

## A Seasonal Rainfall Performance Probability Tool for Famine Early Warning Systems

NICHOLAS S. NOVELLA

*Climate Prediction Center, NOAA/National Centers for Environmental Prediction, College Park, and Innovim, LLC, Greenbelt, Maryland*

WASSILA M. THIAW

*Climate Prediction Center, NOAA/National Centers for Environmental Prediction, College Park, Maryland*

(Manuscript received 9 March 2016, in final form 22 June 2016)

### ABSTRACT

This paper reports on the development of a new statistical tool that generates probabilistic outlooks of seasonal precipitation anomaly categories over Africa. Called the seasonal performance probability (SPP), it quantitatively evaluates the probability of precipitation to finish at predefined percent-of-normal anomaly categories corresponding to below-average (<80% of normal), average (80%–120% of normal), and above-average (>120% of normal) conditions. This is accomplished by applying methods for kernel density estimation (KDE), which compute smoothed, continuous density functions on the basis of more than 30 years of historical precipitation data from the Africa Rainfall Climatology, version 2, dataset (ARC2) for the remaining duration of a monsoon season. Discussion of various parameterizations of KDE and testing to determine optimality of density estimates (and thus performance of SPP for operational monitoring) are presented. Verification results from 2006 to 2015 show that SPP reliably provides probabilistic outcomes of seasonal rainfall anomaly categories by after the early to midstages of rain seasons for the major monsoon regions in East Africa, West Africa, and southern Africa. SPP has proven to be a useful tool by enhancing operational climate monitoring at CPC for its prognostic capability for famine early warning scenarios over Africa. These insights are anticipated to translate into better decision-making in food security, planning, and response objectives for the U.S. Agency for International Development/Famine Early Warning Systems Network (USAID/FEWS NET).


### 1. Introduction

The continuing development of several satellite-derived precipitation-estimator products, in both operational and research capacities, has enabled users to better diagnose and understand the scope of precipitation regimes, extreme events, trends, hydrologic cycles, and climate variability in the most inaccessible regions of the globe. To facilitate the monitoring of precipitation for food security and famine early warning systems, NOAA's Climate Prediction Center (CPC) had developed and operationally deploys two rainfall

products: the Rainfall Estimator (RFE; [Herman et al. 1997](#)) and the African Rainfall Climatology (ARC; [Love et al. 2004](#); [Novella and Thiaw 2013](#)). These products were created in response to the need for higher-resolution operational daily rainfall estimates in support of the U.S. Agency for International Development/Famine Early Warning Systems Network (USAID/FEWS NET). The RFE and ARC products are unique in comparison with many other satellite rainfall estimators because of their high, 0.1° gridded spatial resolution and their ability to blend in situ gauge and satellite information on a near-real-time basis to render daily (0600–0600 UTC) precipitation estimates over the African continent. Since 2001, these products have provided critical, timely early warning for food security preparedness, as they are routinely used in the identification of rainfall conditions over rain-fed agricultural and pastoral regions in parts of Africa ([Verdin et al. 2005](#)).

In 2013, the completion of the ARC, version 2.0 (ARC2), dataset ([Novella and Thiaw 2013](#)) has allowed

---

 Denotes Open Access content.

---

*Corresponding author address:* Nicholas S. Novella, Climate Prediction Center, National Centers for Environmental Prediction, NOAA Center for Weather and Climate Prediction, 5830 University Research Court, College Park, MD 20740.  
E-mail: [nicholas.novella@noaa.gov](mailto:nicholas.novella@noaa.gov)

meteorologists at CPC to utilize more than 30 years of daily, high-resolution satellite precipitation estimates for USAID/FEWS NET activities over Africa. ARC2 has become an effective tool for the climate and famine early warning systems community as satellite records are now long enough to facilitate historical examinations of climate conditions. Because ARC2 has been instrumental for depicting the daily evolution of seasonal rainfall in real time, statistical techniques can also be applied to these data to help to predict the outcome of the rainfall season. In particular, for several locations in Africa that exhibit a well-defined monsoonal cycle, there are periods at which precipitation is climatologically expected to commence, peak, and then weaken throughout the course of the season. Continuous monitoring of these periods during operations allows us to determine how well current seasonal rainfall performs with respect to the daily climatological normal totals as the season progresses. This process, however, does not give us an objective measure of the projection of seasonal rainfall performance starting from a given date—midseason for instance. It is apparent, though, that the certainty of a seasonal outcome invariably increases as the remaining monsoon season grows shorter. We postulate here that this relationship may also be expressed in terms of a quantitative probability measure that converges to 1 toward the end of the season. To quantify probability, we consider an array of hypothetical precipitation rates ( $\text{mm day}^{-1}$ ) from zero to infinity that, when projected onto the current seasonal rainfall total, will satisfy a number of rainfall categories (i.e., below-average, average, and above-average rainfall) by the end of the season. Such an array will then allow for a continuous variable distribution necessary for probability function analysis. In viewing future rainfall rates required to achieve various degrees of anomaly, the following inquiries are prompted: What are the probabilities of such rates to occur according to the ARC2 long-term climatological record? How do we best estimate these probabilities? How prevalent are seasonal rainfall reversals relative to persistence in the climatological monsoon data? Furthermore, how do historical precipitation rates deviate from the climatological normal precipitation rate throughout various instances in the season?

The objective of this paper is to report on the development of a new statistical tool that sheds light on these questions by utilizing the ARC2 long-term precipitation record to generate probabilistic outlooks for seasonal rainfall throughout Africa. Simply named “seasonal performance probability” (SPP), this tool quantitatively evaluates the probability for seasonal precipitation to finish at predefined percent-of-normal anomaly categories corresponding to below-average (<80% of normal),

average (80%–120% of normal), and above-average (>120% of normal) conditions. These seasonal percent-of-normal thresholds (80% and 120%) have been a reliable standard for demarcating anomalous rainfall conditions in operational monitoring at CPC, because they capture hydrometeorological impacts (e.g., drought or flooding hazards) for famine early warning systems over Africa. SPP is accomplished by applying kernel density estimation (KDE) methods, which compute probability density functions (PDFs) and cumulative distribution functions (CDFs) on the basis of the historical performance of precipitation from a given point in a season to the end of season. SPP exclusively relies on ARC2 real-time and climatological precipitation data over Africa, because the high temporal and spatial resolutions of ARC2 are essential for deriving probabilities of seasonal rainfall performance. Unlike numerical weather/climate prediction methods/models (GFS, CFS, ECMWF, etc.), SPP is not so much a traditional forecast but simply a non-parametric way to estimate the probability density of a random variable like precipitation with a complex distribution. SPP output consists of a set of probability maps that are designed to provide the end user with a new measure of the expected outcome of anomalously dry, average, or anomalously wet seasonal conditions during any time in the season. It is anticipated that SPP will be a useful tool by enhancing operational climate monitoring at CPC and also by providing additional guidance for seasonal-outlook scenarios. Coupled with real-time precipitation anomaly maps, the generation of probabilistic outlooks throughout various stages of the rainy season will assist in better decision-making in food security, planning, and response. [Section 2](#) presents an overview of the seasonality of African rainfall. [Section 3](#) describes KDE statistical methods and the parameters selected for SPP over Africa. [Section 4](#) illustrates and discusses SPP output and shows several case and validation studies over monsoonal regions of Africa. The final section provides a summary of SPP and final remarks.

