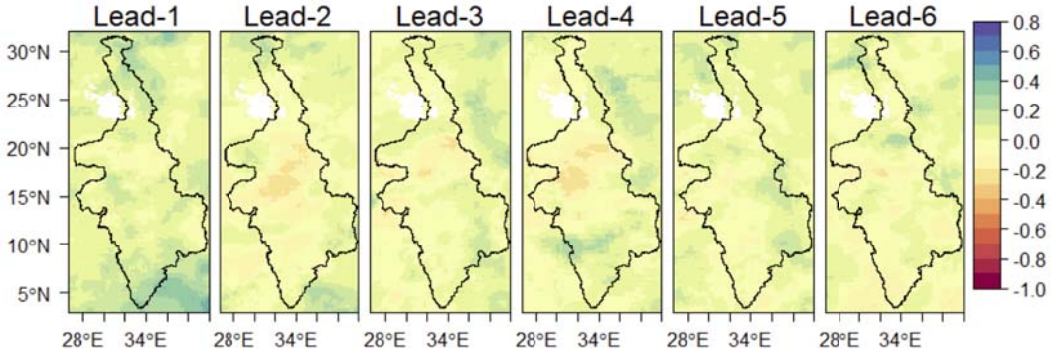


Figure 29. Correlation of the lead 1–6 ensemble mean forecasts of monthly precipitation anomalies for each model with MSWEP.

Figure 30 shows the spatial distribution of correlations for each model, but for specific key seasons (MAM, JJAS, OND). OND is generally better forecast than other seasons, with JJAS being well forecast by some models (e.g. DWD) and poorly simulated by other models (e.g. CMCC). Again, the increasing gradient in skill between the drier north and wetter south is evident.

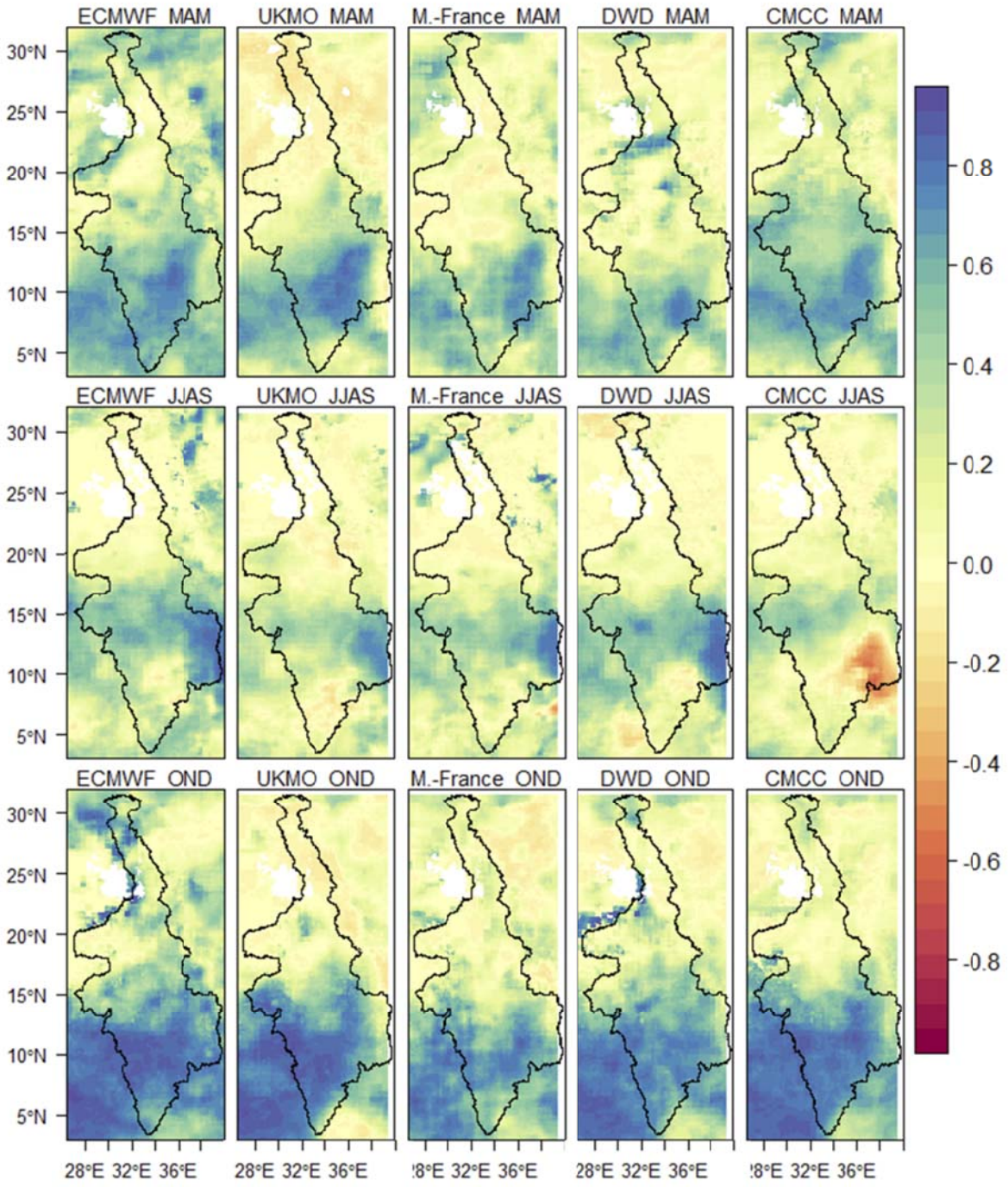


Figure 30. Spatial distribution of correlation between MSWEP and the five forecast models for lead-1, for (a) MAM, (b) JJAS, and (c) OND average precipitation.

Lead-1 forecasts of Tmax and Tmin for all models are shown in Figure 31. The gradient in skill is reversed for these temperature variables, with very high correlations (> 0.9) in the drier northern part of the basin, and values around 0.5–0.7 in the wetter lower part of the basin. Correlations for Tmax are generally slightly higher than for Tmin in the lower part of the basin.

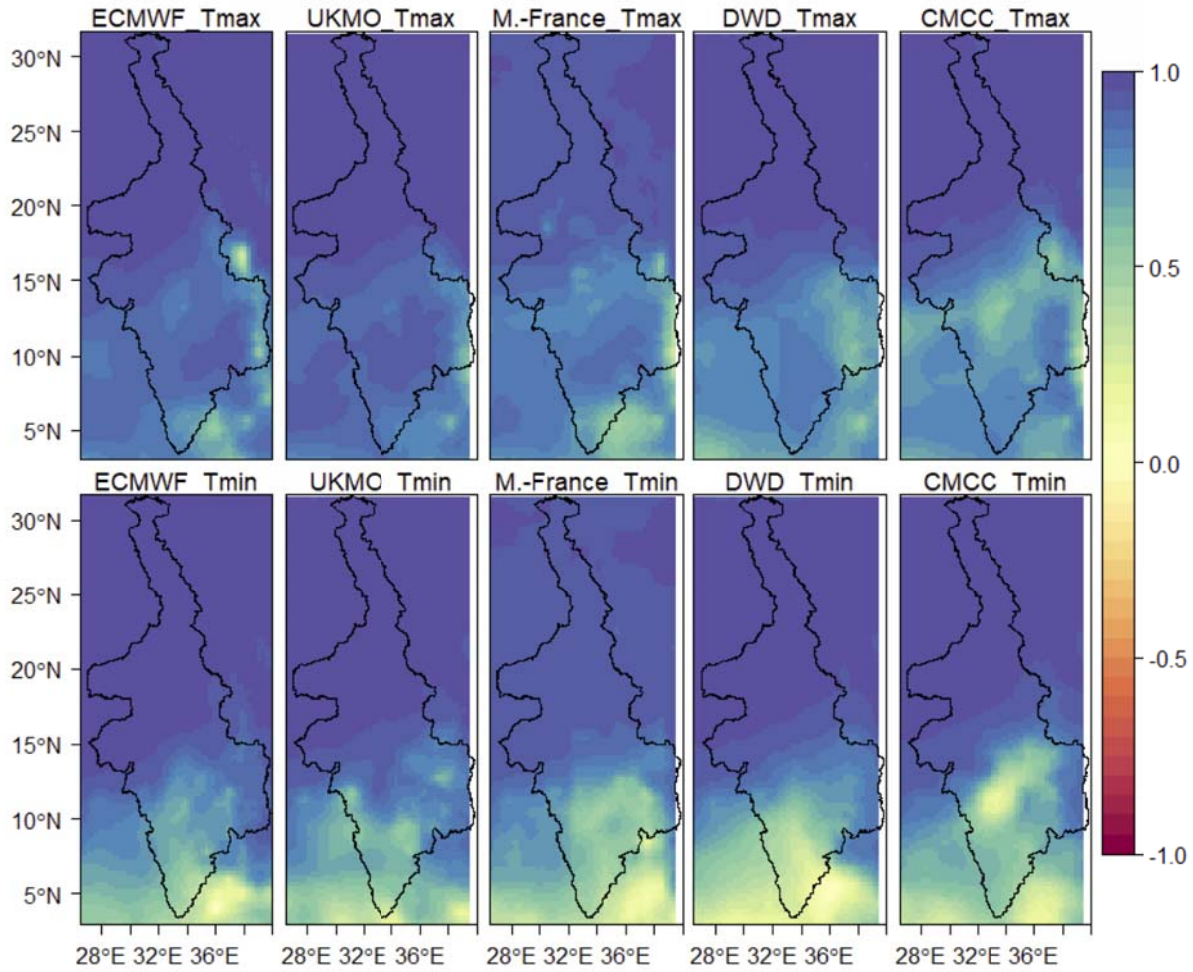


Figure 31. Correlation of forecasts of Tmax and Tmin from the five models with MSMet observations over 1993–2016 for lead-1.

The same analysis is shown in Figure 32 but for the Tmax and Tmin monthly anomalies, which reveals a much more complex pattern of correlation values across the models. Correlation values for the anomalies are much lower than for the actual monthly values, as expected and also shown for precipitation. Some models have correlation values up to 0.6–0.7 in parts of the

region, whilst other models show near zero, or slightly negative correlations. Visually, the best performing model is the ECMWF model, followed by the UKMO model. The Meteo-France model generally shows zero or negative correlation values across the basin.

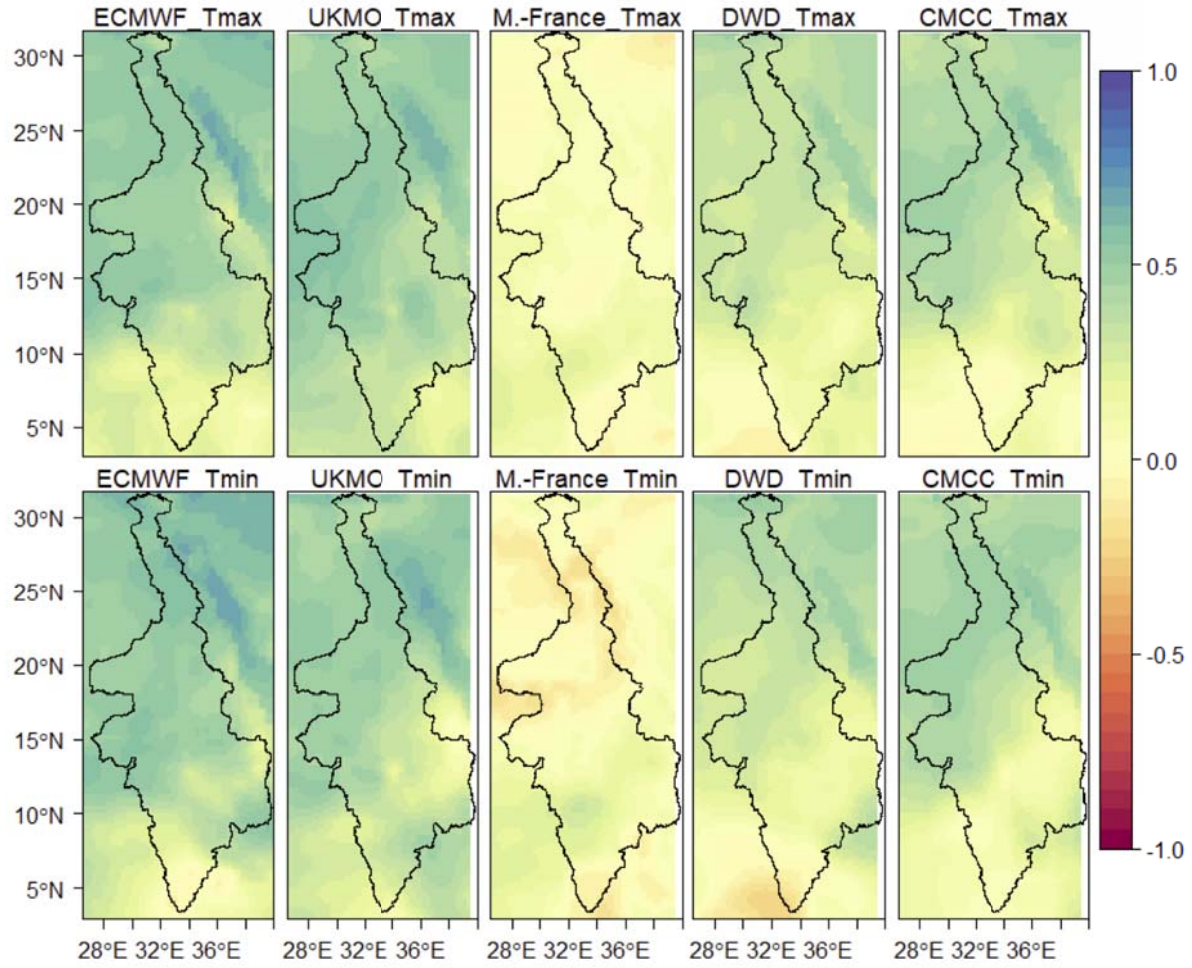
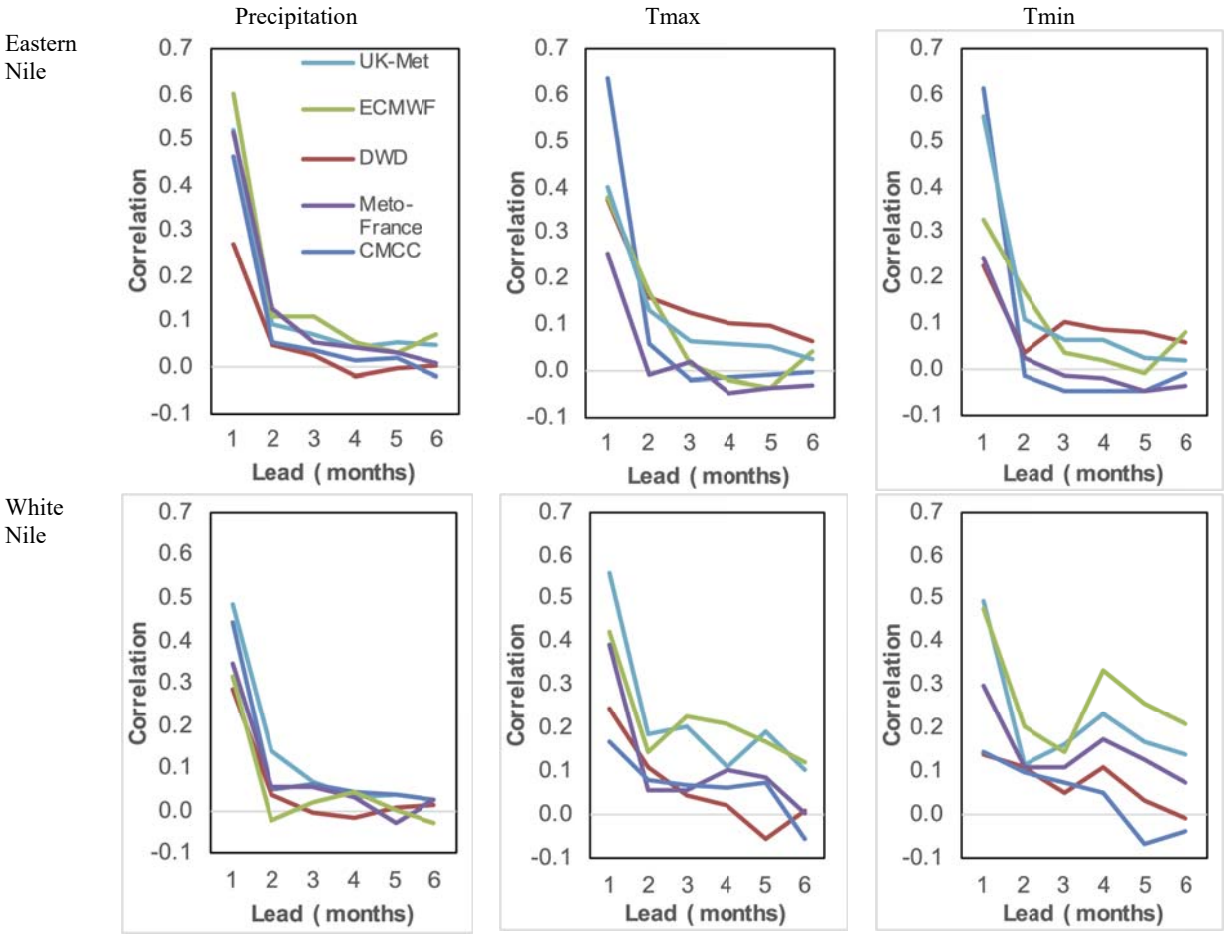


Figure 32. Correlation of forecasts of Tmax and Tmin anomalies from the five models with MSMet observations over 1993–2016 for lead-1.

