NASA's NMME-based S2S Hydrologic Forecast System for Food Insecurity Early Warning in Southern Africa

Abheera Hazra[1,2], Amy McNally[1,3,4], Kimberly Slinski[1,2], Kristi R. Arsenault[1,4], Shraddhanand Shukla[5], Augusto Getirana[1,4], Jossy P. Jacob[1,6], Daniel P. Sarmiento[1,4], Christa Peters-Lidard[1], Sujay V. Kumar[1] and Randal D. Koster[1]

[1]*NASA GSFC, Greenbelt, MD*

[2]*ESSIC, UMD, College Park, MD*

[3]*U.S. Agency for International Development, Washington DC*

[4]*SAIC, Reston, VA*

[5]*Climate Hazards Center, University of Santa Barbara, Santa Barbara, CA*

[6]*Science Systems and Applications Inc, Lanham, MD*

to be Submitted to the *Journal of Hydrology*

~ 2021

Corresponding author:

Abheera Hazra, *ESSIC/NASA GSFC, Greenbelt, MD*

Abstract

In situ hydrologic monitoring over regions most susceptible to food insecurity can be a challenge in current times due to various socio-economic and political issues in combination with environmental factors such as ongoing famine or drought. Hydrologic monitoring and initializing forecasts based on remotely sensed and analyzed data can contribute significantly to early warning in such regions. Routine hydrologic forecasts, as provided by NASA's Hydrologic Forecasting and Analysis System (NHyFAS), are a recent addition to early warning systems. A custom instance of NHyFAS, termed FLDAS-Forecast, is used by FEWS NET's Land Data Assimilation System (FLDAS). The FLDAS-Forecast's dynamic forecasting component was originally set up with Goddard Earth Observing System (GEOS) forecast inputs and has been recently expanded with precipitation forecast forcing from the North American Multi-Model Ensemble (NMME). This paper describes the improvements in seasonal hydrologic forecasts produced with this updated system. Evaluations in this study focus on soil moisture across southern Africa's growing season. Soil moisture forecasts are benchmarked and evaluated relative to climatology-based forecasts and historic runs, which are driven by observation-based meteorological forcing fields, and they are verified with remotely sensed observations of soil moisture and vegetation. Through multiple deterministic and probabilistic skill assessments, we show that using the larger ensemble of NMME precipitation inputs in the forecast system results in higher quality hydrologic forecasts than are allowed by climatology- or GEOS-only-based forecasts. Further, the near-real-time NMME-based rootzone soil moisture forecasts were able to correctly predict developing drought conditions over southern Africa through late 2019 and into early 2020.

1 INTRODUCTION

There is acute food insecurity in many regions around the globe that are also subject to recurring drought conditions (UN 2018). Early warning of drought and response at the right time could help a large number of people, especially in the particularly food-insecure regions of Africa and the Middle East (Vörösmarty et al. 2005; Getirana et al. 2012), which are some of the most food-insecure regions in the world. The Famine Early Warning Systems Network (FEWS NET; https://fews.net, Verdin et al. 2005; Funk et al. 2019) provides objective and evidence-based analyses to help government decision-makers and relief agencies plan for and respond to such humanitarian crises. Forecasting systems such as National Oceanic and Atmospheric Administration's (NOAA)'s Africa-specific sub-seasonal to seasonal (S2S) meteorological forecasts (Thiaw et al. 2015) and the National Aeronautics and Space Administration's (NASA's) multi-model, remote sensing-based Hydrological Forecasting and Analysis System (NHyFAS; Arsenault et. al. 2020) are critical to such efforts.

Seasonal forecasts of meteorological and hydrologic conditions are an important tool for early warning, and also for developing and guiding strategic planning of water resources across different climate-sensitive sectors (Sheffield et al. 2014; Shukla et al. 2014, 2020; De Felice et al. 2015; Viel et al. 2016; Arsenault et al., 2020). Several operational products providing dynamic meteorological seasonal forecasts include NOAA's North American Multi-Model Ensemble's (NMME; Kirtman et al. 2014), NASA SERVIR ClimateSERV's downscaled NMME forecasts (Flores Cordova et al. 2012), Copernicus Climate Change Service (C3S; Buchwitz et al. 2017; http://climate.copernicus.eu/seasonal-forecasts) and World Meteorological Organization's (WMO) long-range forecasts using multi-model ensembles (https://www.wmolc.org). There are operational hydrologic forecast products as well, such as the Global Flood Awareness System (GloFAS; Emerton et al. 2018), which provides probabilistic seasonal forecasts of river flow at up to four-months lead time for a global river network and the Africa Drought and Flood Monitor (http://stream.princeton.edu/; Yuan et al. 2013; Sheffield et al. 2014), which provides hydrologic forecasts out to a week over select regions of Africa using the Canadian Centre for Climate Modeling and Analysis Coupled Climate Model (Merryfield et al. 2013).

However, Wolski et al. (2017) have highlighted the lack of finer than 100km (1^0) spatial resolution, country-scale seasonal hydrologic forecasts of extreme conditions over Africa. Hydrologic forecasts are of particular importance as they reduce the impact of extreme events by identifying high-risk areas; such forecasts, for example, could show how heavy precipitation upstream along a river may affect areas downstream that did not receive above-average precipitation. NHyFAS, developed by Arsenault et al. (2020), fills this gap, producing high-resolution seasonal hydrologic forecasts at 25km (0.25^0) over continental Africa and the Middle East to support food and water security.

NHyFAS has successfully demonstrated the forecasting of both hydrologic drought and flood risks across Africa and the Middle East (Arsenault et al. 2020). It utilizes the data assimilation and modeling capabilities of NASA's Land Information System (LIS; Kumar et al. 2006; Peters-Lidard et al. 2007). A modified version of NHyFAS (without data assimilation) has been providing routine hydrologic forecasts across Africa and the Middle East since 2018 as a part of FEWS NET's Land Data Assimilation System (FLDAS; McNally et al. 2017). Only this custom instance of NHyFAS for FEWSNET, termed as FLDAS-Forecast, is utilized in this study. The FLDAS-Forecast system, was initially set up with a single NMME dynamical model, NASA's Goddard Earth Observing System, version 2 (GEOS; Rienecker et al. 2008; Molod et al. 2012; Borovikov et al. 2019; Molod et al. 2020) sub-seasonal-to-seasonal (S2S) forecast system. This FLDAS-Forecast system is considered version 1 (FFV1). However, various independent studies (Wang 2009; Becker et al. 2014; Krakauer 2017; Wanders et al. 2016; Cash et al. 2019) have shown that NMME's S2S precipitation and temperature forecasts are significantly more skilful than any individual model (including GEOS) in the suite. To leverage this multi-model skill and to thereby improve the early warning system to support proactive drought management efforts, the FLDAS-Forecast was upgraded to incorporate the full suite of NMME precipitation forecasts in combination with GEOS non-precipitation forecast forcing fields. This upgraded system is considered version 2 (FLDAS-Forecast version 2; FFV2). Using precipitation information from the NMME forecast suite is particularly beneficial given that precipitation is the primary driver for surface and ground hydrology.

The FLDAS-Forecast system is a multi-Land Surface Model (LSM) seasonal forecast system that uses NASA's capabilities in modeling and utilizing remote sensing-based products and is set up specifically for continental Africa and the Middle East. This system bias-corrects and downscales all the meteorological forecast fields produced by the dynamical forecast system using well-tested techniques keyed to satellite- and station-based data (Wood et al. 2004, Arsenault et al. 2020, Shukla et al. 2020). The novelty in this study is the incorporation of the full suite of bias-corrected and downscaled NMME precipitation and GEOS non-precipitation meteorological forcings to produce multi-LSM S2S hydrologic forecasts at 0.25^0 spatial resolution over continental Africa and Middle East. One of the main outcomes of implementing NMME into the system's multiple LSMs ensembles is a large increase in the number of ensemble members, which allows us to examine the probabilistic nature of forecasted hydrological extremes.

The hypothesis tested in this study is that, since the ensemble of NMME precipitation forecasts have better seasonal skill than any individual model's forecasts in the NMME suite, the NMME-based hydrological forecasts will also have more skill than those based only on the GEOS precipitation-based forecasts. Various studies have already analyzed the meteorological forecast skill of the different NMME models out to the 6-month forecast lead (Mo and Lyon 2015; Shukla et al. 2016; Setiawan et al. 2017). These studies have shown that using the full

NMME suite provides more accurate forecasts than any single model. Several other studies have highlighted the importance of exploring the ensemble spread as a source of information regarding forecast uncertainty using various probabilistic metrics (Paiva et al. 2012; Schaeybroet and Vannitsem 2016). In this paper, we follow a combined approach of using deterministic analyses and probabilistic analyses of extremes to explore and demonstrate the benefits of using the full NMME ensemble in FFV2 compared to FFV1, and thereby determine whether or not including the full NMME suite is justified.

Implementing NMME into the FLDAS-Forecast system was a substantial undertaking. The main objectives of the present paper are: (1) to describe the framework and methodology for including NMME precipitation forecasts into the FLDAS-Forecast system and (2) to demonstrate that the inclusion of this information improves hydrological forecasts relative to those obtained with GEOS-based forecast information alone. This study compares FFV2 to FFV1 and benchmarks the FLDAS-Forecast versions relative to Ensemble Streamflow Prediction (ESP)-based forecasts. The forecast performance is assessed in terms of:

i) Deterministic (i.e., ensemble mean) forecast skill analysis relative to historic model runs

ii) Deterministic (ensemble mean) comparison of the NMME-based FLDAS-Forecast system with remotely sensed observation; and

iii) Probabilistic forecast skill analysis of extreme conditions relative to historic model runs.

The evaluations focus on southern Africa (SA), which has suffered consecutive droughts in the 1990s and from 2014 through 2020 (Edossa et al. 2014; Nash et al. 2019).

The next section describes the Land Information System (LIS) along with the land surface models and the FLDAS-Forecast framework. Section 3 specifies the data sets used and the analysis's methodological approach. Results are presented in section 4, followed by a discussion of application and limitations of the system in section 5. A summary and a discussion of future developments is provided in section 6.

2 FRAMEWORK

FFV2 routinely produces monthly hydrologic forecasts over continental Africa and the Middle East. Figure 1 describes the schematic for FLDAS-Forecasts which includes: establishing initial conditions for the forecasts using observation- and reanalysis-based meteorological data sets, generating ensemble forecasts of hydrologic conditions based on NMME meteorological forecasts, and producing drought and flood risk analysis products derived from the hydrological forecast data. These steps are described in detail in the following subsections.

2.1 Land Information System (LIS)

NASA's Land Information System (LIS; Kumar et al. 2006; Peters-Lidard et al. 2007) is a high-performance, terrestrial hydrology modeling and data assimilation framework developed in the Hydrological Sciences Laboratory at NASA's Goddard Space Flight Center. LIS is a flexible software framework that can be customized by end-users according to their preference and expanded to meet their changing needs. LIS's Land surface Data Toolkit (LDT; Arsenault et al. 2018) is used to parametrize the hydrological models and pre-process model inputs in NHyFAS and also in its custom instance - FLDAS-Forecast. The LSMs utilized for FLDAS-Forecast through LIS have groundwater schemes and are run in tandem with the Hydrological Modeling and Analysis Platform (HyMAP; Getirana et al. 2012, 2017a). HyMAP provides the river routing scheme and is driven by the total runoff from the LSMs. LIS's Land Verification Toolkit (LVT; Kumar et al. 2012) provides multiple evaluation and drought metrics.

The LIS Framework (LISF) is at the core of FLDAS-Forecast. LISF also supports data assimilation strategies for forecast initialization, which are currently in development. Data assimilation, however, is not currently utilized in FLDAS-Forecasts.

2.2 FLDAS-Forecast Modeling Framework

2.2.1 Land Surface Models and Streamflow Routing

FLDAS-Forecast employs two LSMs through LIS: Noah with multi-parameterizations (Noah-MP; Niu et al. 2011) and NASA's Catchment LSM (CLSM; Koster et al. 2000). It also uses the HyMAP river routing scheme. These two LSMs were also evaluated for water budget over eastern Africa (Jung et al. 2017). The entire system is currently set up over continental Africa and the Middle East at a spatial resolution of 0.25° x 0.25°, and at a 15-minute model time step. The hydrologic output is produced at monthly time step. Both the LSMs have soil moisture profiles, with water contents at surface (NoahMP-10cm, CLSM-2cm; average-6cm), rootzone (1m), and total (2m) depths. Hydrological output includes surface and sub-surface runoffs, terrestrial water, and through the application of HyMAP, streamflow.

2.2.2 FLDAS reanalysis

The FLDAS-Forecast modeling framework generates non-forecast historic simulations from 1982 to present (termed as "reanalysis" or RA) based on observational and reanalysis meteorological data sets, following the setup of McNally et al. (2017). This long period of record (nearly 40 years) is critical for drought and flood risk assessment. The RA uses precipitation inputs from Climate Hazards Center InfraRed Precipitation with Station data, version 2.0 (CHIRPS; Funk et al. 2015). All other requisite meteorological inputs are from NASA's Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2; Bosilovich et al. 2016; Gelaro et al. 2017). The FLDAS-Forecast uses CHIRPS-prelim (Funk et al. 2015; https://data.chc.ucsb.edu/products/CHIRPS-2.0/prelim/) only for near-real-time monitoring and producing forecast initial conditions because the 3-day latency of this product allows model updates close to real time. For all prior time steps, the system uses the CHIRPS-final 6-hourly product (Dinku et al. 2018; http://data.chc.ucsb.edu/products/CHIRPS-2.0/africa_6-hourly/), which has a latency of ~2-3 weeks and is available for continental Africa and part of the Middle East domain. MERRA-2 data have about an internal 10-day latency, which is sufficient for seasonal climate forecast initialization and monitoring. The RA is used to initialize the NMME-ensemble, GEOS-only, and ESP-based forecast model runs and is considered the primary reference data set in this study.

2.2.3 Meteorological forcing data for NMME based hydrologic forecasts

The meteorological forecasts used in setting up FFV2 are dynamic; the meteorological fields within the models that produce them evolve in time in response to imposed dynamical equations in the modeled atmospheric and ocean components and parameterizations of physical processes, such as turbulence and moist convection. FFV2 uses dynamical precipitation forecasts from the NMME suite. NMME provides near-real-time monthly forecasts based on 98 ensemble members and hindcasts (1982-2010) based on 68 ensemble members from the Climate Forecast System, version 2 (CFSv2; Saha et al. 2010, 2014), Geophysical Fluid Dynamics Laboratory's (GFDL; Delworth et al. 2012, Vecchi et al. 2014) forecast-oriented climate model version 2.5, Canadian Coupled Models (CanCM4i and GNEMO; Lin et al. 2020); the NCAR Climate System Model, version 4 (CCSM4; Gent et al. 2011) and NASA Global Modelling and Assimilation Office's (GMAO) GEOS version 2 (GEOSv2; Borovikov et al. 2019), as provided in Table 1. A particular innovation of FFV2 is that it has the flexibility to run any combination of the NMME suite's models. FFV1 is a subset of this option; it uses the meteorological forcings from GEOSv2 alone. The current FFV2 system employs all active NMME models with all the ensemble members (as of 2021; Table 1). Most models have 10 ensemble members, the only exception is CFSv2, for which we use 12 of the 24 ensemble members to aptly handle the processing of the model. For GFDL-Flor, we use all the 24 ensemble members, but average each of the 12 members of its two sub-models to have 12 averaged ensemble members in case of the hindcast period (1982-2010).

Because the NMME does not provide the full complement of required input fields (e.g., temperature, radiation, winds), FFV2 uses non-precipitation meteorological forcings from GEOS seasonal forecasts. The GEOS seasonal forecasts consist of 10 ensemble members for real-time applications and 4 ensemble members for the hindcast period (1982-2010). The smaller number of GEOS non-precipitation ensemble members are in part randomly matched with the NMME-based precipitation ensemble members in the FFV2 setup. For example, GEOS's four non-precipitation ensemble members are first matched twice up to the first eight out of the ten precipitation ensemble members of CCSM4 and then randomly matched to the last two precipitation ensemble members of CCSM4 in the hindcast period. For the near-real-time period, GEOS's non-precipitation ensemble members and CCSM4 precipitation ensemble members are matched one to one.

The meteorological forecasts are then bias-corrected and spatially downscaled (BCSD; Wood et al. 2004) using the RA-based meteorological data sets (CHIRPS and MERRA-2) as the reference data, following the methods outlined in Arsenault et al. (2020). Forecasts from climate models do not usually match the statistical properties of the RA inputs (e.g., due to lead-dependent climate drift [Gupta et al. 2013; Hermanson et al. 2018]); bias correction is used to reduce such errors (Cui et al. 2012). The BCSD method used in this study has been evaluated and verified across Africa for NHyFAS (Arsenault et. al 2020, Shukla et al. 2020).

2.2.4 ESP-based hydrologic forecasts

In addition to NMME, FLDAS-Forecast can also employ the ESP-type forecasting approach (Twedt et al. 1977; Day 1985; Yuan et al. 2015; Li et al. 2009; Yossef et al. 2017). The ESP forecasts are climatology-based and use the same observed inputs (CHIRPS and MERRA-2) as RA. LIS configures the ensemble of meteorological forecast members by assembling individual years from the historical MERRA-2 and CHIRPS meteorological data, taking each historical year as a representation of a potential single "forecast"; thus, the 1982–2011 CHIRPS and MERRA-2 data holdings allow us to produce 30 ensemble members for each initial condition. The skill of the hydrological ESP forecast is derived only from the initial hydrological conditions.

In this study, the ESP-based forecast is used as a benchmark for both versions of FLDAS-Forecast output, as both dynamic forecast inputs are bias-corrected and downscaled relative to the same sets of observed data that go into ESP-based runs. Also, the initial conditions are the same for the ESP-based forecasts and both versions of FLDAS-Forecast.

2.2.5 Summary of FLDAS-Forecast version 2 workflow

The FLDAS-Forecast workflow (illustrated in Figure 2) consists of three main steps: 1) Pre-processing, 2) LIS-based processing, and 3) Post-processing. During the pre-processing step, all NMME precipitation and complementary GEOS non-precipitation fields needed to run the LSMs are gathered, downscaled and bias-corrected relative to CHIRPS and MERRA-2 data sets, respectively, and are then temporally disaggregated to sub-daily time-steps (6-hour intervals). In the LIS-based processing section, the two LSMs (NoahMP and CLSM) are run with the forcing data produced in the previous step, producing a combined model total of 196 near-real-time hydrologic forecast ensemble members and 136 hindcast ensemble members per forecast start date. The FFV1 considered in this study produces a total of 20 ensemble members for near-real-time forecasts and 8 ensemble members for the hindcasts. In this step, FLDAS-Forecast is also configured in ESP mode for the same set of LSMs and initial conditions to produce ESP-based forecasts with a total of 60 ensemble members per forecast start date. In the Post-processing step, we analyze the near-real-time hydrologic forecast outputs for hydrologic extremes, such as droughts and flood-potentials relative to hindcast climatology. These results are updated on our webpage (https://ldas.gsfc.nasa.gov/fldas/models/forecast) by the second week of every month and are provided to our FEWS NET partners and to regional scientists.

