

## Dynamical and statistical downscaling of a global seasonal hindcast in eastern Africa

Grigory Nikulin<sup>a,\*</sup>, Shakeel Asharaf<sup>b,1</sup>, María Eugenia Magariño<sup>c</sup>, Sandro Calmanti<sup>e</sup>, Rita M. Cardoso<sup>d</sup>, Jonas Bhend<sup>g</sup>, Jesús Fernández<sup>c</sup>, María Dolores Frías<sup>c</sup>, Kristina Fröhlich<sup>b</sup>, Barbara Früh<sup>b</sup>, Sixto Herrera García<sup>c</sup>, Rodrigo Manzananas<sup>f</sup>, José Manuel Gutiérrez<sup>f</sup>, Ulf Hansson<sup>a</sup>, Michael Kolax<sup>a</sup>, Mark A. Liniger<sup>g</sup>, Pedro M.M. Soares<sup>d</sup>, Christoph Spirig<sup>g</sup>, Ricardo Tome<sup>d</sup>, Klaus Wyser<sup>a</sup>

<sup>a</sup> Swedish Meteorological and Hydrological Institute, Folkborgsvägen 17, 60176 Norrköping, Sweden

<sup>b</sup> Deutscher Wetterdienst, Frankfurter Straße 135, 63067 Offenbach am Main, Germany

<sup>c</sup> Meteorology Group, Dept. Applied Mathematics and Computer Science, University of Cantabria, Santander, Spain

<sup>d</sup> Instituto Dom Luiz, Faculdade de Ciências da Universidade de Lisboa, 1749-016 Lisboa, Portugal

<sup>e</sup> ENEA, Agenzia Nazionale per le Nuove Tecnologie, l'Energia e lo Sviluppo Economico Sostenibile, Via Anguillarese 301, Roma, Italy

<sup>f</sup> Meteorology Group, Instituto de Física de Cantabria (CSIC-UC), Santander, Spain

<sup>g</sup> Federal Office of Meteorology and Climatology MeteoSwiss, Zurich, Switzerland

### ARTICLE INFO

#### Keywords:

Seasonal forecast

Eastern Africa

Downscaling

Drought early-warning system

### ABSTRACT

Within the FP7 EUPORIAS project we have assessed the utility of dynamical and statistical downscaling to provide seasonal forecast for impact modelling in eastern Africa. An ensemble of seasonal hindcasts was generated by the global climate model (GCM) EC-EARTH and then downscaled by four regional climate models and by two statistical methods over eastern Africa with focus on Ethiopia. The five-month hindcast includes 15 members, initialised on May 1 st covering 1991–2012. There are two sub-regions where the global hindcast has some skill in predicting June–September rainfall (northern Ethiopia – northeast Sudan and southern Sudan - northern Uganda). The regional models are able to reproduce the predictive signal evident in the driving EC-EARTH hindcast over Ethiopia in June–September showing about the same performance as their driving GCM. Statistical downscaling, in general, loses a part of the EC-EARTH signal at grid box scale but shows some improvement after spatial aggregation. At the same time there are no clear evidences that the dynamical and statistical downscaling provide added value compared to the driving EC-EARTH if we define the added value as a higher forecast skill in the downscaled hindcast, although there is a tendency of improved reliability through the downscaling. The use of the global and downscaled hindcasts as input for the Livelihoods, Early Assessment and Protection (LEAP) platform of the World Food Programme in Ethiopia shows that the performance of the LEAP platform in predicting humanitarian needs at the national and sub-national levels is not improved by using downscaled seasonal forecasts.

### Practical Implications

We present work on downscaling a seasonal hindcast in

eastern Africa done in the FP7 EUPORIAS project. The main focus in our activities was on assessing the utility of downscaling techniques to provide seasonal forecasts for impact models in eastern Africa and answering the question “Can downscaling show a higher predictive skill on seasonal time

\* Corresponding author.

E-mail addresses: [grigory.nikulin@smhi.se](mailto:grigory.nikulin@smhi.se) (G. Nikulin), [shakeel.asharaf@jpl.nasa.gov](mailto:shakeel.asharaf@jpl.nasa.gov) (S. Asharaf), [eugenia.magarino@unican.es](mailto:eugenia.magarino@unican.es) (M.E. Magariño), [sandro.calmanti@enea.it](mailto:sandro.calmanti@enea.it) (S. Calmanti), [rmcardoso@fc.ul.pt](mailto:rmcardoso@fc.ul.pt) (R.M. Cardoso), [jonas.bhend@meteoswiss.ch](mailto:jonas.bhend@meteoswiss.ch) (J. Bhend), [jesus.fernandez@unican.es](mailto:jesus.fernandez@unican.es) (J. Fernández), [mariadolores.frias@unican.es](mailto:mariadolores.frias@unican.es) (M.D. Frías), [kristina.froehlich@dwd.de](mailto:kristina.froehlich@dwd.de) (K. Fröhlich), [barbara.fruh@dwd.de](mailto:barbara.fruh@dwd.de) (B. Früh), [sixto.herrera@unican.es](mailto:sixto.herrera@unican.es) (S.H. García), [rmanzananas@ifca.unican.es](mailto:rmanzananas@ifca.unican.es) (R. Manzananas), [manuel.gutierrez@unican.es](mailto:manuel.gutierrez@unican.es) (J.M. Gutiérrez), [ulf.hansson@smhi.se](mailto:ulf.hansson@smhi.se) (U. Hansson), [michael.kolax@smhi.se](mailto:michael.kolax@smhi.se) (M. Kolax), [mark.liniger@meteoswiss.ch](mailto:mark.liniger@meteoswiss.ch) (M.A. Liniger), [pmssoares@fc.ul.pt](mailto:pmssoares@fc.ul.pt) (P.M.M. Soares), [christoph.spirig@meteoswiss.ch](mailto:christoph.spirig@meteoswiss.ch) (C. Spirig), [rtome@fc.ul.pt](mailto:rtome@fc.ul.pt) (R. Tome), [klaus.wyser@smhi.se](mailto:klaus.wyser@smhi.se) (K. Wyser).

<sup>1</sup> Present address: Jet Propulsion Laboratory, 4800 Oak Grove Drive, MS 233-301A, Pasadena, CA 91109, USA.

<https://doi.org/10.1016/j.cliser.2017.11.003>

Received 8 January 2017; Received in revised form 5 November 2017; Accepted 10 November 2017

Available online 01 December 2017

2405-8807/ © 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

scales comparing to its global driving seasonal forecast?” In particular, the Drought Early-Warning System – LEAP of the World Food Programme (WFP) was used to predict humanitarian needs at the national and sub-national levels taking global and downscaled hindcasts as input data.

At the beginning of the EUPORIAS project after consultations with WFP, it was decided to focus on the Kiremt rainy season (June–September, JJAS) in Ethiopia using a seasonal hindcast initialised in May, which can be used as input to the LEAP system. While the potential predictability of rainfall in eastern Africa has been known for a relatively long-time, the orography of Ethiopia is complex and it was considered important to assess the possibility of improving the accuracy of forecast large-scale rainfall patterns over this particular area at seasonal time scales. This was also a trade-off between user needs, more keen on rainy season forecasts, when impacts of water deficits on agriculture are larger, and forecast skill, which peaked south of Ethiopia in November–January, associated to the El Niño–Southern Oscillation (ENSO) variability. We finally opted for addressing the end-user needs, focusing on JJAS.

A five-month global seasonal hindcast of 15 members was generated using the EC-EARTH model for the 1991–2012 period at about 80 km resolution and then downscaled over eastern Africa by four regional climate models at about 25 km resolution and by two statistical methods at about 50 km resolution (limited by observations). Applying a number of deterministic and probabilistic verification metrics we found two regions in eastern Africa where some predictive skill is evident in EC-EARTH: northern Ethiopia – North-East Sudan and southern Sudan – northern Uganda. In general, both dynamical and statistical downscaling are able to capture and reproduce the predictive signal evident in the global EC-EARTH hindcast with different level of accuracy. However, on average, the downscaled hindcasts show no added value as compared to the driving model if we define the added value as a higher skill in predicting future seasonal anomalies. There is some tendency of improved reliability through the downscaling but predictive skill is mainly sensitive to forecast resolution and increase in reliability does not correspond to an actual gain in information. Instead the probabilistic forecasts reflect the probability of occurrence more accurately. Therefore, an improvement in reliability can benefit end users.

The LEAP platform driven by the global and downscaled hindcasts also shows that predicting humanitarian needs at the national and sub-national levels is not improved by using the downscaled seasonal forecasts. There is, however, indication that statistical downscaling may slightly improve forecasts of rainfall intensity, with forecasts of precipitation frequency (number of wet days) unaffected by downscaling.

The experimental setup was not perfect in all aspects and outcomes do not meet the initial expectation on possible improvement of a global seasonal hindcast by downscaling in eastern Africa. Nevertheless, sharing our experience from the EUPORIAS project can help climate services working with applications of seasonal forecasting. We should also note that our findings are only for the June–September season in Ethiopia and for a limited number of parameters and tools (models and statistical methods) and therefore cannot be generalised for other regions, seasons and seasonal forecasting tools.

## 1. Introduction

In the last decades a significant progress has been achieved in the prediction of seasonal mean states of weather and, therefore, seasonal forecasting has become an operational activity in a number of national weather services worldwide (Graham et al., 2011). Global seasonal prediction systems are being used increasingly, operating at a 50–200 km range of resolution, while many users require seasonal forecast at impact-relevant regional to local scales. A common approach for providing high-resolution climate information in the future climate projection framework is to supplement global models by empirical-statistical or dynamical downscaling techniques (ESD or DD respectively). To derive regional climate information ESD applies a statistical relationship between information from global models (predictors) and local-scale processes (predictants) (e.g. Maraun et al., 2010) while DD uses regional climate models (RCMs) driven by global models (e.g. Rummukainen, 2010). ESD is a relatively fast and computationally efficient approach but strongly depends on the availability of observations and is limited to a few variables. In practice, this is not a real limitation in seasonal forecasting as most of the needs are temperature and precipitation. Dynamical downscaling using RCMs is computationally expensive delaying the provision of the forecasts and requires much more resources than ESD (e.g. saving a wealth of driving boundary conditions from GCMs). However, in contrast to the ESD approach, RCMs can provide a larger number of variables in a physically consistent way, including regional and local feedbacks which can be important in seasonal forecasting. Due to its simplicity, ESD is applied in seasonal forecasting more often than RCMs but still, running ESD and/or RCMs for operational seasonal forecast production is not a common practice.

At the same time there are many experimental studies (not operational activities) on applying RCMs for downscaling of seasonal forecasts (e.g. Díez et al., 2011; Castro et al., 2012; Diro et al., 2012; Cheneka et al., 2016). The main question in such studies is whether downscaling can provide the added value to global forecasts or not. The definition of the added value is a complex topic and, even after applying RCMs for downscaling climate projections over the last two decades, the added value issue is still debated in the climate downscaling community (Di Luca et al., 2015). There is no unique way to define the added value, which depends on many factors, such as different spatial and time scales, variables and processes and usually includes higher-order statistics, namely: local details in a region with complex topography and land-sea contrast, extreme events, sub-daily variability etc. (Di Luca et al., 2015; Rockel, 2015; Rummukainen, 2016). A downscaling methodology showing the added value in one region and/or season does not necessarily provide similar added value in other regions and seasons. In this study, the added value of downscaling in the seasonal prediction framework is defined as more skilful seasonal forecast compared to its driving global prediction system. The reduction of systematic biases may be also defined as the added value of downscaling and can be important for impact modelling but are not considered in this work (see e.g. Manzananas et al., 2017). Such reduction does not lead to increased knowledge of future seasonal anomalies, and systematic biases can always be dealt with by means of bias correction techniques.