3.3.5. Summaries of Skill for the EN and Sub-Basins

The correlations are summarized for all models for precipitation, Tmax and Tmin anomalies in Figure 33. For the EN basin, the ECMWF model has the highest correlation for precipitation at lead-1 as mentioned above, followed by Meteo-France, CMCC, UKMO and DWD. Overall the UKMO model performs the best for Tmax and Tmin, whilst Meteo-France does well for Tmin at lead-1. For the sub-basins, the correlations decrease slightly for precipitation at all leads.

For temperature, the correlations for the sub-basins are generally weaker for lead-1, but actually increase for longer leads. For Tmin and some models, the correlations are above 0.3 for leads out to four months..



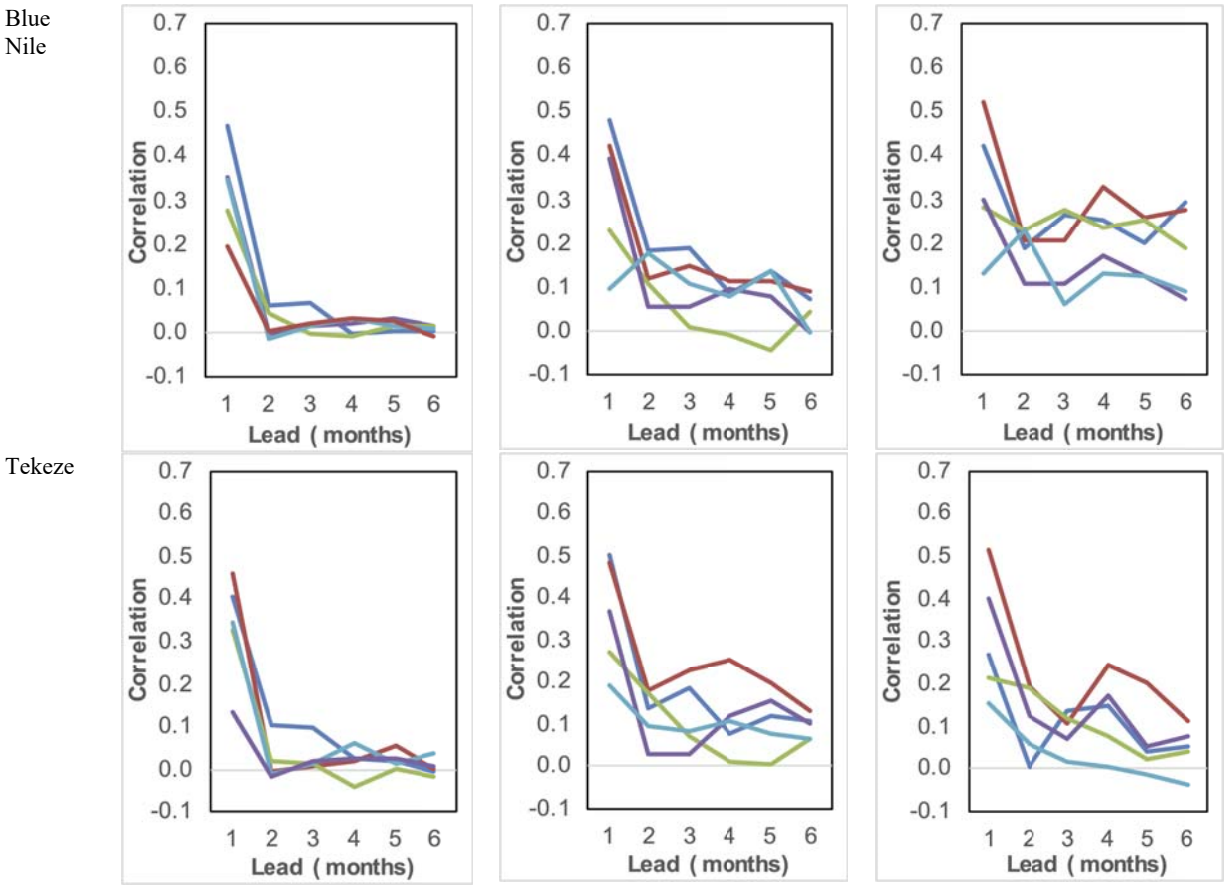


Figure 33. Summary of correlation between time series of forecast and observations of monthly precipitation, Tmax and Tmin anomalies for the EN for 1993–2016, and lead 1–6 months.

3.3.6. Evaluation of Drought Forecasts based on SPI

Time series of EN basin average SPI for MSWEP and the five models for lead-1 are shown in Figures 34–37. SPI values below zero (red shading) are considered to be drier than normal, and above zero (blue shading) wetter than normal. A value of −1.0 is often considered to be a threshold for drought in the literature.

The differences in the variability as represented by the different SPI time scales is evident, with high frequency variability in the SPI-1 (Figure 34) through to long-term decadal variability in the SPI-12 (Figure 37). Historic drought events are clearly shown in the MSWEP dataset for SPI-3 and longer, highlighting drought events in the early 2000s, 2009, 2011/2012 and 2015, but also wet events in 1995, 1997/98 and 2007. The climate models can represent many of these events, but do fail to capture the magnitude and timing of some large drought events or even

identify droughts when there was none (e.g. UKMO model forecasts a large drought in 2007 whilst the MSWEP shows wet conditions).

The corresponding correlation values for each of the models and each SPI time scale are shown in Figure 38. The correlations peak at about 0.6 and are highest for the ECMWF model, which is consistent with the monthly precipitation evaluations, but are also high for the Meteo-France model, despite its skill being relatively low for monthly precipitation. The other models have slightly lower correlation values. Correlations generally increase with increasing SPI time scale, particularly for the DWD model, which is expected as the forecast anomalies are smoothed out with increasing time scale. The UKMO model alone shows a decrease in correlation values with increasing SPI time scale.

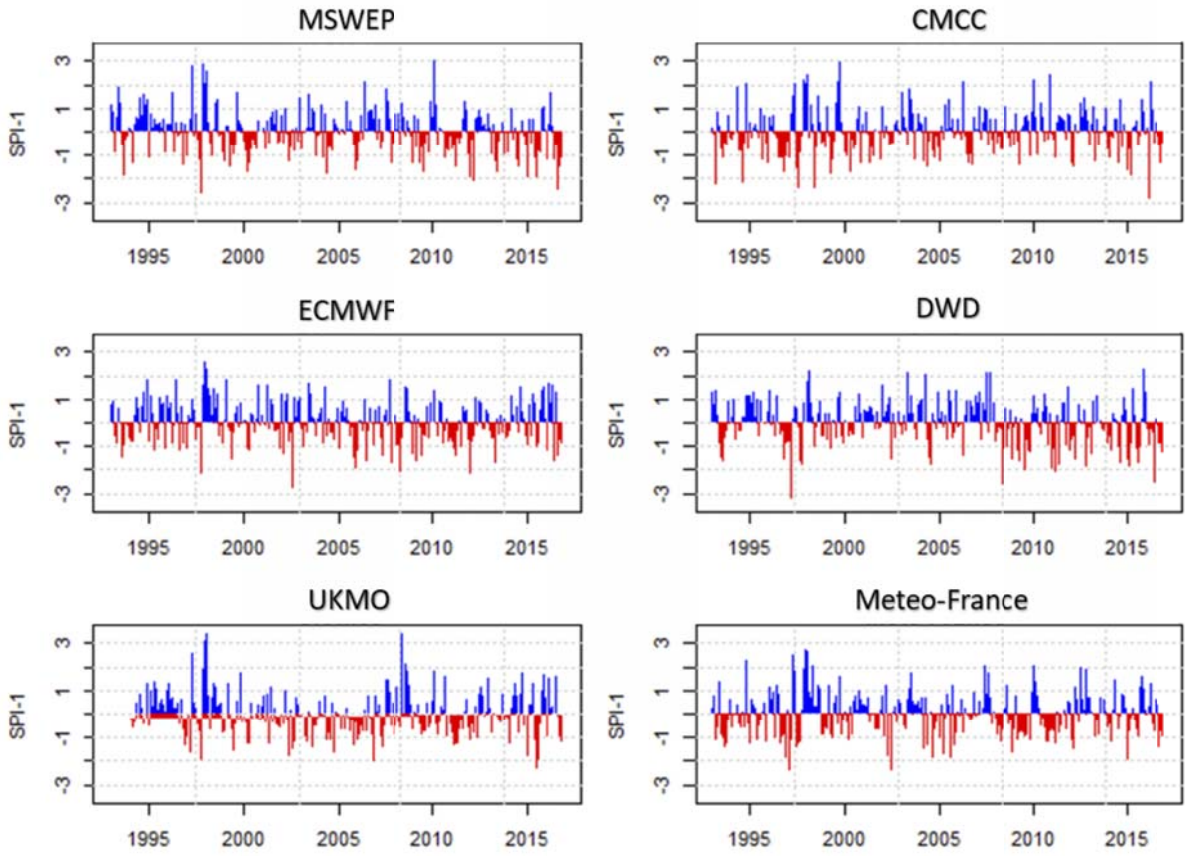


Figure 34. Time series of EN basin average SPI-1 for MSWEP and time series of lead-1 forecasts from the five models.

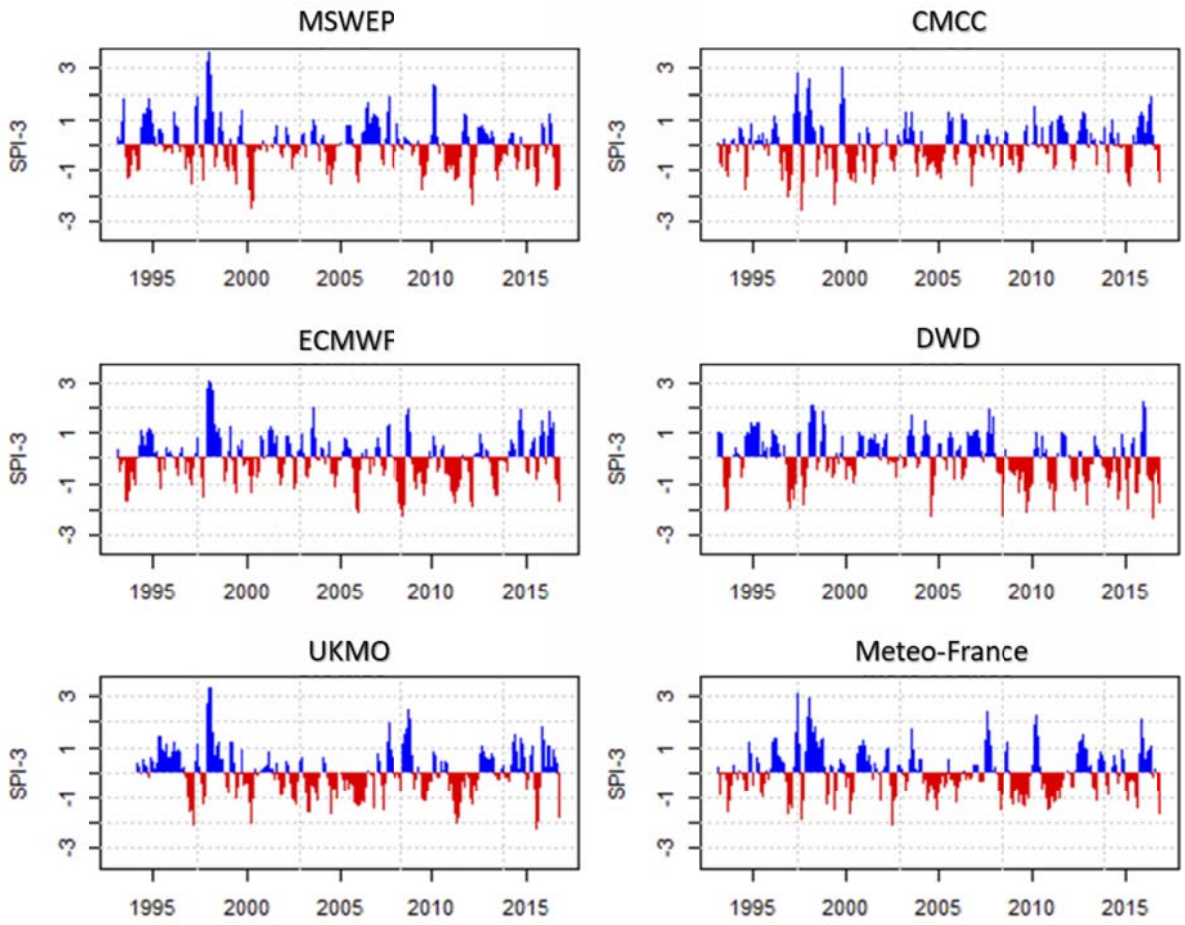


Figure 35. Same as Figure 34 but for SPI-3.

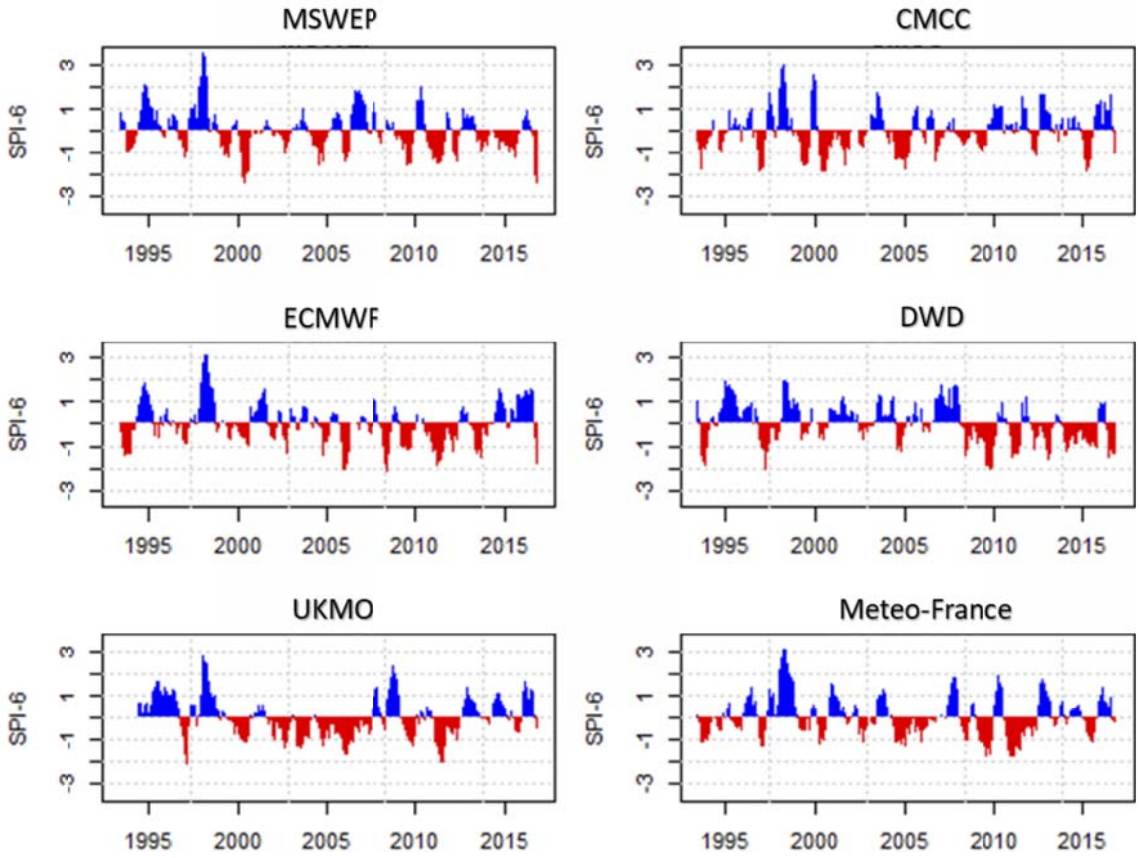


Figure 36. Same as Figure 34 but for SPI-6.