2.2.6 Derived products

Derived products enhance FEWS NET's early warning capabilities by allowing regional experts to visualize the potential hydrologic impacts of forecasted climate. FLDAS forecast-based real-time ESP forecasts over Africa are used to produce probabilistic tercile maps for rootzone soil moisture (RZSM) to indicate the likelihood of "above-normal" (greater than 67th percentile), "normal" (between 33rd to 67th percentile), and "below-normal" (less than 33rd percentile) conditions. These conditions, especially "below normal" (a potentially drought-like condition in hydrology), forms the basis of various categorical probabilistic evaluations in this study. The ESP forecast percentiles are derived from the RA-based output, and the tercile maps provide estimates of hydrologic conditions when the input meteorological forcings are considered to be climatologically average for forecast initial conditions.

Similarly, NMME RZSM forecast-based probabilistic tercile maps are also produced over Africa and the Middle East, where percentiles are based on the NMME RZSM hindcasts. The terciles describe the likelihood of the three categories based on the ensemble members. Additionally, FFV2 routinely provides the ensemble median of NMME-based forecasted soil moisture percentile, anomaly, standardized anomaly, and percent saturation, along with 3-month aggregates of soil moisture percentile, anomaly, and standardized anomaly (https://ldas.gsfc.nasa.gov/fldas/models/forecast).

3 Methodology

The impact of implementing NMME in FLDAS-Forecast was evaluated by assessing improvements in deterministic forecast skill, ensemble spread, and probabilistic forecast skill. All the assessments are focused on the forecasts covering the wet season over SA, as it is the critical growing season for crops (Laux et al. 2009; Sultan et al. 2010; Mubaya et al. 2012; Trambauer et al. 2015; Seibert et al. 2017). Hence, forecasts initialized with hydrologic conditions covering the start of the wet season –September, October, and November (SON)– are considered in this study. The following subsections describe the evaluation data set, as well as our analysis approach.

3.1 Dataset

For all the analyses in this study, except those using remotely sensed observations, the ESP, FFV1, and FFV2 forecasts and the RA are first standardized by converting them to percentiles. For the analyses using remotely sensed observations, all the data sets are standardized by converting them to standardized anomalies due to the observation's relatively short temporal extent. Unless otherwise noted, all the forecasts (ESP, FFV1, and FFV2) are then averaged over three months of data: data at zero lead (i.e., over the first month of the forecast), at 1-month-lead (over the 2nd month of the forecast), and so on out to the 5-month lead. For simplicity, we will refer to the 0-month-lead through 5-month-lead averages as OND (for October, November, and December start dates), NDJ, DJF, JFM, FMA, and MAM, respectively. The initial condition (IC) period covers the start of the rainy season (SON) and is provided by the RA. Usually, skill is lost with increasing forecast lead; hence, averaging the forecasts starting from the three different IC months helps isolate, for each lead, the forecast signal from the noise. Since, RA is used as the "truth" to evaluate the seasonal forecasts, it is also averaged over the OND, NDJ, DJF, JFM, FMA, and MAM periods. We will hereafter use "hindcast" interchangeably with "forecast" for the sake of clarity.

3.2 Deterministic forecast skill analysis

For all deterministic skill assessments of RZSM (top 1m) forecasts, we consider the ensemble mean, as it is the "best" deterministic forecast (Ehrendorfer 1997). We calculate the anomaly correlation (AC) and the root mean squared error (RMSE) between the ensemble mean of ESP, FFV1, and FFV2 seasonal RZSM forecast percentiles with respect to the RA ("perfect run") RZSM percentiles, which are used here as the "truth" for validating the forecasts. The anomalies used in AC are computed by subtracting the seasonally varying climatologies from the individual monthly values (forecasts or RA). Both metrics are assessed over the period of 1982-2010; the equations used for computing them are provided in Appendix 1 (Equations 1 and 2). The AC is chosen as one of the evaluation metrics, as it shows how well the variabilities of the forecast anomalies match the variability of the RA anomalies. It does not, however, provide any information about the magnitude of forecast error, which is determined using RMSE. Statistical significance for AC is determined using the Fisher transformation. The Fisher transformation ensures Gaussian distribution of the AC coefficients, and the Z-score provides the probability-value (p-value) of the correlation coefficient for ESP, FFV1, and FFV2 with RA. For both metrics, we also determine the differences between values produced by FFV1 and FFV2. The significance of the difference of AC between the FLDAS-Forecast versions is also determined using Fisher transformation, the statistically significant values for AC and AC differences are provided at the 95% confidence interval (Appendix 1 [Equations 1a. and 1b.]).

Further, to estimate the best forecast's deterministic accuracy relative to observations, we calculate the correlation (Appendix 1 [Equation 3]) between the seasonal RA and ensemble mean of the forecast's surface soil moisture (SSM; top 6cm) standardized anomalies with respect to remotely sensed soil moisture and vegetation standardized anomalies. Both the remotely sensed observations are first spatially upscaled to 25km before being standardized. We map the correlation between the SSM and the NASA Soil Moisture Active Passive (SMAP) mission standardized anomalies between 2015-2020. SMAP provides high-quality soil moisture estimates posted at 9 km spatial and daily temporal resolution (Entekhabi et al. 2016). In addition to correlations with SMAP, we also compute the correlation between model SSM and Global Inventory Monitoring and Modeling System's (GIMMS) Normalized Difference Vegetation Index (NDVI; Spruce et al. 2016; Tucker 1978) standardized anomalies between 2010-2018. The GIMMS vegetation index is available at 0.25 km spatial and 8-day temporal resolution. To correlate the seasonal RA and forecasts relative to both of these remotely sensed observations, the observations are also averaged over the seasonal periods of OND, NDJ, DJF, JFM, FMA, and MAM. For the correlation with remotely sensed observations, the Fisher transformation is used to test the significance which we provide at the 95% confidence interval.

3.3 Ensemble spread analysis

Before studying the probabilistic features of the ensembles in the seasonal forecasts, we analyze the ensemble spreads of FFV2 and FFV1, as ensemble spread is often related to uncertainty (Grimit and Mass 2007; Schaeybroeck and Vannitsem 2014; Hopson 2014). We compare the ensemble spreads of the RZSM forecasts over 1982 to 2010 through the sixth forecast lead between the FLDAS-Forecast versions for each initial condition (IC) month (September, October, and November). The ensemble spread is defined as the difference between the maximum and minimum values across each forecast ensemble (Fortin et al. 2014; Appendix 1 [Equation 4]).

3.4 Probabilistic forecast skill analysis for hydrologic drought

We use different probability metrics to examine how often the forecasts are accurate, and this requires categorizing certain forecast events. This in fact fits in well with our interest in categories of soil moisture (drought; below normal). Unless otherwise noted, here too we average the forecasts covering the period 1982-2010 over the different lead periods (OND, NDJ, DJF, JFM, FMA, and MAM) and over the IC months (SON), which represents the start of the rainy season. The probability metrics are based on all the ensembles present in each of the three forecasts.

First, we compute the Rank Probability Skill Score (RPSS; Müller et al., 2005), which describes the quality of categorical probabilistic forecasts for drought, both at the seasonal and sub-seasonal scales. For this calculation, drought conditions are categorized as an RZSM less than the 33^{rd} percentile. RPSS is based on the Rank Probability Skill (RPS), which is the cumulative squared probability error. The ratio of the difference between the climatological RPS and forecast RPS to the climatological RPS is the RPSS (Appendix 1 [Equation 5]). RPSS rewards a forecast for the number of ensemble members that fall within the observed category—the larger that number is, the higher the RPSS. An RPSS greater than 0 is considered skilful in comparison to the climatological forecast, and an RPSS of 1 indicates a perfect forecast. For the RPSS, we adopted the threshold of above five percent (5%) as representing "good skill" in a forecast, following Goddard and Dilley (2005). To support the results of RPSS, we also compute the forecast Hit Rate for drought conditions by creating a contingency table and using the ratio of forecasted drought events (RZSM less than 33rd percentile) and observed drought events (Appendix 1[equation 6]).

Reliability Diagrams representing probability of seasonal forecast events relative to observations are also included in the probabilistic analysis. These are graphs of the observed frequency of a categorical event plotted against the forecast probability of that event during our hindcast period of study from 1982 to 2010 over SA (Brocker and Smith 2007; Wilks 1995; Jolliffe and Stephenson 2003). Similar to the RPSS analysis, we define drought category using RZSM percentiles. This diagram is useful for decision-making purposes, as it tells users how often a given forecast probability of a certain category (drought) matches the frequency of the event in observations. A perfectly reliable forecast system would have a 1:1 association with the observed frequency (e.g., a forecast probability of 50% will be associated with observed frequency of 50%). In this study, the reliability diagram is plotted by dividing the forecast probabilities (0–1.0) into 10 bins and estimating observed frequency of a given category for each of those bins.

Finally, we compute the Relative Operating Characteristics (ROC; Mason and Graham 1999, Hogan and Mason 2012, Siegmund et al 2015), a complementary metric to the reliability diagram. ROC is used in forecast verification to measure the ability of the forecasts to distinguish an event from a non-event; it utilizes plots of the probabilistic Hit Rate (HR) to

False Alarm Rate (FAR) for a given category (drought). It is plotted by constructing a contingency table for drought, with HR probabilities (Y-axis) plotted against the FAR probabilities (X-axis) by dividing the respective probabilities of 0 to 1.0 into 10 bins. A perfect ROC curve would have high HR probability relative to the FAR probabilities, lending further confidence to the forecast accuracy.

4 RESULTS AND DISCUSSION

4.1 Deterministic Analysis

Figure 3 and row-1 in Table 2 present forecast skill as represented by anomaly correlation coefficient (ACC) of the RZSM forecast percentiles relative to the RA percentiles. The ensemble mean seasonal forecasts are averaged over OND, NDJ, DJF, JFM, FMA, and MAM following the SON ICs. The first row of Figure 3 is the skill (ACC) of the seasonal hydrological ESP forecast, where skill is derived from the initial hydrological conditions alone. Moderate skill is seen at lead 1 and lead 2 at 0.3-0.6 ACC with localized areas of 0.8 ACC in the Democratic Republic of Congo. RZSM skill for both FFV1 and FFV2, shown in rows 2 and 3 (Figure 3), respectively, is higher than the ESP-based seasonal forecast skill (row 1). In terms of area averages, the skills over southern Africa (Table 2, row 1) for both FFV1 and FFV2 are about 0.1-0.15 ACC larger than that for ESP skill in the early part of the season.

The overall ACC of all the seasonal forecasts decreases with increasing lead. However, the ACC of both the dynamical versions of the seasonal forecasts lasts longer than that of the ESP seasonal forecasts. This is expected, though not assured, given that the dynamical models are deriving skill from both initial conditions and the forecasted meteorology, with the latter providing skill at longer leads. Importantly, FFV2 has higher ACC than FFV1; as seen in the last row of Figure 3, many regions show improvements of more than 0.2 ACC with FFV2 relative to FFV1, a difference that is statistically significant, as determined by the Fisher Transformation (Appendix 1.b). The figure also shows that the difference between the skill of

FFV2 and FFV1 is relatively smaller in the first lead because of greater influence of IC than in the longer leads. Although there are some areas in SA that show FFV2 with lower skill than FFV1, those areas are relatively small. These results are similar to those shown by Becker et al. (2014) for precipitation forecasts over the northern hemisphere (23°–75°N), where ACC for the entire NMME suite-based forecast was found to be significantly more skilful than that for the GEOS model alone.

In addition to ACC, we compared the root mean squared error (RMSE) of the three forecast percentiles (Figure 4) relative to the RA percentiles. Here too we provide results for the ensemble means at the six forecast leads (OND, NDJ, DJF, JFM, FMA, and MAM). The results are similar to those for ACC. The RMSE values gradually increase with lead for all three forecasts; as expected, the RMSE values at 0-month lead are the lowest, given the proximity of this lead to the IC. RMSEs for both FFV1 and FFV2 (rows 2 and 3, respectively) are smaller than those for ESP (row 1), consistent (in terms of skill) with what was found in Figure 3 for ACC. Also, outside of a few small areas, FFV2 shows higher skill (lower RMSEs) than FFV1. The areally-averaged seasonal forecast RMSEs tabulated in Table 2 (row 2) confirm that RMSEs in all forecasts increase with lead and that FFV2 produces the lowest RMSEs, the latter further highlighting the improvements obtained with the NMME-based FLDAS-Forecast.

4.2 Comparison to Observations

We show above that FFV2 provide forecasts that agree best with the RA model products. Here, we evaluate the FFV2 forecasts directly against observations. We compare the correlation (R) between the seasonal RA and FFV2 ensemble mean (top 6 cm) SSM with seasonal SMAP soil moisture (top 5 cm) over SA between 2015 to 2020. There is generally a good agreement between the RA and SMAP values in row 1 of Figure 5, where a large section over the eastern part of southern Africa shows more than 0.9R, and with most of the other regions showing significant values of more than 0.6R. Even the FFV2-based ensemble-mean seasonal forecasts (row 2 of Figure 5) show significantly high correlations (R>0.6) with SMAP over large areas of SA out to the 2-month lead (DJF), with gradually decreasing values thereafter.

Results of the correlation between RA and FFV2 ensemble mean SSM with GIMMS NDVI over SA between 2010 to 2018 are similar (Figure 6). The seasonal RA SSM (row 1) shows good agreement with NDVI, with significant values of more than 0.7R over large sections of eastern and central SA through the DJF season; the correlation tapers off during JFM and FMA but recovers in the MAM season. The reason for this behaviour may be related to the fact that the timing of peak seasonal precipitation differs from the timing of peak seasonal NDVI. By MAM (the harvest season), both SSM and NDVI reduce in similar manner, explaining the high correlation. FFV2 SSM also shows significant correlation values of more than 0.5R over considerable regions of eastern and central SA out to the third lead month (DJF), after which it tapers off. However, NDVI-based results show lower correlations than the SMAP-based results.

The areally-averaged correlations between SMAP/NDVI and RA, ESP, FFV1, and FFV2 SSM are provided in Table 2 [row4 (SMAP) and row5 (NDVI)]. As was found in the RA analysis, comparisons against the remotely sensed data show that FFV2 is more skilful than either ESP or FFV1. However, the skill of the forecasts relative to the observations is much smaller, as are the differences in skill between ESP, FFV1, and FFV2. Both SMAP- and NDVI-based forecast correlations show particularly lower correlation values in the fourth (JFM) and fifth (FMA) seasonal leads compared to the sixth seasonal lead (MAM). This dip in correlation of forecasted SSM can be attributed to the SSM forecasts' inability to correctly predict the magnitude of moisture at the peak period of the wet season (JFM, FMA) at longer leads, which produces lower SSM variability in forecasts than in the observations. When the seasonal SSM forecast lead is towards the end of the wet season (MAM), the variabilities are captured better, and therefore have higher correlation values than the previous two leads.

The deterministic analyses evaluated the ensemble mean of the dynamical and ESP forecasts, showing that the ensemble mean of FFV2 has higher skill and accuracy than the FFV1 and ESP ensemble mean forecasts at the seasonal scale over SA. However, these evaluations do not consider the probabilistic nature of the forecasts, which is the main reason for updating the version of FLDAS-Forecast with a much larger number of ensemble members. The next section of this study evaluates the probabilistic skill of these forecasts over SA.

4.3 Probabilistic Analysis of Droughts

The ensemble spreads of FFV2-based and FFV1-based RZSM forecasts through the monthly leads of each (SON) IC are illustrated in Figure 7. The September through November ICs' forecasts ensemble spread are stacked in the figure, and the FFV2 monthly RZSM ensemble forecasts show a larger spread by 0.02-0.1m^3/m^3 as compared to the FFV1-based forecasts for all ICs over large regions of SA, except over the very dry regions of Namibia, South Africa, and the Democratic Republic of Congo, where the difference in spread is less than 0.03m^3/m^3. The ensemble spread is seen to increase through the wet season, peaking in February and reducing thereafter for both the forecast version and the three initial conditions over SA.

The forecasts are now evaluated for extremes (droughts) by using the Ranked Probability Skill Score (RPSS). This metric describes the quality of categorical probabilistic forecasts. Figure 8 shows the RZSM forecast RPSS for conditions when RZSM is less than the 33rd percentile (categorized as "drought" hereafter). As shown in Figure 8, FFV2 (row 3) exhibits RPSS above 50% over a larger region of SA compared to the FFV1 and ESP forecasts (row 2 and row 1). We find both the dynamic models' forecast RPSS to be better than that of ESP, and we find that the FFV2 forecast RPSS is 10-20% better than that of FFV1 (row 4).

RPSS decreases with increasing lead for all three forecasts, but not as rapidly as it does for the deterministic skill, retaining significant RPSS of over 5% until the sixth seasonal forecast lead. These results are also confirmed by areally-averaged RPSS for the seasonal leads (Table 2, row 3). Additionally, more areas over SA show higher RPSS in FFV2 relative to FFV1 than found for the deterministic skill assessment.

The NMME forecasts are also known to have good skill in the sub-seasonal or monthly scale (Wanders and Wood 2016; Cash et al. 2019). The FFV2 RZSM RPSS values in the drought category are evaluated separately for each of the early season's monthly ICs over SA to further evaluate the sub-seasonal probabilistic skill in RZSM. Figure 9 shows monthly forecast RPSS through the sixth lead for the September to November ICs. The forecast RPSS values are progressively higher as the wet season advances from September to November IC. The October IC forecasts have an average of ~2% higher RPSS than the September IC forecasts. The November IC forecasts have on average another ~2% higher skill score than the October IC forecasts during the first three forecast leads. This increase in skill score can be attributed to the availability of more moisture in the initial condition with the advancing wet season. The September IC is relatively dry, and the dryness persists through the hydrologic forecasts. Hence, forecasts made with September ICs are more likely to have more false alarms for drought conditions than the forecasts initialized during later months that have higher soil moisture content, leading to lower RPSS for the September IC forecasts. To support this argument, FFV2 RZSM hit rate (HR) for drought conditions in Figure 10 shows FFV2 forecasts initialized in September have an average of ~2% lower HR than those initialized in October and November. These results are also reflected in the areally-averaged monthly FFV2 RPSS and HR values (Table 3, rows 1-3). Hence, it can be concluded that hydrologic forecasts made further into the early wet season have better drought-based RPSS and HR in the first three to four forecast lead months than those made "too" early in the season.

Figure 11 presents reliability diagrams for the ESP, FFV1, and FFV2 seasonal forecasts (SON ICs) over SA in the upper row. Reliability diagrams diagnose categorical forecast quality by plotting the observed frequency of a categorical event against the forecast probability of that event (in this case, drought). The highest reliability is shown by the forecast closest to the diagonal. FFV2 shows better drought forecast reliability than ESP and FFV1 through all the forecast seasonal leads, though the overall reliability decreases with increasing lead. The lower row in Figure 11 presents the Relative Operating Characteristics (ROC) curve for the same set of seasonal forecasts. ROCs, which are used to measure the ability of the forecasts to distinguish an event from a non-event, are shown here as plots of the probabilistic hit rate (HR) to false alarm rate (FAR) for a given category, taken here to be drought. At low leads, the HR for FFV2 in Figure 10 is initially much larger than the FAR, which brings the trajectory of the FFV2 curve closer to the top-left-hand corner. This trajectory indicates that FFV2 RZSM has a better ROC than ESP and FFV1, that is, a better ability to discriminate events from non-events. The ROC of all forecasts decreases with increasing lead. FFV2 still shows the best ROC across all leads, even at longer leads, showing FFV2 is better at discriminating events (drought condition) from non-events (non-drought conditions).