The FP7 EUPORIAS project (Hewitt et al., 2013) aimed to improve our ability to maximise the societal benefit of seasonal to decadal predictions. One of EUPORIAS activities is the provision of downscaled and/or bias-corrected seasonal forecasts for use in EUPORIAS impact and climate service applications. The first focus area in the downscaling EUPORIAS activities is Europe, where high-quality observations exist and ESD methods are applied (Manzananas et al., 2017). A second focus of the downscaling activities is in eastern Africa, where temperature and precipitation exhibit better predictability at seasonal timescale than in the extra-tropics (e.g. Philippon et al., 2002; Diro et al., 2011; Omondi et al., 2013), potentially allowing to apply seasonal forecast data in

sectors such as: food security, drought early-warning systems and human health. In addition, eastern Africa has a complex orography and thus could provide a test bench for downscaling methodologies. This work aims at summarizing the EUPORIAS eastern Africa activities and outcomes, which assessed the utility of downscaling techniques to provide seasonal forecast data for impact models over the region. In particular, the World Food Programme (WFP) assessed the utility of downscaled data in their Livelihoods, Early Assessment and Protection (LEAP) system in Ethiopia (Hoefsloot and Calmanti, 2012). In pursuing this aim, we built the largest ensemble of downscaling techniques ever applied to a global seasonal forecast over a region. For the sake of the reproducibility of the results shown next, all raw and downscaled (dynamical and statistical) seasonal forecasts will be openly published through the European Climate Observations, Modelling and Services initiative (ECOMS) User Data Gateway (<http://meteo.unican.es/udg-tds>, see Cofiño et al., this issue, for details of this data service).

## 2. Data and methods

### 2.1. Observations

The main focus of the study is on precipitation and there are a large number of gridded precipitation observational datasets covering Africa at various temporal (from hourly to monthly) and spatial resolutions (from 0.0375° to 2°). However, even if the observational data sets agree quite well with respect to the large-scale precipitation pattern, significant deviations across them can occur locally (Nikulin et al., 2012). To estimate observational uncertainties, we include in our analysis a number of gridded precipitation products (Table 1). Three of them (GPCC, CRU and UDEL) are gauge-based-only datasets while the rest (TAMSAT, ARC, FEWS and CHIRPS) are satellite-gauge combinations. The WFDEI dataset is a quasi-observational product and presents the bias-corrected 3-hourly ERA-Interim reanalysis where the CRU or GPCC observations are used as reference for adjustment. As such, the WFDEI dataset provides a number of variables at daily resolution globally and is popular as reference for bias correction, statistical downscaling and impact modelling.

### 2.2. Global seasonal hindcasts

At the beginning of the EUPORIAS project in 2011, after consultations with WFP, it was decided to focus on the Kiremt rainy season (June–September – JJAS) in Ethiopia using seasonal hindcast initialised in May, which can be used as input to the LEAP system. This was a trade-off between user needs, more keen on summer rainy season forecasts in Ethiopia when impacts of water deficits on agriculture are larger, and forecast skill, which is associated to ENSO variability and

peaks in November–January south of Ethiopia. We finally opted for addressing the end-user needs, focusing on JJAS.

The first step was to provide boundary conditions from a global seasonal forecast to regional climate modelling groups for subsequent downscaling. A straightforward approach was to downscale the operational ECMWF System-4 (S4) seasonal hindcast (Molteni et al., 2011) which became available in 2011. An operational model would allow for a potential real-time implementation of an operational downscaling system. However, the S4 model levels necessary for dynamical downscaling were not archived for all members of the seasonal hindcast ensemble and only every second model level is saved.

To provide consistent boundary conditions for dynamical downscaling, the first 15 members of the S4 hindcast, initialised on May 1st, were rerun using a coupled global climate model – EC-EARTH (v. 3.1) in the atmospheric-only mode. This method is called two-tiered seasonal forecasting meaning that an atmospheric-only model was forced with forecasted SST from another coupled seasonal prediction system. EC-EARTH is a consortium model (<http://www.ec-earth.org>), which contributed to the CMIP5 activities and is based on the ECMWF Integrated Forecast System (Hazeleger et al., 2010). One additional advantage of using EC-EARTH is that this model can be configured to use exactly the same horizontal resolution (T255) and the same 91 vertical levels as in S4. Since the drift in seasonal forecasts in the tropics is mainly related to sea surface temperature (SST), a bias correction replacing the S4 monthly mean SST climatology with the ERA-Interim reanalysis SST climatology (but preserving anomalies) has been applied. The members of the EC-EARTH hindcast ensemble only differ from each other in the SST. Atmospheric initial conditions (ICs) on May 1st were generated at ECMWF using a methodology similar to the one applied in S4. The same atmospheric ICs are used for all 15 members since the influence of the atmospheric initial conditions is small beyond two weeks and does not impact significantly the skill of seasonal forecasts. Temperature and soil moisture ICs are taken directly from S4 and both are the same for all 15 members as in S4. Following this approach, a 5-month (May–September) global seasonal ensemble hindcast has been generated by EC-EARTH taking the above initial conditions and the bias-corrected SST from the S4 hindcast. The ensemble includes 15 members, initialised on May 1st and covers the 22-year period 1991–2012.

### 2.3. Regional climate models

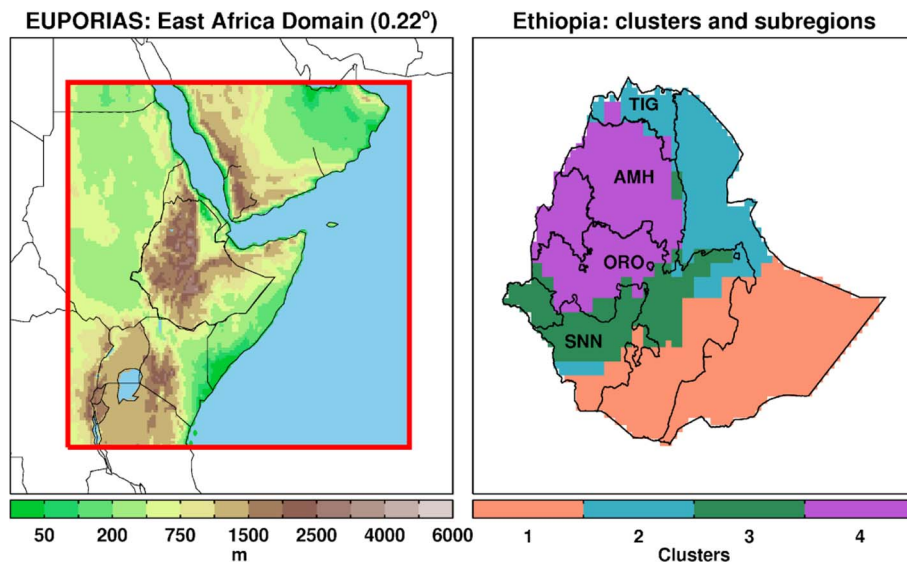
Five research groups have contributed by downscaling the EC-EARTH hindcast using four different RCMs listed in Table 2. Two groups used the Weather Research and Forecasting Model (WRF) but applying different versions and configurations originated in the Euro-CORDEX activities. The WRF Euro-CORDEX community shows that an

**Table 1**  
List of gridded precipitation datasets and their details.

Dataset	Dataset acronym	Version	Resolution	Source	References
Climate Research Unit Time-Series	CRU	3.23	0.5°	Gauges	Harris et al. (2014)
Global Precipitation Climatology Centre	GPCC	7	0.5°	Gauges	Schneider et al. (2014); [dataset] Schneider et al. (2015)
University of Delaware	UDEL	4.01	0.5°	Gauges	Legates and Willmott (1990)
Tropical Applications of Meteorology using SATellite data and ground-based observations	TAMSAT	2.0	0.0375°	Satellites and gauges	Maidment et al. (2014); Tarnavsky et al. (2014)
African Rainfall Climatology	ARC	2.0	0.1°	Satellites and gauges	Novella and Thiaw (2013)
African Rainfall Estimation Algorithm (from the Famine Early Warning System)	FEWS	2.0	0.1°	Satellites and gauges	Novella and Thiaw (2013)
Climate Hazards Group InfraRed Precipitation with Stations	CHIRPS	2.0	0.05°	Satellites and gauges	Funk et al. (2015)
WATCH-Forcing-Data-ERA-Interim	WFDEI	N/A	0.5°	Bias corrected ERA-Interim reanalysis (CRU and GPCC as reference)	Weedon et al. (2014)
ECMWF Interim Reanalysis	ERAINT	N/A	0.75°	Reanalysis	Dee et al. (2011)

**Table 2**  
List of RCMs and their details.

Institution (acronym)	RCM (short name)	Stream	Full domain	References
Deutscher Wetterdienst (DWD)	CCLM4-8-21 (CCLM4)	Full	24°E – 65°E 9°S – 27°N	Rockel et al. (2008)
Swedish Meteorological and Hydrological Institute (SMHI)	RCA4 (RCA4)	Full	– 24°E – 65°E – 14°S – 38°N	Strandberg et al. (2015)
Italian National Agency for New Technologies, Energy (ENEA)	RegCM-4-3 (RegCM4)	Subset	10°S – 23°N 19°E – 68°E	Giorgi et al. (2012)
Universidad de Cantabria (UCAN)	WRF341I (WRF341)	Subset	24°E – 64°E 9°S – 27°N	Skamarock et al. (2008); Katragkou et al. (2015)
Universidade de Lisboa (UL-IDL)	WRF381D (WRF381)	Subset	24°E – 64°E 9°S – 27°N	Skamarock et al. (2008); Katragkou et al. (2015)



**Fig. 1.** The EUPORIAS East Africa domain at 25-km resolution, with topography (left) and clustering according to the average seasonal cycle of the monthly cumulated rainfall based on WFDEI-GPCC (right). Four administrative regions used in the LEAP platform are also shown: Tigray (TIG), Amhara (AMH), Oromia (ORO) and SNNPR (SNN).

ensemble of WRF simulations with different combinations of parameterizations generates a spread across the ensemble member similar to a multi-RCM ensemble (Katragkou et al., 2015; García-Díez et al., 2015). A number of domain configurations for eastern Africa, different in size and resolution, have been tested. Considering the computational costs, a common domain at 0.22° resolution was finally set up (Fig. 1a) and used by all groups. Two streams of downscaling, depending on resources available, have been defined. These are as follows:

- (i) Full hindcast – all 15 members and all years.
- (ii) A subset of the full hindcast – 15 members for four preselected years (two wet years – 2006/2007 and two dry years – 2002/2009 in Ethiopia) and the first three members for all years in order to establish the downscaled hindcast climatology.

The full hindcast was downscaled by DWD and SMHI and the subset by ENEA, UCAN and UL-IDL. No nudging toward EC-EARTH was applied in any of the RCMs within the model domain.

#### 2.4. Empirical-statistical downscaling

In addition to the dynamical downscaling, two ESD methods were applied to downscale the full stream of the EC-EARTH hindcast. The first ESD method (AN1) is based on the popular Analog technique (Lorenz, 1969), which estimates the local downscaled values corresponding to a particular atmospheric configuration. Analog-based methods have been applied in several previous studies to downscale precipitation in the context of seasonal forecasting (e.g. Frías et al., 2010; Shao and Li, 2013). In spite of its simplicity, the analog technique performs as well as other more sophisticated ones (Zorita and von

Storch, 1999) and it is one of the most widely used. The second ESD method is based on Generalised Linear Models (GLMs) (Nelder and Wedderburn, 1972) which are an extension of the classical linear regression which allows modelling the expected value for non-normally distributed variables. Despite being widely used for statistical downscaling of climate change scenarios (e.g. Manzananas et al., 2015), GLMs have been rarely applied to seasonal forecasts. Both ESD methods were first calibrated in Perfect Prognosis conditions (e.g. San-Martín et al., 2017). To this end, daily precipitation from WFDEI (GPCC-calibrated version) was used as predictand and a combination of circulation (zonal wind at 850 hPa) and thermodynamic (specific humidity and temperature at 500 hPa) daily mean variables from the ERA-Interim re-analysis were used as predictors. These variables, taken here over the domain (26°E–56°E, 3°S–21°N), have been typically used for statistical downscaling of precipitation and provide a good representation of the synoptic phenomena affecting the climate of the study region.