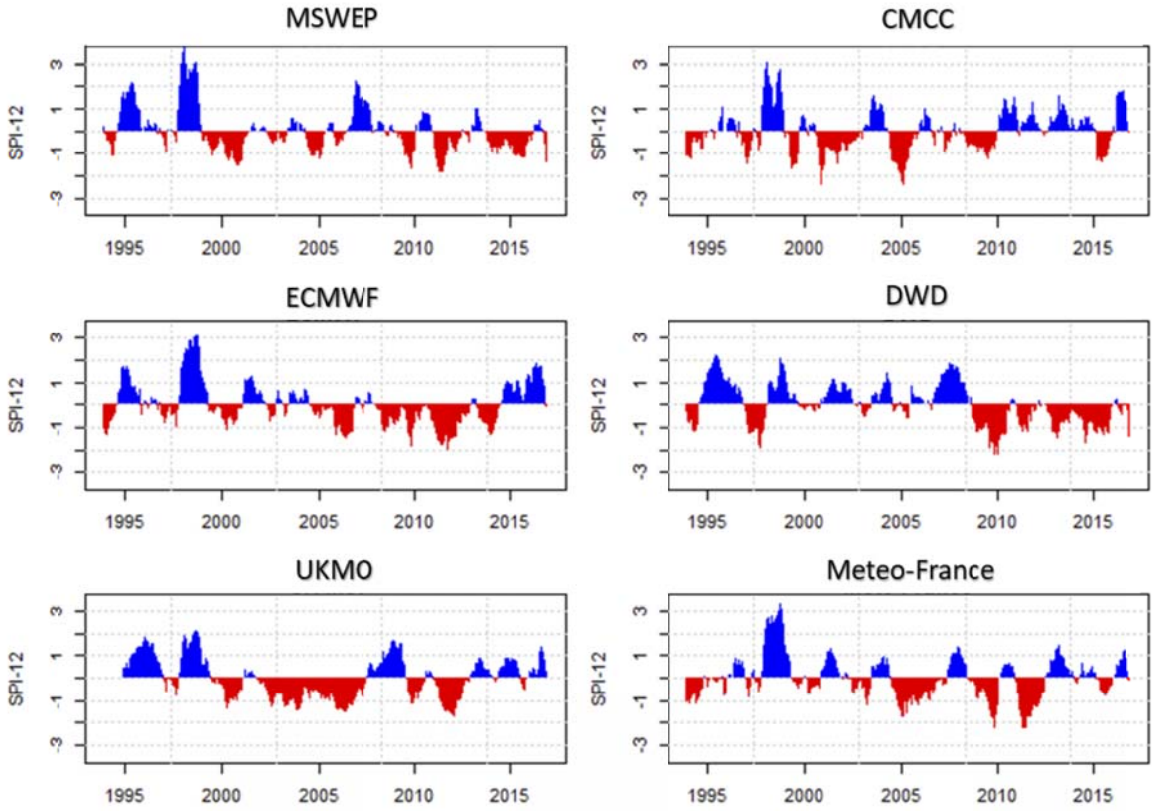


Figure 37. Same as Figure 34 but for SPI-12.

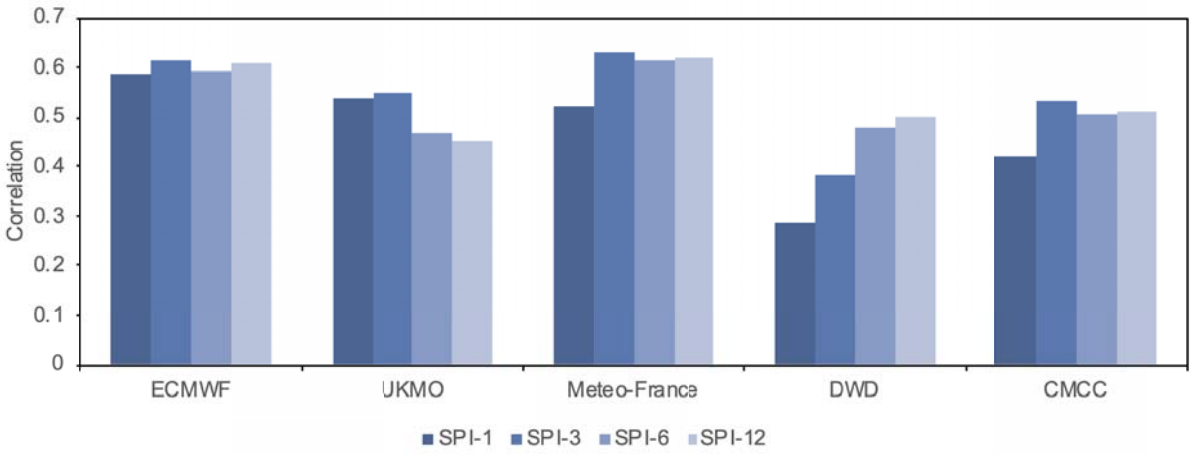


Figure 38. Correlation of MSWEP and the five models for lead-1 forecasts of SPI-1, -3, -6, and -12.

3.3.7. Evaluation of Individual Drought Events

Here we evaluate the forecast skill of the models for SPI for individual drought events. We use the UKMO and ECMWF models to show examples of skill assessments for SPI-1 and then summarize across models and basins in Table 6. Results are shown for different drought severities: moderate drought (SPI <−1.0), severe drought (−1.0 > SPI >−1.5) and extreme drought (SPI <−2.0). Figure 39 shows the number of lead-1 ensembles of the UKMO model that forecast correctly an observed drought month based on SPI-1. A perfect forecast for the UKMO model would show all seven ensemble members in drought, as this is the ensemble size of this modelfor its hindcasts. Note that this analysis only looks at *hits* when the model forecasts the observation correctly, and does not consider *false alarms* when the model forecasts a drought when there was none in the observations (all statistics are shown later in Table 6). The model forecasts identify drought in at least one ensemble member for about 50% of the drought months, meaning that it fails to identify drought at all in the other 50%. Of those drought events that are forecast by at least one ensemble member, the total number ranges from 1 to 4, with a mean of 1.1 (16%) members for all observed events. For severe and extreme drought, the number of observed events decreases as expected, but the mean number of members that forecast any drought increases to 1.7 (24%) and 2.4 (34%), respectively. Figure 40 shows the results for the ECMWF model, which forecasts moderate drought in at least one member in 93% of events, although ECMWF has a much higher ensemble size (25) in its hindcasts. The mean percentage of members that forecast moderate, severe, and extreme drought is 19, 14, and 22%, respectively, which is generally lower than for the UKMO model.

Summary statistics for all models are provided in Table 6 in terms of the hit rate and false alarm rate, and Brier score. The models generally have a hit rate of about 0.34-0.44 for the EN basin, and this drops slightly for the sub-basins, especially for the DWD and CMCC models, which have rates of less than 0.25 for the Tekeze-Setit-Atbara. The ECMWF and UKMO models show the highest rates overall. The false alarm ratios show a similar distribution across models and basins, with lower ratios (less drought events mistakenly forecast) for the EN basin and ECMWF and UKMO models. These results are consistent with the correlation values for the precipitation monthly values and anomalies, indicating that the ECMWF and UKMO models do better than the others for precipitation variability and drought, and that skill improves slightly over larger domains, as one might expect.

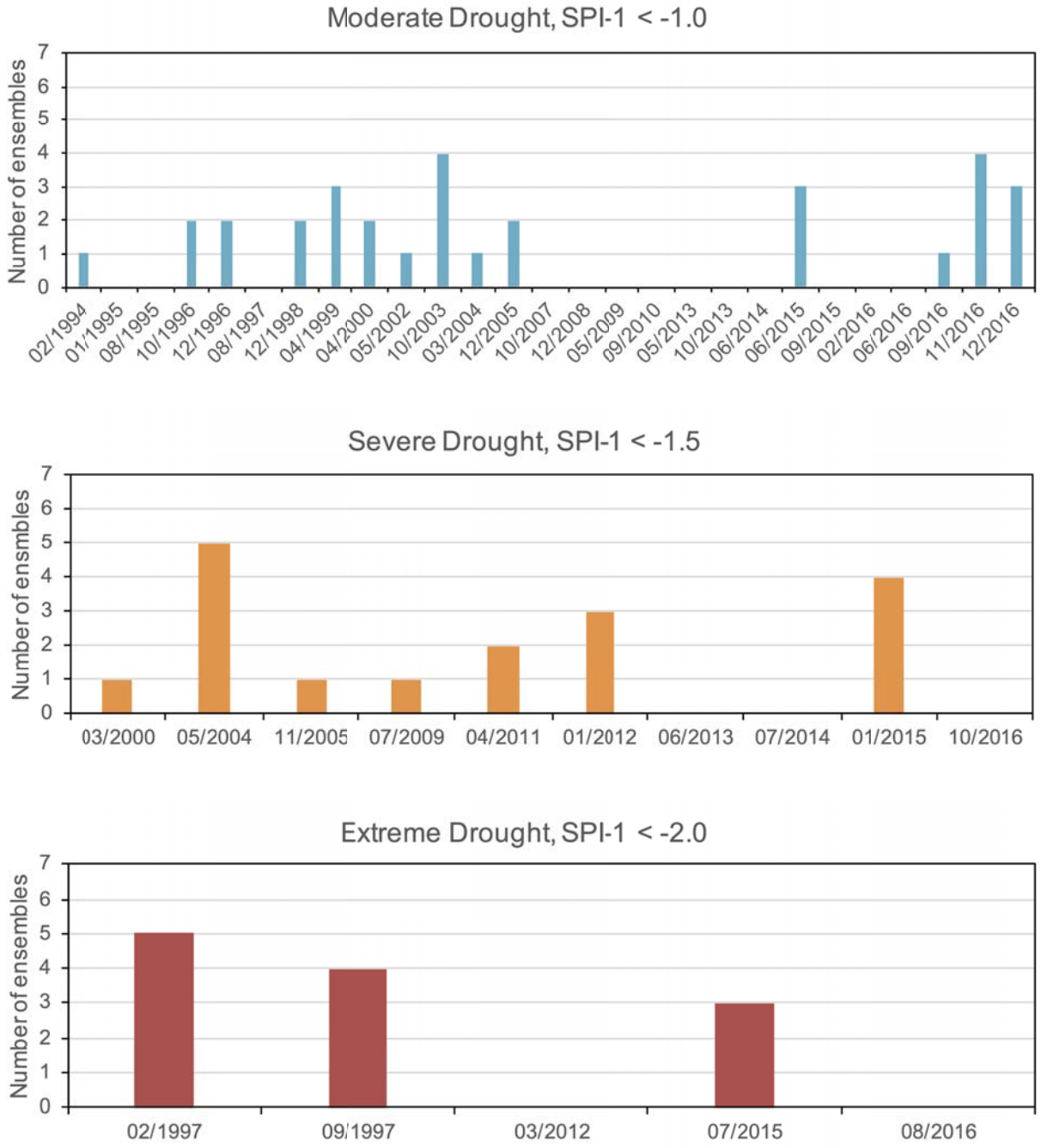
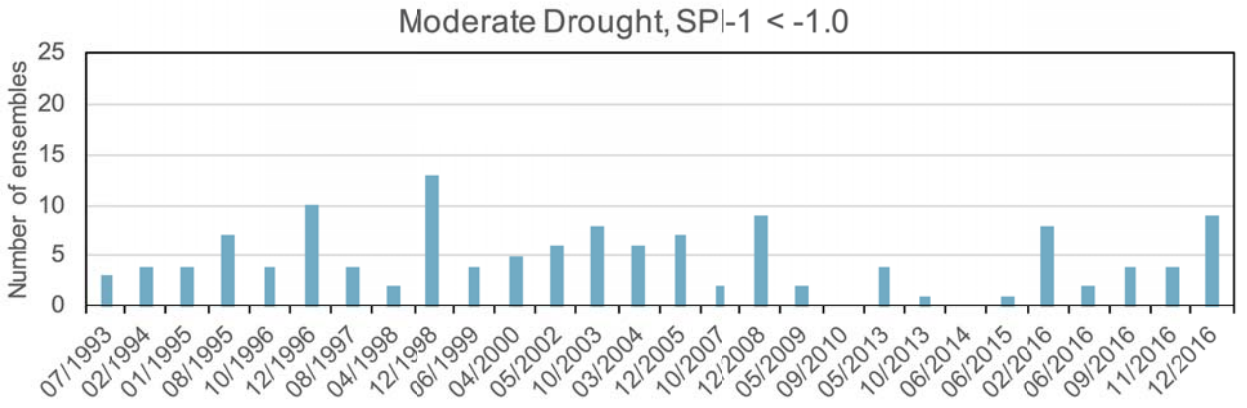
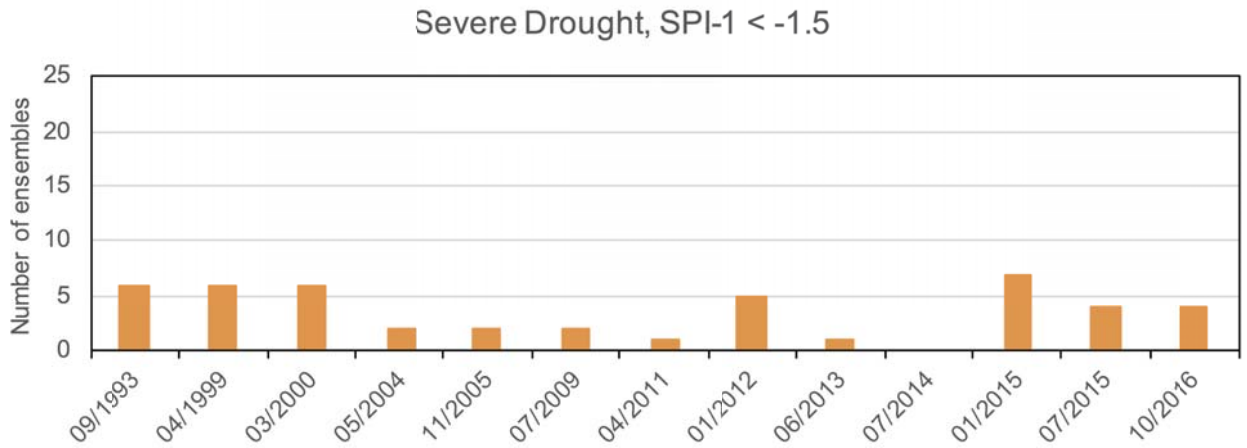


Figure 39. Number of ensembles of the UKMO model which correctly forecast observed moderate, severe and extreme drought events, based on the SPI-1 meteorological drought index for the EN basin. Only months that were in drought in the historic observational record are

shown. Note that for some months, there was a drought in the historic record, but no ensembles forecast the drought.





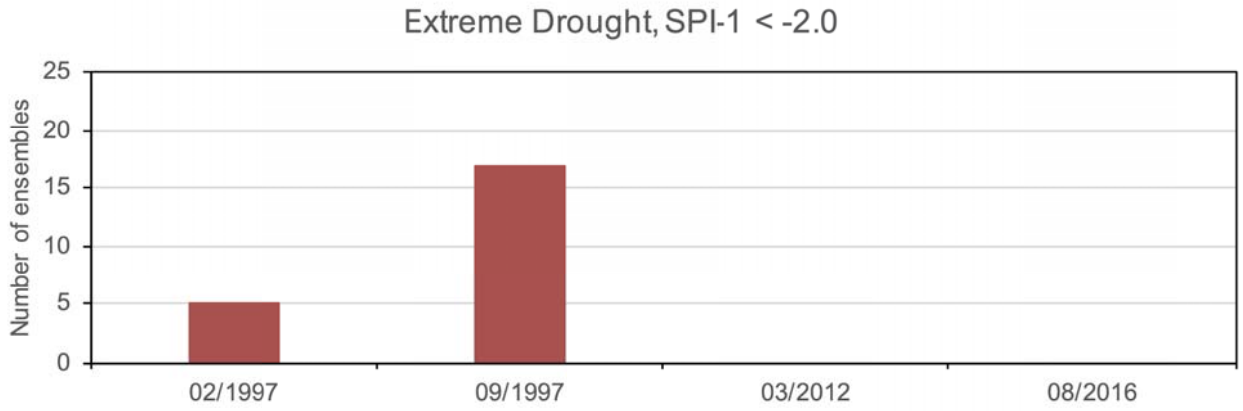


Figure 40. As Figure 39, but for the ECMWF model.

Table 6. Summary statistics for lead-1 forecasts of SPI-1, moderate (SPI <−1) drought events. Hit rates and false alarm ratios are averaged over all ensemble members.