5 APPLICATION AND LIMITATIONS

The results thus far indicate that FFV2 forecasts can serve as a useful tool in the early reporting of extreme events. As an example, the near-real-time RZSM forecast percentiles with five monthly leads for October 2019 IC (Figure 12.a) are able to aptly forecast severe drought conditions over parts of the SA region, as corroborated by news reports in NASA's Earth Observatory (Carlowicz and Dauphin 2019), Relief-web (2019), and The New Humanitarian (Anyadike 2019). The FLDAS-Forecast-based RZSM percentiles use a drought severity category scale similar to that of the Climate Prediction Center (CPC) and United States Drought Monitor (USDM), which uses five classifications to characterize a soil moisture percentile in the context of drought: (i) abnormally dry (D0 21-30 percentile), showing areas that may be going into or are coming out of drought; (ii) moderate drought (D1 10-20 percentile); (iii) severe drought (D2 6-10 percentile); (iv) extreme drought (D3 3-5 percentile); and (v) exceptional drought (D4 0-2 percentile). In Figure 12.a, the November 2019 through March 2020 FFV2 RZSM forecasts of extreme to severe drought conditions in SA is found consistent with the crop monitor alert (GEOGLAM 2019) and other alerts, like the Joint Call For Action's World Food Programme (WFP 2019). Further, Figure 12.b provides the probability of below-normal, normal, and above-normal RZSM conditions over the region. The below-normal conditions highlight the likelihood of potentially severe drought-like conditions. The probability of below-normal conditions (i.e., drought) was found to have a probability of over 60% over parts of South Africa, Botswana, and Namibia.

Results presented in this study are limited by the initialization of a single season and a relatively small study domain. This work is intended to support remote early warning in the vulnerable regions of the world, where there are few in situ data sets to support this type of analysis. Other limitations include the absence of complementary meteorological fields for driving land surface and hydrology models, like wind, radiation, humidity and surface pressure, from the different NMME modeling centers' forecast data sets. This leads to the less optimal option of combining the NMME precipitation forecasts with the GMAO GEOS forecasts for the other meteorological fields, which may lead to potential inconsistencies, such as high solar radiation during times of high rainfall. We currently have no way to quantify the impact of this limitation. Further improvements to the system inputs could also be made by using higher resolution data sets such as the European Centre for Medium-Range Weather Forecasts' (ECMWF) reanalysis version 5 (ERA5) and the Integrated Multi-satellite Retrievals for Global

Precipitation Measurement (IMERG) or other relevant reanalyses and/or satellite-based products for the bias-correction and downscaling procedures.

6 CONCLUSIONS

This study evaluates the NMME-based FLDAS soil moisture forecasts relative to an observation-based forcing driven reanalysis (RA) through deterministic and probabilistic analyses. The entire NMME suite of precipitation forecasts is now used in FLDAS-Forecast, as it has higher skill than any of the individual forecast products in the suite. Because the original version of FLDAS-Forecast (FFV1) used only GEOS-based inputs, the updated FLDAS-Forecast (FFV2) is shown to perform better. Here, we demonstrate and quantify this improved performance by examining FFV1 and FFV2 forecasts over southern Africa, using the same simulation design as used for producing routine real-time hydrologic forecast products.

Results show that the FFV2-based RZSM forecasts have higher deterministic skill than FFV1. Both versions of FLDAS-Forecast are also found to have better deterministic skill and probabilistic accuracy than the ESP-based forecasts through the wet season over southern Africa, which is also the main growing season in that region. Additionally, the RA SSM shows excellent correlation with remotely sensed observations, like SMAP and NDVI, through the wet season over southern Africa. The NMME-based SSM forecasts also show good correlation with these observations until the third seasonal forecast lead, which includes the peak of the wet season; hence, FFV2 shows promise for forecasting the main part of this season well.

Further, FFV2-based RZSM forecasts yield higher ensemble spread through the wet season compared to FFV1-based forecasts. Probabilistically, the FFV2-based seasonal RZSM forecasts also have higher RPSS for the drought category. A monthly probabilistic analysis of the wet season showed that, for the FFV2-based forecasts, RPSS and HR for drought over southern Africa improves with later initializations during the wet season. In addition, the probabilistic drought study over the region shows that the FFV2-based RZSM forecasts have better reliability and ROC scores than either the FFV1 or ESP forecasts, highlighting an improved ability to discriminate events (droughts) from non-events. The probabilistic analyses also show that both versions of FLDAS-Forecast have higher forecast skill and quality than ESP-based forecasts through the wet season over southern Africa. In presenting a past operationally based case, near-real-time FFV2-based RZSM percentiles for October 2019 ICs successfully predicted drought-like conditions over southern Africa. Having a larger number of ensemble members, we could provide highly reliable probabilities of drought conditions through the forecast lead months.

This work shows that updating FLDAS-Forecast with the entire NMME S2S suite allows for more skilful hydrologic forecasts. Future work will aim to integrate NMME S2S-based surface temperature along with precipitation into the system and will expand the domain to cover the globe. The setup may also expand to improve initial conditions by including multivariate data assimilation, ingesting, for example, assimilating Leaf Area Index (LAI; Kumar et al. 2019) and soil moisture information (Draper and Reichle 2019; Rahman et al. 2022; Sabater et al. 2008). The near-real-time products have been very useful in providing potential drought forecasts. These drought-based products will continue to provide support for USAID's FEWS NET, as well as for other ongoing early warning systems.

Acknowledgements:

- This work was supported by the NASA Center for Climate Simulation (NCCS) computing resources, NASA ROSES, USAID-NASA PAPA and NASA Harvest. We would like to thank the Climate Hazard Center's Technical Writer & Editor Juliet Way-

Henthorne for providing professional editing. Her editing contributions are supported by USAID (Award # 72DFFP19CA00001).

REFERENCES

Anyadike, O, 2019. Drought in Africa leaves 45 million in need across 14 countries [WWW Document]. URL https://www.thenewhumanitarian.org/analysis/2019/06/10/drought-africa-2019-45-million-in-need

Arsenault, K.R., S. Shukla, A. Hazra, A. Getirana, A. McNally, S. Kumar, R.D. Koster, C.D. Peters - Lidard, B.F. Zaitchik, H. Badr, H.C. Jung, B. Narapusetty, M. Navari, S. Wang, D.M. Mocko, C. Funk, L. Harrison, G.J. Husak, A. Adoum, G. Galu, T. Magadzire, J. Roningen, M. Shaw, J. Eylander, K. Bergaoui, R.A. McDonnell, J.P Verdin, et al., 2020, The NASA hydrological forecast system for food and water security applications. Bull. Am. Meteorol. Soc.. 10.1175/BAMS-D-18-0264.1

Becker, E., den Dool, H. v., & Zhang, Q., 2014. Predictability and Forecast Skill in NMME. Journal of Climate, 27(15), 5891-5906. Retrieved May 11, 2021, from https://journals.ametsoc.org/view/journals/clim/27/15/jcli-d-13-00597.1.xml

Borovikov, A., R. Cullather, R. Kovach, J. Marshak, G. Vernieres, Y. Vikhliaev, B. Zhao, and Z. Li, 2019. GEOS-5 seasonal forecast system. Climate Dyn., 53, 7335–7361, https://doi.org/10.1007/S00382-017-3835-2

Bosilovich, M. G., R. Lucchesi, and M. Suarez, 2016. MERRA-2: File specification. GMAO Office Note 9 (version 1.1). 73 pp., https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich785.pdf

Buchwitz, M., Reuter, M., Schneising-Weigel, O., et al., 2017. Product quality assessment report (PQAR)—main document. Technical report Copernicus Climate Change Service (C3S), available from C3S website https://climate.copernicus.eu/. 20 Oct 2017, p 103

Carlowicz, M, and Dauphin, L., 2019. Drought Threatens Millions in Southern Africa [WWW Document]. URL https://earthobservatory.nasa.gov/images/146015/drought-threatens-millions-in-southern-africa

Cash, B.A., Manganello, J.V. and Kinter, J. L., 2019. Evaluation of NMME temperature and precipitation bias and forecast skill for South Asia. Clim Dyn 53, 7363–7380 (2019). https://doi.org/10.1007/s00382-017-3841-4

Cui, B., Toth, Z., Zhu, Y., & Hou, D., 2012. Bias Correction for Global Ensemble Forecast, Weather and Forecasting. 27(2), 396-410. https://journals.ametsoc.org/view/journals/wefo/27/2/waf-d-11-00011_1.xml

Day, G., 1985. Extended streamflow forecasting using NWSRFS. J. Water Resour. Plann.Manage., 111,157–170,https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157)

Felice, M. D., Alessandri, A., Catalano, F., 2015. Seasonal climate forecasts for medium-term electricity demand forecasting. Applied Energy, Volume 137, Pages 435-444, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2014.10.030

Dinku, T., C. Funk, P. Peterson, R. Maidment, T. Tadesse, H. Gadain, and P. Ceccato, 2018. Validation of the CHIRPS satellite rainfall estimates over eastern Africa. Quart. J. Roy. Meteor. Soc., 144, 292–312, https://doi.org/10.1002/qj.3244

Draper, C., and Reichle, R. H., 2019. Assimilation of Satellite Soil Moisture for Improved Atmospheric Reanalyses. Monthly Weather Review, 147(6), 2163-2188. Retrieved Nov 24, 2021, from https://journals.ametsoc.org/view/journals/mwre/147/6/mwr-d-18-0393.1.xml

Emerton, R., and Coauthors, 2018. Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-seasonal v1.0. Geosci. Model Dev., 11, 3327–3346, https://doi.org/10.5194/gmd-11-3327-2018

Ehrendorfer, M., 1997. Predicting the uncertainty of numerical weather forecasts: A review. Meteor. Z., 6, 147–183.

Entekhabi, D., N. Das, E. G. Njoku, J. T. Johnson, and J. Shi., 2016. SMAP L3 Radar/Radiometer Global Daily 9 km EASE-Grid Soil Moisture, Version 3. [Indicate subset used]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/7KKNQ5UURM2W

Flores Cordova, A. I., E. R. Anderson, D. Irwin, and E. A. Cherrington, 2012. Contributions of SERVIR in promoting the use of space data in climate change and disaster management. Proc. 63rd Int. Astronautical Congress, Naples, Italy, International Astronautical Federation, IAC-12-E3.2.3 x14456, https://iafastro.directory/iac/archive/browse/IAC-12/E3/2/14456/.

Fortin, V., Abaza, M., Anctil, F., & Turcotte, R., 2014. Why Should Ensemble Spread Match the RMSE of the Ensemble Mean? Journal of Hydrometeorology, 15(4), 1708-1713. Retrieved Jul 8, 2021, from https://journals.ametsoc.org/view/journals/hydr/15/4/jhm-d-14-0008_1.xml

Funk, C., and Coauthors, 2015. The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. Sci. Data, 2, 150066, https://doi.org/10.1038/sdata.2015.66

— , and Coauthors, 2019. Recognizing the Famine Early Warning Systems Network (FEWS NET): Over 30 years of drought early warning science ad- vances and partnerships promoting global food security. Bull. Amer. Meteor. Soc., 100, 1011–1027, https://doi.org/10.1175/BAMS-D-17-0233.1

Gelaro, R., and Coauthors, 2017. The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). J. Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z., & Zhang, M., 2011. The Community Climate System Model Version 4, Journal of Climate, 24(19),4973-4991. https://journals.ametsoc.org/view/journals/clim/24/19/2011jcli4083.1.xml

GEOGLAM, 2019. Special Report on southern Africa [WWW Document]. URL https://www.preventionweb.net/files/69061_69061specialreportsouthernafrica201.pdf

Getirana, A. Boone, D. Yamazaki, B. Decharme, F. Papa, and N. Mognard, 2012. The Hydrological Modeling and Analysis Platform (HyMAP): Evaluation in the Amazon basin. J. Hydrometeor., 13, 1641–1665, https://doi.org/10.1175/JHM-D-12-021.1

— , C. Peters-Lidard, M. Rodell, and P. D. Bates, 2017a. Trade-off between cost and accuracy in large-scale surface water dynamic modelling. Water Resour. Res., 53, 4942–4955, https://doi.org/10.1002/2017WR020519

Goddard, Lisa and M. Dilley, 2005. El Niño: Catastrophe or Opportunity. J. Climate, 18, 651–665.

Grimit, E. P., & Mass, C. F., 2007. Measuring the Ensemble Spread–Error Relationship with a Probabilistic Approach: Stochastic Ensemble Results. Monthly Weather Review, 135(1), 203-221, https://journals.ametsoc.org/view/journals/mwre/135/1/mwr3262.1.xml

Gupta, A. S., Jourdain, N. C., Brown, J. N., & Monselesan, D., 2013. Climate Drift in the CMIP5 Models. Journal of Climate, 26(21), 8597-8615. Retrieved Nov 24, 2021, from https://journals.ametsoc.org/view/journals/clim/26/21/jcli-d-12-00521.1.xml

Hermanson, L., Ren, HL., Vellinga, M. et al., 2018. Different types of drifts in two seasonal forecast systems and their dependence on ENSO. Clim Dyn 51, 1411–1426 (2018). https://doi.org/10.1007/s00382-017-3962-9

Hopson, T. M., 2014. Assessing the Ensemble Spread–Error Relationship, Monthly Weather Review, 142(3), 1125-1142. https://journals.ametsoc.org/view/journals/mwre/142/3/mwr-d-12-00111.1.xml

Jung, H. C., A. Getirana, F. Policelli, A. McNally, K. R. Arsenault, S. Kumar, T. Tadesse, and C. D. Peters-Lidard, 2017. Upper Blue Nile basin water budget from a multi-model perspective. J. Hydrol., 555, 535–546, https://doi.org/10.1016/j.jhydrol.2017.10.040

Kirtman, B., and Coauthors, 2014. The North American Multi-Model Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. Bull. Amer. Meteor. Soc., 95, 585–601, https://doi.org/10.1175/BAMS-D-12-00050.1

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000. A catch-ment based approach to modeling land surface processes in a general circula- tion model: 1. Model structure. J. Geophys. Res., 105, 24 809–24 822, https://doi.org/10.1029/2000JD900327

Krakauer, N.Y., 2019. Temperature trends and prediction skill in NMME seasonal forecasts. Clim Dyn 53, 7201–7213 (2019). https://doi.org/10.1007/s00382-017-3657-2

Kumar, S., and Coauthors, 2006. Land information system: An interoperable framework for high resolution land surface modeling. Environ. Modell. Software, 21, 1402–1415, https://doi.org/10.1016/j.envsoft.2005.07.004

— , C. D. Peters-Lidard, J. Santanello, K. Harrison, Y. Liu, and M. Shaw, 2012. Land surface Verification Toolkit (LVT) - A generalized framework for land surface model evaluation. Geosci. Model Dev., 5, 869–886, https://doi.org/10.5194/gmd-5-869-2012

— , M. Mocko, D., Wang, S., Peters-Lidard, C. D., & Borak, J., 2019. Assimilation of Remotely Sensed Leaf Area Index into the Noah-MP Land Surface Model: Impacts on Water and Carbon Fluxes and States over the Continental United States, Journal of Hydrometeorology, 20(7), 1359-1377. Retrieved Feb 5, 2021, from https://journals.ametsoc.org/view/journals/hydr/20/7/jhm-d-18-0237_1.xml

Laux P, Jäckel G, Tingem RM, Kunstmann H., 2009. Onset of the rainy season and crop yield in sub-Saharan Africa—tools and perspectives for Cameroon. In: Ecohydrology of surface and groundwater dependent systems: concepts, methods and recent developments, vol 328. Joint IAHS & IAH Convention, Hyderabad

Li, H., L. Luo, E. F. Wood, and J. Schaake, 2009. The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. J. Geophys. Res., 114, D04114, https://doi.org/10.1029/2008JD010969

Lin, H., Merryfield, W. J., Muncaster, R., Smith, G. C., Markovic, M., Dupont, F., Roy, F., Lemieux, J., Dirkson, A., Kharin, V. V., Lee, W., Charron, M., & Erfani, A., 2020. The Canadian Seasonal to Interannual Prediction System Version 2 (CanSIPSv2), Weather and Forecasting,35(4),1317-1343. https://journals.ametsoc.org/view/journals/wefo/35/4/wafD190259.xml

McNally, A., and Coauthors, 2017. A land data assimilation system for sub-Saharan Africa food and water security applications. Sci. Data, 4, 170 012, https://doi.org/10.1038/sdata.2017.12

Merryfield, W. J., and Coauthors, 2013. The Canadian seasonal to interannual prediction system. Part I: Models and initialization. Mon. Wea. Rev., 141, 2910–2945, https://doi.org/10.1175/MWR-D-12-00216.1

Mo, K. C., & Lyon, B., 2015. Global Meteorological Drought Prediction Using the North American Multi-Model Ensemble. Journal of Hydrometeorology, 16(3), 1409-1424. https://journals.ametsoc.org/view/journals/hydr/16/3/jhm-d-14-0192_1.xml

Molod, A., L. Takacs, M. Suarez, J. Bacmeister, I.-S. Song, and A. Eichmann, 2012. The GEOS-5 Atmospheric General Circulation Model: Mean Climate and Development from MERRA to Fortuna. Technical Report Series on Global Modeling and Data Assimilation, 28.

Molod, A., Hackert, E., Vikhliaev, Y., Zhao, B., Barahona, D., Vernieres, G., et al., 2020. GEOS- S2S Version 2: The GMAO high- resolution coupled model and assimilation system for seasonal prediction. Journal of Geophysical Research: Atmospheres, 125(5). https://doi.org/10.1029/2019JD031767

Mubaya, Chipo Plaxedes, Jemimah Njuki, Eness Paidamoyo Mutsvangwa, Francis Temba Mugabe, Durton Nanja, 2012. Climate variability and change or multiple stressors? Farmer perceptions regarding threats to livelihoods in Zimbabwe and Zambia, Journal of Environmental Management, Volume 102, Pages 9-17, ISSN 0301-4797, https://doi.org/10.1016/j.jenvman.2012.02.005

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, Journal of Climate, 18(10), 1513-1523. https://journals.ametsoc.org/view/journals/clim/18/10/jcli3361.1.xml

Nash, D.J., Klein, J., Endfield, G.H. et al., 2019. Narratives of nineteenth century drought in southern Africa in different historical source types. Climatic Change 152, 467–485. https://doi.org/10.1007/s10584-018-2352-6

Niu, G.-Y., and Coauthors, 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evalua- tion with local-scale measurements. J. Geophys. Res., 116, D12109, https://doi.org/10.1029/2010JD015139

Paiva, R. C. D., Collischonn, W., Bonnet, M. P., and de Gonçalves, L. G. G., 2012. On the sources of hydrological prediction uncertainty in the Amazon, Hydrol. Earth Syst. Sci., 16, 3127–3137, https://doi.org/10.5194/hess-16-3127-2012, 2012.

Peters-Lidard, C. D., and Coauthors, 2007. High-performance Earth system model- ing with NASA/GSFC's land information system. Innov. Syst. Software Eng., 3, 157–165, https://doi.org/10.1007/s11334-007-0028-x

Rahman, A., Maggioni, V., Zhang, X., Houser, P., Sauer, T., and Mocko, D. M., 2022. The Joint Assimilation of Remotely Sensed Leaf Area Index and Surface Soil Moisture into a Land Surface Model. *Remote Sensing*, *14*(3), 437.