## 2. Seasonality of Africa precipitation

As a vast continent spanning hemispheres, Africa possesses much diversity in its precipitation climate and seasonal-transition regimes. A basic understanding of the regions experiencing major and minor wet seasons and dry periods and of their timing over Africa is where SPP will be most applicable. According to [Novella and Thiaw \(2013\)](#), a continentally averaged annual rainfall maximum typically develops during the March–May time frame when convection is extremely active within the intertropical convergence zone (ITCZ) in the Gulf of Guinea region, Congo, and equatorial eastern Africa.

The annual rainfall minimum that shortly follows this maximum occurs from May to October, when rains shift to the northern tropical belt region ( $0^{\circ}$ – $20^{\circ}$ N) coinciding with the West Africa monsoon. As the West Africa monsoon begins to weaken, a gradual increase in mean rainfall is associated with the onset of rainfall in central Africa, southern Africa, and equatorial eastern Africa. These rains remain spatially robust throughout the Southern Hemisphere and the southwestern Indian Ocean until rains begin to shift northward by March.

A recent study by Liebmann et al. (2012) presented an investigation of African wet-season characteristics, including objectively defined onset and cessation timings, using GPCP precipitation data. A spatial analysis illustrating the ratio of the second to first annual harmonic over Africa nicely details where rainfall is characteristically monsoonal. From Fig. 6 of Liebmann et al. (2012), the annual cycles at selected points offer a closer look at the seasonality of rainfall. In the Sahel region of western Africa, a single well-defined rainy season is evident during the June–September time frame, with little to no mean rainfall received throughout the remainder of the year. Farther south in the Gulf of Guinea region of western Africa, the mean annual cycle of rainfall is bimodal, experiencing a pronounced wet season during the March–June time frame and a weaker wet season during the September–December time frame. A distinct minimum separating these two seasons in August is concurrent with the annual maximum farther north in the Sahel and has been previously described in literature as the West Africa “monsoon jump” (Hagos and Cook 2007). In eastern Africa, the seasonality of precipitation exhibits more variation on regional and local scales because of the complex topography (Dinku et al. 2007) and the influence of convergence associated with the Indian Ocean ITCZ (Williams and Funk 2011; Hutchinson 1992; Bowden and Semazzi 2007). Although comparatively drier, the Horn of Africa generally exhibits two distinct rainfall seasons corresponding to the March–June and September–December time frames. In the northwestern part of the horn, however, the mean annual cycle of rainfall becomes more unimodal as western Ethiopia experiences a single wet season that is more consistent with the June–September time frame experienced in the Sahel. In a more refined seasonality analysis over the greater horn, Herrmann and Mohr (2011) presented a high-resolution map outlining rainfall modality classes in the greater horn with local seasonal names. In southern Africa, the mean annual cycle rainfall is largely unimodal during the October–April time frame, with greater meridional extent when compared with western and eastern Africa. Spatially, monsoonal rainfall migrates and expands southeastward

from the Congo and northwestward from the southeastern coast in South Africa. As a result, wet-season lengths show some variation with respect to latitude as rainfall in the Southern Hemispheric tropical belt ( $17^{\circ}$ – $23^{\circ}$ S) typically experiences a later onset and earlier cessation relative to surrounding continental latitudes in the Southern Hemisphere. After this transition, mean rainfall throughout much of southern Africa reaches its peak in magnitude during the December–February time frame.

For operational monitoring at CPC, meteorologists have designated six seasonal time frames over three main domains in Africa. These include the East Africa domain, encompassing the March–May, June–September, February–September, and October–December time frames, as well as the West Africa and southern Africa domains, covering the May–September and October–May time frames, respectively. These seasonal time frames are used as a general platform for capturing the evolution of monsoon rainfall over Africa and should serve well for SPP. They also cover pertinent agricultural calendars and cropping activities on the ground for famine early warning systems.

### 3. Method

#### a. Kernel density estimation

The main purpose in SPP lies in determining the PDF and CDF of historical precipitation rates from a current point in a season to the end of season. The most elementary means to estimate density functions is found in relative-frequency histogram plots, but these histograms are strongly influenced by the selection of bin centering and widths that can artificially skew and hide the true density of data (Wilks 2011). Instead, SPP applies KDE methods on the ARC2 30+-yr climatological dataset to acquire a more refined, smoother estimate of the PDF. Using a set of observations  $[x(1), x(2), \dots, x(n)]$  from some distribution with an unknown density  $f(\mathbf{x})$ , the KDE is defined as

$$f(\mathbf{x}) = \frac{1}{nh} \sum_{i=1}^n K \left[ \frac{\mathbf{r} - x(i)}{h} \right],$$

where  $n$  is the number of historical observations, the  $x(i)$  are the historical observations, and  $h$  is the bandwidth parameter. The selection of both the bandwidth parameter  $h$  and kernel type  $K$  has a marked effect on the shape of the estimated density, and more discussion on this topic is included later in this section. Unlike a relative-frequency histogram that crudely estimates the density of a distribution, the KDE yields a more smooth, continuous structure in which pointwise contributions to the density estimate are dependent on the distances