#### 2.5. Subregions

Four observational datasets (ARC, TAMSAT, WFDEI and REF), used as input to the LEAP platform, are taken for the definition of geographical sub-regions in Ethiopia. Grid points are clustered according to the Euclidean distance between the associated average seasonal cycle of the monthly cumulated rainfall. The Ward hierarchical clustering with a cutoff at 4 branches of the dendrogram was applied. Depending on the underlying dataset, clusters show different patterns, although large-scale features are about the same (not shown). Taking into account that the WFDEI is scaled by the GPCC and this is used as one of the main reference observational datasets it was decided to use the four clusters (sub-regions) based on the WFDEI-GPCC rainfall (Fig. 1). Seasonal cycle



of rainfall is pretty close for clusters 2, 3 and 4 and, thus, an additional cluster (5) is defined as a combination of clusters 2, 3 and 4. Fig. 1 also shows 4 administrative regions in Ethiopia (Tigray, Amhara, Oromia and SNNPR, the Southern Nations, Nationalities, and Peoples' Region) used to assess performance of the global and downscaled hindcasts in the LEAP platform at sub-national level.

## 2.6. Rainfall indices

In addition to JJAS mean rainfall, we also evaluate the seasonal forecasts of climate information indices derived from daily rainfall time series. Here we show two indices, the Simple Daily Intensity Index (SDII) and the Wet Day Frequency (WDF) that are widely used in the community. Following the WMO Expert Team for Climate Change Detection and Indices (Karl et al., 1999; Peterson et al., 2001) definition, the SDII is defined as the mean rainfall on days with more than 1 mm of rain. Correspondingly, the WDF is defined as the fraction of days with more than 1 mm of rain. Mean rainfall is proportional to the product of SDII and WDF (see e.g. Keller et al., 2015). Indices have been computed on daily rainfall time series of GCM and downscaled forecasts. In contrast to the verification of JJAS mean rainfall, WFDEI-GPCC has been used for verification due to the requirement of daily rainfall time series to compute indices.

## 2.7. Verification metrics

A number of deterministic and probabilistic verification metrics are applied and a short summary of the used metrics is as follows.

### 2.7.1. Interannual correlation

The correlation coefficient is a simple traditional deterministic approach for forecast verification. As the name suggests, interannual correlation here is simply correlation between seasonal mean forecasts and observations. The Pearson correlation method is used for the ensemble mean forecast. It varies from  $-1$  (perfectly bad forecast) to  $+1$  (perfect forecast) and it is close to zero when there is no skill in the forecast.

### 2.7.2. Brier skill score (BSS)

It is based on the Brier Score (BS), which measures the mean squared error of probability forecasts for a binary event. Based on this score and considering the climatology as benchmark forecast, BSS is a relative measure of probabilistic skill (Wilks, 2011). The BSS value ranges from  $-\infty$  to  $+1$ . For a perfect forecast BSS is equal to  $+1$ , while negative BSS indicates poorer skill than the climatology forecast (i.e. issuing the observed event frequency as forecast probability). As the classic BSS is sensitive to ensemble size and negatively oriented for a small ensemble size (Müller et al., 2005; Weigel et al., 2007) a strictly proper fair BSS (Ferro, 2014) was used to overcome this discrepancy.

### 2.7.3. ROC skill score (ROCSS)

This skill score is based on the area beneath the Relative Operating Characteristic (ROC) curve (Jolliffe and Stephenson (2003)). The ROC curve measures forecast discrimination for a binary event. The ROC skill score compares this area with that of a climatological forecast. Numerically, ROCSS ranges from  $1$  (perfect forecast) to  $-1$  (perfectly bad forecast). Zero indicates no skill compared to a forecast based on the climatological frequency of the event. Two binary rainfall events (above-normal and below-normal) were considered based on upper and lower terciles, respectively. Given the small sample size (22 years), all terciles and anomalies were calculated in a 'one-year out cross-validation' mode, i.e. leaving the target year out from the computation in order to avoid the overfitting (Wilks, 2011; Weigel et al., 2008).

### 2.7.4. Attributes or reliability diagram

It is a graphical summary of important statistical-attributes such as

reliability, resolution, uncertainty. It provides useful information about the performance of a prediction system (e.g. Weisheimer and Palmer, 2014). Reliability or so-called the attributes diagram measures how well the forecast probabilities of an event are in line with the equivalent observed frequencies. For instance, a forecast probability of  $0.8$  is called perfectly reliable if and only if the event is true for  $80\%$  of the cases when the predicted probability was  $0.8$ . In the reliability diagram, the diagonal line reflects perfect reliability. Reliability of a forecast is high if it closely follows the diagonal line, inferring a good correspondence between the forecast probabilities and the observed frequencies for a binary event. Ten equidistant bins to discretize the forecast probabilities were used. As suggested by Doblas-Reyes et al. (2008), the maximum number of bins can go up to the number of ensemble members plus one. In a reliability diagram, the data points (i.e. forecast probabilities vs. observed frequencies) usually do not stay along a line and therefore, following Weisheimer and Palmer (2014), a weighted linear regression as a best guess estimate on all data bins was used.

## 3. Observational uncertainties

We first evaluate the spread among the precipitation datasets over eastern Africa to get an estimate of consistency and uncertainty across the observations. Consistency in reproducing interannual variability is even more critical for verification of seasonal hindcasts than reproducing seasonal mean climatology. If, for example, observational datasets do not agree on wet/dry years over a region, the verification result can be very uncertain and depends on observational data set chosen as Ref. Fig. 2 shows interannual correlation for the June–September (JJAS) mean precipitation between GPCC, taken as reference, and other gridded precipitation products. All time series are detrended by removing linear trends at each grid box.

There are large discrepancies in interannual precipitation variability across the observational datasets. The correlation in eastern Ethiopia and Somalia is close to zero. This common no skill feature can be expected, since these regions are very dry in JJAS with almost no precipitation. However, another pronounced feature is low or zero correlation over western Ethiopia and southern Sudan (a climatologically wet region in JJAS) for TAMSAT and ARC and, to a lesser degree, for CRU, UDEL and CHIRPS. Taking other datasets as reference (not shown) reveals that, in general, there are two dataset clusters with similar interannual variability: the satellite-based (TAMSAT, ARC and CHIRPS) and gauge-based (GPCC, CRU and UDEL) products, although large local uncertainties can be found even within both clusters. On average, the best agreement over large part of Ethiopia for two datasets from the different clusters is found for GPCC and CHIRPS.

An illustrative example of the impact of observational uncertainties on the probabilistic verification of a forecast is shown in Fig. 3, where ROCSS for the lower tercile is estimated for the EC-EARTH hindcast using eight individual precipitation data sets as reference. The highest ROCSS, with similar spatial patterns, are found for GPCC, WFDEI-GPCC and CHIRPS and, to some degree, for TAMSAT. Using ARC as reference leads to ROCSS values close to zero over the entire Ethiopia. CRU and WFDEI-CRU are somewhere in between. The large uncertainties in interannual variability of precipitation among the observational datasets in eastern Africa can put serious limitations on our ability to verify seasonal forecasts at grid box scale over this region. For all of the following verification analysis we select GPCC as the reference dataset, although this is, to some extent, a subjective choice as for selecting any other dataset.

## 4. Predictive skill of global and downscaled hindcasts

### 4.1. Interannual correlation

Fig. 4 shows interannual correlation for GPCC vs. global (S4 and EC-EARTH) and downscaled (RCA4, CCLM4, GLM and AN1) hindcasts with

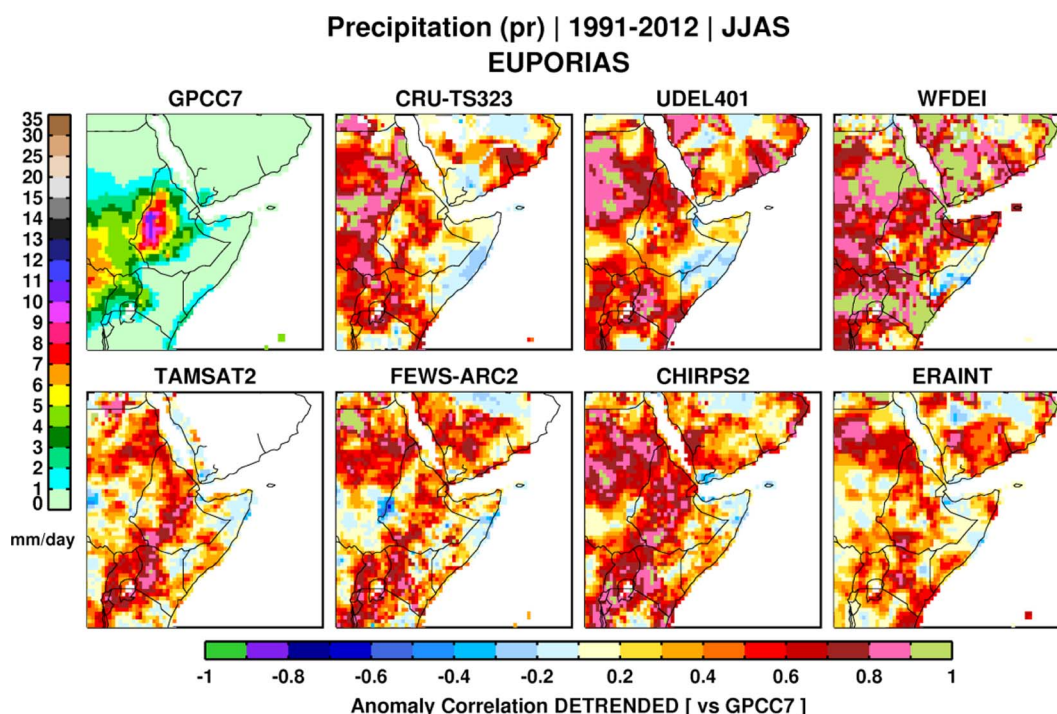


Fig. 2. GPCC mean JJAS rainfall (1991–2012) [upper left] and interannual correlation between other gridded precipitation products and GPCC. All datasets are detrended by removing linear trend.

all 15 members available. In the S4 hindcast the correlation pattern has two distinct regions with higher correlation: northern Ethiopia – northeast Sudan and southern Sudan – northern Uganda. These two spots are pretty well reproduced by the EC-EARTH hindcast, although absolute values of correlation are a bit lower. Both regional models – RCA4 and CCLM4 driven by EC-EARTH also reproduce almost the same correlation pattern. Both statistical downscaling methods – GLM and AN1 lose a part of the signal (weaker correlation) in northern Ethiopia and northeast Sudan, especially the AN1 method. At the same time the spot of high positive correlation in southern Sudan – northern Uganda is a robust feature in all global and downscaled hindcasts.

The same verification for only the first three members but including all RCMs is shown in Fig. 5. As expected, taking only the first three members of the hindcasts makes the correlation patterns noisier due to the smaller ensemble size. The two regions with higher correlation became weaker and they are not evident in all hindcasts. The signal in northern Ethiopia and northeast Sudan is not pronounced in S4 compared to the full ensemble while EC-EARTH still shows two spots of higher correlations similar to the full ensemble. All RCMs can reproduce the correlation pattern of their driving hindcast with different level of accuracy. An interesting feature is an improved forecast skill in the RegCM4 hindcast over most of Ethiopia, in regions where both S4 and EC-EARTH do not show any forecast signal. Similar to the full ensemble verification, both ESD methods lose a part of the signal in the correlation pattern.