Basin	Model	Hit rate	False alarm ratio	Brier score
Eastern Nile	ECMWF	0.44	0.55	0.08
	UKMO	0.38	0.59	0.12
	Meteo-France	0.34	0.65	0.11
	DWD	0.36	0.64	0.19
	CMCC	0.36	0.62	0.12
Baro-Akobo-Sobat-White Nile	ECMWF	0.34	0.61	0.14
	UKMO	0.33	0.59	0.14
	Meteo-France	0.29	0.72	0.15
	DWD	0.28	0.65	0.15
	CMCC	0.31	0.69	0.15
Abay-Blue Nile	ECMWF	0.30	0.70	0.12
	UKMO	0.37	0.54	0.12
	Meteo-France	0.31	0.71	0.13
	DWD	0.24	0.73	0.14
	CMCC	0.24	0.75	0.13
Tekeze-Setit-Atbara	ECMWF	0.34	0.67	0.11
	UKMO	0.35	0.66	0.10
	Meteo-France	0.30	0.72	0.11
	DWD	0.22	0.76	0.11
	CMCC	0.24	0.75	0.12

3.3.8. Evaluation of Multi-model Ensemble

We first show the improvement in skill from an equally weighted multi-model mean (MMM, unweighted mean), and then show how this can be improved further by taking into account the

skill of individual models (MMWM, weighted mean). The weights for lead-1 precipitation forecasts are shown in Figure 41. These weights can be calculated for each lead time and target month, but for simplicity we assume that the weights are unchanging. The ECMWF and UKMO models have the highest weights over the entire EN basin (Table 7), and are generally higher in the southern part, with the DWD having lower weight in this area. Otherwise the weights are reasonably similar among the models, ranging from about 0.1 (10 %) to 0.25 (25 %). The largest differences among models are in the drier north, where the UKMO weights are up to 50 % and Meteo-France and DWD have weights close to zero.

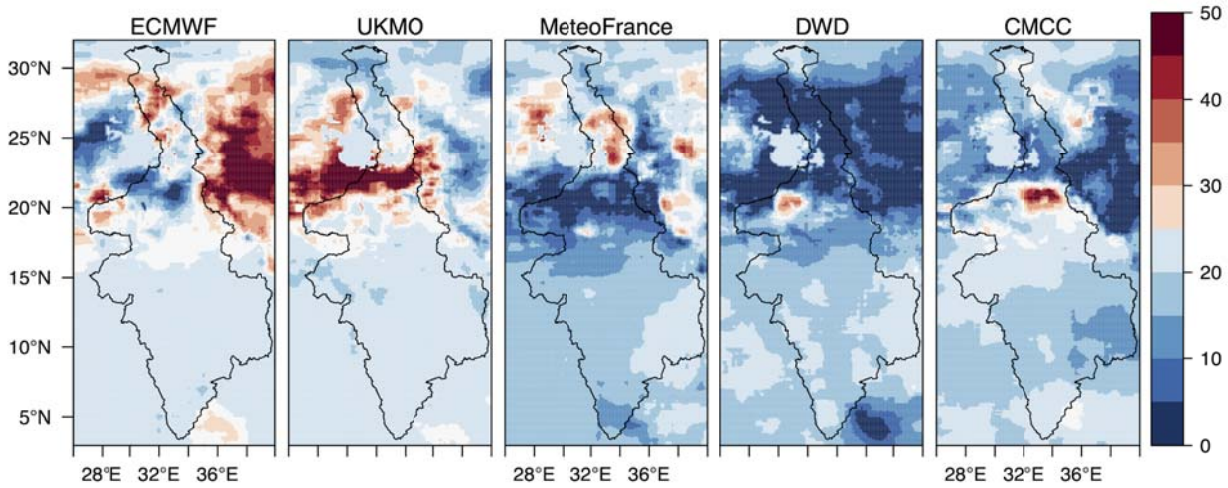
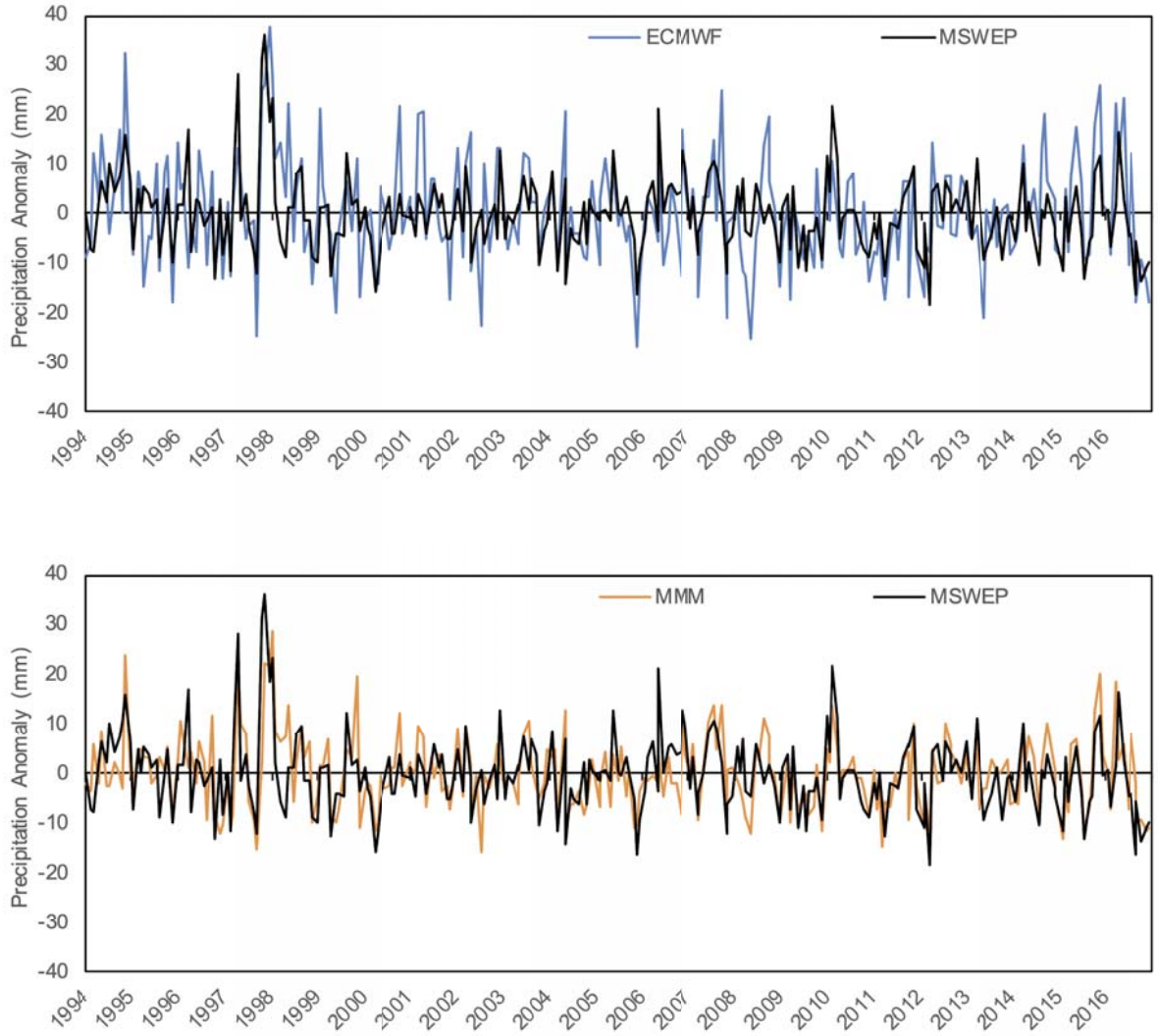


Figure 41. Spatial distribution of model weights (%) for monthly precipitation. The weights are calculated based on the covariability of the lead-1 forecasts from each model with the observational data.

Table 7. Model weights based on forecast skill for lead-1 monthly precipitation. Weights are averaged over the EN basin.

Model	EN basin average weight
ECMWF	0.25
UKMO	0.24
Meteo-France	0.18
DWD	0.15
CMCC	0.18

Figure 42 shows the EN basin average monthly precipitation anomalies for MSWEP, the ECMWF model (as an example of one of the better performing models) and the multi-model ensemble unweighted (MMM) and weighted mean (MMWM) for lead-1. The MMM and MMWM follow the observation (MSWEP) better than the ECMWF model, as expected. The difference between the MMM and MMWM is visually small. The equivalent correlation values are given in Table 8.



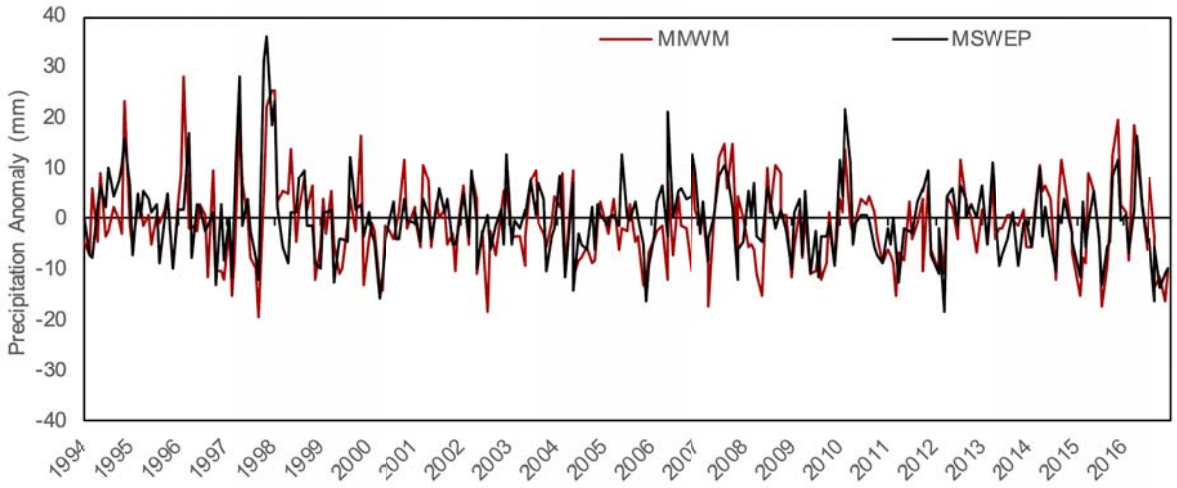


Figure 42. Time series of EN basin average monthly precipitation anomalies for 1994–2016 of the MSWEP observational dataset compared to the ECMWF model (top), the multi-model unweighted mean (MMM) (middle), and the multi-model weighted mean (MMWM) (bottom).

Table 8. Correlation of the lead-1, ensemble mean precipitation (absolute and anomaly) forecasts for each model and the multi-model unweighted (MMM) and weighted (MMWM) means.

Model	Monthly	Monthly anomaly
ECMWF	0.89	0.60
UKMO	0.90	0.52
Meteo-France	0.81	0.51
DWD	0.75	0.27
CMCC	0.83	0.46
MMM	0.92	0.64
MMWM	0.93	0.64

Figure 43 shows the spatial distribution of correlation values for the 5 models, the MMM and the MMWM for MAM, JJAS and OND. This is similar to Figure 30 but includes the MMM and MMWM results, which again show the higher skill in the lower part of the basin, but a slightly more robust spatial pattern of higher correlation values compared to the better performing models, such as ECMWF, and CMCC for MAM. This is also shown in Figure 44 for Tmax and Tmin for all months. As the weights for the MMWM are based on the skill values for all months pooled together, there may be opportunities to improve the skill of the MMWM by applying the weights separately for each month and lead time.

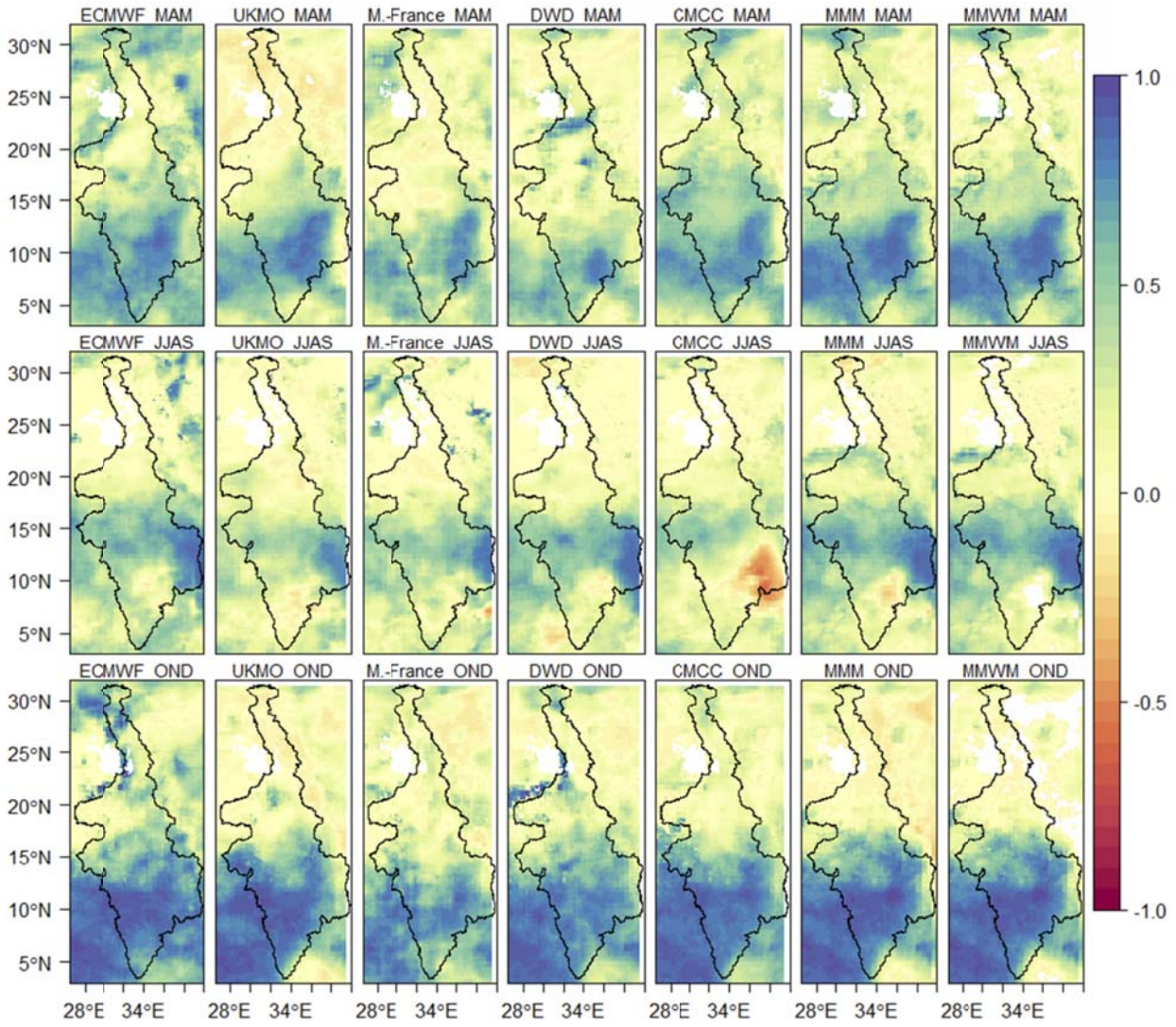


Figure 43. Spatial distribution of correlation between MSWEP and the five models, the unweighted multi-model mean, and the weighted multi-model mean, for (a) MAM, (b) JJAS, and (c) OND precipitation anomalies.

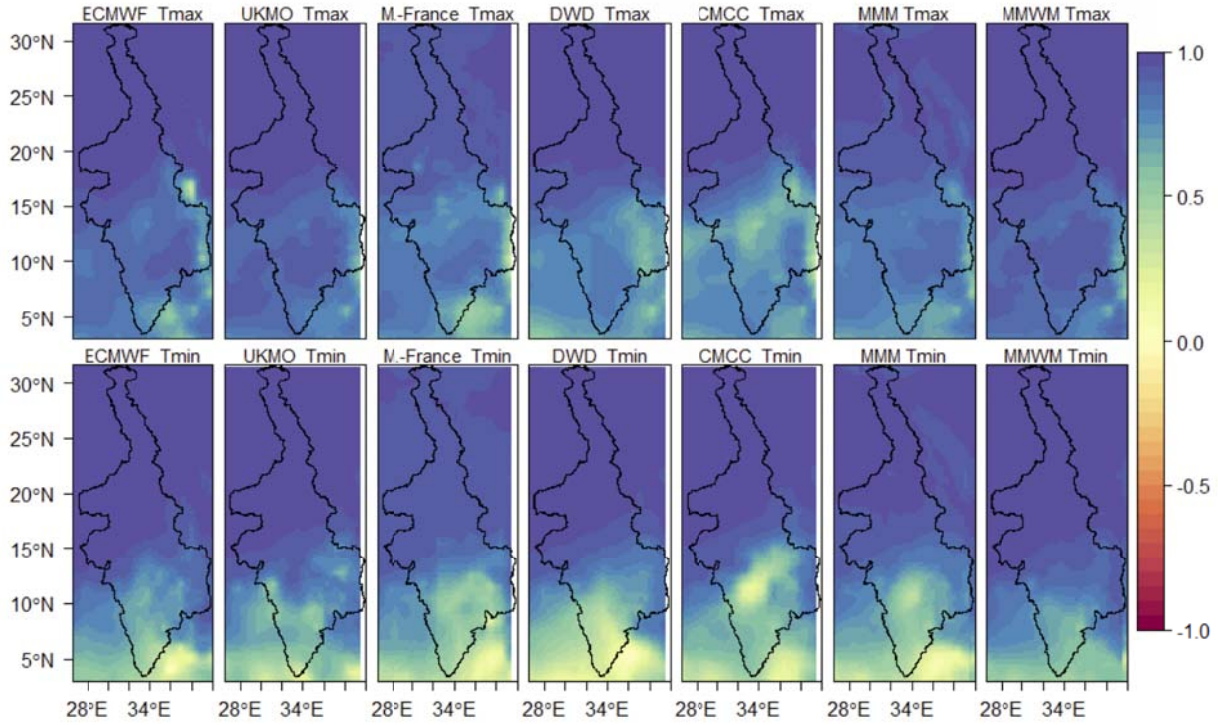


Figure 44. Spatial distribution of correlation between MSWEP and the five models, the unweighted multi-model mean, and the weighted multi-model mean, for (top) Tmax and (bottom) Tmin anomalies.