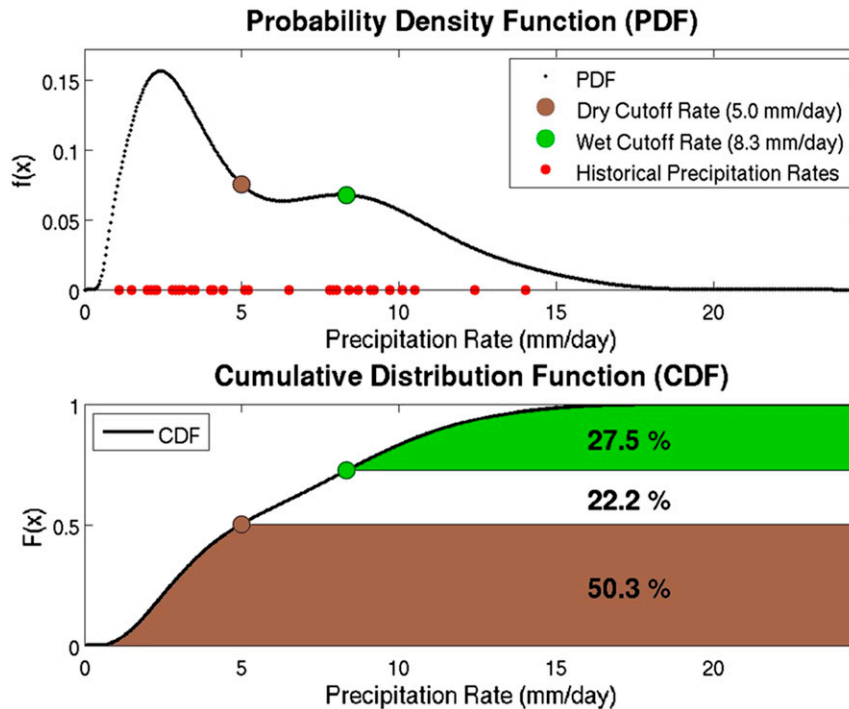


FIG. 1. Example of the (top) PDF and (bottom) CDF with derived SPP probabilities estimated from KDE from a sample set of historical precipitation rates during the remainder of the season.

between  $\mathbf{r}$  and  $x(i)$ . The process of summing the individual kernels centered on each point and dividing this sum by the product  $nh$  will satisfy the condition that the area under the entire PDF curve integrates to 1. If we let  $\mathbf{r}$  be an array of hypothetical precipitation rates (from 0 to  $\infty$ ) required to satisfy an array of percent-of-normal rain categories by the end of season, then it will not only render a smoother estimate of the density function  $f(\mathbf{x})$ , but it will allow the probabilities for each hypothetical precipitation rate responsible for the end-of-season anomaly to be determined. Taking the integral of  $f(\mathbf{x})$  results in the CDF  $F(\mathbf{x})$ , and it is here where probabilities within specified intervals along  $F(\mathbf{x})$  can be ascertained and plotted to render a probability value for each point in space.

To illustrate, let us suppose the following for a given location where 1) the current seasonal accumulated total is 100 mm at time  $T_{\text{current}}$ , 2) the current seasonal climatological normal total is 150 mm at  $T_{\text{current}}$ , 3) the end-of-season climatological normal total is 500 mm at  $T_{\text{final}}$ , and 4) the number of days remaining in the season is 60. Given that the current seasonal percent-of-normal anomaly is well below average at 66%, we would therefore find that threshold precipitation rates of 5.00, 6.66, and 8.33 mm day $^{-1}$  are required for the remainder of season to finish at least 80%, 100%, and 120% of

normal, respectively. Using a sample set of historical precipitation rates (i.e., observations)  $x(i)$  over the last 30+ years from  $T_{\text{current}}$  to  $T_{\text{final}}$  as well as an array of hypothetical precipitation rates  $\mathbf{r}$  required in the future, plotting the below-average (brown) and above-average (green) threshold-rate points along the  $x$  axis on both the PDF and CDF curves (as illustrated in Fig. 1) shows that the highest probability ( $\sim 50\%$ ) exists for seasonal rainfall to remain in the below-average category ( $< 80\%$  of normal) by the end of the season. Also evident is the second-highest probability ( $\sim 28\%$ ) for seasonal rains to finish in the above-average category ( $> 120\%$  of normal), and the lowest probability ( $\sim 22\%$ ) for seasonal rains to finish in the average category ( $\geq 80\%$  and  $\leq 120\%$  of normal) by the end of the season. In this example, the persistence of below-average rainfall by the end of the season is climatologically favored over a seasonal recovery according to ARC2, because historical precipitation from  $T_{\text{current}}$  to  $T_{\text{final}}$  exhibits a bimodal distribution (Fig. 1) with greater density located in the first mode and lesser density in the second mode. Of interest is that the mean of all historical precipitation rates ( $5.83 \text{ mm day}^{-1}$ ) is centered between the two modes in a local minimum of the PDF, indicating that there have been fewer years in which rainfall for the remainder of the season has been close to this value.

This example of KDE demonstrates how the ARC2 climatological dataset can depict a more insightful rainfall outlook than would simply using the historical mean as a projection for the remainder of the season.

*b. Parameterization background: Kernel type and bandwidth selection*

In KDE, the kernel type determines the shape of the individual weighting function centered for each point in  $x(i)$ . The summation of the kernels as reflected in KDE for the expression results in “stacking”: the sum of all kernel heights contributing to the smoothed density curve at each given value in  $\mathbf{r}$  (Wilks 2011). Of all kernel types available, initial testing of historical precipitation rates showed that the Gaussian kernel was an attractive choice because of its infinite support and intrinsic breadth in density-curve estimation, which was believed to lead to more equitable probabilities. With respect to using KDE with daily precipitation data, studies by Rajagopalan et al. (1993, 1997) have referenced the implementation of the Epanechnikov kernel. These authors noted that, because precipitation is a bounded (nonnegative), continuous variable, a kernel with an inherent bounded support was preferred over one with infinite support to minimize potential boundary effects. These studies also showed that boundary issues are ameliorated through the use of a logarithmic transformation within the kernel because it prevents any “leakage” of the probability mass extending beyond the boundary (Rajagopalan et al. 1997). This is most evident in cases in which there is a high concentration of density estimates near the origin of the sample space. Regardless of the kernel type, the log transformation was considered to be necessary for SPP so as to properly handle the fixed lower bound of precipitation so that  $f(\mathbf{x})$  still integrates to 1. Furthermore, log transforms with support of  $[0, \infty]$  can be applied to any kernel.

For bandwidth selection, literature has indicated that the choice of kernel type is usually less important than selection of the bandwidth  $h$  because this parameter is critical to the shape of density estimates (Wilks 2011; Lall et al. 1993). A bandwidth value that is too low may lead to curve undersmoothing, which potentially causes artificial modes in the density estimates, whereas a bandwidth that is too high may cause oversmoothing, which likely masks the finescale structure of the density estimates. Several authors have outlined bandwidth-selection methods with discussion of their strengths and weaknesses relative to capturing the true density of sample data. The bandwidth method that is most commonly referenced in literature is Silverman’s rule of thumb (Silverman 1986). An appealing feature of this method for precipitation data is that it uses an adaptive

estimate of the spread by taking into account the minimum of either the standard deviation or the interquartile range divided by a constant. In doing so, this method has been found to counteract extreme historical outliers that may inflate the bandwidth, helping to prevent oversmoothing of the density estimates (Wand and Jones 1995). Other studies have suggested that this method may not be aptly suited for multimodal distributions, however, and underperformance has been linked to its heavy reliance on assumptions of the underlying distribution (Rajagopalan et al. 1997). As an alternative, the “plug in” or recursive method of Sheather and Jones (1991; hereinafter this method is referred to as SJ) has also been widely described in KDE-associated literature. Although computationally intensive, the SJ bandwidth method has been shown to frequently outperform parametric-referenced and other cross-validation bandwidth-selection methods. In using the log transformation, SJ has been recommended because of its ability to handle boundary effects and local adaptation in a natural way given the erratic nature of precipitation (Rajagopalan et al. 1997; Sheather 2004).