Relief-web, 2019. Sothern Africa: Drought – 2018-201 [WWW Document]. URL https://reliefweb.int/disaster/dr-2018-000429-zwe#overview

Rienecker, M.M., M.J. Suarez, R. Todling, J. Bacmeister, L. Takacs, H.-C. Liu, W. Gu, M. Sienkiewicz, R.D. Koster, R. Gelaro, I. Stajner, and J.E. Nielsen, 2008. The GEOS-5 Data Assimilation System - Documentation of Versions 5.0.1, 5.1.0, and 5.2.0. Technical Report Series on Global Modeling and Data Assimilation, 27.

Sabater, J. M., Rüdiger, C., Calvet, J. C., Fritz, N., Jarlan, L., and Kerr, Y., 2008. Joint assimilation of surface soil moisture and LAI observations into a land surface model. *Agricultural and forest meteorology*, *148*(8-9), 1362-1373.

Saha, S. and Coauthors, 2010. The NCEP climate forecast system reanalysis. Bull. Amer. Meteor. Soc., 91, 1015-1057.

Saha, S., and Coauthors, 2014. The NCEP Climate Forecast System version 2. J. Climate, 27, 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1

Schaeybroeck, V. and Bert; Vannitsem, Stéphane, 2014. Ensemble post- processing using member- by- member approaches: theoretical aspects. , Quarterly Journal of the Royal Meteorological Society, Vol. 141, Issue 688, 807, Wiley Online Library.

Seibert, M., B. Merz, and H. Apel, 2017. Seasonal forecasting of hydrological drought in the Limpopo Basin: A comparison of statistical methods. Hydrol. Earth Syst. Sci., 21, 1611–1629, https://doi.org/10.5194/hess-21-1611-2017

Setiawan, A. M., Y Koesmaryono, A Faqih and D Gunawan, 2017. IOP Conf. Ser.: Earth Environ. Sci. 58 012035

Sheffield, J., and Coauthors, 2014. A drought monitoring and forecasting system for sub-Sahara African water resources and food security. Bull. Amer. Meteor. Soc., 95, 861–882, https://doi.org/10.1175/BAMS-D-12-00124.1

Shukla, S., A. McNally, G. Husak, and C. Funk, 2014. A seasonal agricultural drought forecast system for food-insecure regions of East Africa. Hydrol. Earth Syst. Sci., 18, 3907–3921, https://doi.org/10.5194/hess-18-3907-2014

— , Roberts, J., Hoell, A. et al., 2016. Assessing North American multimodel ensemble (NMME) seasonal forecast skill to assist in the early warning of anomalous hydrometeorological events over East Africa. Clim Dyn 53, 7411–7427 (2019). https://doi.org/10.1007/s00382-016-3296-z

— , and Coauthors, 2020. Improving early warning of drought-driven food insecurity in southern Africa using operational hydrological monitoring and forecasting products. Nat. Hazards Earth Syst. Sci., 20, 1187–1201, https://doi.org/10.5194/nhess-20-1187-2020

Spruce, J.P., G.E. Gasser, and W.W. Hargrove, 2016. MODIS NDVI Data, Smoothed and Gap-filled, for the Conterminous US: 2000-2015. ORNL DAAC, Oak Ridge, Tennessee, USA. http://dx.doi.org/10.3334/ORNLDAAC/1299

Sultan, B., B. Barbier, J. Fortilus, S. M. Mbaye, and G. Leclerc, 2010. Estimating the potential economic value of seasonal forecasts in West Africa: A long- term ex-ante assessment in Senegal. Wea. Climate Soc., 2, 69–87, https://doi.org/10.1175/2009WCAS1022.1

XThiaw, W. M., and V. B. Kumar, 2015. NOAA's African desk: Twenty years of developing capacity in weather and climate forecasting in Africa. Bull. Amer. Meteor. Soc., 96, 737–753, https://doi.org/10.1175/BAMS-D-13-00274.1

Trambauer, P., M. Werner, H. C. Winsemius, S. Maskey, E. Dutra, and S. Uhlenbrook, 2015. Hydrological drought forecasting and skill assessment for the Limpopo River basin, southern Africa. Hydrol. Earth Syst. Sci., 19, 1695–1711, https://doi.org/10.5194/hess-19-1695-2015

Twedt, T. M., J. C. Schaake Jr., and E. L. Peck, 1977. National Weather Service extended streamflow prediction. Proc. 45th Annual Western Snow Conference, Albuquerque, NM, Western Snow Conference, 52–57.

United Nations, 2018. The State of Food Security and Nutrition in the World 2018: Building Climate Resilience for Food Security and Nutrition. FAO, 181 pp.

Vecchi, G. A., Delworth, T., Gudgel, R., Kapnick, S., Rosati, A., Wittenberg, A. T., Zeng, F., Anderson, W., Balaji, V., Dixon, K., Jia, L., Kim, H.-S., Krishnamurthy, L., Msadek, R., Stern, W. F., Underwood, S. D., Villarini, G., Yang, X., and Zhang, S., 2014. On the Seasonal Forecasting of Regional Tropical Cyclone Activity. Journal of Climate, 27(21), 7994-8016. https://journals.ametsoc.org/view/journals/clim/27/21/jcli-d-14-00158.1.xml

Verdin, J., C. Funk, G. Senay, and R. Choularton, 2005. Climate science and famine early warning. Philos. Trans. Roy. Soc. London, 360B, 2155–2168, https://doi.org/10.1098/rstb.2005.1754

Viel, C., Beaulant, A.-L., Soubeyroux, J.-M., and Céron, J.-P., 2016. How seasonal forecast could help a decision maker: an example of climate service for water resource management, Adv. Sci. Res., 13, 51–55, doi: 10.5194/asr-13-51-2016, 2016.

Vörösmarty, C. J., E. M. Douglas, P. A. Green, and C. Revenga, 2005. Geospatial indicators of emerging water stress: An application to Africa. AMBIO: J. Hum. Environ., 34, 230–236, https://doi.org/10.1579/0044-7447-34.3.230

Wanders, N., & Wood, E. F., 2016. Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations. Environmental Research Letters, 11(9), 094007. https://doi.org/10.1088/1748-9326/11/9/094007

Wanders, N., S. Thober, R. Kumar, M. Pan, J. Sheffield, L. Samaniego, and E. F. Wood, 2019. Development and evaluation of a pan-European multimodel seasonal hydrological forecasting system. J. Hydrometeor., 20, 99–115, https://doi.org/10.1175/JHM-D-18-0040.1

Wang, A., T. J. Bohn, S. P. Mahanama, R. D. Koster, and D. P. Lettenmaier, 2009. Multimodel ensemble reconstruction of drought over the continental United States. J. Climate, 22, 2694–2712, https://doi.org/10.1175/2008JCLI2586.1

Wolski, P., C. Lennard, C. Jack, and M. Tadross, 2017. Seasonal hydrological forecasting—Current state of play. Water Wheel, 16, 30–33.

Wood, A. W., L. R. Leung, V. Sridhar, and D. P. Lettenmaier, 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Climatic Change, 62, 189–216, https://doi.org/10.1023/B:CLIM.0000013685.99609.9e

World Food Programme, 2019. Joint Call for Action to Address the Impacts of Climate Change and a Deepening Humanitarian Crisis in Southern Africa [WWW Document]. URL https://reliefweb.int/sites/reliefweb.int/files/resources/Joint%20Call%20for%20Action%20to%20Address%20the%20Impacts%20of%20Climate%20Change%20and%20a%20Deepening%20Humanitarian%20Crisis%20in%20Southern%20Africa.pdf

Yossef, N. C., R. van Beek, A. Weerts, H. Winsemius, and M. F. P. Bierkens, 2017. Skill of a global forecasting system in seasonal ensemble streamflow prediction. Hydrol. Earth Syst. Sci., 21, 4103–4114, https://doi.org/10.5194/hess-21-4103-2017

Yuan, X., E. Wood, N. Chaney, J. Sheffield, J. Kam, M. Liang, and K. Guan, 2013. Probabilistic seasonal forecasting of African drought by dynamical models. J. Hydrometeor., 14, 1706–1720, https://doi.org/10.1175/JHM-D-13-054.1

— , J. K. Roundy, E. F. Wood, and J. Sheffield, 2015. Seasonal forecasting of global hydrologic extremes: System development and evaluation over GEWEX basins. Bull. Amer. Meteor. Soc., 96, 1895–1912, https://doi.org/10.1175/BAMS-D-14-00003.1

FIGURES

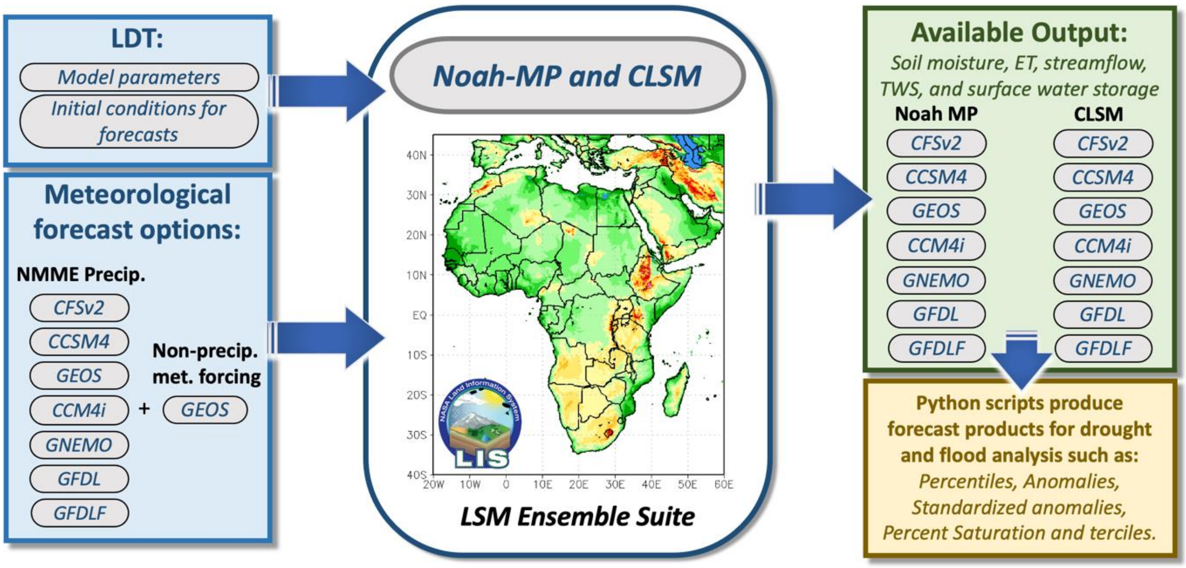


Figure 1. Schematic of the FLDAS-Forecast version 2 (FFV2).

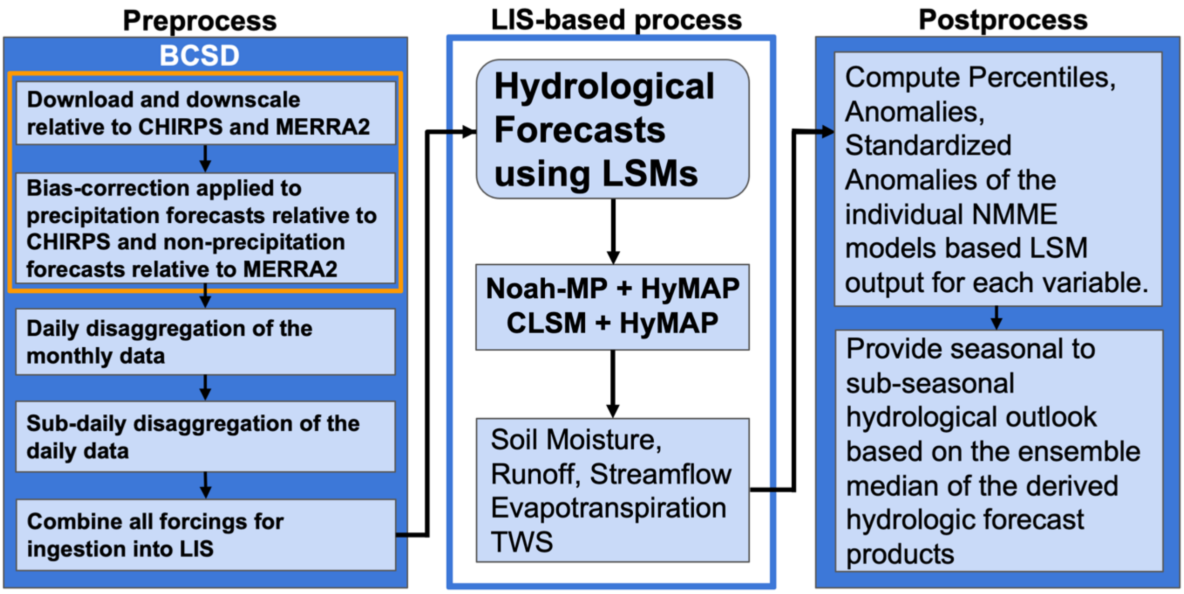


Figure 2. FFV2 workflow includes pre-processing, LIS-based processing and post-processing. The orange box highlights the bias-correction and spatial downscaling (BCSD) steps.

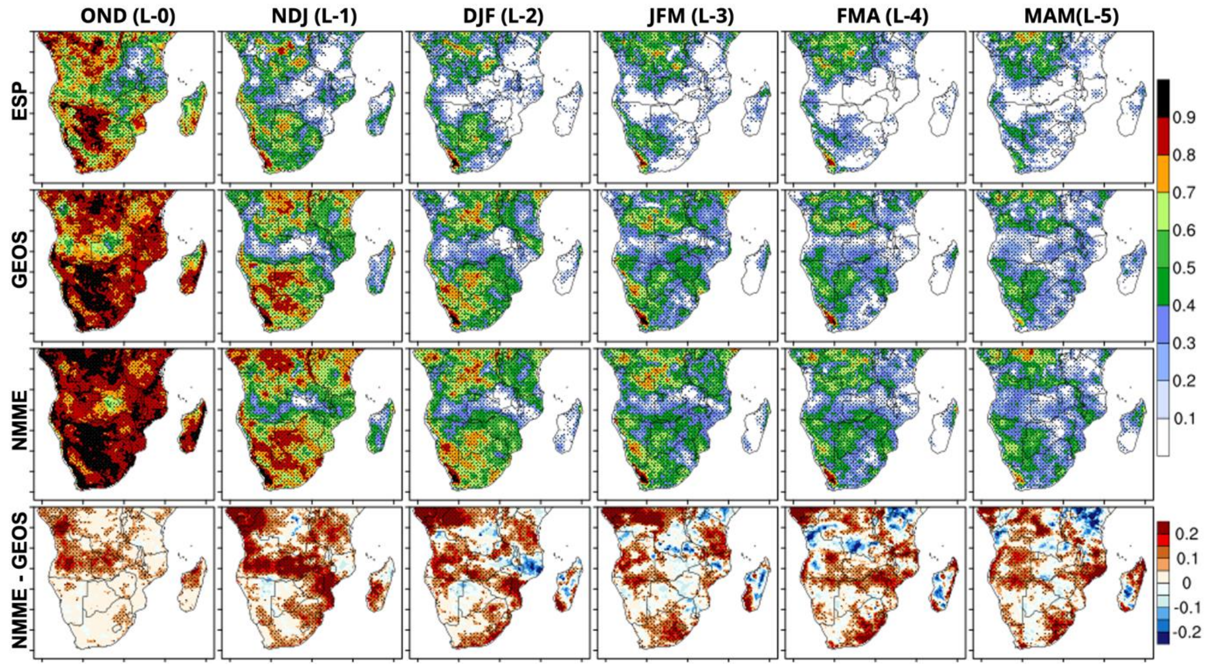


Figure 3: Skill (anomaly correlation) between RA RZSM and ESP (row 1), GEOS/FFV1 (row 2) and NMME/FFV2 (row 3) based seasonal RZSM forecast percentiles between 1982-2010 for September, October and November ICs. Row 4 is the difference in skill between NMME/FFV2-based and GEOS/FFV1-based RZSM forecast percentiles. Correlation coefficient values less than 0.1 are in white and significant values are stippled.

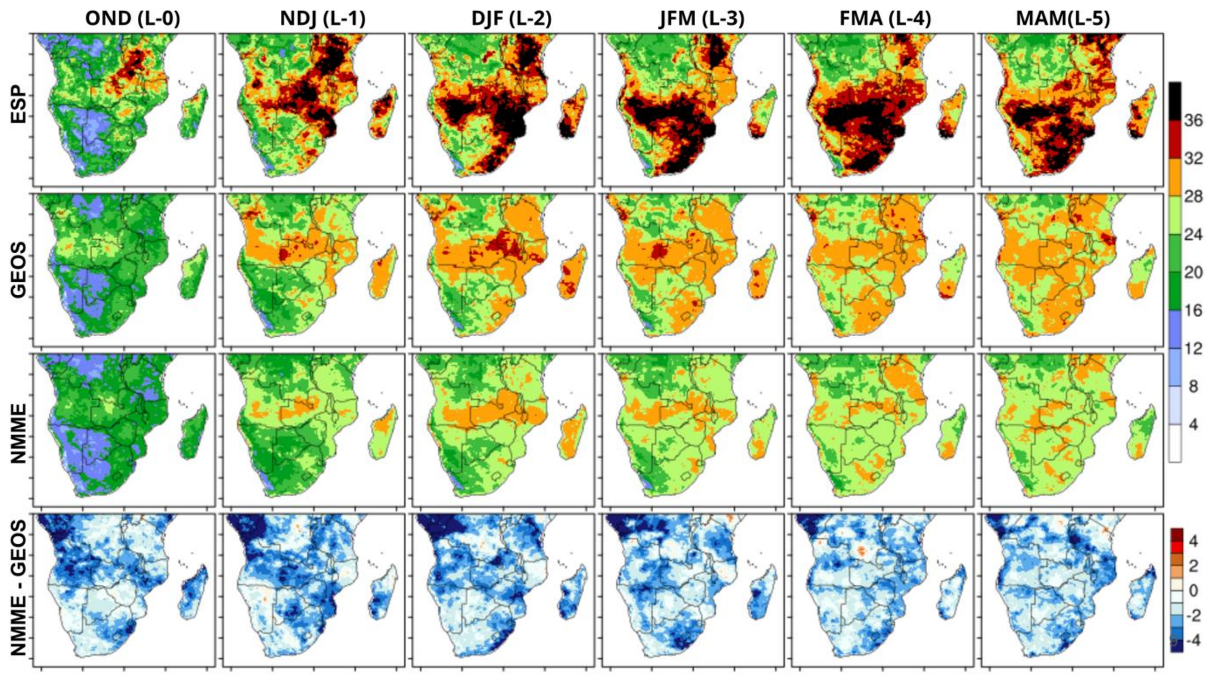


Figure 4: Root mean square error (RMSE) between RA RZSM and ESP (row 1), GEOS/FFV1 (row 2), and NMME/FFV2 (row 3) based seasonal RZSM forecast percentiles between 1982-2010 for September, October, and November ICs. Row 4 is the difference in RMSE between NMME/FFV2-based and GEOS/FFV1-based RZSM forecast percentiles. RMSE values less than 4 percentile are in white.