3.4. Evaluation of Hydrological forecasts

Evaluation of the hydrological forecasts is done in a similar way to the climate forecasts, focusing on monthly correlations at different lead times for the different models and the multi-model mean. Soil moisture is evaluated at the grid cell scale, and streamflow is evaluated at all river reaches in the full Nile basin, and for 30 selected forecast points of interest to ENTRO (Table 9).

Table 9. Locations and type of the 30 selected forecast points in the EN basin.

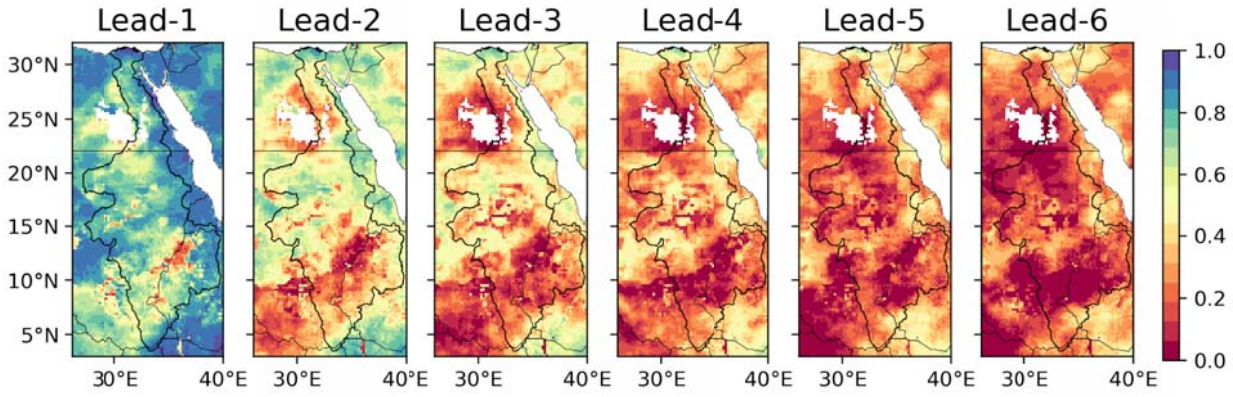
ID	Location	Region	Country	Lon (°)	Lat (°)	Purpose	Infrastr.
1	Baro-at-Gambella	Baro-Akobo	Ethiopia	34.65	8.23	eflow; irrigation	BAS_Gambella Dam
2	DS-Pibor-and-Akobo-confl	Baro-Akobo	South Sudan	32.98	7.89	dam operation; water supply; eflow;	BAS_Pibor Irrigation

						irrigation	
3	DS-Pibor-Baro-and-Gilo-confl	Baro-Akobo	South Sudan	33.15	8.49	eflow; irrigation	No
4	Sobat-at-Huleit-Duleib	Baro-Akobo	South Sudan	31.62	9.31	eflow; dam operation; irrigation	No
12	White-Nile-DS-Jebel-Awlia	White Nile	Sudan	32.50	15.58	reservoir operation; water supply; eflow; dam operation; irrigation	Jebel Awlia Dam
13	White-Nile-Malakal	White Nile	South Sudan	31.65	9.52	navigation; eflow; dam operation; irrigation	No
14	White-Nile-Melut	White Nile	South Sudan	32.20	10.43	navigation; eflow; dam operation; irrigation	No
25	TK5	Atbara	Ethiopia	38.75	13.34	dam operation; Irrigation	dam; irrigation
26	Hamadiet	Atbara	Ethiopia; Sudan	36.55	14.27	dam operation	DCUA at Setit; Atbara
27	Basalam	Atbara	Ethiopia	36.47	13.75	dam operation	DCUA at Setit; Atbara
28	Elassera	Atbara	Ethiopia; Sudan	36.15	12.98	dam operation	DCUA at Setit; Atbara
29	DCUA	Atbara	Sudan	35.91	14.26	dam operation	TA_RumelaBurdana Dam
30	Khashm-Girba-Dam	Atbara	Sudan	35.91	14.94	dam operation	KhashmElgirba Dam
31	Sedon	Atbara	Sudan	35.35	16.32	eflow	No
32	Ribb	Blue Nile	Ethiopia	37.66	12.05		No
33	Dirma	Blue Nile	Ethiopia	36.95	12.09		No
34	Bahir-Dar	Blue Nile	Ethiopia	37.37	11.58	dam operation	No
35	Karadobi	Blue Nile	Ethiopia	36.76	10.29	dam operation	No
36	Fincha	Blue Nile	Ethiopia	36.12	10.14	dam operation	No

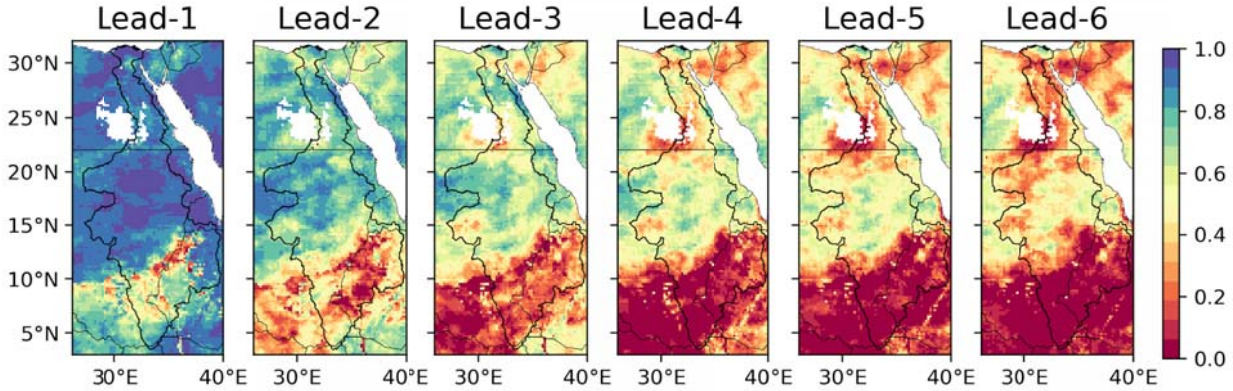
37	Beko-Abo	Blue Nile	Ethiopia	36.50	10.47	dam operation	No
38	Didessa	Blue Nile	Ethiopia	35.69	9.94	dam operation	BN_Didessa Dam
39	GERD/Eddiem	Blue Nile	Ethiopia	35.09	11.22	dam operation	GERD
41	Rosires-Dam	Blue Nile	Sudan	34.39	11.80	dam operation	Roseires Dam; Gezira Scheme
42	Singa	Blue Nile	Sudan	33.95	13.17	dam operation	Sinnar Dam; Gezira Scheme
43	Sennar-Dam	Blue Nile	Sudan	33.63	13.55	dam operation	Sinnar Dam; Gezira Scheme
44	Ombegara	Blue Nile	Sudan	33.97	13.73	eflow; dam operation; irrigation	Karsh Elfil at Rahad River
45	Karsh-Elfil-Rahad	Blue Nile	Sudan	34.07	13.91	eflow; dam operation; irrigation	Om Bagra at Dinder River
49	Merowe-dam	Main Nile	Sudan	32.06	18.68	dam operation	Merowe Dam
51	Kokka	Main Nile	Sudan	30.59	20.02	eflow	Kajbar Dam
52	High-Aswan-Dam	Main Nile	Egypt; Sudan	32.90	23.82	dam operation	High Aswan Dam

Figure 45 shows the skill of the HBV model forecasts for monthly soil moisture anomalies as forced by the downscaled and bias-corrected precipitation and temperature data from the climate models and ESP. The lead-1 correlations are generally above 0.8 for all models, except for in parts of the southern part of the EN basin. The ECMWF forced values are slightly higher in the wetter south than other models, but lower in the north The correlations drop quickly with lead time for all models in the south, but persist in the north for most models and the ESP, with values remaining above 0.6 in the middle part of the EN basin. This is expected to some extent because of the inherent persistence in soil moisture, but the slightly higher skill from the climate models compared to ESP is encouraging. The sub-basin correlations are summarized in Figure 46 for ECMWF and ESP, confirming that the climate model forced forecasts of soil moisture anomalies from ECMWF slightly outperform the ESP, especially for the Baro-Akabo-Sobat White Nile basin.

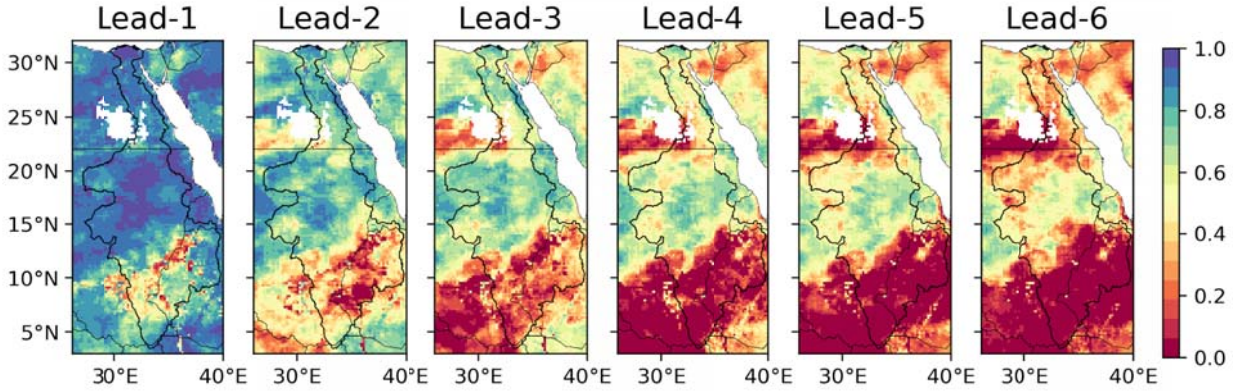
(a) ECMWF



(b) UKMO



(c) Meteo-France



(d) ESP

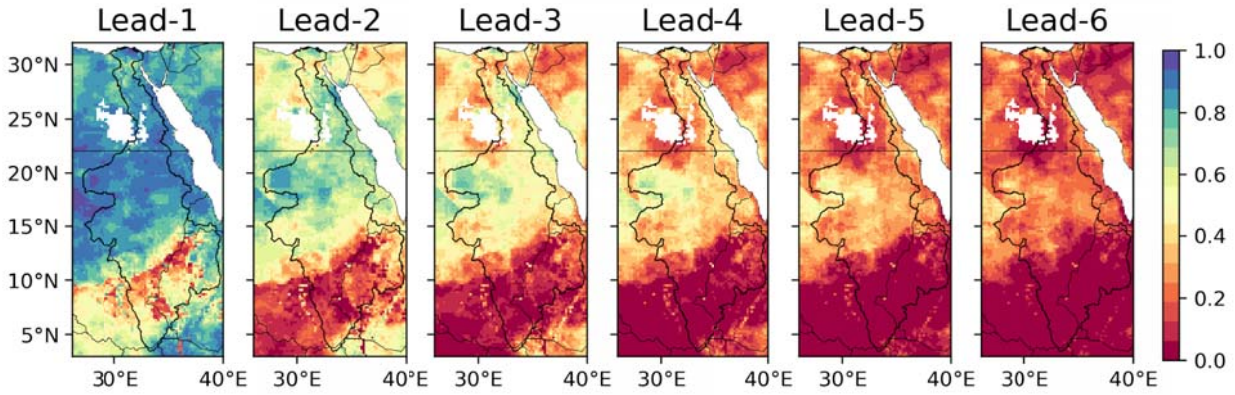
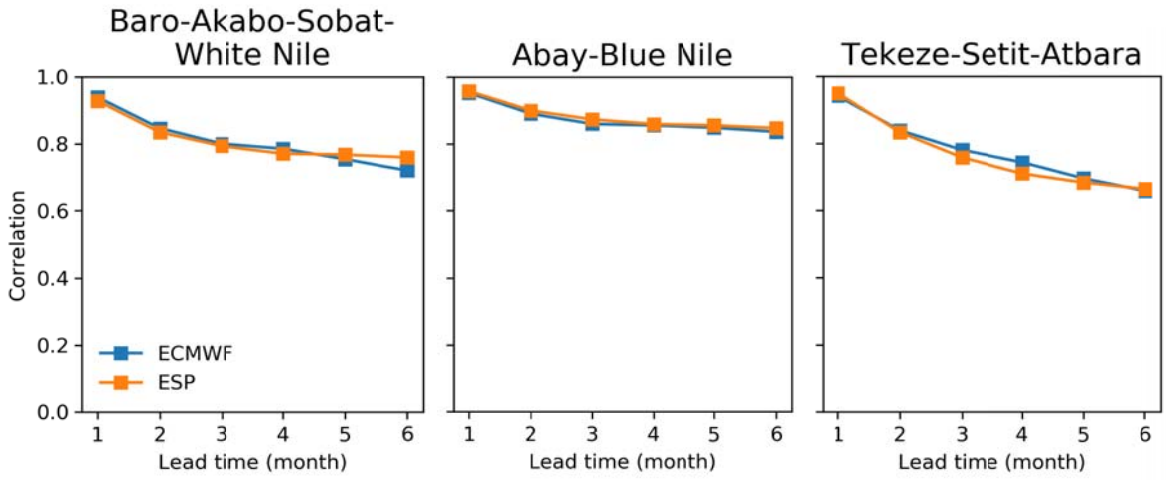


Figure 45. Spatial distribution of correlation of monthly soil moisture anomalies between the observational historic simulation and those forced by the ECMWF, UKMO, Meteo-France and ESP climate forecasts.



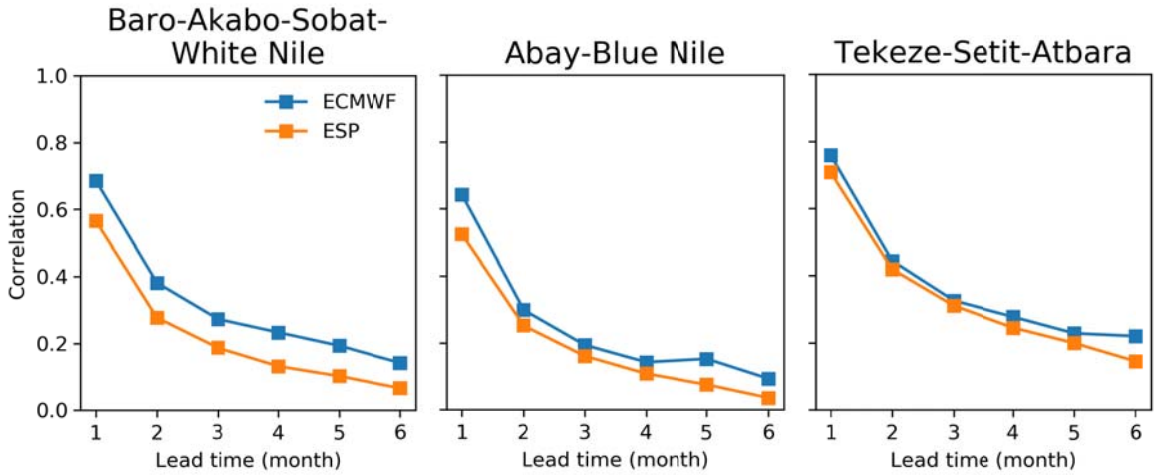
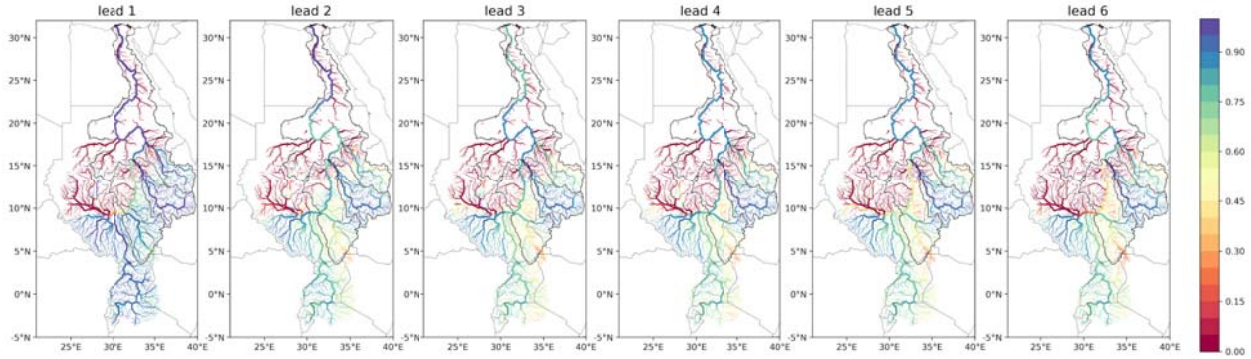


Figure 46. For the three main sub-basins of the lower part of the EN basin, mean correlations of monthly soil moisture (top) and anomalies (bottom) between the observational historic simulation and those forced by the ECMWF and ESP climate forecasts.