In light of all findings related to the kernel type and bandwidth methods, log-transformed Gaussian and Epanechnikov kernels, as well as the Silverman and SJ bandwidth methods, were evaluated in verification analyses to determine optimality for SPP (section 4). With these different KDE parameterizations, however, additional conditions needed to be implemented in the SPP algorithm to maintain continuity in processing. To be specific, historical precipitation rates equal to 0 were replaced with an arbitrary small value (e.g.,  $1.00 \times 10^{-7} \text{ mm day}^{-1}$ ). The reasoning behind this replacement was to ensure that the log transform returns real numbers (i.e., not  $-\infty$ ), because very small precipitation rates allow SPP to capture historical years with dry outcomes while not compromising the integrity of the density estimate. In addition, in cases in which a high number of historical precipitation rates were equal to 0 and/or were not well defined in ARC2, SPP probabilities were automatically adjusted to 100% in accordance with the current seasonal percent-of-normal category, but only when both the mode and mean of historical precipitation rates were equal to 0 and  $<1$ , respectively. This adjustment was considered to be the most sensible one, since these cases often occur when there are very few days remaining in a seasonal period and therefore when there is little to no time for any change in the current precipitation anomaly fields. Last, a climatological dry mask was implemented in which SPP is also only processed over areas that receive more than  $1 \text{ mm day}^{-1}$  for the respective seasonal period. This mask was intended to ensure that SPP be performed over regions where

TABLE 1. Averaged HHP scores from 2006 to 2015 using various KDE parameterizations in SPP for all seasons and regions (identified in the date rows).

Kernel	Bandwidth	Dates within the season or HHP scores				
East Africa (March–May)		15 Mar	1 Apr	15 Apr	1 May	15 May
Gaussian	Silverman	0.5522	0.6223	0.7217	0.8059	0.8883
Gaussian	SJ	0.5534	0.6235	0.7216	0.8048	0.8876
Epanechnikov	Silverman	0.5521	0.6228	0.7219	0.8056	0.8884
Epanechnikov	SJ	0.5537	0.6236	0.7221	0.8046	0.8873
West Africa (July–September)		15 Jul	1 Aug	15 Aug	1 Sep	15 Sep
Gaussian	Silverman	0.6103	0.6588	0.7315	0.8162	0.8824
Gaussian	SJ	0.6102	0.6593	0.7308	0.8165	0.8826
Epanechnikov	Silverman	0.6101	0.6588	0.7313	0.8162	0.8826
Epanechnikov	SJ	0.6097	0.6590	0.7305	0.8164	0.8828
East Africa (October–December)		15 Oct	1 Nov	15 Nov	1 Dec	15 Dec
Gaussian	Silverman	0.5404	0.6177	0.7006	0.7849	0.8633
Gaussian	SJ	0.5369	0.6143	0.6986	0.7831	0.8610
Epanechnikov	Silverman	0.5399	0.6170	0.7009	0.7845	0.8633
Epanechnikov	SJ	0.5363	0.6133	0.6984	0.7833	0.8606
Southern Africa (December–February)		15 Dec	1 Jan	15 Jan	1 Feb	15 Feb
Gaussian	Silverman	0.5723	0.6213	0.6923	0.7804	0.8708
Gaussian	SJ	0.5726	0.6211	0.6934	0.7816	0.8710
Epanechnikov	Silverman	0.5728	0.6213	0.6925	0.7803	0.8709
Epanechnikov	SJ	0.5731	0.6215	0.6939	0.7818	0.8709

rainfall is seasonally active, with sufficient signal to circumvent any errors in the processing.

#### 4. Results and discussion

##### a. Historical reprocessing and probabilistic verification

In determining the optimal KDE parameters for SPP during operational monitoring, the SPP algorithm was reprocessed, using the kernels and bandwidth methods discussed in section 3, over several key monsoonal periods and regions in eastern, southern, and western Africa from 2006 to 2015. No reprocessing prior to 2006 was performed, since SPP still requires a number of historical years (1983–2005) to generate reliable density estimates. For this exercise, verification consisted of calculating the Heidke hit proportion (HHP) scores for probabilistic forecasts (International Research Institute for Climate and Society 2013) and relative operating characteristic (ROC) for each of the three SPP probability categories (Mason 1982). These metrics were regarded as the most straightforward and relevant in measuring how well SPP outlooks correspond to the later observed below-average, average, and above-average seasonal rainfall categories. HHP awards credit (a hit) in cases in which the highest categorical SPP probability matches the observed category by the

end of the season. If more than one category shares the highest probability, partial credit is awarded. Hits were then summed and divided by the total number of forecasts in the domain space. ROC was also regarded as a useful measure because of its ability to show how well SPP discriminates between seasonal outcomes. ROC verification consisted of computing SPP hit and false-alarm rates over varying probability thresholds and where scores were computed from the area under the plotted curves (scores of  $>0.5$  indicate skill). Other verification skill scores, such as the Heidke and Brier skill scores, were considered but were not used here, since their methods imply some relative reference to forecasting performance over that of the climatological normal and SPP outlooks are already climatologically based.

Table 1 shows the averaged HHP scores from 2006 to 2015 using various parameterizations in SPP for all seasons and regions. These parameters include the kernel type  $K$  (either Gaussian or Epanechnikov) and bandwidth  $h$  method (either Silverman or SJ). The most salient observation from Table 1 is that there does not appear to be any distinct advantage in using a particular kernel or a particular bandwidth method in terms of improved HHP verification scores. To be specific, the differences in HHP scores between kernel types and bandwidth methods appear to be negligible at certain