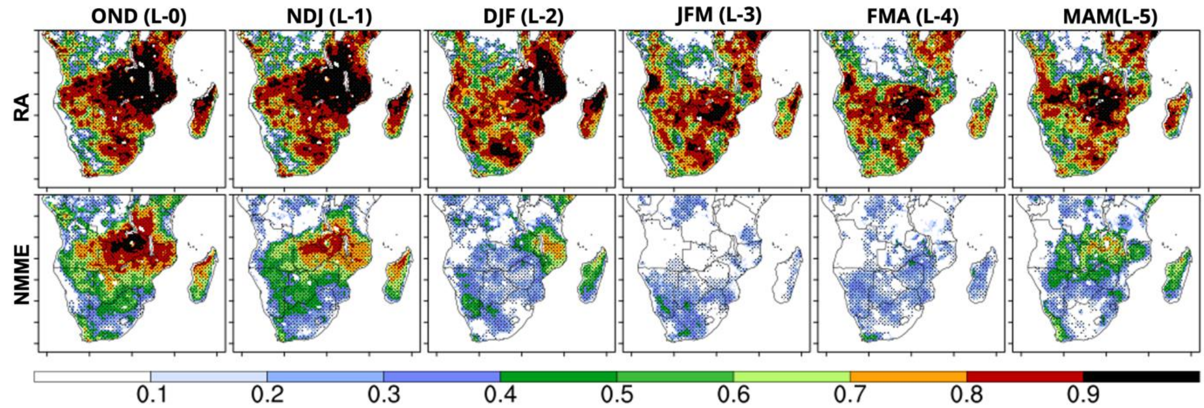


Figure 5: Correlation between SMAP and RA SSM (row 1) and SMAP and NMME SSM forecasts (row 2) between 2015-2020 for September, October and November ICs. Correlation coefficients less than 0.1 are in white and significant correlation coefficients are stippled.

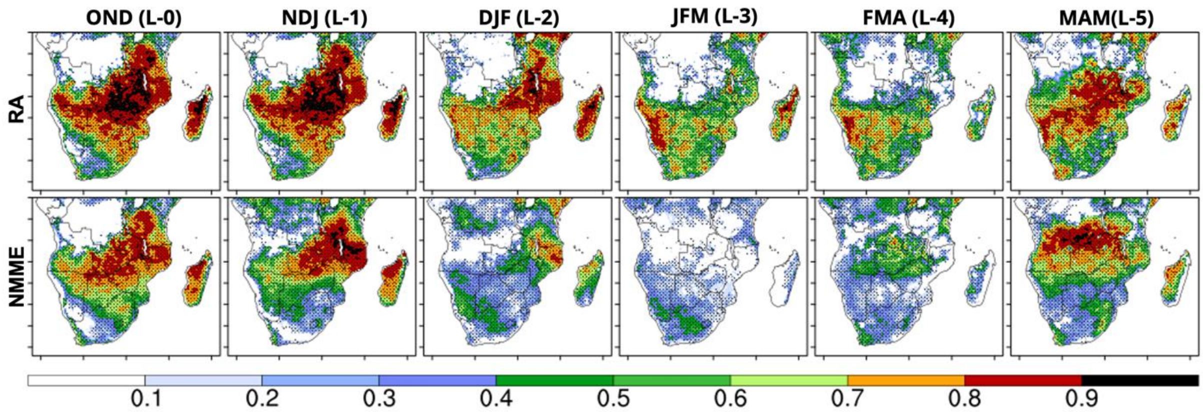


Figure 6: Correlation between GIMMS NDVI and RA (row1) and GIMMS NDVI and NMME based SSM forecasts (row2) between 2010-2018 for September, October and November ICs. Correlation coefficients less than 0.1 are in white and significant correlation coefficients are stippled.

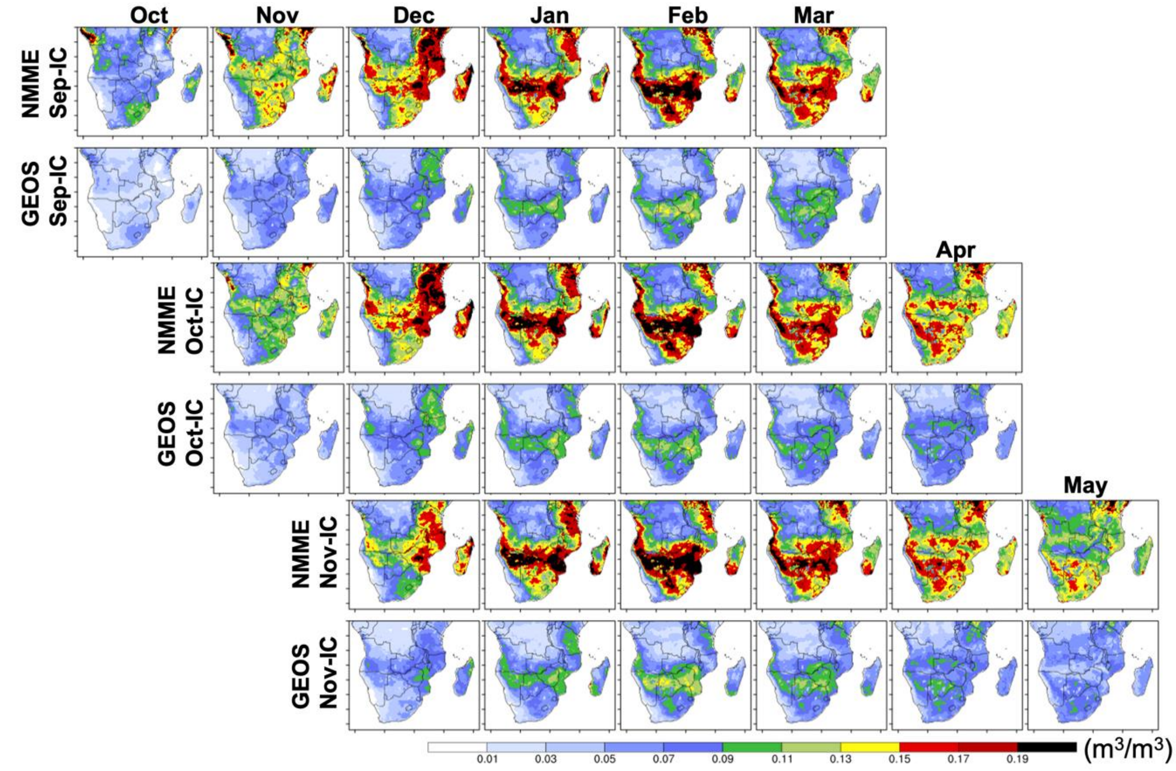


Figure 7: Ensemble spread [(Ens_{max}-Ens_{min}) of FFV2 and FFV1 September (rows 1, 2), October (rows 3, 4) and November (rows 5, 6) ICs based RZSM forecasts.

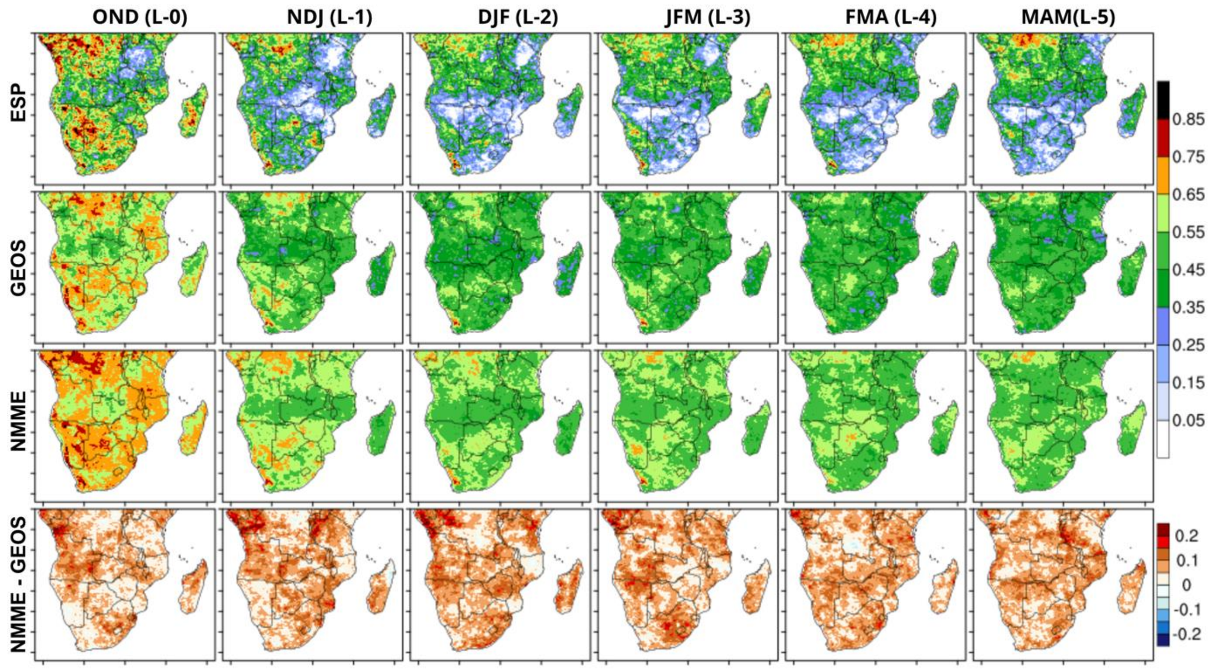


Figure 8: Seasonal RPSS between RA RZSM and ESP (row 1), GEOS/FFV1 (row 2) and NMME/FFV2 (row 3) based seasonal RZSM forecast percentiles between 1982-2010 September, October and November ICs, for categorically drought conditions (<33percentile). Row 4 is the difference between NMME/FFV2 and GEOS/FFV1 RPSS. RPSS less than 0.05 are in white, which is also the cut-off. The RPSS cut-off values are small enough to be considered insignificant.

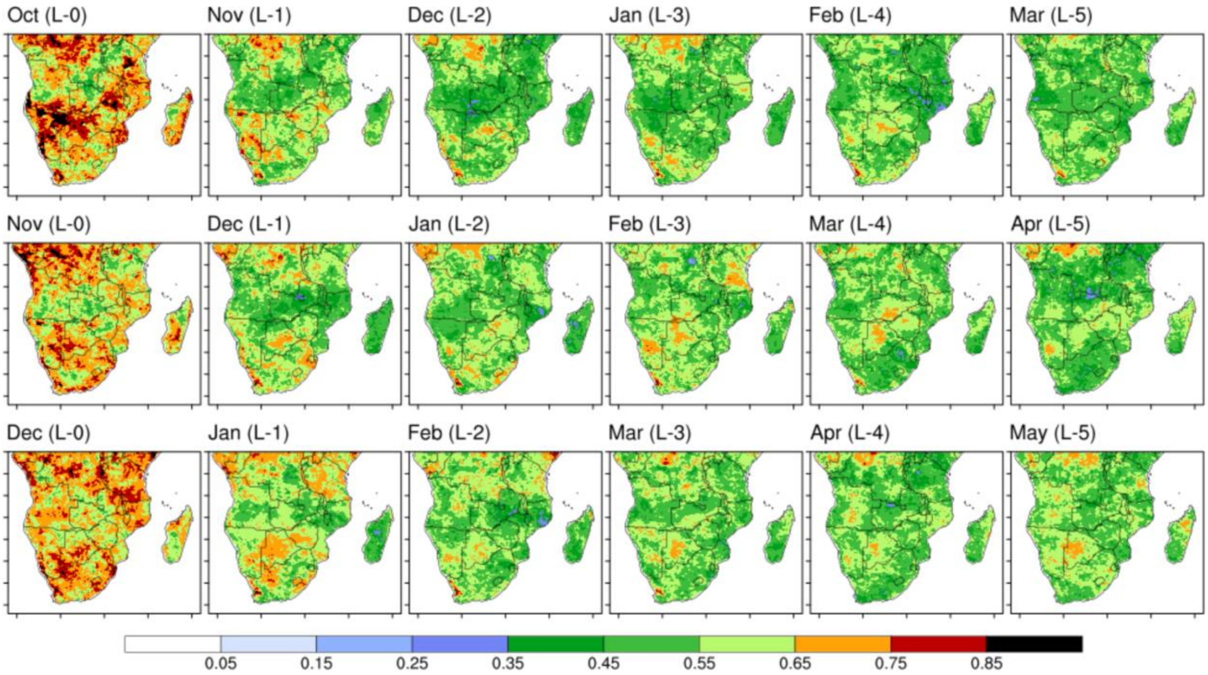


Figure 9: Monthly RPSS (for drought) of FFV2 RZSM. Row 1: September IC forecasts; Row 2: October IC forecasts; Row 3: November IC forecasts.

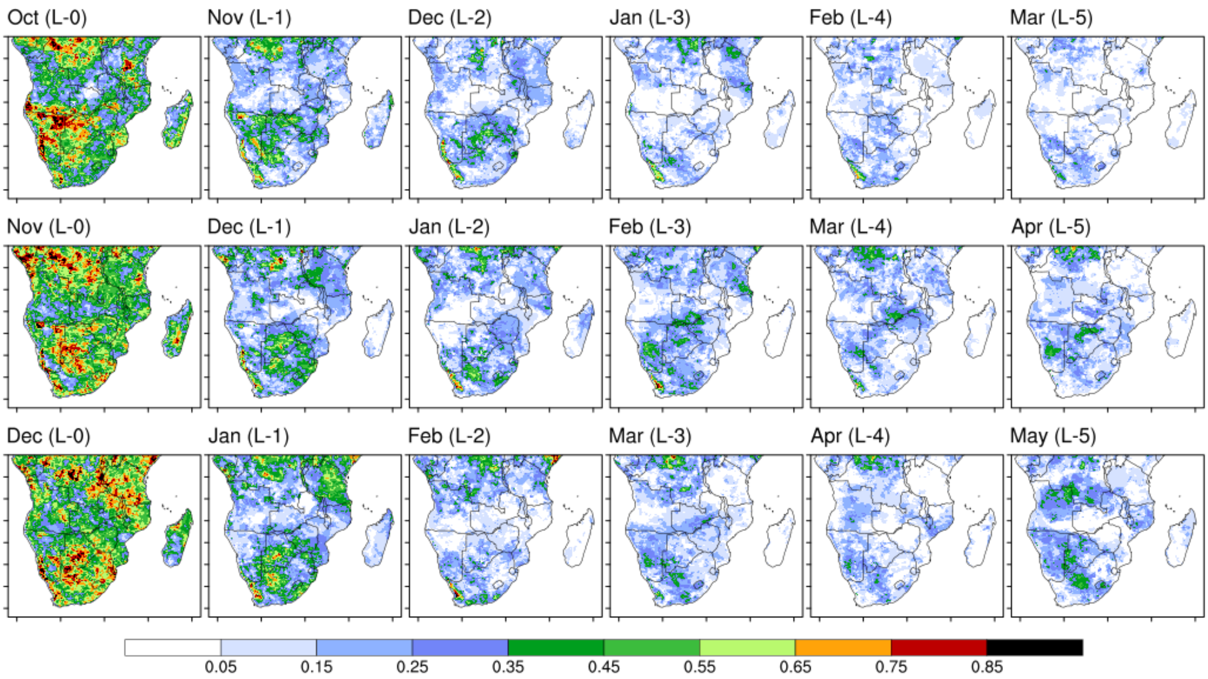


Figure 10: Monthly HR (for drought) of FFV2 RZSM. Row 1: September IC forecasts; Row 2: October IC forecasts; Row 3: November IC forecasts.

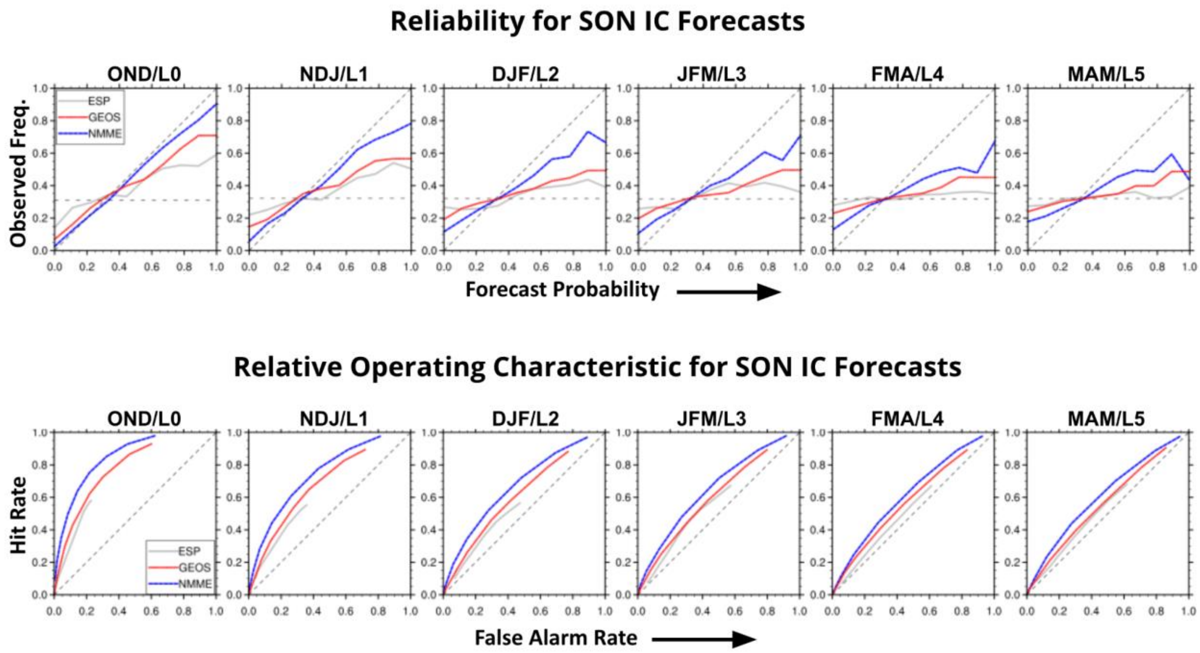


Figure 11: Reliability Diagram (row 1) and ROC (row 2) of ESP (grey), GEOS-only (red) and NMME (blue) based forecasts for SON ICs for drought.