Fig. 6 shows interannual correlation for time series averaged over the sub-regions (clusters in Fig. 1). All hindcasts have low correlation in the first (dry) cluster especially both ESD methods but surprisingly the RegCM4 hindcast has really high correlation (about 0.8) outperforming all other hindcasts including its driving EC-EARTH. In general, RegCM4 almost always shows the highest correlation in all five clusters. In clusters 2 and 3, the global and downscaled hindcasts perform similarly, while in clusters 4 and 5, the ESD methods and RegCM4 outperform all other RCMs. Spatial aggregation has a pronounced effect on the verification results. For example, both ESD methods lose a part of the signal in the correlation spatial pattern (see Figs. 4 and 5) but provide a higher skill in clusters 4 and 5, after spatial aggregation.

Similar results are obtained for all 15-member hindcasts, indicating that the spatial aggregation can effectively reduce the noise seen in the correlation patterns.

#### 4.2. Probabilistic verification

The previous results considered the ensemble mean as a deterministic forecast. In this section, the full potential of ensemble forecasts is exploited and probabilistic skill scores are considered.

ROCSS is shown in Fig. 7 illustrating how the individual hindcasts can discriminate seasonal mean rainfall falling in the upper and lower tercile. Since many members are necessary for such probabilistic verification here we present only the 15 member (full stream) hindcasts. Similar to the interannual correlation results the strongest signal (the highest ROCSS values) in S4 are found over northwest Ethiopia with a similar spatial pattern for both upper and lower terciles. The EC-EARTH hindcast has lower ROCSS and a different spatial pattern for the below normal seasons and almost completely losing the signal for the above normal seasons. Both RCMs and ESD methods show some improvement compared to EC-EARTH but spatial patterns vary a lot.

In Fig. 8, a tercile plot (Manzanas et al., 2014) is used to visualize the forecast skill for one particular region (cluster 5). In it, the probabilistic forecast for each year, computed as the number of members falling within each category for that particular year, is represented. White circles correspond to the observed precipitation tercile. Moreover, the ROCSS of the spatial mean is also shown for each tercile. All hindcasts correctly predict a few dry years in 90's, namely: 1991, 1992 and 1997 (strong El Niño). Focusing on the four years, chosen as reference of dry (2002 and 2009) or wet years (2006 and 2007) one can see that all hindcasts are able to predict dry 2009 and wet 2007 but all miss wet 2006. Both global system S4 and EC-EARTH cannot predict the dry 2002 summer. However, all downscaled hindcasts correctly forecast the event.

Another verification metric – the fair BSS calculated for the five clusters and for the 15-member hindcasts is shown in Fig. 9. All hindcasts have negative skill in the first and third clusters indicating a poor forecast skill over these two regions. In agreement with the above

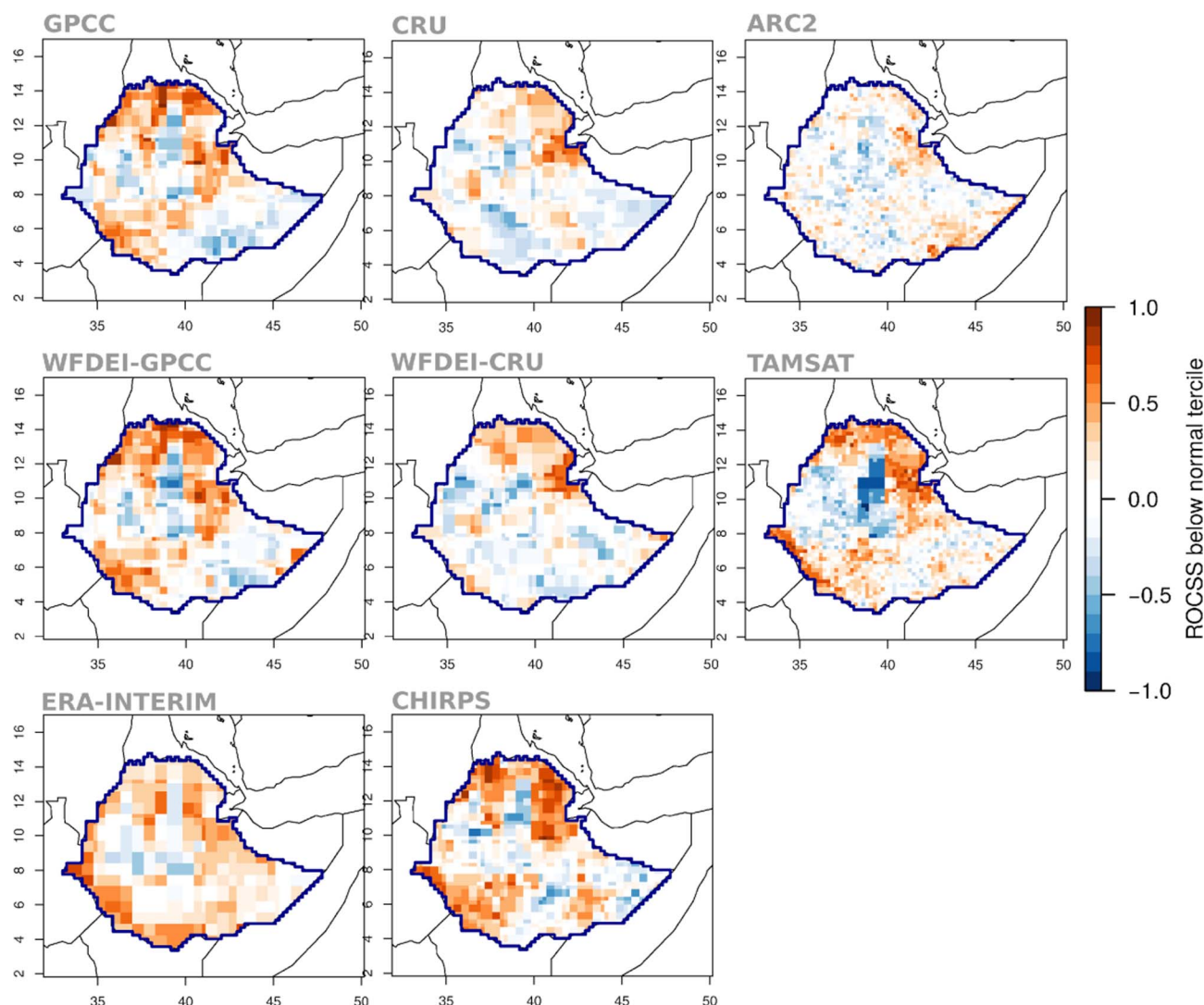


Fig. 3. ROCSS maps for the EC-EARTH hindcasted rainfall in JJAS (below normal) computed considering eight different datasets as verifying observations over Ethiopia.

results (interannual correlation and ROCSS) the maximum predictive skill is found in the second cluster that contains mostly the Ethiopian Highland regions. All hindcasts have positive skill here for both above and below normal terciles, with the exception of RCA4 for the below normal one (a small negative BSS). Performance of the hindcasts is mixed in the fourth and fifth clusters and the S4 hindcast (always positive BSS) outperforms all other hindcasts. An interesting detail is that even if two RCMs have negative BSS their multi-model ensemble (MME) always shows positive BSS (clusters 4 and 5). Here the MME was produced by a simple ‘pool’ approach, i.e., pooling the RCMs (RCA4 and CCLM4) ensembles together, considering each member has equal weight (Weigel et al., 2008). The probabilistic score for the MME was computed afterwards.

Fig. 10 presents the attribute or so-called reliability diagrams for the upper and lower terciles for the five cluster regions. The results that fall in the BSS area are quite consistent with the fair BSS results of Fig. 9. For the first and third clusters and for both lower and upper tercile cases, all models show overconfidence, i.e., they are below the diagonal line. All hindcasts are above the “no skill” line in the second cluster, indicating the predictive skill. Only S4 and RCM ensemble hindcasts have the skill for both terciles in the fourth and fifth clusters. The multi-model combination, i.e., MME, often outplays the driving GCM model and the individual RCMs. The better performance of MME can be explained by the fact that the RCMs were mostly overconfident, and their

fusion led to an increase in reliability (Weigel et al., 2008), bringing the MME closer to the diagonal line. Both ESD methods are above the diagonal line for the below normal events in the fourth cluster showing some improvement of their predictor EC-EARTH. Finally, we should note that there is a common tendency of improved reliability through the downscaling. The EC-Earth line (green) is worse in most cases than most of the downscaling methods.

#### 4.3. Verification for rainfall indices

In Fig. 11, we show the distribution of ROCSS for JJAS rainfall, SDII and WDF over Ethiopia. The rainfall results (leftmost group of box-and-whiskers in Fig. 11) correspond to the spatial maps of ROCSS for rainfall shown in Fig. 7, with the only difference that WFDEI-GPCC has been used for verification in Fig. 11 for consistency with the verification of index forecasts.

Both dynamical and statistical downscaling slightly improve the discrimination, as measured by the ROCSS, of SDII forecasts (middle group of box-and-whiskers in Fig. 11) compared with the host model EC-EARTH. The largest improvements are found for the statistical downscaling and the GLM 15PC method in particular. In contrast, forecast quality of WDF forecasts is barely affected by downscaling (both statistical and dynamical). Also, as shown for mean rainfall, S4 outperforms all other models for WDF forecasts and forecasts of the



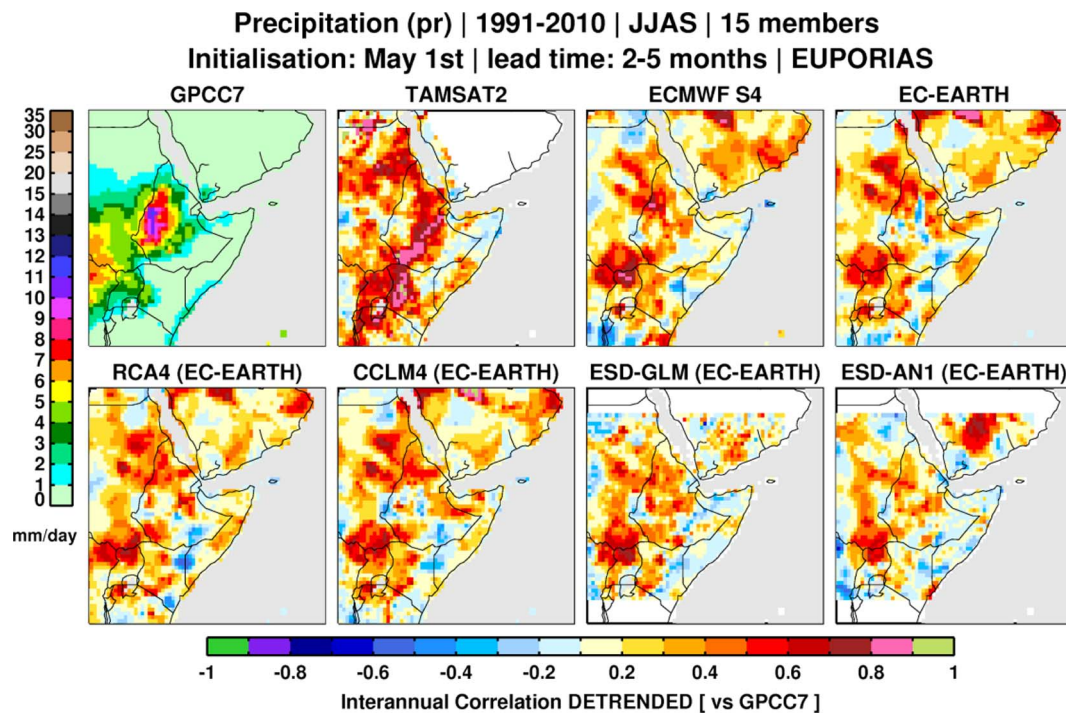


Fig. 4. GPCC mean JJAS rainfall (1991–2010) [upper left] and interannual correlation between other datasets and GPCC. Hindcast members 1–15.