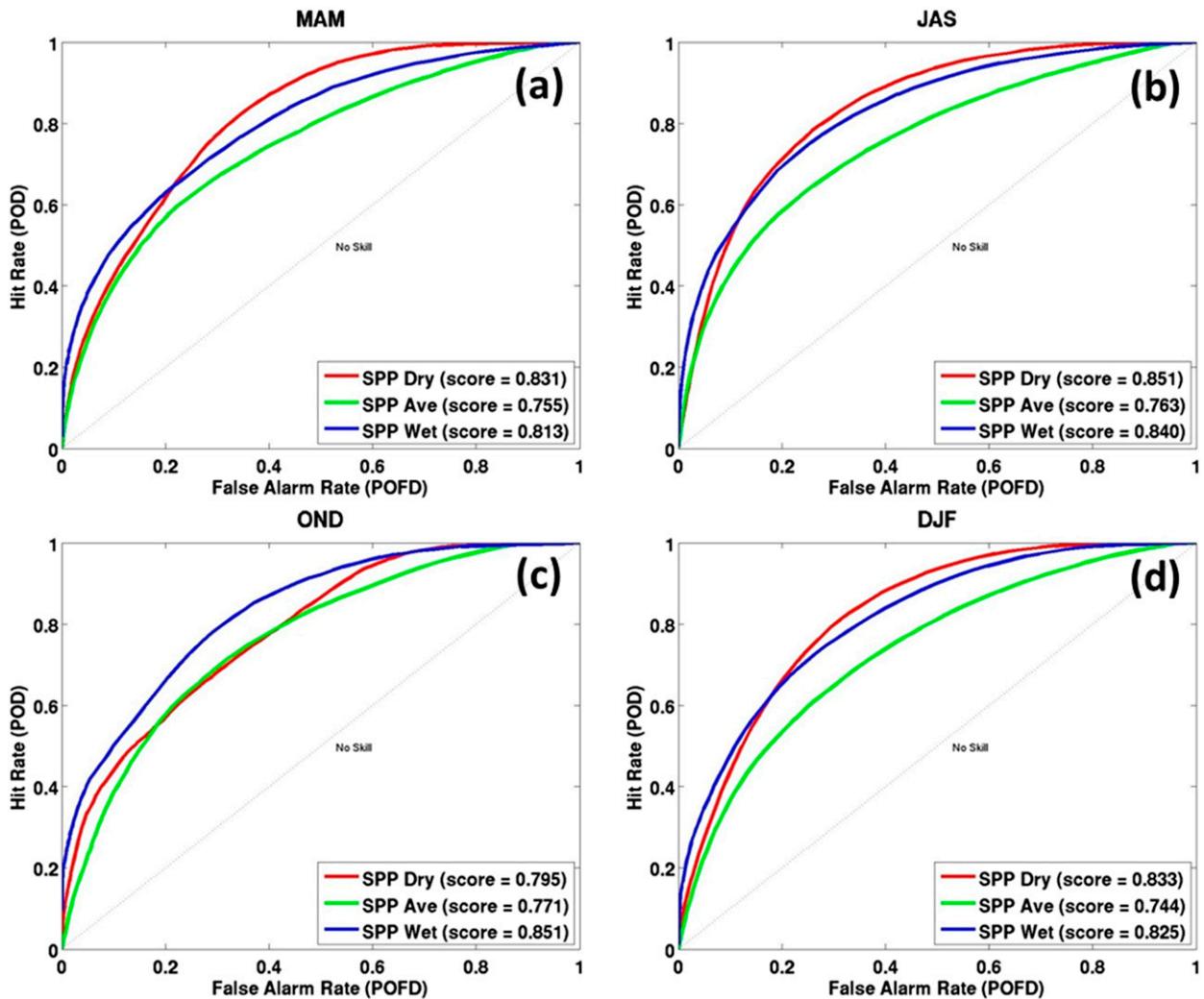


FIG. 2. Averaged ROC verification plots and scores from 2006 to 2015 for each of the three SPP categories by the end of the first month of each season for (a) March–May (label MAM) seasonal rainfall over East Africa, (b) July–September (label JAS) seasonal rainfall over West Africa, (c) October–December (label OND) seasonal rainfall over East Africa, and (d) December–February (label DJF) seasonal rainfall over southern Africa.

points and do not exhibit any discernible pattern across all seasonal stages and all regions. Otherwise, we find that HHP scores increase as the season progresses, but this result is obvious since the lead time shortens and probabilities are converging to 1 by the end of the season. In addition, ROC scores were also found to exhibit little to no variation between the different parameterizations, again suggesting that no kernel or bandwidth method offers a distinct advantage over any of the others (ROC scores for each rainfall category are not shown).

On the basis of these verification results, the Gaussian kernel and Silverman’s bandwidth method were selected for operational SPP implementation purely for the purposes of computational efficiency in daily processing. This KDE parameterization was also used to

take a closer look at two selected seasonal SPP case studies in [section 4b](#). In terms of the overall SPP performance using this parameterization, we find that, for all seasons and regions, on average SPP has the ability to perform consistently well. By the end of the first month and through the midpoint of each season, HHP scores range between 0.6 and 0.7, indicating that at least 60% of the SPP probability fields correctly “verified” in their respective anomaly category ([Table 1](#)). In our verification using ROC, [Figs. 2a–d](#) illustrate the averaged ROC curves and scores from 2006 to 2015 for each SPP category by the end of the first month of each season. Like HHP, ROC scores also increase to 1 as the season progresses (not shown), but we find that all four seasons and SPP categories yield scores of greater than 0.5,

which indicate a moderate degree of skill in SPP during this early time stage for all seasons and domains. It was also encouraging to find that, for all seasons/domains, ROC scores for the dry and wet anomaly categories are slightly higher than those for the average category, since these SPP probabilities are of greater interest in operational monitoring for identifying potential ground impacts. Although not perfectly reliable and discriminatory, both HHP and ROC verification scores suggest a level of confidence in SPP for operational monitoring to provide reasonable guidance of a seasonal rainfall outcome to users before halfway through the season.

### *b. SPP case studies*

Perhaps the most well-known drought case study in recent years was the severe drought that devastated East Africa from 2010 to 2011. This drought had been characterized by the international community as the worst in the last 60 years, triggering famine and the displacement of thousands of people. In our monitoring of precipitation conditions, ARC2 accurately depicted the onset of the drought during the October–December rainfall season and captured the extent of worsening dryness conditions due to poor rains during the following March–May rainfall season in the same region (Novella and Thiaw 2013). Figures 3a–d show the reprocessed SPP for the October–December 2010 rainfall season in East Africa. After one month (1/3) into the season, the percent-of-normal ARC2 rainfall category on 1 November 2010 (Fig. 3a) begins to show developing dryness throughout much of southern Somalia, southwestern Ethiopia, eastern Kenya, and across much of Tanzania. For areas that experienced rapidly developing dryness, SPP probabilities are highest in the below-average category (Fig. 3c), with 80%–90% probabilities over local areas where, from a climate perspective, lesser amounts of rainfall are expected for the remainder of the season, thus reflecting the increased likelihood of drought development and persistence before the end of the season. SPP for above-average rainfall depicts relatively low probabilities (<40%) throughout the entire East Africa domain, with the exception of a few local anomalously wet areas exceeding 90% along the climatological-dry-mask boundary. SPP for average rainfall depicts very high probabilities in eastern Congo and northern South Sudan. Although rainfall in eastern Congo is slightly below average by the end of one-third of the season, this region is climatologically one of the wettest in Africa, suggesting ample opportunity for moisture recovery before the end of December. Analysis of the final seasonal percent-of-normal rainfall categories (Fig. 3b) and HHP verification score map

(Fig. 3d) on 1 November 2010 indicates that nearly 70% of the seasonably active areas in East Africa had SPP probabilities that correctly verified in the respective anomaly categories. For decision-makers in food security, it is interesting to note that the majority of correctly verified hits occurred before midseason over agricultural and pastoral areas in Somalia, Kenya, Ethiopia, and Tanzania, which heavily rely on seasonal rainfall.