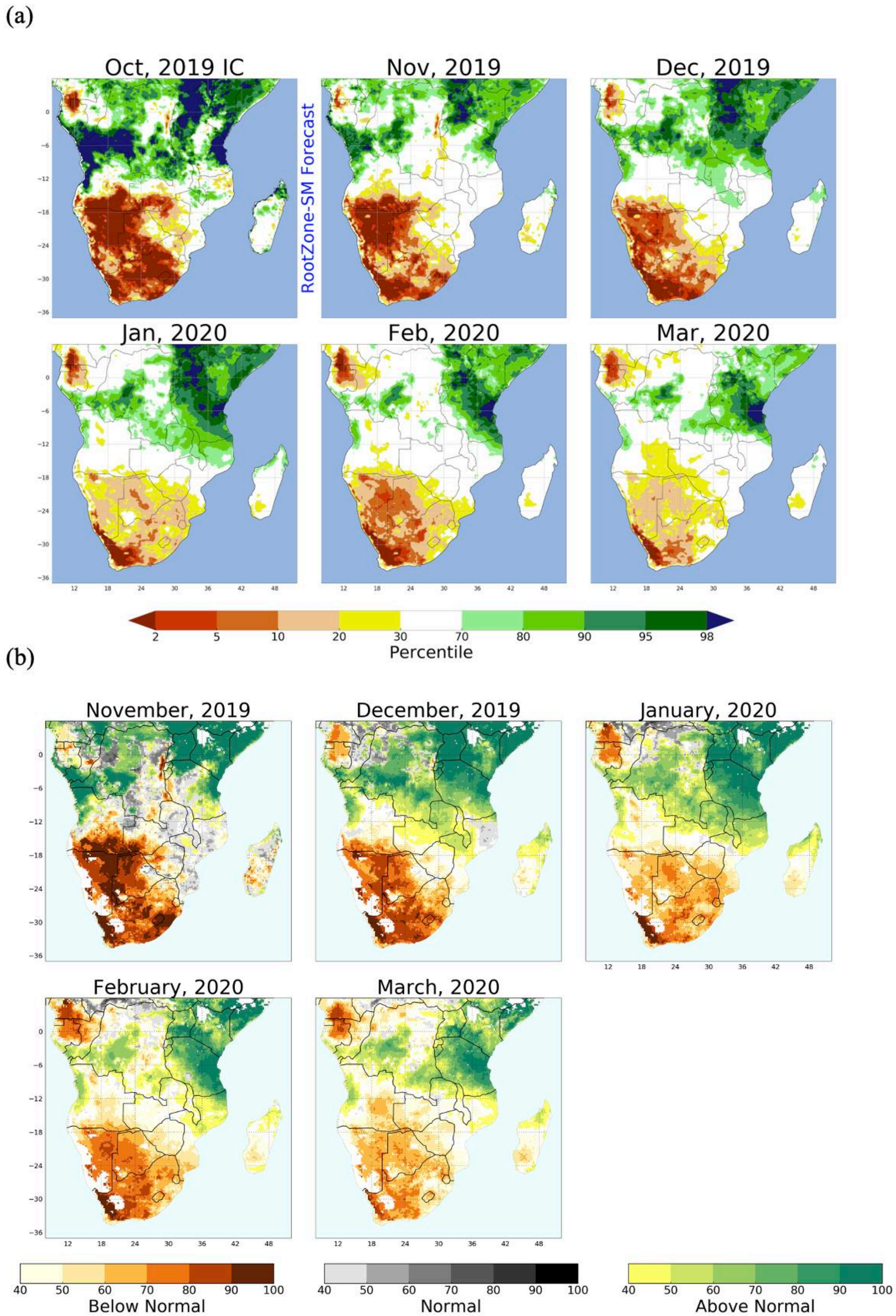


Figure 12: October 2019 IC, FFV2 RZSM forecast (a) percentiles, and (b) forecast probabilities, based on 1982-2010 climatology.

3 TABLES

Table 1. NMME Dataset, numbers in brackets are the number of ensemble members used in the FLDAS-Forecast setup and processing.

Models	**Centers**	**Hindcast Ensemble Members (1982-2010)**	**Forecast Ensemble Members (2011-present)**
CFSv2	NOAA/NCEP	24(12)	24
GEOS	NASA	4	10
CanCM4i	Environment Canada	10	10
GEM-NEMO	Environment Canada	10	10
CCSM4	NCAR	10	10
GFDL	GFDL	10	10
GFDL-Flor	GFDL	24(12)	24

Table 2. Spatially averaged 3-monthly anomaly correlation coefficient, RMSE, RPSS between ESP/FFV1/FFV2 forecast and RA (rows 1-3), and 3-monthly correlation coefficient between RA/ESP/FFV1/FFV2 forecast and SMAP (rows 4), and GIMMS NDVI (rows 5)

Metric	**Lead 0**	**Lead 1**	**Lead 2**	**Lead 3**	**Lead 4**	**Lead 5**
Anomaly Correlation	**OND** 0.74/0.85 /0.89	**NDJ** 0.54/0.65 /0.74	**DJF** 0.48/0.54 /0.63	**JFM** 0.44/0.47 /0.57	**FMA** 0.37/0.41 /0.52	**MAM** 0.34/0.38 /0.46
RMSE	**OND** 16.5/10.4 /9.2	**NDJ** 22.7/15.2 /13.7	**DJF** 23.8/17.1 /15.9	**JFM** 24.3/18.0 /16.9	**FMA** 25.8/18.6 /17.4	**MAM** 27.1/20.1 /18.6
RPSS	**OND** 0.65/0.69 /0.74	**NDJ** 0.53/0.59 /0.65	**DJF** 0.51/0.56 /0.62	**JFM** 0.5/0.54 /0.61	**FMA** 0.47/0.52 /0.58	**MAM** 0.44/0.50 /0.57
Correlation (SMAP)	**OND** 0.73/0.51/ 0.54/0.55	**NDJ** 0.73/0.40/ 0.44/0.45	**DJF** 0.74/0.22/ 0.27/0.30	**JFM** 0.66/0.01/ 0.08/0.09	**FMA** 0.62/0.02/ 0.1/0.1	**MAM** 0.66/0.22/ 0.23/0.24
Correlation (NDVI)	**OND** 0.57/0.49/ 0.46/0.48	**NDJ** 0.57/0.48/ 0.45/0.51	**DJF** 0.50/0.28/ 0.31/0.38	**JFM** 0.41/0.08/ 0.14/0.16	**FMA** 0.35/0.23/ 0.25/0.29	**MAM** 0.55/0.50/ 0.51/0.54

Table 3. Spatially averaged monthly RPSS/HR between RA and FFV2 forecast (rows 1-3) over southern Africa

Metric	Lead 0	Lead 1	Lead 2	Lead 3	Lead 4	Lead 5
RPSS/HR	**O** 0.47/0.27	**N** 0.57/0.12	**D** 0.62/0.1	**J** 0.63/0.06	**F** 0.64/0.05	**M** 0.65/0.04
RPSS/HR	**N** 0.46/0.29	**D** 0.58/0.14	**J** 0.62/0.11	**F** 0.61/0.1	**M** 0.63/0.08	**A** 0.65/0.07
RPSS/HR	**D** 0.45/0.31	**J** 0.57/0.16	**F** 0.61/0.11	**M** 0.62/0.09	**A** 0.64/0.1	**M** 0.61/0.1

Appendix 1

1. Anomaly Correlation Coefficient (ACC) =

$$\frac{\sum_{i=1}^{n}(f_i - \bar{f}_c)(o_i - \bar{o}_c)}{\sqrt{\sum_{i=1}^{n}(f_i - \bar{f}_c)^2 \sum_{i=1}^{n}(o_i - \bar{o}_c)^2}}$$

where n is the number of samples, f_i and o_i are the forecasts and validation data, and $\overline{f_c}$ and $\overline{o_c}$ are climatological mean of the forecasts and validation data, respectively.

1.a Statistical Significance of ACC using Fisher transformation:

$$F = 0.5ln\frac{(1+\rho)}{(1-\rho)}$$

Where ρ is the ACC, and then compute confidence interval at 95% as follows:

$$Z_L = F - 1.96/\sqrt{N-3} \quad ; \qquad Z_U = F + 1.96/\sqrt{N-3}$$

$$\rho_L = \frac{e^{2Z_L - 1}}{e^{2Z_L + 1}} \quad ; \qquad \rho_U = \frac{e^{2Z_U - 1}}{e^{2Z_U + 1}}$$

Where, N is the number of samples (N=29, seasons from 1982 through 2010) for both the forecast versions. ρ_L and ρ_U indicate confidence interval at 95%.

1.b Statistical significance of the differences in ACC using Fisher transformation,

$$F_{FFV2} = 0.5ln\frac{(1+\rho_{FFV2})}{(1-\rho_{FFV2})} \quad ; \qquad F_{FFV1} = 0.5ln\frac{(1+\rho_{FFV1})}{(1-\rho_{FFV1})}$$

Where ρ_{FFV2} and ρ_{FFV1} are the ACC of the two versions of forecasts and the test statistic (Z_{diff}) is computed for the differences between the Fisher transformation of the ACC of the two forecast versions as follows:

$$Z_{diff} = \frac{F_{FFV2} - F_{FFV1}}{\sqrt{\frac{\sigma^2_{FFV2}}{N_{FFV2}} + \frac{\sigma^2_{FFV1}}{N_{FFV1}}}}$$

Where, σ^2 is the variance of F and N is the number of samples (N=29, seasons from 1982 through 2010) for the forecast versions. Values of Z greater than 1.96 indicate that the difference between ACC is statistically significant at 95% confidence level.

2. Root Mean Square Error =

$$\sqrt{\frac{\sum_{i=1}^{n}(f_i - o_i)^2}{n}}$$

where n is the number of samples, f_i and o_i are the forecasts and validation data.

3. Correlation Coefficient =

$$\frac{\sum_{i=1}^{n}(f_i - \bar{f})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^{n}(f_i - \bar{f})^2 \sum_{i=1}^{n}(o_i - \bar{o})^2}}$$

where n is the number of samples, f_i and o_i are the forecasts and validation data and $\bar{f}$ and $\bar{o}$ are the mean of the forecasts and validation data.

4. Ensemble Spread = $Ens_{max} - Ens_{min}$

Where Ens_{max} and Ens_{min} are the maximum and minimum values respectively of the ensemble members.

5. Rank Probability Skill Score, RPSS = $1 - \frac{RPS}{RPS_{clim}}$

where Rank Probability Skill, RPS =

$$\left[\left(\sum_{k=1}^{m} p_k\right) - \left(\sum_{k=1}^{m} o_k\right)\right]^2$$

where p_k is the probability of forecast category *k* and o_k is an indicator (0=no, 1=yes) for the observations category *k*. *k* is considered as drought or below normal category in this study.

6. Hit Rate = $\frac{\text{Hits}}{\text{Hits+Misses}}$

Where Hits and Misses are from the contingency table below,

Drought forecasts	**Drought Observed**	
	Yes	No
Yes	Hits	False Alarms
No	Misses	Correct non-event

Models	Centers	Hindcast Ensemble Members (1982-2010)	Forecast Ensemble Members (2011-present)
CFSv2	NOAA/NCEP	24(12)	24
GEOS	NASA	4	10
CanCM4i	Environment Canada	10	10
GEM-NEMO	Environment Canada	10	10
CCSM4	NCAR	10	10
GFDL	GFDL	10	10
GFDL-Flor	GFDL	24(12)	24

Metric	Lead 0	Lead 1	Lead 2	Lead 3	Lead 4	Lead 5
Anomaly Correlation	**OND** 0.74/0.85 /0.89	**NDJ** 0.54/0.65 /0.74	**DJF** 0.48/0.54 /0.63	**JFM** 0.44/0.47 /0.57	**FMA** 0.37/0.41 /0.52	**MAM** 0.34/0.38 /0.46
RMSE	**OND** 16.5/10.4 /9.2	**NDJ** 22.7/15.2 /13.7	**DJF** 23.8/17.1 /15.9	**JFM** 24.3/18.0 /16.9	**FMA** 25.8/18.6 /17.4	**MAM** 27.1/20.1 /18.6
RPSS	**OND** 0.65/0.69 /0.74	**NDJ** 0.53/0.59 /0.65	**DJF** 0.51/0.56 /0.62	**JFM** 0.5/0.54 /0.61	**FMA** 0.47/0.52 /0.58	**MAM** 0.44/0.50 /0.57
Correlation (SMAP)	**OND** 0.73/0.51/ 0.54/0.55	**NDJ** 0.73/0.40/ 0.44/0.45	**DJF** 0.74/0.22/ 0.27/0.30	**JFM** 0.66/0.01/ 0.08/0.09	**FMA** 0.62/0.02/ 0.1/0.1	**MAM** 0.66/0.22/ 0.23/0.24
Correlation (NDVI)	**OND** 0.57/0.49/ 0.46/0.48	**NDJ** 0.57/0.48/ 0.45/0.51	**DJF** 0.50/0.28/ 0.31/0.38	**JFM** 0.41/0.08/ 0.14/0.16	**FMA** 0.35/0.23/ 0.25/0.29	**MAM** 0.55/0.50/ 0.51/0.54

Metric	Lead 0	Lead 1	Lead 2	Lead 3	Lead 4	Lead 5
RPSS/HR	**O** 0.47/0.27	**N** 0.57/0.12	**D** 0.62/0.1	**J** 0.63/0.06	**F** 0.64/0.05	**M** 0.65/0.04
RPSS/HR	**N** 0.46/0.29	**D** 0.58/0.14	**J** 0.62/0.11	**F** 0.61/0.1	**M** 0.63/0.08	**A** 0.65/0.07
RPSS/HR	**D** 0.45/0.31	**J** 0.57/0.16	**F** 0.61/0.11	**M** 0.62/0.09	**A** 0.64/0.1	**M** 0.61/0.1