The skill for streamflow is summarized in Figure 47 for ECMWF and ESP, and some example forecasts for selected points (as listed in Table 9) are shown in Figure 48, including the ensemble members and ensemble mean. Figure 47 shows the diversity of skill across the basin which is dictated by the variation in the skill of climate forcing spatially and how that skill is propagated through the river network. Overall, the spatial variation in skill is similar between the two, but the ECMWF forecasts are less skilful than the ESP in the western part of the basin. ECMWF tends to slightly outperform ESP in other regions such as the southeast, but the differences are small and dependent on the lead time. Of note is the persistence of correlation values in certain regions, even out to 6 months, although this is partly due to the strong seasonality.



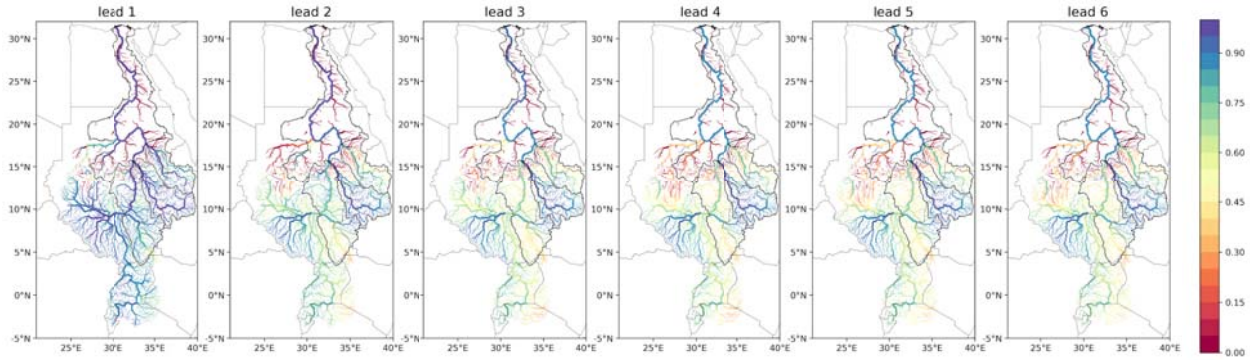


Figure 47. Correlation values of monthly streamflow at the 69,744 river reaches between the observational historic simulation and the simulations forced by the ECMWF (top) and ESP (bottom) climate forecasts.

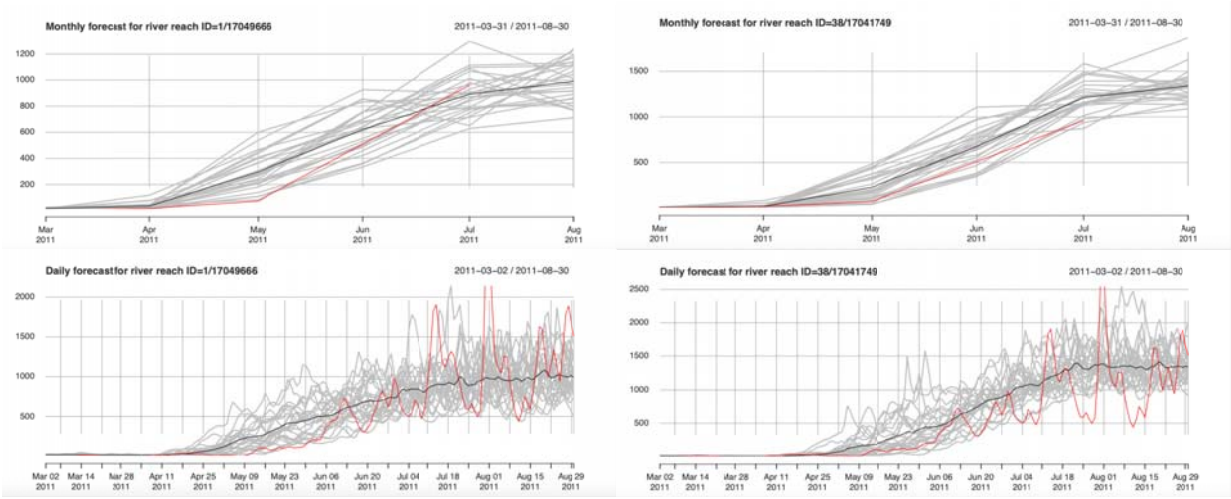


Figure 48. Examples of streamflow forecasts derived from the ECMWF climate forecasts, initialized on 1-March-2011, for (left) the Baro-at-Gambella Dam in Ethiopia (ID=1) and (right) the Didessa Dam on the Blue Nile in Ethiopia (ID=38), at daily and monthly time scales. The grey lines are the individual ensemble members, the black line if the ensemble mean, and the red is the observational estimate.

3.5. Online Dashboard

The Eastern Nile Drought Monitor system is accessible from a web-based application, which is made up of three parts:

1) a “dashboard” tab showing a summary of current conditions and the latest forecast;
2) a series of “analysis” tabs, which show more detailed information; and
3) an“About/Help” tab, which provides background information on the system and help on how to use the web app interface. More details on each of these tabs is provided next.

Dashboard Tab

The dashboard tab shows a summary of current conditions and the latest forecast (Figure 49). It consists of three main parts:

- A summary of drought in the Nile basin in three warning boxes which give a visual summary of the date of the latest update of the system, the area in drought averaged over the past week, and the change in drought area since the week before. These drought statistics are color-coded to indicate different levels of warning. For the drought area: green < 10%; 10% <= yellow < 30%; 30% <= orange < 50%; 50% < red. For the change in drought area: red indicates an increase and green indicates a decrease.
- A map showing current conditions for different hydrological variables and drought indices. Overlain on these are vector layers for context (e.g. dams, cities, river network) and to extract more detailed information for selected polygons (Nile basin, Nile sub-basins) and points of interest (river network locations). The map layers can be downloaded into standard GeoTiff format for use in other software such as a Geographical Information System (GIS).
- A time series of past conditions for the last 6 months and forecasts for the next 6 months relative to the date of the latest update. The data shown can be updated by clicking on the different polygons/points in the main map. Again, the data can be download to standard csv format. The time series shows past conditions for a range of hydrological variables, plus a drought index and the area in drought. The future conditions are shown just for the drought index and are presented as the mean of the forecast ensemble and the 5th and 95th percentiles of its distribution. Summary statistics of the time series information is given, showing the current area in drought, recent rainfall accumulation etc.

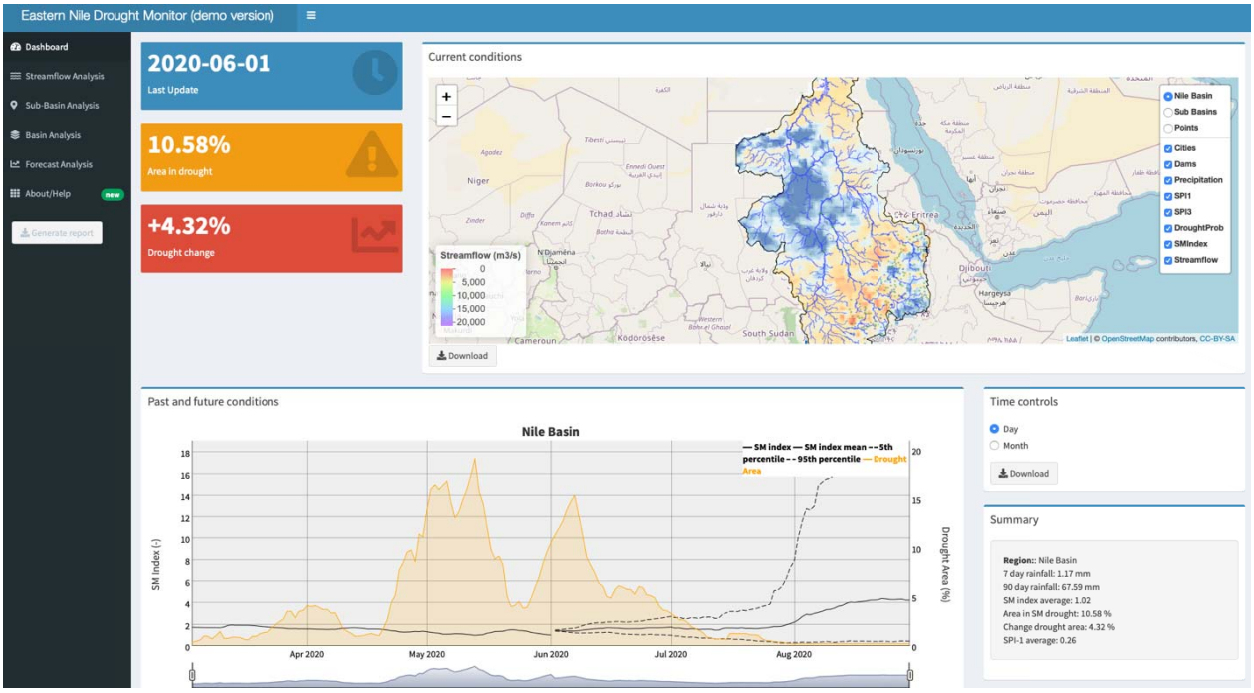


Figure 49. Screenshot of the EN drought monitoring system dashboard.

Streamflow Analysis Tab

This analysis tab (Figure 50) shows more detailed information on current and forecast conditions, with a focus on streamflow at a series of points of interest.On the left of the tab is are two time series charts showing the past 6 months and future 6 months forecast of streamflow for two locations of interest.The full list of locations is given in the table on the right, which also lists the current streamflow conditions (m3/s) and the equivalent percentile. The user can compare pairs of forecast points by selecting locations from the drop down list or clicking on the table rows.

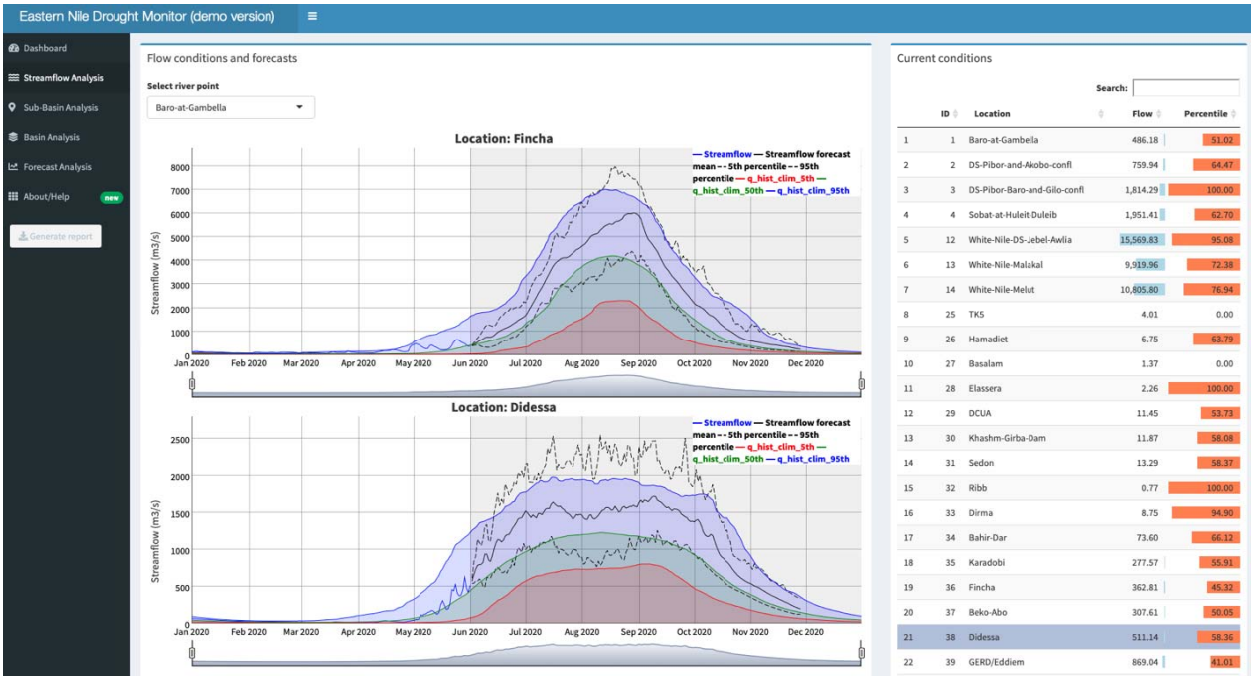


Figure 50. Screenshot of the EN drought monitoring system showing the analysis page for selected streamflow forecast points.

Sub-Basin Analysis Tab

This analysis tab shows more detailed information on current and forecast conditions for the sub-basins, with a focus on drought indices and water budget components.The full list of sub-basins is given in the table at the top, which also lists several statistics about recent rainfall and current drought indices. The user can sort the columns of the table by clicking on the arrows at the top of each column, and filter the list using the search box.

Below are three time series showing soil moisture drought conditions, SPI drought conditions and the water budget for the past 6 months and future 6 months forecast. The user can update which sub-basin is shown in each of the time series charts by clicking on a row of the table.

Basin Analysis Tab

This analysis tab shows more detailed information on current and forecast conditions for the EN basin, with a focus on maps of drought indices, and water budget components.

Forecast Analysis Tab

This analysis tab shows more detailed information on the forecasts (Figure 51). A series of information boxes shows information on the system models, forecast statistics, and forecast status. A series of maps showing the probability of drought for soil moisture and SPI for different lead times (1-month, 2-months, ...).Skill maps showing the long-term skill of the forecasts. These can be used to identify where forecasts are likely to be more robust.

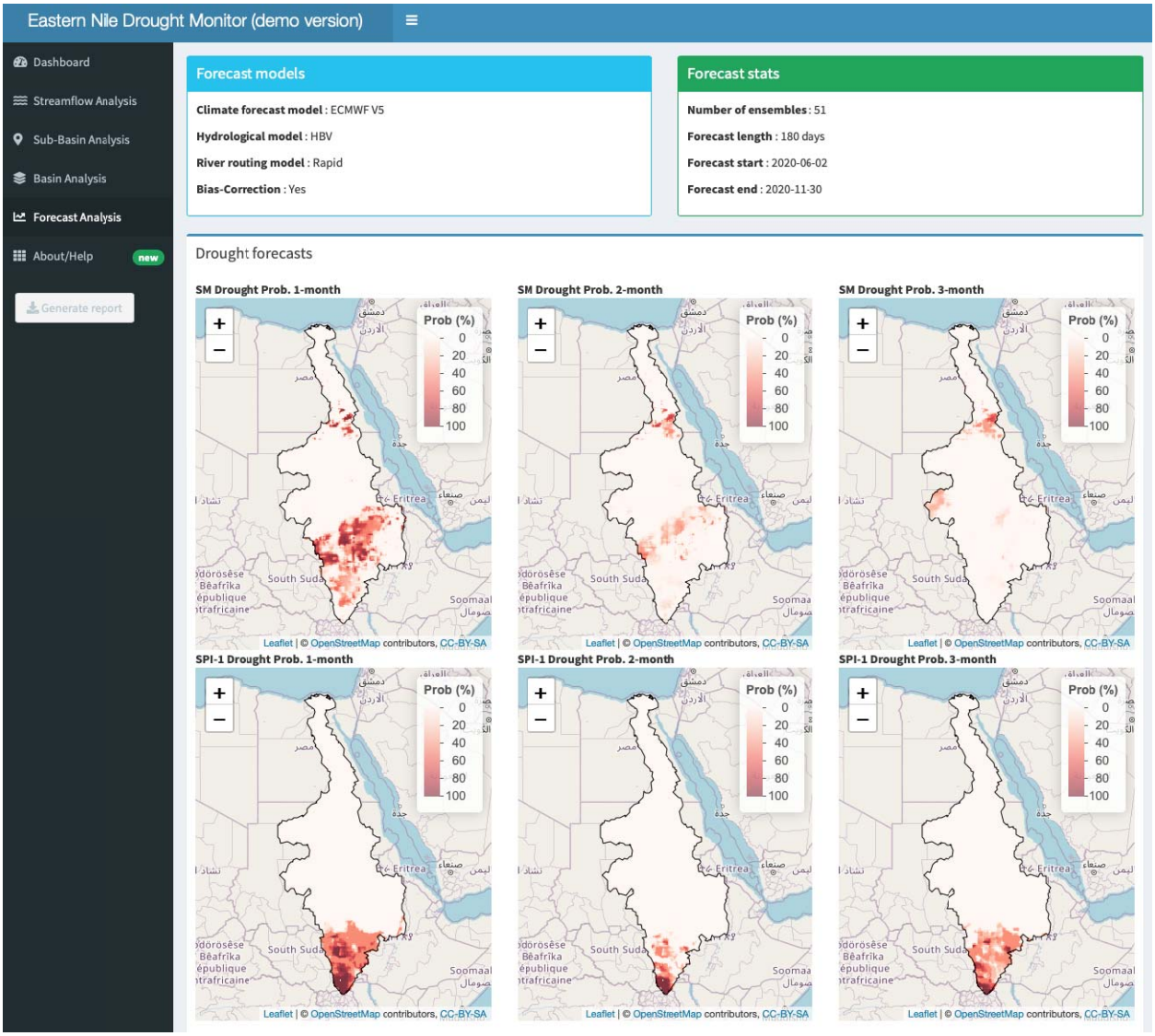


Figure 51. Screenshot of the EN drought monitoring system showing the forecasts analysis page, which presents statistics on the current forecasts and maps of the probability of drought for future time periods.