In a more recent case study, the core of the rainfall season in southern Africa (December 2014–February 2015) was characterized as being poor and highly erratic. This presented a greater challenge to SPP during operational monitoring because of unusual reversals in the monsoon circulation that were observed throughout the course of the season. By the end of February, a dipole anomaly pattern emerged with the southeastern portion of the African continent having experienced well above-average rainfall and below-average moisture conditions prevailing throughout much of southwestern Africa (Fig. 4b). The evolution of this dipole was not straightforward or gradual, despite what one might expect. In the middle of December of 2014, much of southeastern Africa (i.e., eastern Zambia, Malawi, and western Mozambique) had experienced a delayed onset of the monsoon, raising concerns of anomalous dryness persisting into the season. SPP probabilities for below-average December–February rainfall began to increase and expand throughout the region until extreme rains fell in late December, which led to an abrupt reversal in the SPP probabilities between the above- and below-average anomaly categories. By early January of 2015, SPP probabilities over much of southern Angola, northern Namibia, and the Caprivi Strip did not correctly verify as being below average. Only after an extended dry spell had occurred in January in the region did SPP point to a high probability for below-average rainfall by the end of the season. In Figs. 4c and 4d, we see the SPP probabilities and HHP hit map illustrating that nearly 70% of the seasonably active areas in southern Africa had SPP probabilities that correctly verified in the respective anomaly categories by 15 January. In comparing with the previous case study, it is seen that HHP verification over eastern Africa reached a similar score earlier during the October–December 2010 season.

As seen in this southern Africa case study, there are some areas that experienced seasonal rainfall accumulations that either met or surpassed the climatological normal total well before the end of the season. As a result, SPP depicts 100% probability between average or above-average rainfall categories for these areas, since seasonal anomalies become “locked in” and negative rainfall cannot exist. These instances are numerically



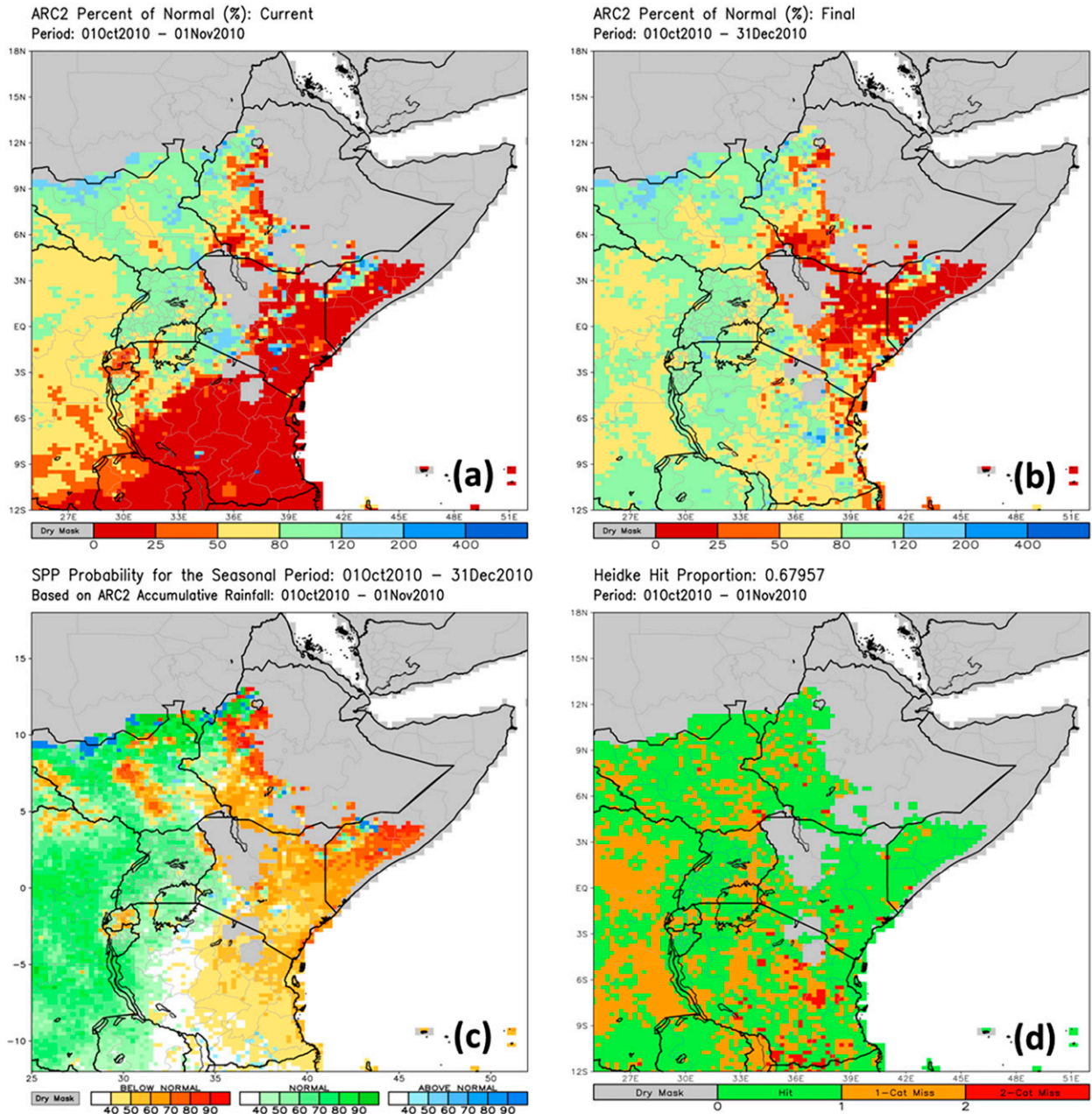


FIG. 3. East Africa spatial maps of (a) percent of normal rainfall on 1 Nov 2010, (b) the percent of normal rainfall at the end of the season on 31 Dec 2010, (c) SPP probabilities for below-average (<80% of normal), average (80%–120% of normal), and above-average (>120% of normal) rainfall reprocessed on 1 Nov 2010, and (d) HHP of verified hits (green) and both misses by 1 category (yellow) and 2 categories (red). All gray areas denote seasonal dry masking.