Figure1

Click here to access/download;Figure;Figure1.png

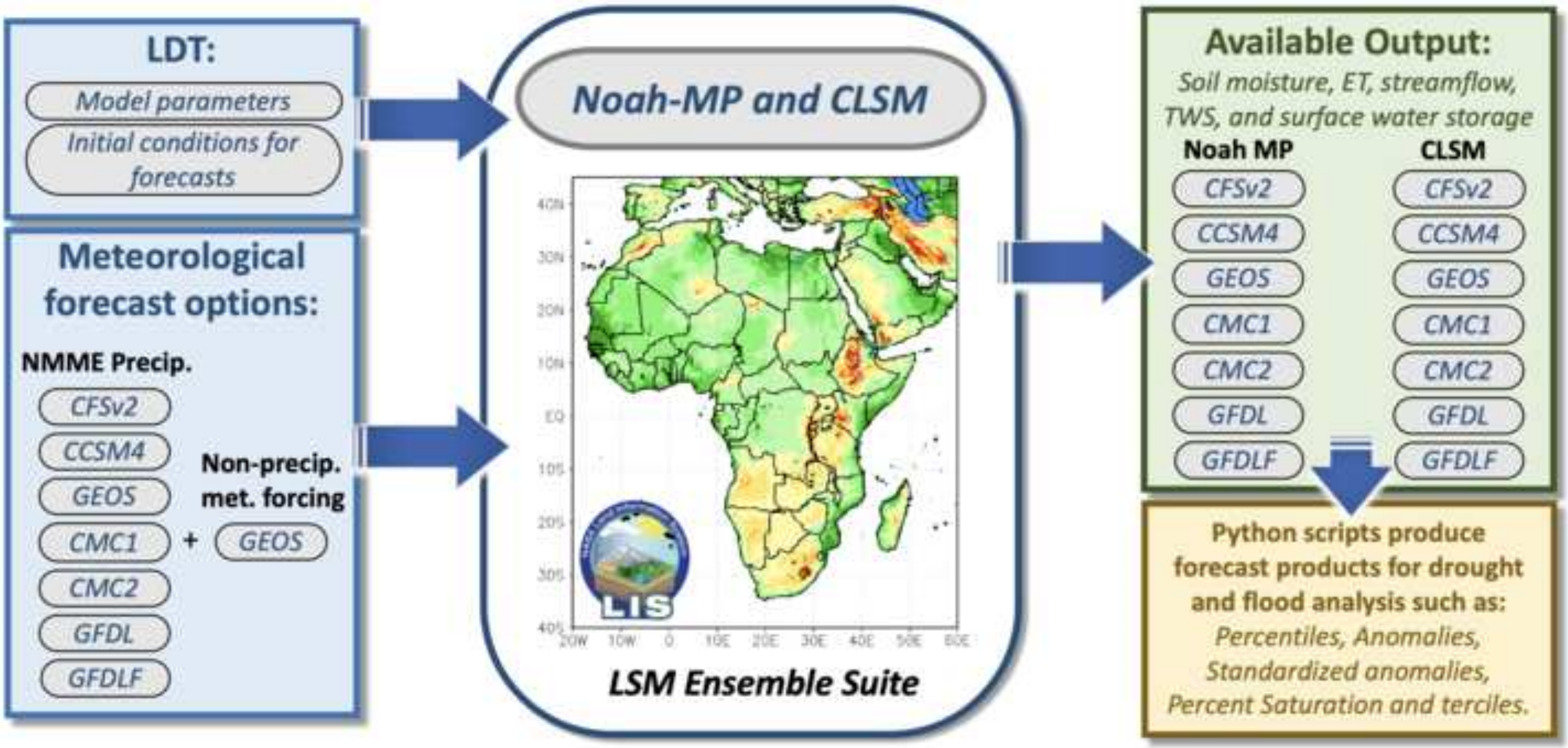


Figure2 Click here to access/download;Figure;Figure2.png

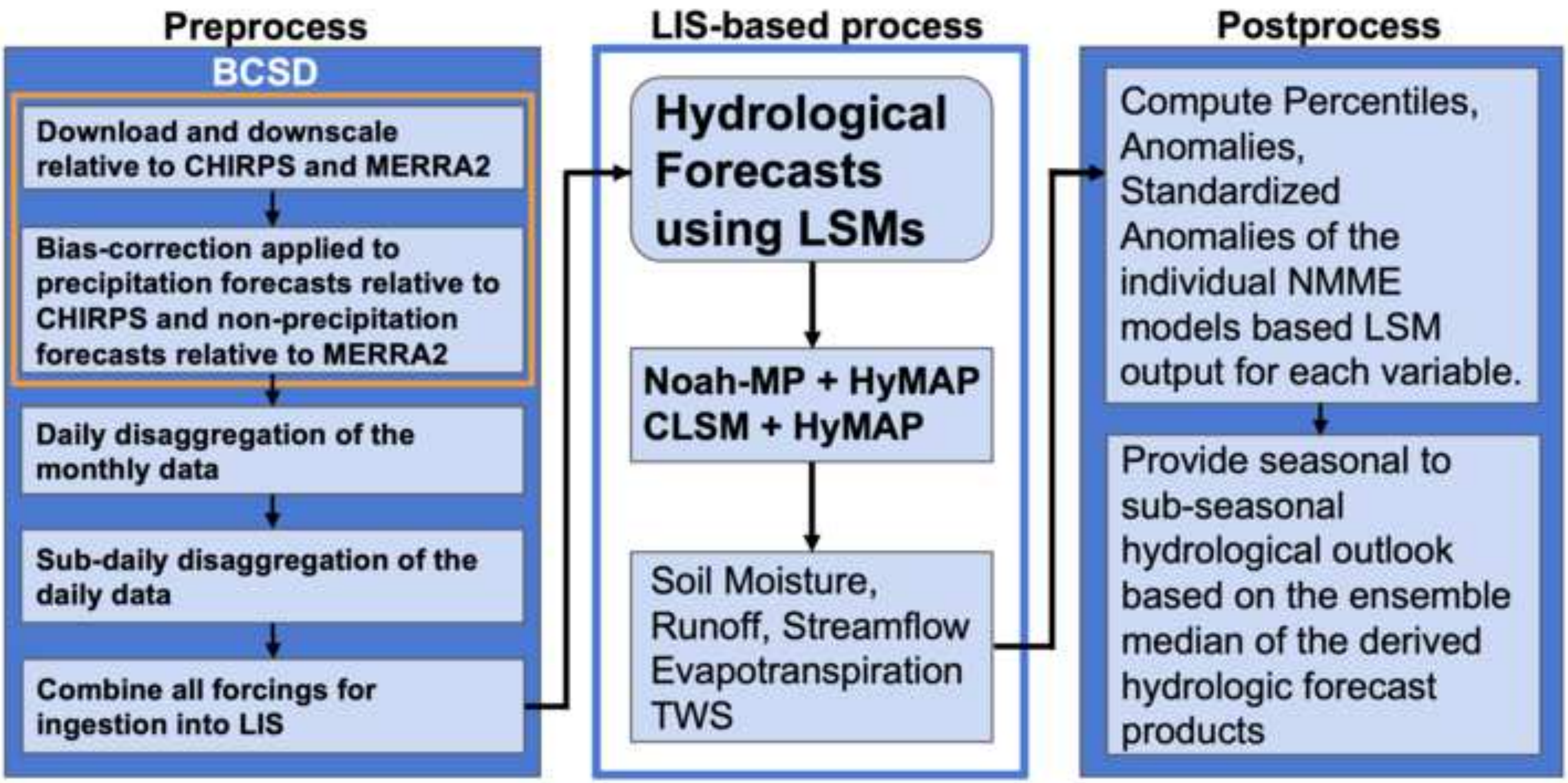


Figure3

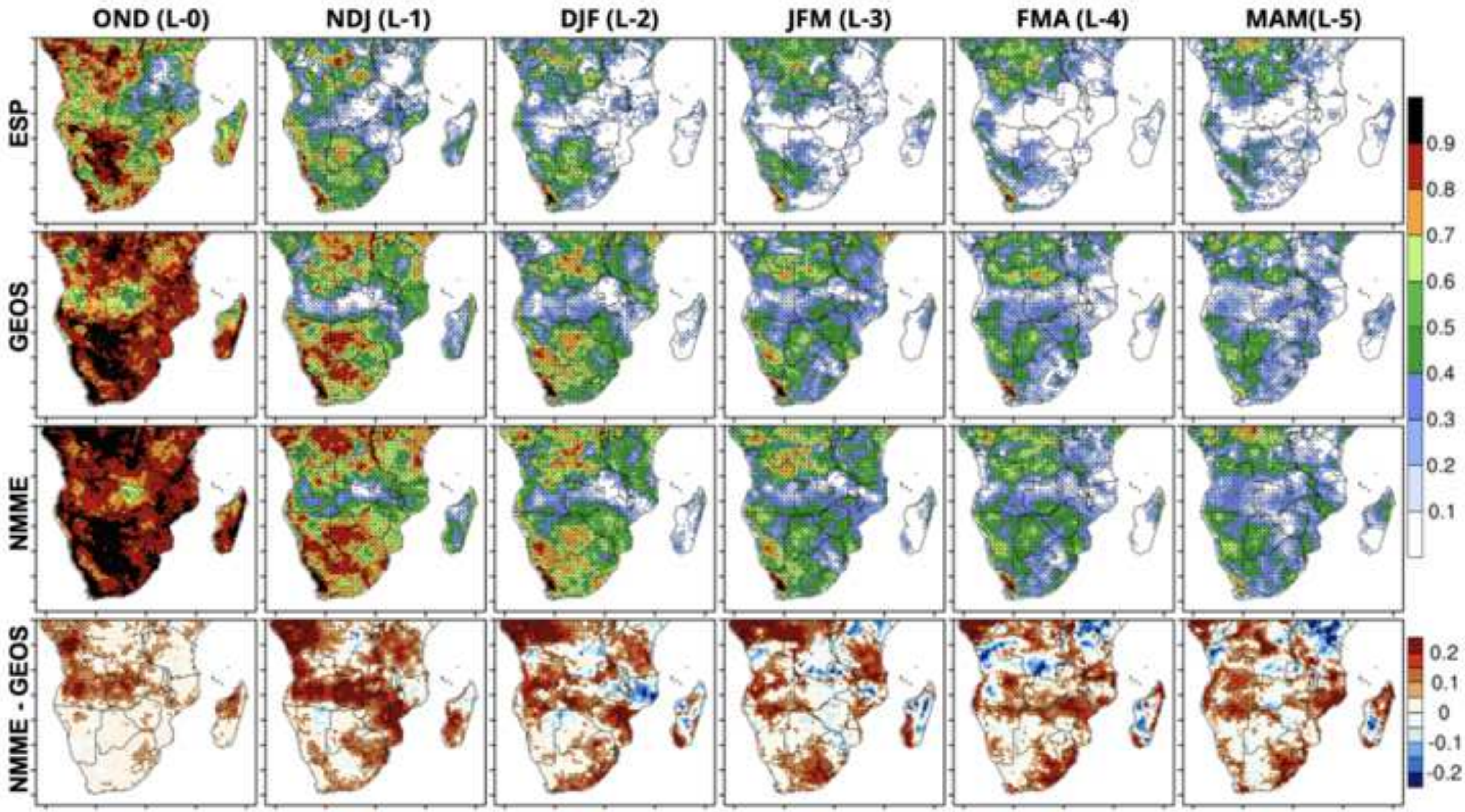

OND (L-0)
NDJ (L-1)
DJF (L-2)
JFM (L-3)
FMA (L-4)
MAM(L-5)
ESP
GEOS
NMME
NMME - GEOS
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.2
0.1
0
-0.1
-0.2

Figure4

Click here to access/download;Figure;Figure4.png

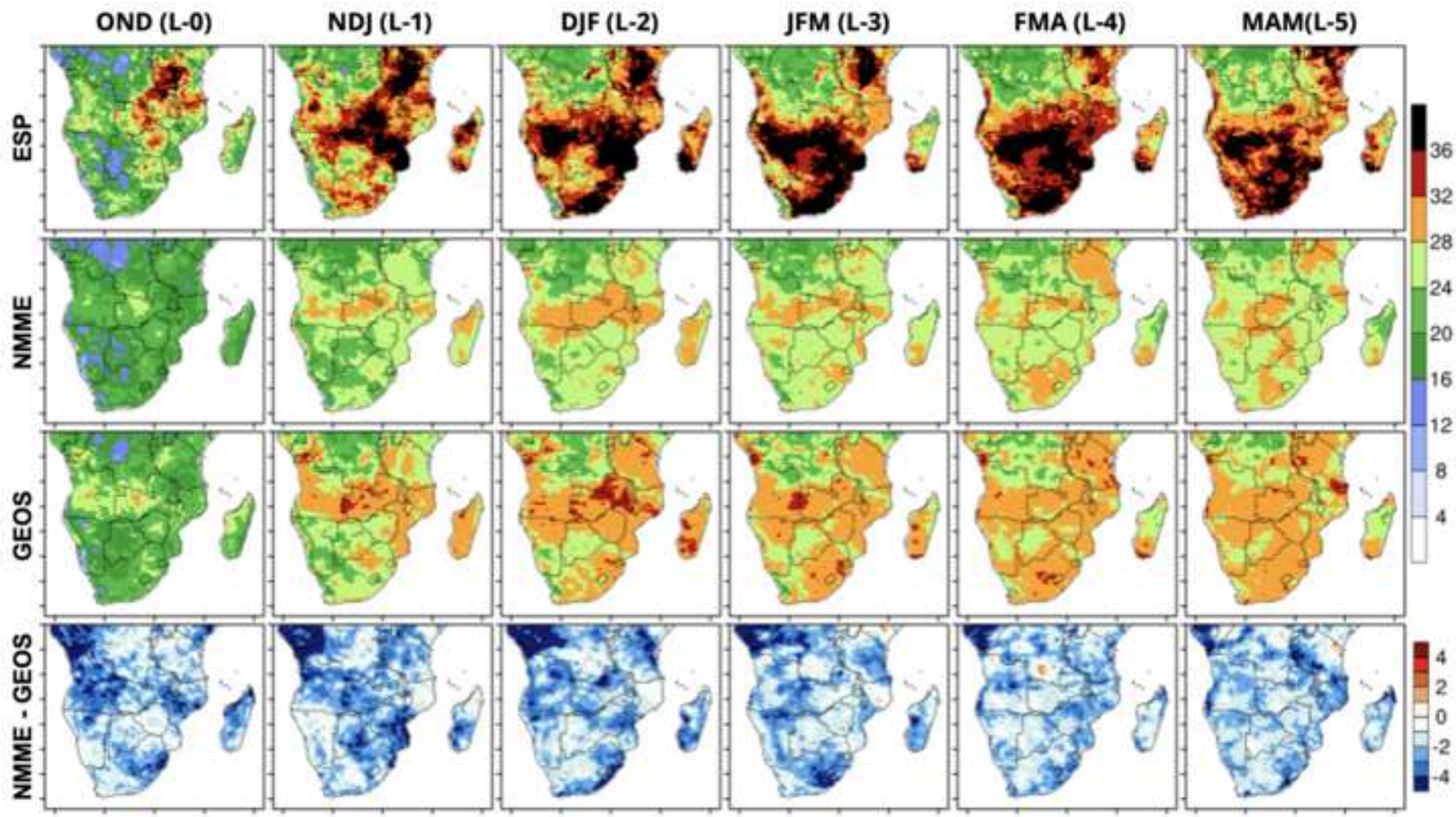


Figure5

Click here to access/download;Figure;Figure5.png

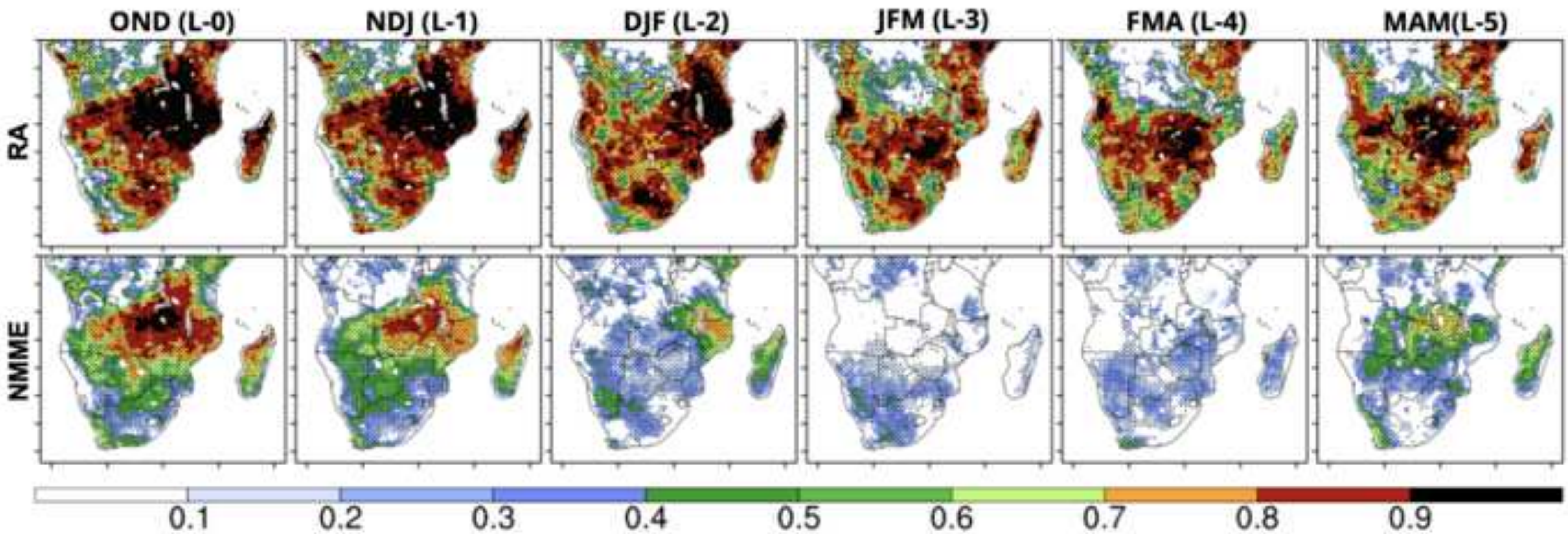


Figure6

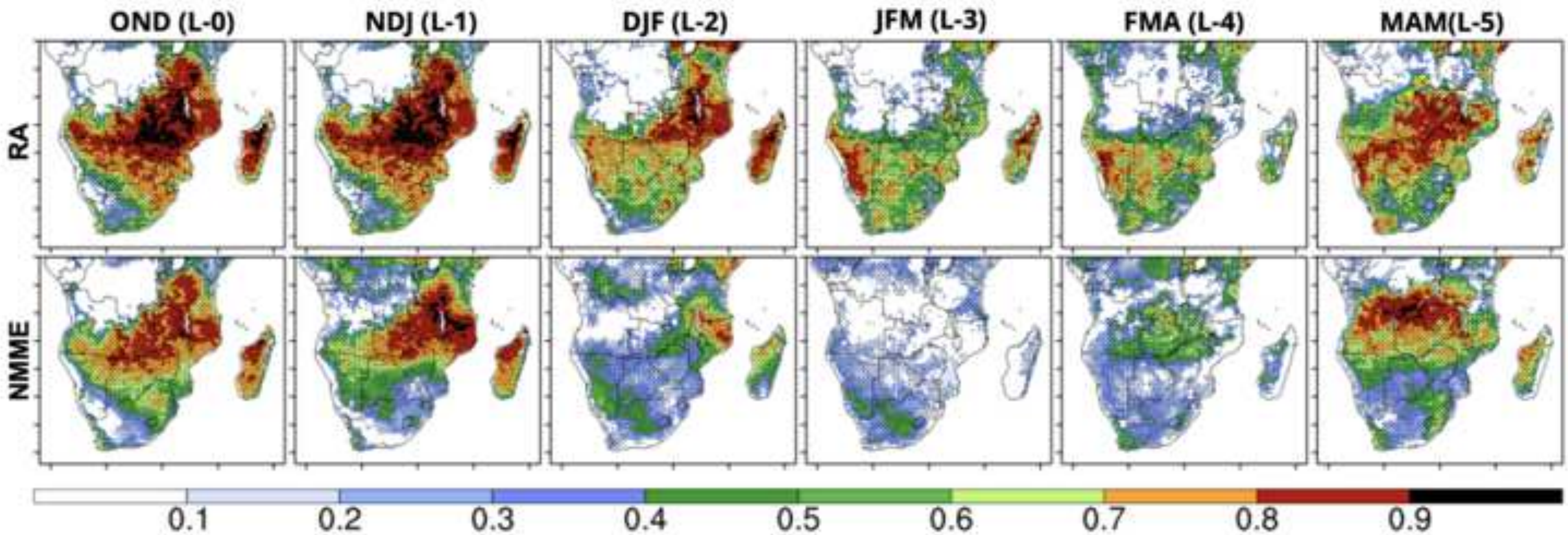

OND (L-0)
NDJ (L-1)
DJF (L-2)
JFM (L-3)
FMA (L-4)
MAM(L-5)
RA
NMME
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

Figure7

Click here to access/download;Figure;Figure7.png

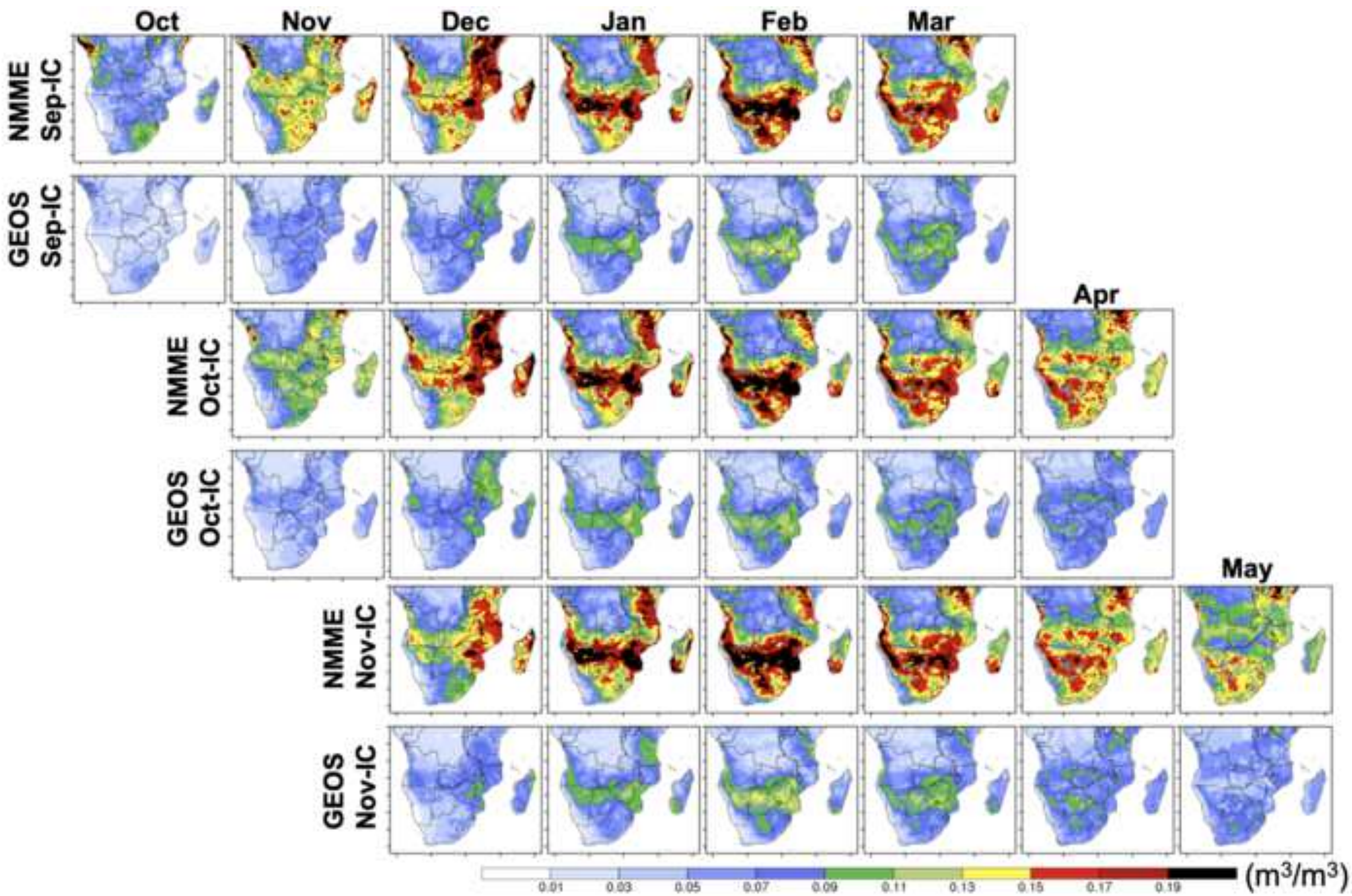
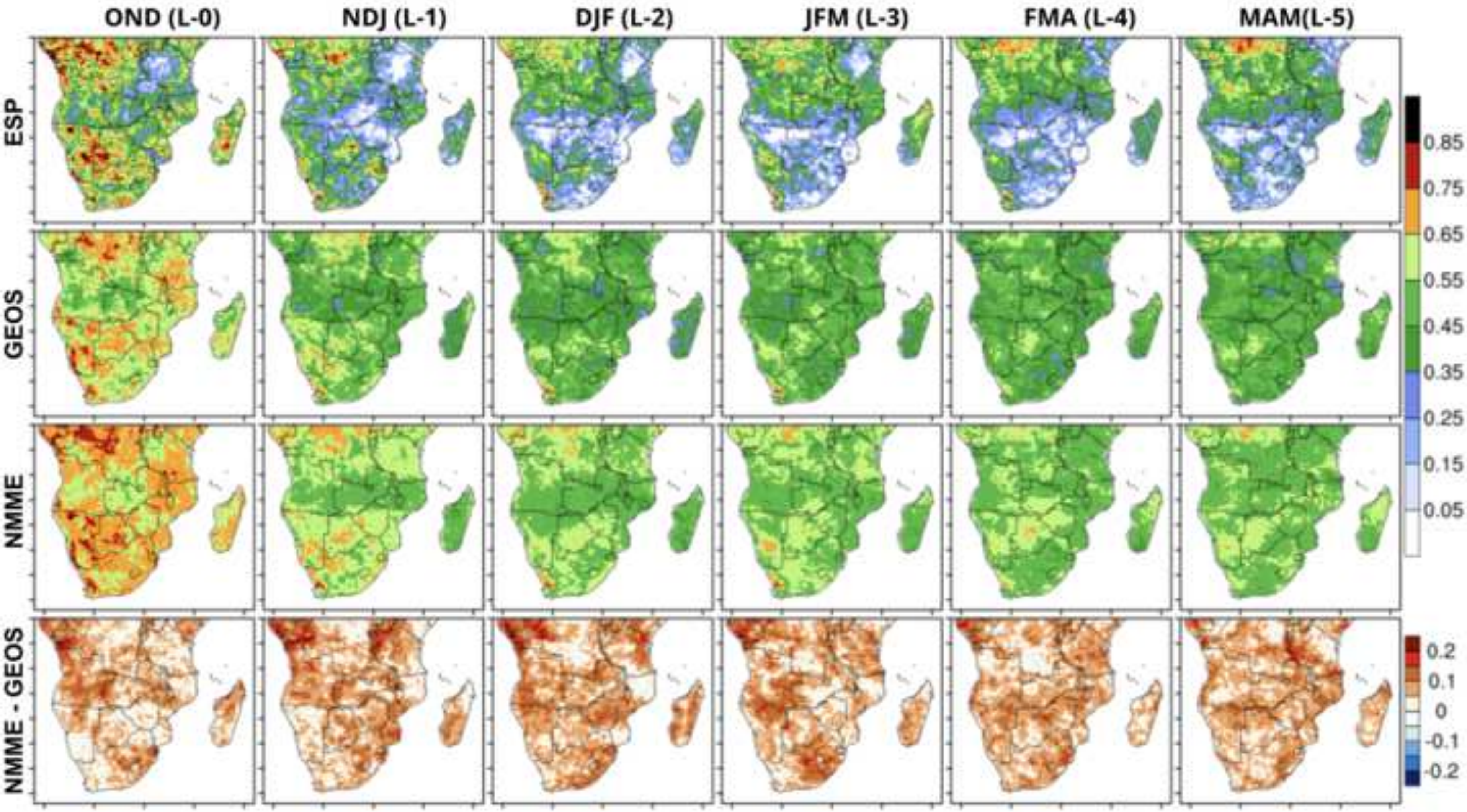