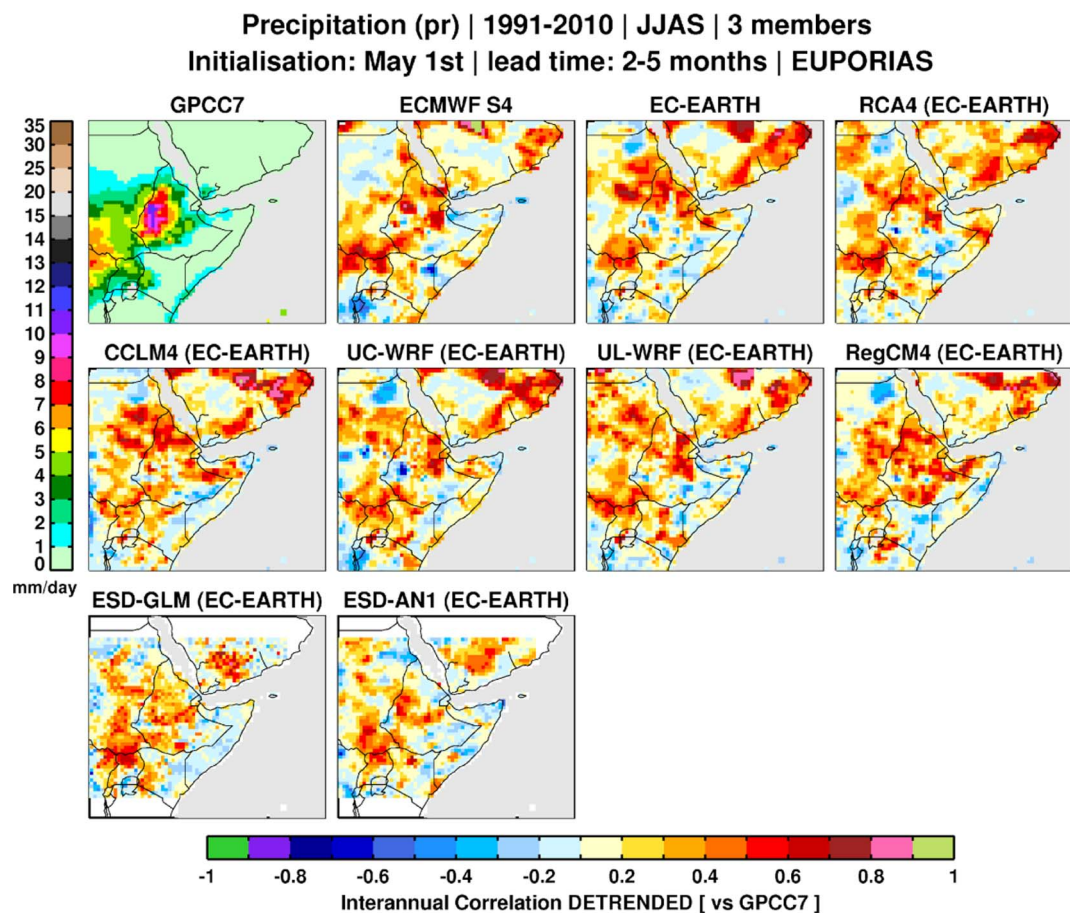
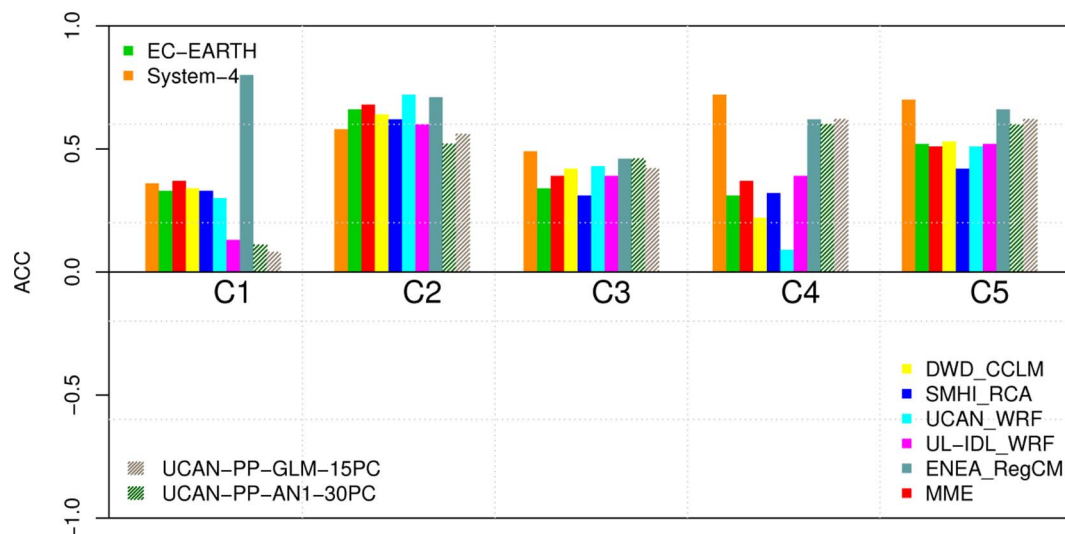


Fig. 5. GPCC mean JJAS rainfall (1991–2010) [upper left] and interannual correlation between other datasets and GPCC. Hindcast members 1–3.

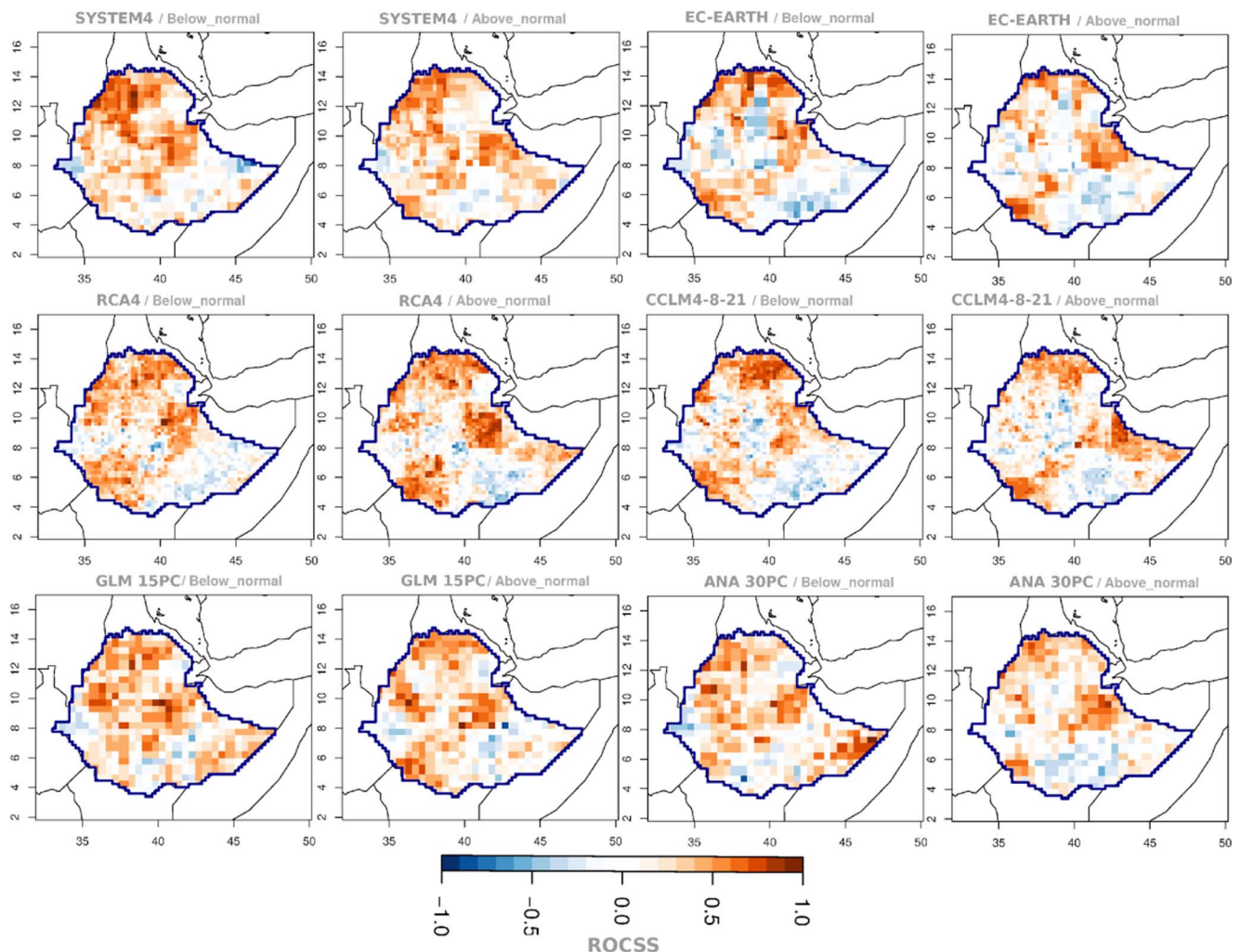




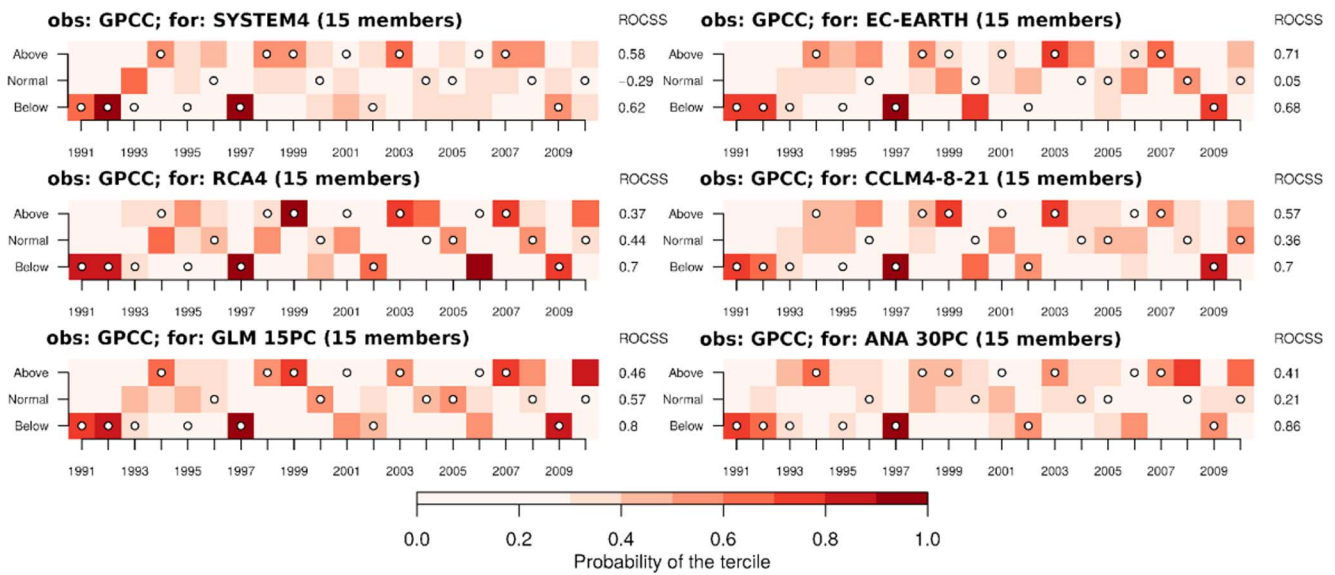
**Fig. 6.** Interannual correlation for the global and downscaled hindcasts with respect to GPCP rainfall in JJAS (1991–2010). Time series were averaged in each cluster as shown in Fig. 1 (right) plus cluster 5, aggregating cluster 2, 3 and 4.

lower tercile of SDII. The above differences in ROCSS are also reflected in the Brier skill score and the anomaly correlation (not shown). These results indicate that while statistical downscaling does not necessarily improve forecast quality of mean rainfall forecasts, there is the

potential for better forecasts of rainfall intensity (SDII) using statistical downscaling and the GLM 15PC method in particular. Applications that are sensitive to rainfall intensity may therefore benefit from statistical downscaling. To what extent statistical downscaling may be used to



**Fig. 7.** ROCSS maps for the global and downscaled hindcast rainfall in JJAS computed considering GPCP as verifying observations (1991–2010).