sound, but such instances may be misleading with respect to the temporal distribution of rains and potential ground conditions relevant to famine early warning activities. For example, if a location receives a large percentage of its normal seasonal rainfall total over a brief period early in the season but then experiences a prolonged dry spell or little subsequent rainfall for the remainder of the season, SPP will return high probabilities

for average to above-average rainfall favorable for ground conditions as based on season-to-date totals. In reality, however, a situation of torrential rains over a short period followed by an irregularly dry distribution of rainfall is likely to lead to early season flooding, followed by more adverse ground impacts such as failed crops, livestock losses, and depleted water availability and possible displacement of local populations. This had

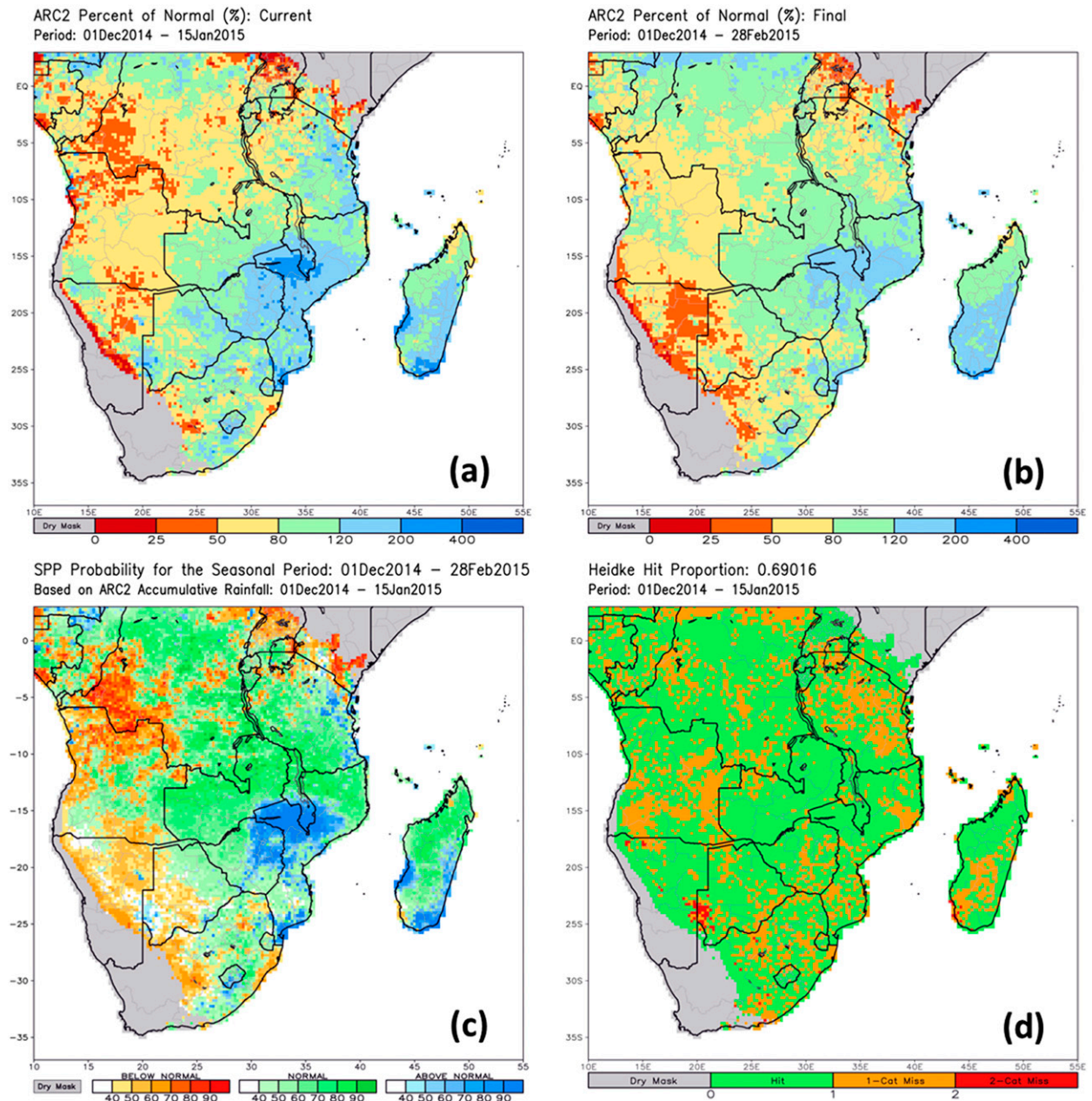


FIG. 4. Southern Africa spatial maps of (a) percent of normal rainfall on 15 Jan 2015, (b) the percent of normal rainfall at the end of the season on 28 Feb 2015, (c) SPP probabilities for below-average (<80% of normal), average (80%–120% of normal), and above-average (>120% of normal) rainfall reprocessed on 15 Jan 2015, and (d) HHP of verified hits (green) and both misses by 1 category (yellow) and 2 categories (red). All gray areas denote seasonal dry masking.

been the unfortunate case for many local areas in eastern Zimbabwe (near Mutare, Zimbabwe) where heavy, flood-inducing rains in December of 2014 were followed by significant suppression in rainfall during January and February of 2015, which resulted in major reductions in crop production. SPP probabilities for average to above-average December–February rainfall had remained very high in the region, and SPP did not point to this dryness

until later stages of the season. While we note that these cases demonstrate some of the weaknesses associated with SPP, they also underscore the need to examine SPP output on the basis of monthly time scales to better capture short-term anomalies in our operational monitoring. Monthly SPP analyses are made available by CPC and are outlined in the following operational output section.

### c. SPP real-time operational output

Consistent with the real-time, daily maps and time series produced at CPC, the SPP algorithm consists of generating probabilistic output for every gridded pixel every day over Africa. To further relax an intense computational environment, the resulting SPP spatial fields are aggregated from  $0.1^\circ$  resolution to  $0.25^\circ$  resolution. Lowering the spatial resolution is not expected to impede SPP's ability to capture regional anomalies and trends. To best cover the seasonality of precipitation over Africa, users will have several options in selecting SPP output of interest. They will be able to choose any base period spanning 1–3 months of ARC2 accumulated rainfall and a probabilistic outlook period ranging from the end of the current month out to 2 subsequent months. This temporal format encompasses not only the key monsoons but also smaller-scale and bimodal rainfall cycles over Africa. All SPP output consists of a set of maps corresponding to the probabilities for below-average ( $<80\%$  of normal), average ( $\geq 80\%$  and  $\leq 120\%$  of normal), and above-average ( $>120\%$  of normal) rainfall for the end of every projection period. As previously mentioned, SPP output maps also include dry masks so that users can easily focus only on seasonally active regions.