Figure8



OND (L-0)
NDJ (L-1)
DJF (L-2)
JFM (L-3)
FMA (L-4)
MAM(L-5)
ESP
GEOS
NMME
NMME - GEOS
0.85
0.75
0.65
0.55
0.45
0.35
0.25
0.15
0.05
0.2
0.1
0
-0.1
-0.2

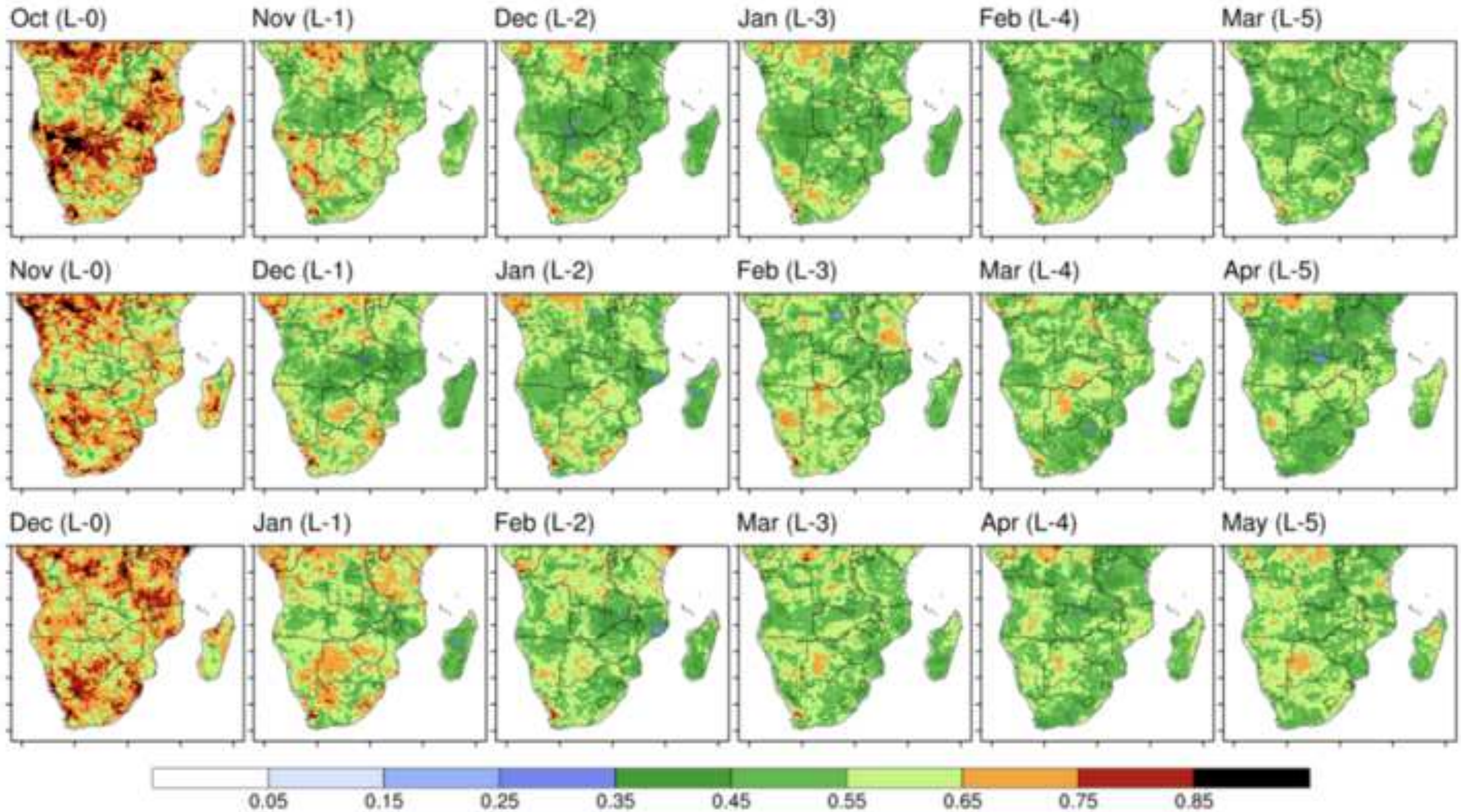

Oct (L-0)
Nov (L-1)
Dec (L-2)
Jan (L-3)
Feb (L-4)
Mar (L-5)
Nov (L-0)
Dec (L-1)
Jan (L-2)
Feb (L-3)
Mar (L-4)
Apr (L-5)
Dec (L-0)
Jan (L-1)
Feb (L-2)
Mar (L-3)
Apr (L-4)
May (L-5)
0.05
0.15
0.25
0.35
0.45
0.55
0.65
0.75
0.85

Figure10

Click here to access/download;Figure;Figure10.png

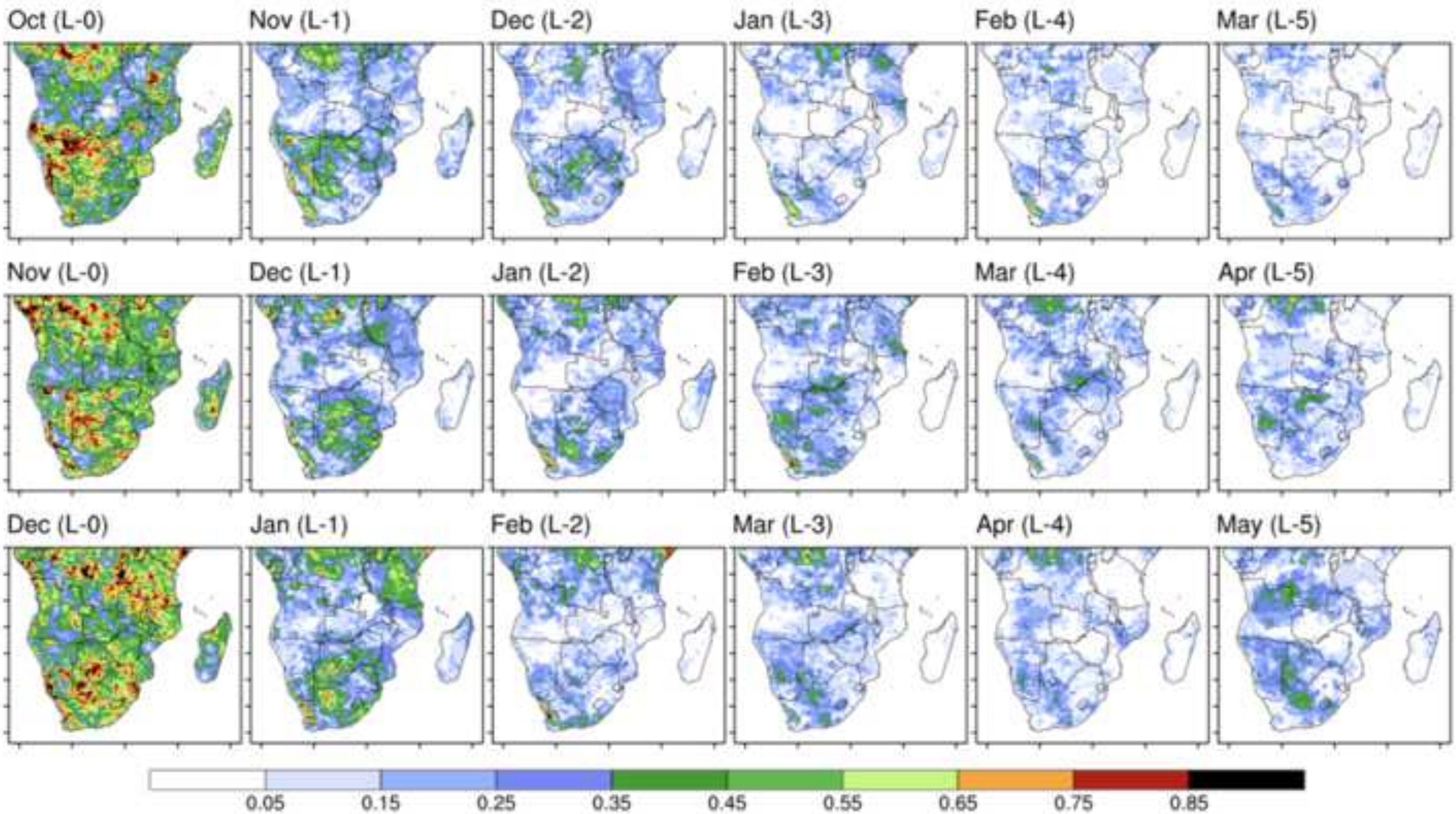


Figure11

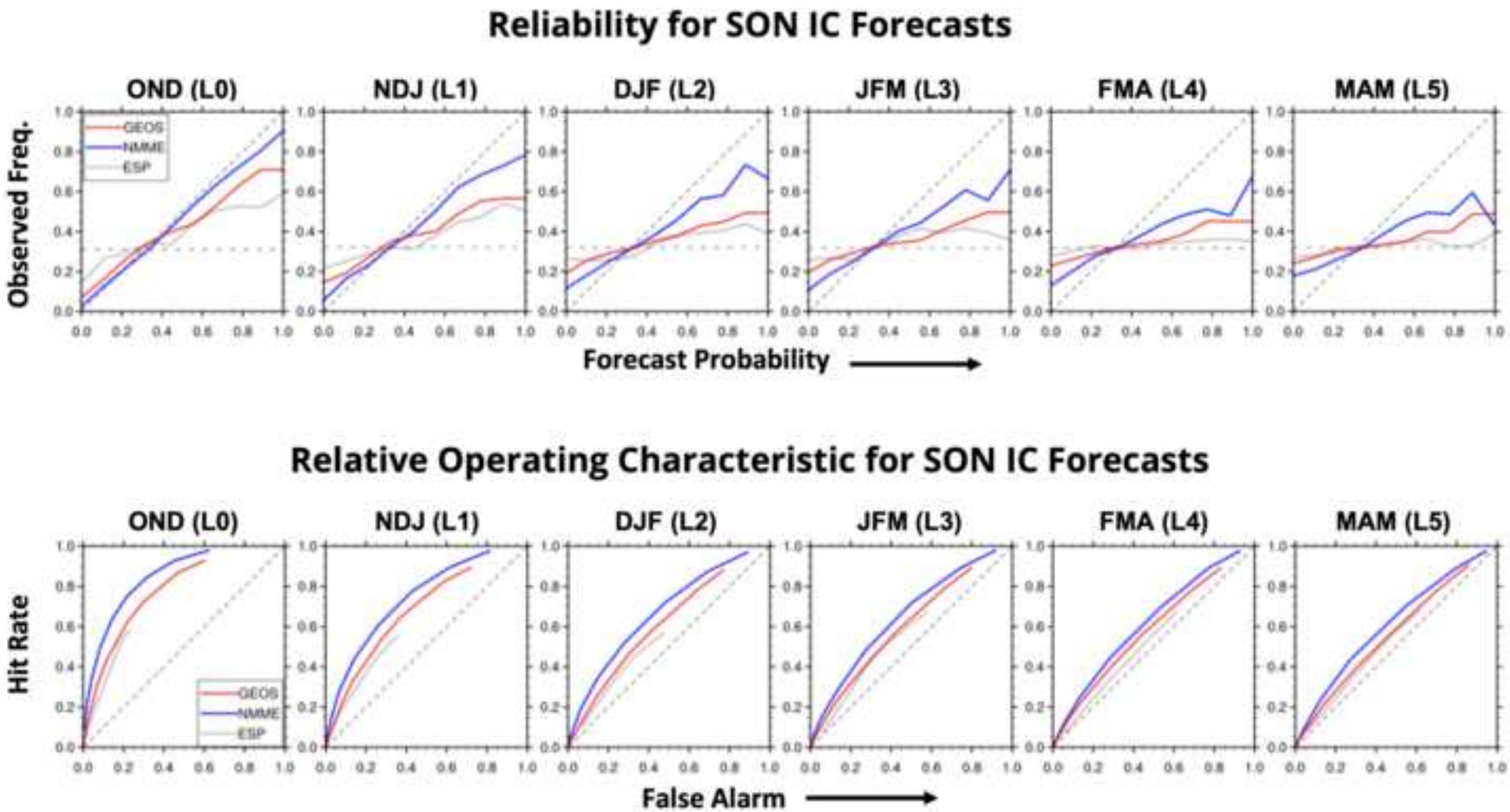
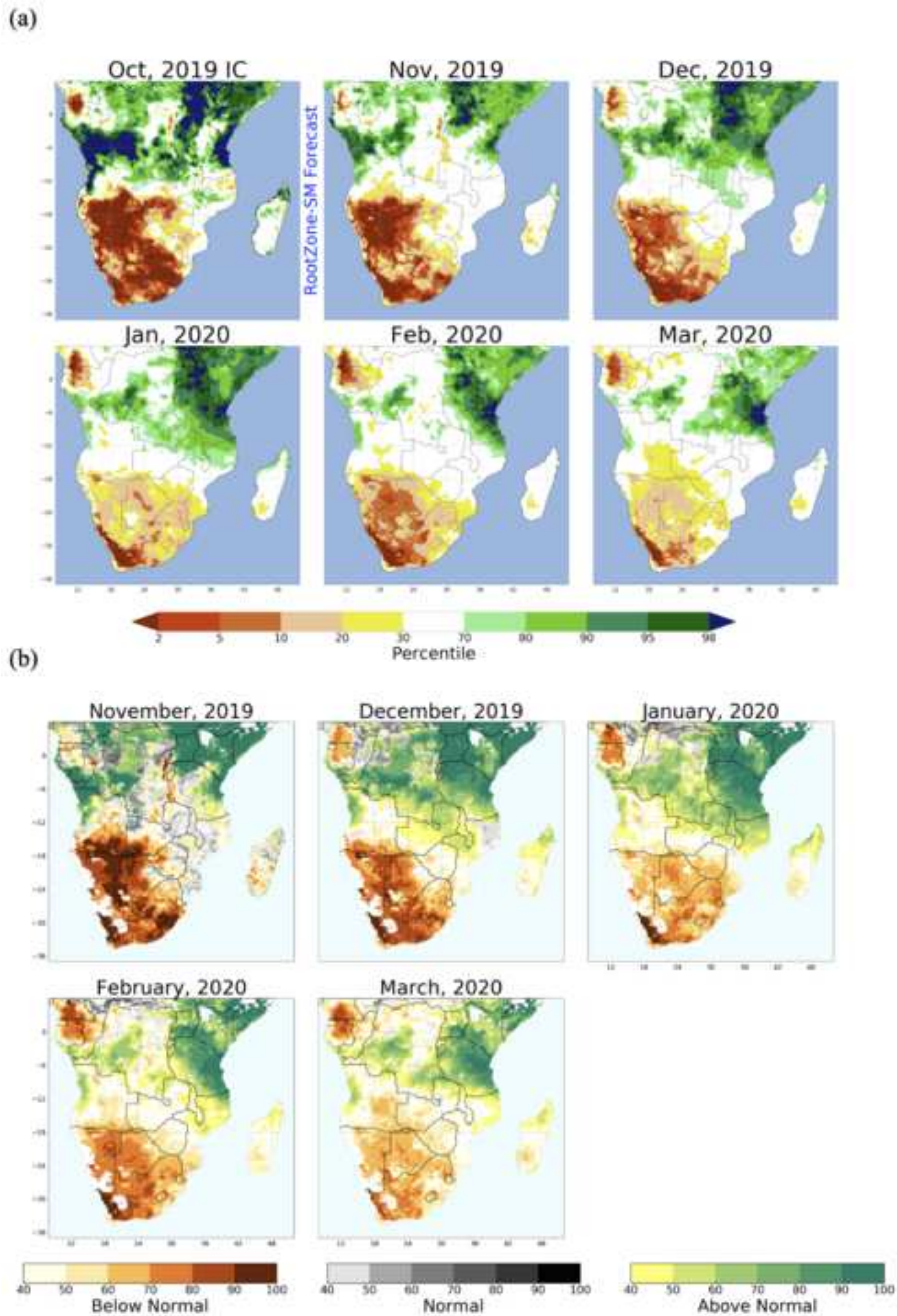

Reliability for SON IC Forecasts
OND (L0)
NDJ (L1)
DJF (L2)
JFM (L3)
FMA (L4)
MAM (L5)
Observed Freq.
GEOS
NMME
ESP
Forecast Probability
Relative Operating Characteristic for SON IC Forecasts
OND (L0)
NDJ (L1)
DJF (L2)
JFM (L3)
FMA (L4)
MAM (L5)
Hit Rate
GEOS
NMME
ESP
False Alarm

Figure12 Click here to access/download;Figure;Figure12.png



CRediT author statement

Abheera Hazra: Conceptualization, Writing, Original draft preparation, Visualization, Investigation
Amy McNally: Supervision, Original draft preparation, Reviewing, Editing
Kimberly Slinski: Supervision, Original draft preparation, Reviewing, Editing
Kristi R. Arsenault: Original draft preparation, Reviewing, Editing
Shraddhanand Shukla: Methodology Software, Reviewing, Editing
Augusto Getirana: Software, Science Reviewing
Jossy P. Jacob: Data curation, Reviewing, Editing
Daniel P. Sarmiento: Software, Reviewing, Editing
Christa Peters-Lidard: Supervision, Science Reviewing, Editing
Sujay V. Kumar: Software, Science Reviewing, Editing
Randal D. Koster: Science Reviewing, Editing