Report Download

The user can generate a downloadable report (in html or pdf format) that summarizes the monitoring and forecast data shown in the dashboard. The report includes textual summaries of current conditions (e.g. total area in drought, change in drought area, etc), and maps of current and future forecast conditions for the EN basin, as well as a timeseries of drought conditions averaged over the whole basin. Table summaries and timeseries of sub-basin average conditions (hydrology, drought metrics) are also provided. Finally a short summary of the background of the system is provided to give context.

4. Conclusions and Recommendations

A demonstration hydrological forecast system has been developed and implemented for the EN basin, focused on drought monitoring and forecasting. The system combines ground, satellite and model data to provide estimates of climate and hydrological variables at high resolution for the recent past, current conditions and forecasts out to 6 months. The forecasts are based on climate model forecasts from the ECWMF set of seasonal models, which are analyzed based on their set of hindcasts for their ability to forecast the evolution of seasonal climate and specific drought events.The climate forecasts are downscaled and bias corrected, and used to force a hydrological model (HBV) and river routing model (RAPID) to produce hydrological forecasts of key variables including soil moisture and streamflow. The skill of the hydrological forecasts is analyzed relative to climatological based forecasts. The system is demonstrated operationally for recent forecasts times, accessible from an online interface that includes a dashboard of current and forecast conditions across the basin, and a set of more detailed analysis functions.

4.1. Conclusions

1. The reconstruction of historic climate and hydrology provides a high-resolution and state-of-the-art dataset for use in hydroclimate studies and water resources assessments, and forms the basis of the EN drought monitoring system. The dataset is based on the MSWEP precipitation dataset which has been shown to perform well over the basin, and outperforms other similar gridded large-scale datasets here and in most other regions globally. Because of this, the estimated hydrological variables likely provide a good representation of the hydrological cycle over the basin. Comparisons with observed streamflow records for a limited set of locations in the basin show reasonable performance to evidence this. Previous evaluations against streamflow and soil moisture observations in other locations globally confirm this.
2. The system demonstratesreasonable skillfor the climate variables (precipitation and temperature) with better skill for temperature, as expected. There is reasonable skill in estimating the evolution of seasonal climate at short lead times (1-2 months), and modest skill in predicting specific drought events. The five evaluated climate models show a range of skill with some models outperforming the others in general (e.g. ECMWF) although this is not the case for all metrics, seasons and lead times.
3. Improved skill is found by merging models together into a multi-model ensemble, which matches or outperforms any individual modelfor the skill metrics evaluated, which is expected. The weighted multi-model ensemble improves the skill slightly but further work is needed to analyze performance and refine the model skill weightings. Because of this, and to simplify the running of

the system, the best performing model, ECWMF, was chosen to provide the climate forecasts operationally.

4. The skill of the hydrological forecasts is generally better than for the climate forecasts (despite being driven by these) because of hydrological persistence, especially for soil moisture and for streamflow in larger rivers and downstream. There is a diversity of skill across the basin, which is a function of the climate forcing skill but also how skill is propagated through the river network, and this generally outperforms the climatologically based ESP forecasts.
5. The system has beendemonstrated operationally for recent forecasts times (June, July, August, 2020), which confirms that the data processing and modeling chain works and can be run automatically with very little human interaction.
6. Overall, the system provides useful predictive skill for hydrological monitoring and forecasting with a focus on drought, and there is an expectation that it can help support information services in ENTRO and decision making in the NBI and its member states.

4.2. Recommendations

1. The system provides high-resolution estimates at 5km for the gridded climate and hydrology data, and on a vector-based river network for 10,000s of river reaches. It is recommended that gridded data and associated modeling be increased to, say, 1km resolution, to represent the heterogeneity of the landscape, especially in the wetter and more topographically diverse regions of the basin. Although this is constrained by the lack of fine spatial resolution rainfall data, there may be benefits to focusing on higher resolutionfor certain parts of the basin.
2. Further validation of the historic estimates and forecasts using more available ground observations as well as more use of satellite estimates of hydrologic variables. Currently, validation has been done for a small set of stations in Ethiopia with reasonable length of records (9 stations) and it would be useful to expand this to more stations across the wetter part of the basin. Furthermore, given the availability of satellite remote sensing data of many aspects of the water cycle (soil moisture, evapotranspiration, water levels) it would be useful to make comparisons spatially, with a focus on spatial metrics of how well the patterns of variability are presented.
3. The operational system is based on a single hydrological model (HBV), which despite being carefully implemented and evaluated, has errors in the representation of hydrology in the basin.This is partially due to errors in the climate forcing, and due to parameter uncertainties despite the regionalization of parameters. It is recommended that other models are considered to capture the structural uncertainties in the modeling approach. For example, the HBV model has simplified representation of certain processes in the sub-surface, and could be compared to models which have a more sophisticated approach (e.g. VIC). There is scope to provide an ensemble of hydrological estimates based on multiple models as well as via incorporation of parameter uncertainties within those models.
4. Similarly, a single climate forecast model (ECMWF) is implemented in the operational system and there would be benefits to further develop and implement a multi-model ensemble to take advantage of individual climate model skill. Further analysis is needed to understand where and when the models are skillful, and therefore how they could be combined better into a weighted multi-model ensemble. Other forecast approaches could also be considered based on climatology

and conditional sampling, as well as statistical models for certain regions/locations and specific forecast variables.

5. Further stakeholder and user feedback is needed to refine the system and its outputs. In particular, further discussions would be useful to understand how the system can be used operationally (technically and for decision making)and how the system can be better integrated into ENTRO's technical operations. In terms of decision making, further discussions are required to understand how the system could be used and how it can be better tailored to provide decision relevant information and data, including refinements to the online dashboard.

5. References

Abtew, W., A. M. Melesse, and T. Dessalegne, 2009: El Niño Southern Oscillation link to the Blue Nile River Basin hydrology. *Hydrol. Process.*, 23, 3653–3660.

Amani et al., 2019. Strengthening flood and drought risk management tools for the Lake Chad basin, in *Water resource management in the climate change context in Africa*, E. S. Diop, P. Scheren and A. Niang (eds), in press.

Awange, J. L., E. Forootan, M. Kuhn, J. Kusche, and B. Heck, 2014: Water storage changes and climate variability within the Nile Basin between 2002 and 2011, *Advances in Water Resources*, 73, 1-15, https://doi.org/10.1016/j.advwatres.2014.06.010.

Bastola, S., Ishidaira, H., & Takeuchi, K. Regionalisation of hydrological model parameters under parameter uncertainty: A case study involving TOPMODEL and basins across the globe. Journal of Hydrology, 357 (3), 188–206, 2008.

Bayissa, Y., S. Maskey, T. Tadesse, S. J. Van Andel, S. Moges, A. Van Griensven and D. Solomatine, 2018: Comparison of the Performance of Six Drought Indices in Characterizing Historical Drought for the Upper Blue Nile Basin, Ethiopia. *Geosciences* 2018, *8*(3), 81; https://doi.org/10.3390/geosciences8030081

Beck, H.E., E.F. Wood, M. Pan, C.K. Fisher, D.G. Miralles, A.I. van Dijk, T.R. McVicar, and R.F. Adler, 2019a: MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment. Bull. Amer. Meteor. Soc., 100, 473–500, https://doi.org/10.1175/BAMS-D-17-0138.1, 2019a.

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., van Dijk, A. I. J. M., Huffman, G. J., Adler, R. F., and Wood, E. F.: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS, Hydrol. Earth Syst. Sci., 23, 207–224, https://doi.org/10.5194/hess-23-207-2019, 2019b.

Beck, H.E., E.F. Wood, T.R. McVicar, M. Zambrano-Bigiarini, C. Alvarez-Garreton, O.M. Baez-Villanueva, J. Sheffield, and D.N. Karger, 0: Bias correction of global high-resolution precipitation climatologies using streamflow observations from 9372 catchments. J. Climate, 0, https://doi.org/10.1175/JCLI-D-19-0332.1, 2019c.

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A.: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, Hydrol. Earth Syst. Sci., 21, 589–615, https://doi.org/10.5194/hess-21-589-2017, 2017a.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F.: Global-scale evaluation of 22 precipitation

datasets using gauge observations and hydrological modeling, Hydrol. Earth Syst. Sci., 21, 6201–6217, https://doi.org/10.5194/hess-21-6201-2017, 2017b.

Beck, H.E., M. Pan, P. Lin, J. Seibert, A.I.J.M. van Dijk, and E.F. Wood, in review: Global fully-distributed parameter regionalization based on observed streamflow from 4229 headwater catchments, available on request.

Berg, A., and J. Sheffield, 2018: Climate Change and Drought: the Soil Moisture Perspective.*CurrClim Change Rep*4, 180–191. doi:10.1007/s40641-018-0095-0

Buchhorn, M., B. Smets, L. Bertels, M. Lesiv, N.-E. Tsendbazar, M. Herold, and S. Fritz, 2019: Copernicus Global Land Service: Land Cover 100m: epoch 2015: Globe. Dataset of the global component of the Copernicus Land Monitoring Service 2019. DOI 10.5281/zenodo.3243509

Camberlin P, Philippon N (2002) The East African march–may rainy season: associated atmospheric dynamics and predictability over the 1968–97 period. *J. Climate*, 15:1002–1019. doi: 10.1175/1520-0442(2002)015<1002:TEAMMR>2.0.CO;2

Diro, G. T., E. Black, and D. I. F. Grimes, 2008: Seasonal forecasting of Ethiopian spring rains. *Meteorol. Appl.*, 15:73–83. doi: 10.1002/met.63

Fortin, F., F. De Rainville, M. Gardner, M. Parizeau, and C. Gagne (2012), DEAP: Evolutionary algorithms made easy, J. Mach. Learn. Res., 13, 2171–2175.

Gebrehiwot, S.G.; Seibert, J.; Gärdenäs, A.I.; Mellander, P.E.; Bishop, K. Hydrological change detection using modeling: Half a century of runoff from four rivers in the Blue Nile basin. Water Resour. Res. 2013, 49, 3842–3851.

Gleixner, S., Keenlyside, N., Viste, E. et al., 2017: The El Niño effect on Ethiopian summer rainfall.*Clim. Dyn.*, 49: 1865. https://doi.org/10.1007/s00382-016-3421-z

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 370, 80–91, 2009.

Hargreaves, G. H. (1994). Dening and using reference evapotranspiration. Journal of Irrigation and Drainage Engineering, 120 (6), 1132–1139.

Huang, B., P. W. Thorne, V. F. Banzon, T. Boyer, G. Chepurin, J. H. Lawrimore, M. J. Menne, T. M. Smith, R. S. Vose, and H. Zhang, 2017: Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons.*J. Climate*, 30, 8179–8205, https://doi.org/10.1175/JCLI-D-16-0836.1

Hundecha, Y., and Bardossy, A. Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. Journal of Hydrology, 292, 1(4), 281–295, 2004.

Kebede, A., U. Jaya Prakash, D. Koricha, and M. Nigussie, 2019: Characterizing and Monitoring Drought over Upper Blue Nile of Ethiopia with the Aid of Copula Analysis. 10.9734/IJECC/2018/45102.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, J. Hydrol., 424–425, 264–277, https://doi.org/10.1016/j.hydrol.2012.01.011, 2012.

Mishra, A. K., Singh, V.P. A review of drought concepts. J. Hydrol. (2010), doi:10.1016/j.jhydrol.2010.07.012

Rientjes, T.; Perera, B.; Haile, A.; Reggiani, P.; Muthuwatta, L. Regionalisation for lake level simulation-the case of Lake Tana in the upper Blue Nile, Ethiopia. Hydrol. Earth Syst. Sci. 2011, 15, 1167–1183.

Segele, Z. T., P. J. Lamb, and L. M. Leslie, 2009: Large-scale atmospheric circulation and global sea surface temperature associations with Horn of Africa June–September rainfall. *Int. J. Climatol.* , 29:1075–1100

Seleshi, Y., and U. Zanke, 2004: Recent changes in rainfall and rainy days in Ethiopia. *International Journal of Climatology*, 24: 973–983.

Sheffield J., and E. F. Wood, 2008: Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations, *Climate Dynamics*, 13 (1), 79-105, doi:10.1007/s00382-007-0340-z.

Sheffield, J., and E. F. Wood, 2011: Drought: Past Problems and Future Scenarios, Earthscan, UK, pp 192.

Sheffield, J., E. F. Wood, N. Chaney, K. Guan, S. Sadri, X. Yuan, L. Olang, A. Amani, A. Ali, and S. Demuth, 2014; A Drought Monitoring and Forecasting System for Sub-Sahara African Water Resources and Food Security. Bull. Am. Met. Soc., 95, 861–882. doi: http://dx.doi.org/10.1175/BAMS-D-12-00124.1

Seibert, J. and Vis, M. J. P.: Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, Hydrol. Earth Syst. Sci., 16, 3315–3325, https://doi.org/10.5194/hess-16-3315-2012, 2012.

Van Lanen, H.A.J., M. Fendeková, E. Kupczyk, A. Kasprzyk, and W. Pokojski, 2004. Flow generating processes. In: Tallaksen, L.M., and H. A. J. Van Lanen (Eds.), Hydrological drought - Processes and estimation methods for streamflow and groundwater. Developments in Water Science, 48, Elsevier Science B.V., pp. 53-96.

WorldPop (www.worldpop.org - School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). https://dx.doi.org/10.5258/SOTON/WP00645

Worqlul, A.W.; Yen, H.; Collick, A.S.; Tilahun, S.A.; Langan, S.; Steenhuis, T.S. Evaluation of CFSR, TMPA 3B42 and ground-based rainfall data as input for hydrological models, in data-scarce regions: The upper Blue Nile Basin, Ethiopia. Catena 2017, 152, 242–251.