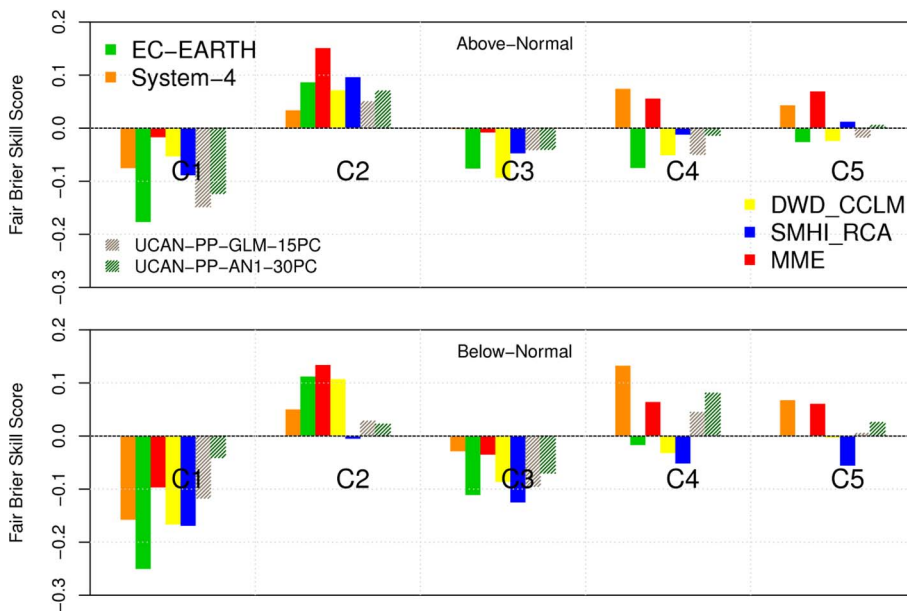
**Fig. 8.** Tercile plots for the cluster 5 (combination of 2, 3 and 4 in Fig. 1) showing the probability of each JJAS rainfall tercile forecast by each model (red shades) along with the observed tercile (white circles). The ROCSS of the spatial mean is also shown for each tercile and dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

improve forecasts of heavy and extreme rainfall events, however, remains to be analyzed.

Discrimination of SDII forecasts is similar or slightly reduced compared to the ROCSS of the mean rainfall forecasts. In contrast, WDF forecasts slightly outperform the mean rainfall forecasts for all dynamical models. Qualitatively similar results are found for other metrics of forecast quality including the Brier skill score and anomaly correlation (not shown). We conclude that the frequency of rainfall events, as measured by the WDF, is slightly better predictable than the intensity, as measured by the SDII. Comparable results have been found in climate change studies, where there is indication of better agreement on projections of rainfall frequency compared to rainfall intensity (Mtongori et al., 2016). The results clearly show that for each method, the seasonal mean precipitation skill is consistently higher than the skill for SDII and WDF confirming the more theoretical considerations in the study of Bhend et al. (2016).

## 5. LEAP platform

Finally, the rainfall data derived from the different hindcasts (S4, RCA4 and CCLM4) are adopted as input for the LEAP platform to hindcast humanitarian needs (Fig. 12). Consistent with the predictive skill of the corresponding rainfall forecast, the performance of LEAP in predicting humanitarian needs at the national level is not improved by using downscaled seasonal forecasts (Table 3). In particular, the correlation between historical needs, derived from observed rainfall data, and the ensemble average of hindcast needs is 0.80 for the case of S4; 0.79 for RCA4 and 0.78 for DWD-CCLM4. However, by reducing the range of ensemble spread, forecasts based on downscaled data have a higher probability of a complete mismatch with observed data (see for example the case of 2002 and 2009 in Fig. 12). There are also instances, such as in 2006, when observations are missed even if the spread is largely increased by the downscaled hindcasts. Also, in 1997, the downscaling forecast is closer to the observations, in a year with a clear overestimation of humanitarian needs by S4. However, unless these



**Fig. 9.** Fair BSS for the lower and upper JJAS rainfall terciles for the four clusters (Fig. 1, right) plus cluster 5, aggregating cluster 2, 3 and 4. The multi-model ensemble (MME) combines the CCLM and RCA ensembles.

## Reliability diagram

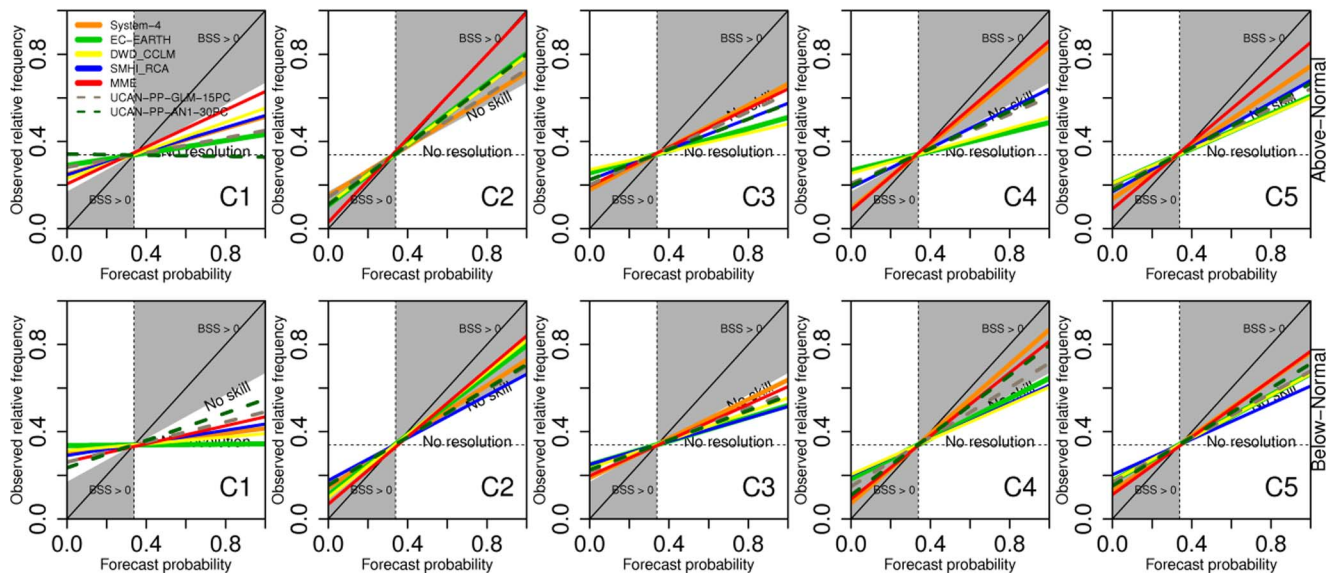


Fig. 10. Reliability diagram for the upper and lower JJAS rainfall terciles for the five clusters.

specific details can be explained (e.g. linked to ENSO), they cannot be used to improve the forecast. The only systematic behaviour that emerges is the ability of S4 in bracketing the observed value more than two thirds of the time.

At the sub-national level, for the four regions shown in Fig. 1, the use of higher resolution dynamically downscaled rainfall data also does not produce significant improvements in the forecast as well, see Table 3. Small differences exist between the correlation obtained with S4 and the corresponding correlation obtained with the RCM down-scaled data. However, such differences cannot be considered as statistically significant owing to limited length of the historical series.

While the potential predictability of rainfall in East Africa has been known for a relatively long-time, the orography of Ethiopia is complex and it was considered important to assess the possibility of improving the accuracy of large-scale rainfall patterns over this particular area. The expected impact of the improvement of rainfall pattern over the complex orography of Ethiopia is the possibility to produce more reliable information at the sub-national level. However, the use of the downscaled hindcast does not meet the initial expectation on possible enhancements of the early warning system.

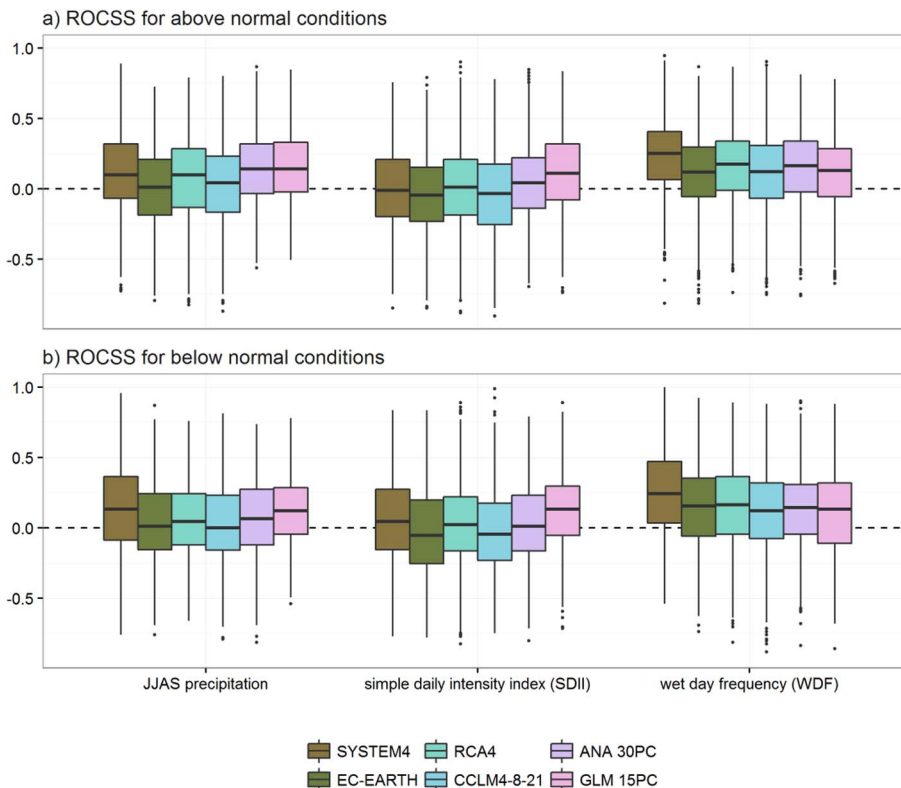


Fig. 11. Distribution of ROCSS for the global and down-scaled rainfall, SDII, and WDF forecasts in JJAS computed considering WFDEI-GPCC as verifying observations (1991–2010).



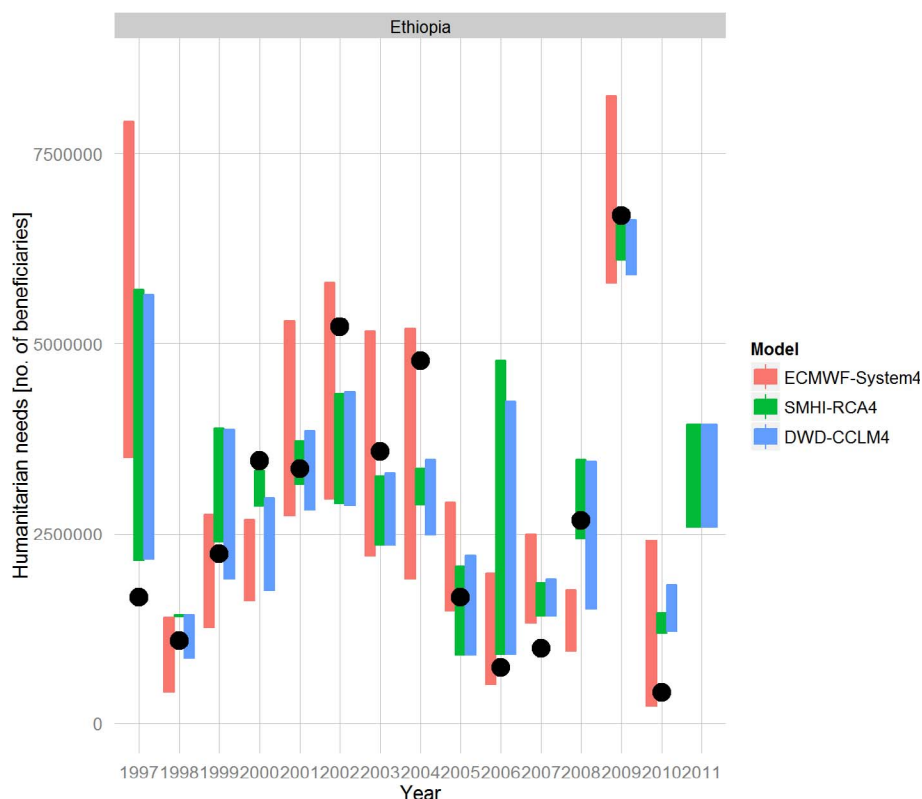


Fig. 12. Needs for humanitarian interventions in Ethiopia based on historical rainfall estimates (WFDEI-GPCC, black circles) and seasonal forecast products (bars).

Table 3

Linear correlation between the national and sub-national historical needs and the corresponding rainfall data derived from different hindcasts for the 1997–2010 period.

	ECMWF System4	SMHI-RCA4	DWD-CCLM4
Ethiopia	0.80	0.79	0.78
Amhara	0.50	0.43	0.45
Oromia	0.76	0.78	0.73
SNNPR	0.89	0.91	0.87
Tigray	0.74	0.63	0.71

## 6. Summary and conclusions

A 5-month seasonal hindcast generated by the ECMWF System-4 (S4) was rerun by a coupled global climate model, EC-EARTH in the atmospheric-only mode in order to provide boundary conditions for dynamical downscaling. The hindcast ensemble includes 15 members, initialised on May 1st and covers the 22-year period 1991–2012. The EC-EARTH seasonal hindcast has been downscaled over eastern Africa by five RCMs from about 80 km resolution to 25 km. In addition to the dynamical downscaling, two statistical downscaling methods were applied to the EC-EARTH hindcast. The EUPORIAS downscaled seasonal hindcast ensemble is the largest ensemble ever downscaled over eastern Africa in a consistent and coordinated way. We also should note that multi-model approach (pooling several GCMs together) has shown better and more reliable seasonal forecasts (e.g. Stockdale et al., 2009) but for practical reason we used only one GCM downscaled by different RCMs.

Both global and downscaled seasonal hindcast ensembles were analysed and verified using a number of different observational datasets and a number of deterministic and probabilistic metrics. As a last step one global and two RCM hindcasts were adopted as input for an early warning system (the LEAP platform) to hindcast rainfall impacts on humanitarian needs.

Here we summarise our main findings:

- (i) *Observational uncertainties.* There are large discrepancies in inter-annual precipitation variability across the different observational datasets at regional scale (e.g. Somalia and eastern Ethiopia). Applying probabilistic verification metrics, such as ROCSS, using different observational datasets as reference shows that the ROCSS verification results are very sensitive to the dataset chosen as reference. The large uncertainties in interannual variability of precipitation among the observational datasets in eastern Africa can be a serious limitation for verifying seasonal forecasts at grid box scale.
- (ii) *Global forecast systems.* Both global hindcasts, S4 and EC-EARTH, show almost the same interannual correlation pattern in East Africa when deterministic verification metric as anomaly correlation is used. There are two distinct regions with higher correlation: northern Ethiopia – North-East Sudan and southern Sudan – northern Uganda while there is no skill elsewhere. Probabilistic metrics (ROCSS and BSS) applied only over Ethiopia also show that the global seasonal forecasts have some predictability skill in northern Ethiopia. However, the signal in northern Ethiopia is sensitive to observational datasets chosen and most pronounced if GPCC or WFDEI-GPCC are used as reference. Compared to EC-EARTH, System4 seems to better capture the signal.
- (iii) *Dynamical downscaling.* When the full hindcast (15 members) is downscaled, RCMs (CCLM4 and RCA4) are able to capture the EC-EARTH signal reasonably well. The anomaly correlation pattern shows the same two regions with high correlation as in the driving EC-EARTH hindcast. Probabilistic metrics reveal similar behaviour of the global hindcasts and RCMs in Ethiopia. All models can correctly predict dry (1991, 1992, 1997, 2002 and 2009) and wet (2007) years over the Ethiopian Highlands. However, all the different systems seem to have missed the wet 2006 summer. Including the RCMs downscaling the subset of the full hindcast, in general, supports the above findings, although some results can be noisier due to the only 3 member ensemble, anomaly correlation patterns for example.

- (iv) *Statistical Downscaling*. Statistical downscaling can capture the predictive signal evident in the global hindcasts and in general shows a similar performance, although losing a bit the predictive skill at grid-box scale. The ability of ESD methods to predict seasonal rainfall anomalies also strongly depends on the region and verification metrics applied.
- (v) *Reliability*. Even if both dynamical and statistical downscaling does not show the added value in terms of a higher predictive skill there is a tendency to improved reliability through the downscaling. Increase in reliability does not correspond to an actual gain in information (e.g. the order of the forecasts is not changed), but the probabilistic forecasts reflect the probability of occurrence more accurately. Therefore, an improvement in reliability is a benefit for end users.
- (vi) *Rainfall indices*. Forecast quality of precipitation intensity is comparable to the forecast quality of mean rainfall, forecasts of precipitation frequency are slightly more skilful. As with seasonal mean precipitation, downscaling has little effect on forecast skill of rainfall frequency. Forecasts of rainfall intensity, on the other hand, seem to benefit from statistical downscaling, which offers an interesting perspective for applications sensitive to precipitation intensity.
- (vii) *The early warning system (LEAP platform)*. Consistent with the predictive skill of the corresponding rainfall forecasts, the performance of the LEAP platform in predicting humanitarian needs at the national level is not improved by using downscaled seasonal forecasts. At the sub-national level, the use of higher resolution dynamically downscaled rainfall data also does not produce significant improvements in the forecast as well. While the limited potential use of dynamically downscaled data does not meet the initial expectation on possible enhancements of the early warning system, the significant skill achieved with the coarser resolution global forecast is already conducting to important potential developments of the existing early warning platform.

We can conclude that the RCMs and ESD methods are able to capture and reproduce the existent signal in the driving EC-EARTH seasonal hindcast over northern Ethiopia in June–September showing about the same performance as their driving GCM. However, on average, RCM hindcasts show no added value compared to the driving GCM, if we define the added value as a higher skill in the RCM hindcast, although a tendency for an improvement in reliability is found. In agreement with the above finding, assessing the utility of the downscaled hindcasts in the Drought Early-Warning System (LEAP) of the World Food Programme shows that prediction of humanitarian needs at the national and sub-national levels in Ethiopia is not benefited by using downscaled seasonal forecasts.

We should also note that these conclusions are only for Ethiopia in the June–September season and for a limited set of global and regional hindcasts. The conclusions cannot be generalised for other regions, seasons and prediction systems. Additionally, large observational uncertainties can potentially prevent us from accurate verification of the high-resolution downscaled hindcast.

## Acknowledgements

This work was done in the EUPORIAS project that received funding from the European Union Seventh Framework Programme (FP7) for Research, under grant agreement 308291. The authors thank the European Centre for Medium-Range Weather Forecasts (ECMWF), the Global Precipitation Climatology Centre (GPCC), the British Atmospheric Data Centre (BADC), the University of East Anglia (UEA), the University of Delaware, the University of Reading, the University of California, the Climate Prediction Center (CPC), the US Agency for International Development's Famine Early Warning Network (FEWS NET) and the WATCH project for providing data. For the WRF

simulations, the UCAN group acknowledges Santander Supercomputacion support group at the University of Cantabria, who provided access to the Altamira Supercomputer at the Institute of Physics of Cantabria (IFCA-CSIC), member of the Spanish Supercomputing Network. DWD wants to thank ECMWF for the support during the CCLM4 simulations which have been carried out at the ECMWF computing system. The SMHI RCA4 simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at National Supercomputer Centre (NSC) and the PDC Center for High Performance Computing (PDC-HPC).

## References

- Bhend, J., Mahlstein, I., Liniger, M.A., 2016. Predictive skill of climate indices compared to mean quantities in seasonal forecasts. *Q. J. R. Meteorol. Soc.* 1477–1870. <http://dx.doi.org/10.1002/qj.2908>.
- Castro, C.L., Chang, H., Dominguez, F., Carrillo, C., Schemm, J.K., Juang, H.M.H., 2012. Can a regional climate model improve the ability to forecast the North American monsoon? *J. Climate* 25, 8212–8237. <http://dx.doi.org/10.1175/JCLI-D-11-00441.1>.
- Cheneka, B.R., Brienens, S., Fröhlich, K., Asharaf, S., Fröh, B., 2016. Searching for an added value of precipitation in downscaled seasonal hindcasts over East Africa: COSMO-CLM forced by MPI-ESM. *Adv. Meteorol.* 2016. <http://dx.doi.org/10.1155/2016/4348285>. Article ID 4348285.
- Cofiño, A.S., Bedia, J., Iturbide, M., Vega, M., Herrera, S., Fernandez, J., Frias, M. D., Manzanaras, R., Gutierrez, J. M., 2017. The ECOMS User Data Gateway: Towards seasonal forecast data provision and research reproducibility in the era of Climate Services. Submitted to Climate Services.
- Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., et al., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.* 137, 553–597. <http://dx.doi.org/10.1002/qj.828>.
- Di Luca, A., de Elia, R., Laprise, R., 2015. Challenges in the quest for added value of regional climate dynamical downscaling. *Curr. Clim. Change Rep.* 1, 10–21. <http://dx.doi.org/10.1007/s40641-015-0003-9>.
- Diez, E., Orfila, B., Frías, M.D., Fernández, J., Cofiño, A.S., Gutiérrez, 2011. Downscaling ECMWF seasonal precipitation forecasts in Europe using the RCA model. *Tellus A* 63 (4), 757–762. <http://dx.doi.org/10.3402/tellusa.v63i4.15857>.
- Diro, G.T., Grimes, D.I.F., Black, E., 2011. Teleconnections between Ethiopian summer rainfall and sea surface temperature: part I - observation and modelling. *Clim. Dyn.* 37, 103–119. <http://dx.doi.org/10.1007/s00382-010-0837-8>.
- Diro, G.T., Tompkins, A.M., Bi, X., 2012. Dynamical downscaling of ECMWF ensemble seasonal forecasts over East Africa with RegCM3. *J. Geophys. Res.* 117, D16103. <http://dx.doi.org/10.1029/2011JD016997>.
- Doblas-Reyes, F.J., Coelho, C.A.D.S., Stephenson, D.B., 2008. How much does simplification of probability forecasts reduce forecast quality? *Meteorol. Appl.* 15, 155–162. <http://dx.doi.org/10.1002/met.50>.
- Ferro, C.A.T., 2014. Fair scores for ensemble forecasts. *Q. J. Roy. Meteor. Soc.* 140, 1917–1923. <http://dx.doi.org/10.1002/qj.2270>.
- Frías, M.D., Herrera, S., Cofiño, A.S., Gutiérrez, J.M., 2010. Assessing the skill of precipitation and temperature seasonal forecasts in Spain: Windows of opportunity related to ENSO events. *J. Clim.* 23 (209), 220. <http://dx.doi.org/10.1175/2009JCLI2824.1>.
- Funk, C.C., Peterson, P., Landsfeld, M., Pedreros, D.H., Verdin, J.P., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations – a new environmental record for monitoring extremes. *Sci. Data* 2, 150066. <http://dx.doi.org/10.1038/sdata.2015.66>.
- García-Díez, M., Fernández, J., Vautard, R., 2015. An RCM multi-physics ensemble over Europe: multi-variable evaluation to avoid error compensation. *Clim. Dyn.* 45 (11–12), 3141–3156. <http://dx.doi.org/10.1007/s00382-015-2529-x>.
- Giorgi, F., Coppola, E., Solmon, F., Mariotti, L., Sylla, M.B., Bi, X., 2012. RegCM4: model description and preliminary tests over multiple CORDEX domains. *Clim. Res.* 52, 7–29. <http://dx.doi.org/10.3354/cr01018>.
- Graham, R.J., Yun, W.T., Kim, J., Kumar, A., Jones, D., Bettio, L., Gagnon, N., Kolli, R.K., Smith, D., 2011. Long-range forecasting and the Global Framework for climate services. *Clim. Res.* 47, 47–55. <http://dx.doi.org/10.3354/cr00963>.
- Harris, I., Jones, P.D., Osborn, T.J., Lister, D.H., 2014. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *Int. J. Climatol.* 34, 623–642. <http://dx.doi.org/10.1002/joc.3711>.
- Hazeleger, W., Severijns, C., Semmler, T., Ștefănescu, S., Yang, S., Wang, X., Wyser, K., et al., 2010. EC-Earth: a seamless earth-system prediction approach in action. *Bull. Amer. Meteor. Soc.* 91, 1357–1363. <http://dx.doi.org/10.1175/2010BAMS2877.1>.
- Hewitt, C., Buontempo, C., Newton, P., 2013. Using climate Predictions to better serve society's needs. *Eos Trans. Am. Geophys. Union* 94, 105–107. <http://dx.doi.org/10.1002/2013EO110002>.
- Hoefsloot, P., Calmanti, S., 2012. LEAP version 2.61 for Ethiopia. <http://hoefsloot.com/new/software=leap-development> (accessed 2016.12.19).
- Jolliffe, I.T., Stephenson, D.B., 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley, NY.
- Karl, T.R., Nicholls, N., Ghazi, A., 1999. CLIVAR/GCOS/WMO workshop on indices and indicators for climate extremes Workshop Summary. *Clim. Change* 42, 3–7. <http://dx.doi.org/10.1023/A:1005491526870>.
- Legates, D.R., Willmott, C.J., 1990. Mean seasonal and spatial variability in global surface air temperature. *Theor. Appl. Climatol.* 41, 11–21.