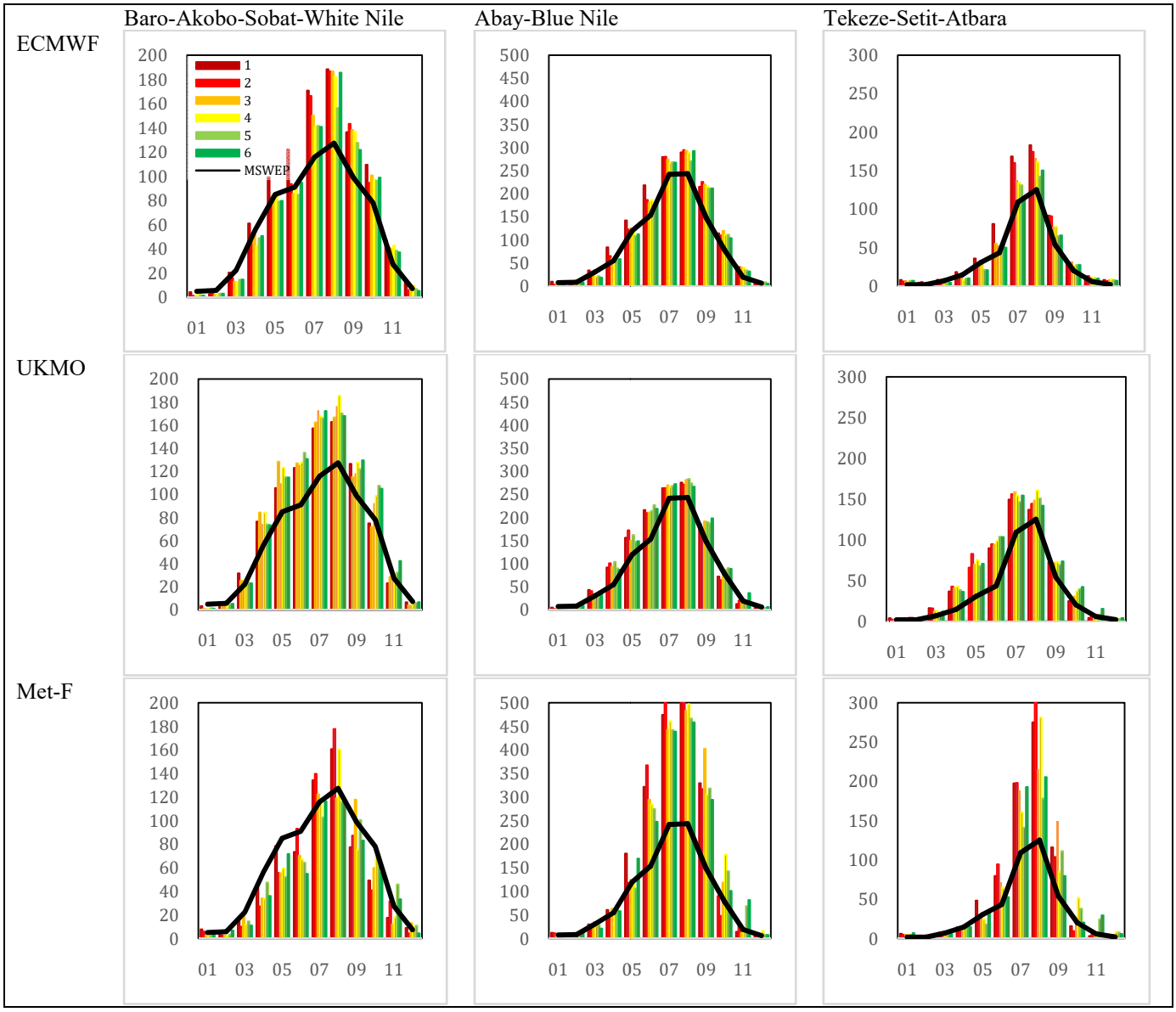
Yamazaki D., D. Ikeshima, R. Tawatari, T. Yamaguchi, F. O'Loughlin, J.C. Neal, C.C. Sampson, S. Kanae& P.D. Bates, 2017: A high accuracy map of global terrain elevations. Geophysical Research Letters, 44, 5844–5853, 2017 doi: 10.1002/2017GL072874.

Yamazaki D., D. Ikeshima, J. Sosa, P.D. Bates, G.H. Allen, T.M. Pavelsky, 2019: MERIT Hydro: A high-resolution global hydrography map based on latest topography datasets. Water Resources Research, doi: 10.1029/2019WR024873

Yokoo, Y., S. Kazama, M. Sawamoto, and H. Nishimura (2001), Regionalization of lumped water balance model parameters based on multiple regression, J. Hydrol., 246(1–4), 209–222.

Zargar, A., R. Sadiq, B. Naser, and F. I. Khan, 2011: A review of drought indices. *Environmental Reviews,19*, 333-349.

Appendix


Baro-Akobo-Sobat-White Nile
Abay-Blue Nile
Tekeze-Setit-Atbara
ECMWF
UKMO
Met-F
1
2
3
4
5
6
MSWEP
01 03 05 07 09 11

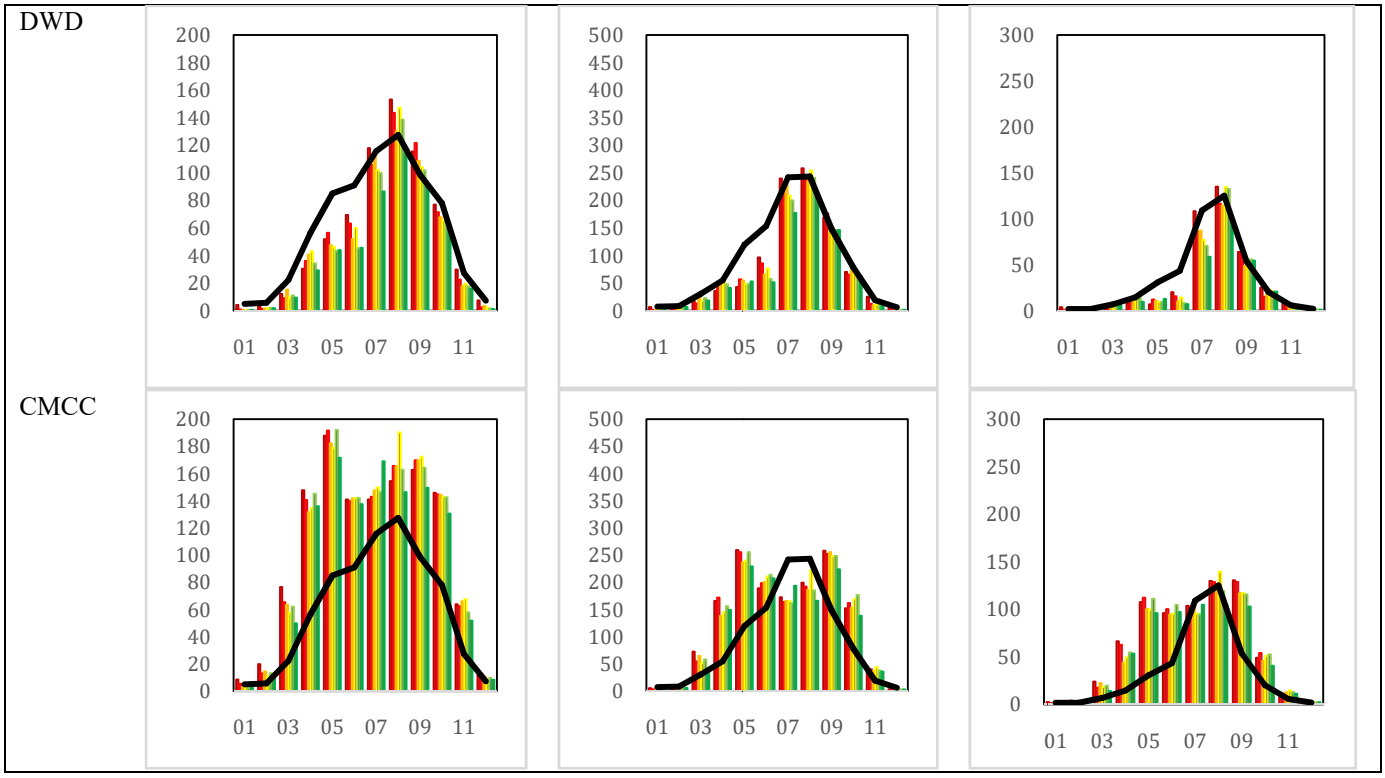


Figure A1. Precipitation monthly average climatological mean (1993–2016) for the sub-basins for MSWEP and the five climate models, for leads 1–6. Units are mm/month.

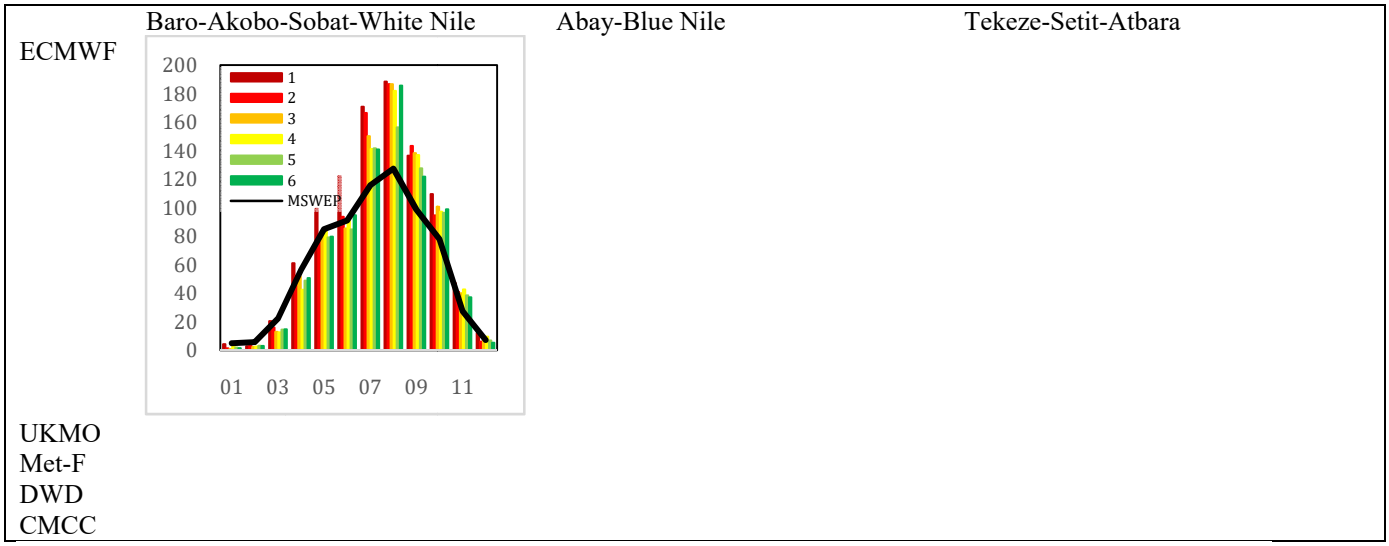
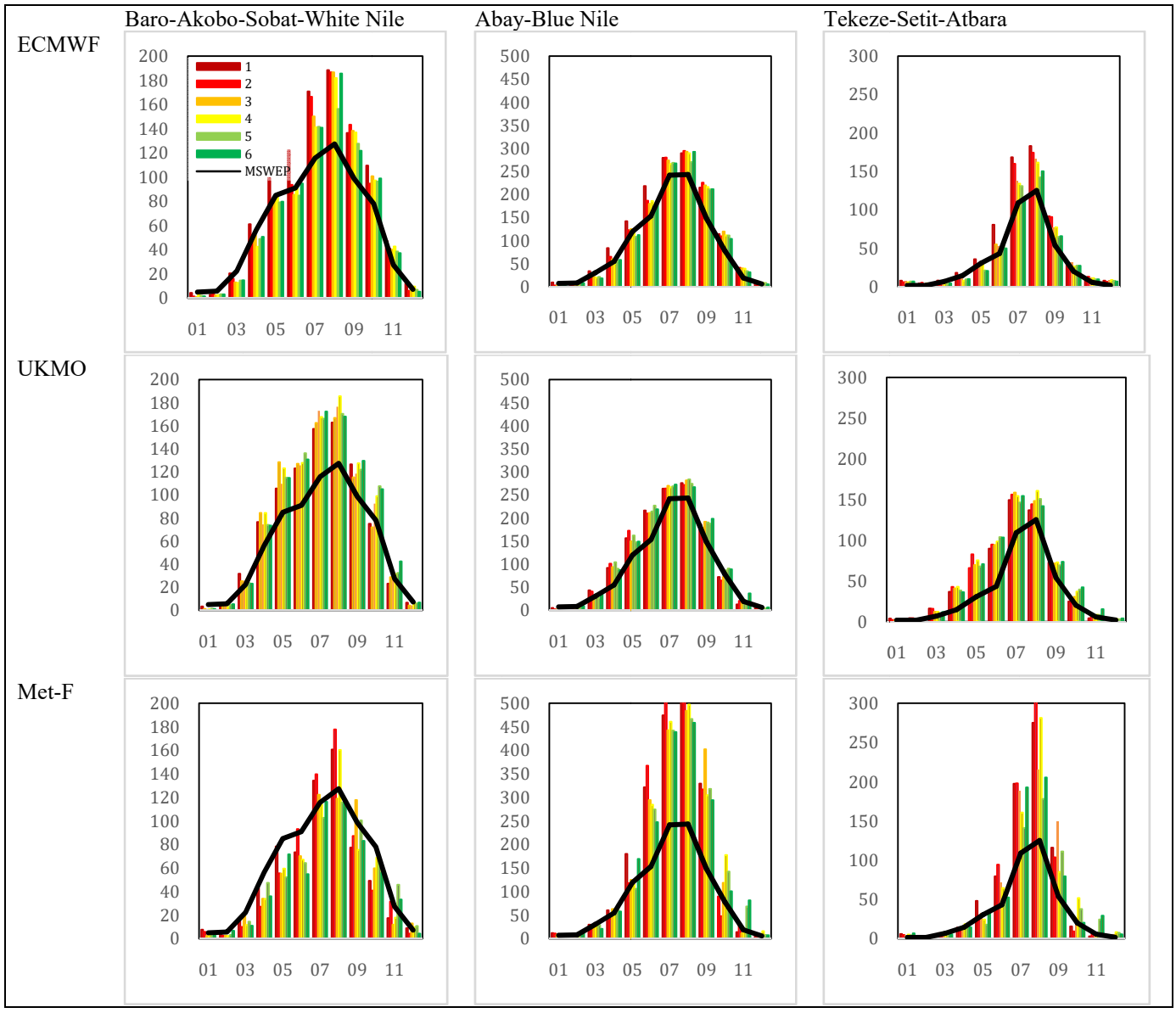


Figure A2.Tmax monthly average climatological mean (1993–2016) averaged over theEN basin for MSMet and the five climate models, for leads 1–6 months.



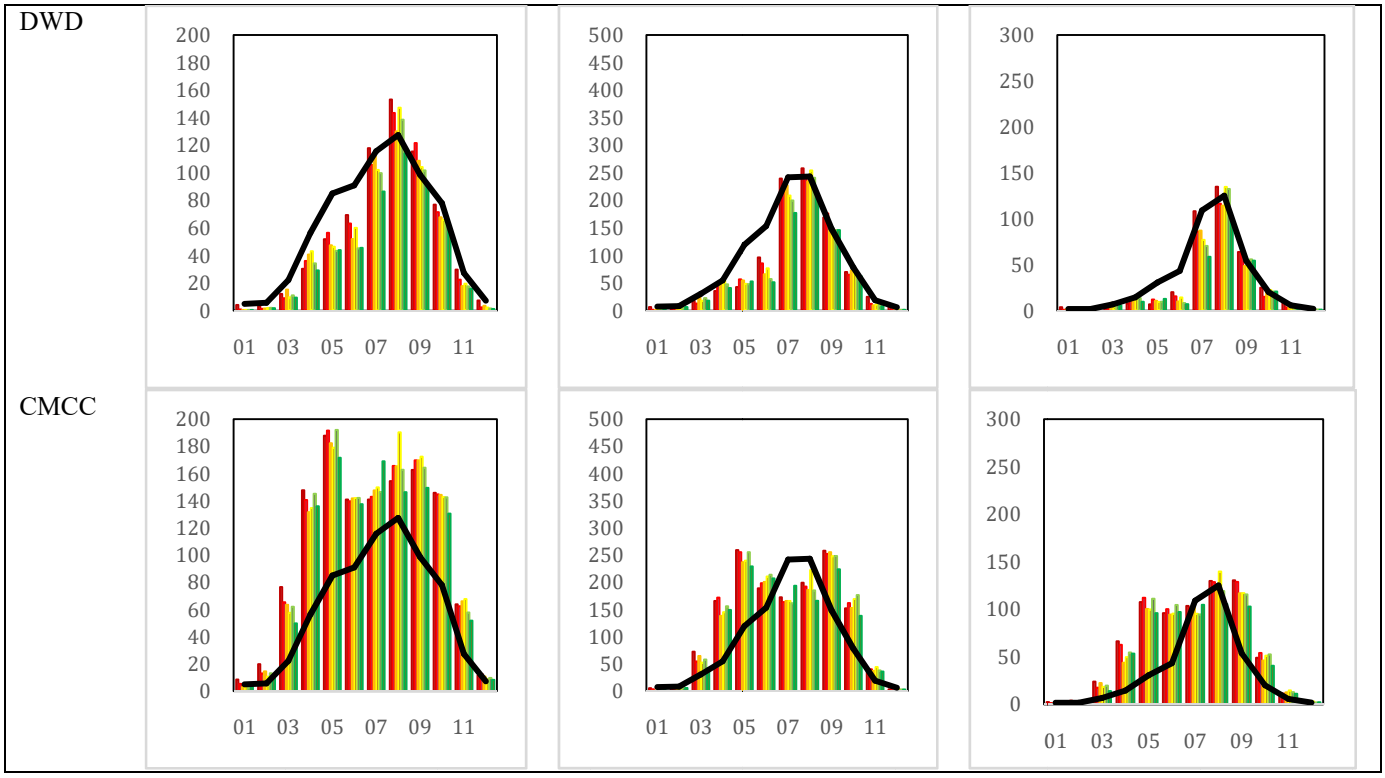


Figure A3.Tmin monthly average climatological mean (1993–2016) averaged over theEN basin for MSMet and the five climate models, for leads 1–6 months.