- Lorenz, E.N., 1969. Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.* 26, 636–646. [http://dx.doi.org/10.1175/1520-0469\(1969\)26<636:APARBN>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2).
- Katragkou, E., García-Díez, M., Vautard, R., Sobolowski, S., Zanis, P., Alexandri, G., et al., 2015. Regional climate hindcast simulations within EURO-CORDEX: evaluation of a WRF multi-physics ensemble. *Geosci. Model Dev.* 8, 603–618. <http://dx.doi.org/10.5194/gmd-8-603-2015>.
- Keller, D.E., Fischer, A.M., Frei, C., Liniger, M.A., Appenzeller, C., Knutti, R., 2015. Implementation and validation of a Wilks-type multi-site daily precipitation generator over a typical Alpine river catchment. *Hydrol. Earth Syst. Sci.* 19, 2163–2177. <http://dx.doi.org/10.5194/hess-19-2163-2015>.
- Maidment, R.I., Grimes, D., Allan, R.P., Tarnavsky, E., Stringer, M., Hewison, T., Roebeling, R., Black, E., 2014. The 30 year TAMSAT African rainfall climatology and time series (TARCAT) data set. *J. Geophys. Res. Atmos.* 119, 10619–10644. <http://dx.doi.org/10.1002/2014JD021927>.
- Manzanas, R., Frías, M.D., Cofiño, A.S., Gutiérrez, J.M., 2014. Validation of 40 year multimodel seasonal precipitation forecasts: The role of ENSO on the global skill. *J. Geophys. Res. Atmos.* 119 (4), 1708–1719. <http://dx.doi.org/10.1002/2013JD020680>.
- Manzanas, R., Brands, S., San-Martín, D., Lucero, A., Limbo, C., Gutierrez, J.M., 2015. Statistical downscaling in the tropics can be sensitive to reanalysis choice: A case study for precipitation in the Philippines. *J. Clim.* 28 (10), 4171–4184. <http://dx.doi.org/10.1175/JCLI-D-14-00331.1>.
- Manzanas, R., Gutiérrez, J.M., Fernández, J., van Meijgaard, E., Magariño, M., Cofiño, A.S., Herrera, S., 2017. Dynamical and statistical downscaling of seasonal temperature forecasts in Europe: added value for sectorial applications. *Clim. Serv.* <http://dx.doi.org/10.1016/j.cliser.2017.06.004>. (in press).
- Maraun, D., Wetterhall, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themeßl, M., Venema, V.K.C., Chun, K.P., Goodess, C.M., Jones, R.G., Onof, C., Vrac, M., Thiele-Eich, I., 2010. Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.* 48, RG3003. <http://dx.doi.org/10.1029/2009RG000314>.
- Molteni F., Stockdale T., Balmaseda M., Balsamo G., Buizza R., Ferranti L., Magnusson L., Mogensen K., Palmer T.N., Vitart F., 2011. The new ECMWF seasonal forecast system (System 4). ECMWF Technical Memorandum, N 656. <http://www.ecmwf.int/publications/library/do/references/show?id=90277> (accessed 2016.12.19).
- Mtongori, H.I., Stordal, F., Benestad, R.E., 2016. Evaluation of empirical statistical downscaling models' skill in predicting Tanzanian rainfall and their application in providing future downscaled scenarios. *J. Clim.* 29 (9), 3231–3252. <http://dx.doi.org/10.1175/JCLI-D-15-0061.1>.
- Müller, W.A., Appenzeller, C., Doblas-Reyes, F.J., Liniger, M.A., 2005. A debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes. *J. Clim.* 18, 1513–1523. <http://dx.doi.org/10.1175/JCLI3361.1>.
- Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized linear models. *J. R. Stat. Soc. A* 135, 370–384. <http://dx.doi.org/10.2307/2344614>.
- Novella, N.S., Thiaw, W.M., 2013. African rainfall climatology version 2 for famine early warning systems. *J. Appl. Meteor. Climatol.* 52, 588–606. <http://dx.doi.org/10.1175/JAMC-D-11-0238.1>.
- Nikulin, G., Jones, C., Giorgi, F., Asrar, G., Buchner, M., Cerezo-Mota, R., Christensen, O.B., Déqué, M., Fernandez, J., Hänsler, A., van Meijgaard, E., Samuelsson, P., Sylla, M.B., Sushama, L., 2012. Precipitation climatology in an ensemble of CORDEX-Africa regional climate simulations. *J. Clim.* 25, 6057–6078. <http://dx.doi.org/10.1175/JCLI-D-11-00375.1>.
- Omondi, P., Ogallo, L.A., Anyah, R., Muthama, J.M., Ininda, J., 2013. Linkages between global sea surface temperatures and decadal rainfall variability over Eastern Africa region. *Int. J. Climatol.* 33, 2082–2104. <http://dx.doi.org/10.1002/joc.3578>.
- Peterson, T.C., Folland, C., Gruza, G., Hogg, W., Mokssit, A., Plummer, N., 2001. Report on the activities of the Working Group on Climate Change Detection and Related
- Rapporteurs 1998–2001. World Meteorological Organization, Geneva, Switzerland.
- Philippon, N., Camberlin, P., Fauchereau, N., 2002. Empirical predictability study of October–December East African rainfall. *Q.J.R. Meteorol. Soc.* 128, 2239–2256. <http://dx.doi.org/10.1256/qj.01.190>.
- Rockel, B., Will, A., Hense, A., 2008. Regional climate modeling with COSMO-CLM (CCLM). *Meteor. Z.* 17, 347–348. <http://dx.doi.org/10.1127/0941-2948/2008/0309>.
- Rockel, B., 2015. The regional downscaling approach: a brief history and recent advances. *Curr. Clim. Change Rep.* 1, 22–29. <http://dx.doi.org/10.1007/s40641-014-0001-3>.
- Rummukainen, M., 2010. State-of-the-art with regional climate models. *WIREs Clim. Change* 1, 82–96. <http://dx.doi.org/10.1002/wcc.8>.
- Rummukainen, M., 2016. Added value in regional climate modeling. *WIREs Clim. Change* 7, 145–159. <http://dx.doi.org/10.1002/wcc.378>.
- San-Martín, D., Manzanas, R., Brands, S., Herrera, S., Gutiérrez, J.M., 2017. Reassessing model uncertainty for regional projections of precipitation with an ensemble of statistical downscaling methods. *J. Clim.* 30, 203–223. <http://dx.doi.org/10.1175/JCLI-D-16-0366.1>.
- Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., Rudolf, B., 2014. GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Climatol.* 115, 15–40. <http://dx.doi.org/10.1007/s00704-013-0860-x>.
- Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Ziese, M., 2015. GPCC Full Data Reanalysis Version 7.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data. [https://doi.org/10.5676/DWD.GPCC/FD\\_M\\_V7\\_050](https://doi.org/10.5676/DWD.GPCC/FD_M_V7_050).
- Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Wang, W., Powers, J. G., 2008. A Description of the Advanced Research WRF Version 3, NCAR/TN-475 + STR, [http://www2.mmm.ucar.edu/wrf/users/docs/arw\\_v3.pdf](http://www2.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf) (accessed 2016.12.19).
- Stockdale, T., Doblas-Reyes, F., Ferranti, L., 2009. EURO-SIP: multi-model seasonal forecasting. *ECMWF Newsl.* 118, 10–16. <http://dx.doi.org/10.21957/7wc0nybvir>.
- Strandberg, G., Barring, L., Hansson, U., Jansson, C., Jones, C., Kjellström, E., et al., 2015. CORDEX scenarios for Europe from the Rossby Centre regional climate model RCA4. Report Meteorology and Hydrology (RMK) 116, <http://www.smhi.se/publikationer/publikationer/cordex-scenarios-for-europe-from-the-rossby-centre-regional-climate-model-rca4-1.90272> (accessed 2016.12.19).
- Tarnavsky, E., Grimes, D., Maidment, R., Black, E., Allan, R., Stringer, M., Chadwick, R., Kayitakire, F., 2014. Extension of the TAMSAT satellite-based rainfall monitoring over Africa and from 1983 to present. *J. Appl. Meteor. Climatol.* 53, 2805–2822. <http://dx.doi.org/10.1175/JAMC-D-14-0016.1>.
- Weedon, G.P., Balsamo, G., Bellouin, N., Gomes, S., Best, M.J., Viterbo, P., 2014. The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resour. Res.* 50, 7505–7514. <http://dx.doi.org/10.1002/2014WR015638>.
- Weigel, A.P., Liniger, M.A., Appenzeller, C., 2007. The discrete Brier and ranked probability skill scores. *Mon. Weather Rev.* 135, 118–124. <http://dx.doi.org/10.1175/MWR3280.1>.
- Weigel, A.P., Liniger, M.A., Appenzeller, C., 2008. Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Q. J. Roy. Meteor. Soc.* 134, 241–260. <http://dx.doi.org/10.1002/qj.210>.
- Weisheimer, A., Palmer, T.N., 2014. On the reliability of seasonal climate forecasts. *J. R. Soc. Interface* 11 (96). <http://dx.doi.org/10.1098/rsif.2013.1162>.
- Wilks, D.S., 2011. *Statistical Methods in the Atmospheric Sciences*. Academic Press.
- Shao, Q., Li, M., 2013. An improved statistical analogue downscaling procedure for seasonal precipitation forecast. *Stoch. Env. Res. Risk A* 27, 819–830. <http://dx.doi.org/10.1007/s00477-012-0610-0>.
- Zorita, E., von Storch, H., 1999. The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *J. Clim.* 12, 2474–2489. [http://dx.doi.org/10.1175/1520-0442\(1999\)012<2474:TAMAA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(1999)012<2474:TAMAA>2.0.CO;2).