## 5. Summary and final remarks

This paper describes a new statistical tool, called SPP, which computes spatial probability maps for seasonal precipitation to finish at rainfall anomaly categories corresponding to below-average ( $<80\%$  of normal), average ( $80\%$ – $120\%$  of normal), and above-average ( $>120\%$  of normal) over Africa. These computations are achieved through the use of kernel density estimation methods that yield probability density functions and cumulative distribution functions that are based on 30+ years of historical ARC2 precipitation for the remaining duration of a monsoon season. The daily, real-time availability of ARC2 used in operational monitoring also permits SPP output to be disseminated to users on the same basis.

In our examination of KDE parameterizations, it was found that using a log transformation was required within a kernel because it prevents issues in density estimation that have to do with bounded properties of a continuous variable like precipitation. It was also found that the implementation of different kernel types and bandwidth methods did not offer any discernible advantage for SPP in terms of improving the probabilities of seasonal outcomes. This result was evidenced in a historical reprocessing of SPP over the major monsoon

regions of eastern, western, and southern Africa from 2006 to 2015 in which the differences in HHP and ROC scores using these parameterizations were negligible. This reprocessing showed that HHP and ROC scores and SPP probabilities converge to 1 by the end of the season, demonstrating how the certainty of a seasonal precipitation outcome for operational monitoring can be statistically quantified. In terms of SPP performance, HHP scores averaged over all regions/seasons ranged between 0.6 and 0.7 after the end of the first month and through the midpoint of each season. At least 60% of the SPP probability fields were correctly verified in their respective anomaly category by one-third of the way into the season, suggesting that there is a reliable degree of confidence in SPP for providing the outcome of seasonal rainfall during operational monitoring. Such information is expected to translate into better decision-making in food security, planning, and response objectives for USAID/FEWS NET.

Some of the weaknesses found in SPP showed that relatively low HHP scores by early to midseason were associated with unusual shifts in the monsoon circulation, as was the case during the December 2014–February 2015 rainy season throughout southern Africa. Somewhat misleading probabilities in the SPP output were linked to untimely and prolonged dry spells that were not captured well in the historical precipitation climatological dataset. On this note, perhaps the greatest weakness in SPP is that it exclusively relies on the quantity/magnitude of accumulated seasonal rainfall and provides no insight on the frequency of seasonal rainfall. Seasonal rainfall frequency and the temporal distribution of rains are significant, because a high number of rain days are required for adequate crop development and pastoral conditions. As a consequence, an extreme-rainfall event that may occur very early in the season is likely to inflate probabilities for above-average rainfall by the end of the season, but it masks what happens following the extreme-rainfall event throughout the remainder of the season. To help alleviate these weaknesses, intraseasonal SPP maps are made available to better highlight short-term precipitation anomalies.

Despite these weaknesses, the main value in SPP lies in its ability to take advantage of the long-term ARC2 climatological dataset to provide simple yet insightful prognostic information for users. Another added value in SPP lies in its flexibility for users. The KDE method also allows probabilities to be determined for any anomaly threshold of interest. We presented here SPP output maps for below-average ( $<80\%$  of normal), average ( $80\%$ – $120\%$  of normal), and above-average ( $>120\%$  of normal), although work is under way to

generate additional SPP output maps for varying degrees of below-average rainfall (e.g., <50%, <25%, and 10% of normal). Doing so further targets the likelihood for conditions of drought or severe drought, which is likely to be of particular interest for FEWS NET. Some potential future work to complement and enhance SPP should focus on how to produce probabilities for the temporal distribution of seasonal rainfall, which would likely include density estimation on the basis of discrete (noncontinuous) variables such as rain days or extreme-rainfall events rather than on season-to-date rainfall totals. It is also conceivable to ingest week-1 and week-2 GFS quantitative precipitation forecast fields within SPP to further enhance probabilistic outlooks. It is expected that SPP performance will improve in the future, since a higher number of years in the ARC2 climatological dataset will help to continue to better define the KDE and its probability estimates.

*Acknowledgments.* Dr. Wassila Thiaw acknowledges the support of the U.S. Agency for International Development (Grant AID-FFP-T-16-00001).

#### REFERENCES

- Bowden, J. H., and F. H. M. Semazzi, 2007: Empirical analysis of intraseasonal climate variability over the Greater Horn of Africa. *J. Climate*, **20**, 5715–5731, doi:10.1175/2007JCLI1587.1.
- Dinku, T., P. Ceccato, E. Grover-Kopec, M. Lemma, S. J. Connor, and C. F. Ropelewski, 2007: Validation of satellite rainfall products over East Africa's complex topography. *Int. J. Remote Sens.*, **28**, 1503–1526, doi:10.1080/01431160600954688.
- Hagos, S. M., and K. H. Cook, 2007: Dynamics of the West African monsoon jump. *J. Climate*, **20**, 5264–5284, doi:10.1175/2007JCLI1533.1.
- Herman, A., V. B. Kumar, P. A. Arkin, and J. V. Kousky, 1997: Objectively determined 10-day African rainfall estimates created for famine early warning systems. *Int. J. Remote Sens.*, **18**, 2147–2159, doi:10.1080/014311697217800.
- Herrmann, S., and K. Mohr, 2011: A continental-scale classification of rainfall seasonality regimes in Africa based on gridded precipitation and land surface temperature products. *J. Appl. Meteor. Climatol.*, **50**, 2504–2513, doi:10.1175/JAMC-D-11-024.1.
- Hutchinson, P., 1992: The Southern Oscillation and prediction of “Der” season rainfall in Somalia. *J. Climate*, **5**, 525–531, doi:10.1175/1520-0442(1992)005<0525:TSAPO>2.0.CO;2.
- International Research Institute for Climate and Society, 2013: Descriptions of the IRI climate forecast verification scores. Columbia University International Research Institute for Climate and Society Tech. Doc., 27 pp. [Available online at <http://iri.columbia.edu/wp-content/uploads/2013/07/scoredescriptions.pdf>.]
- Lall, U., Y. Moon, and K. Bosworth, 1993: Kernel flood frequency estimators: Bandwidth selection and kernel choice. *Water Resour. Res.*, **29**, 1003–1015, doi:10.1029/92WR02466.
- Liebmann, B., I. Blade, G. N. Kiladis, L. M. V. Carvalho, G. B. Senay, D. Allured, S. Leroux, and C. Funk, 2012: Seasonality of African precipitation from 1996 to 2009. *J. Climate*, **25**, 4304–4322, doi:10.1175/JCLI-D-11-00157.1.
- Love, T. B., V. Kumar, P. Xie, and W. Thiaw, 2004: A 20-year daily Africa precipitation climatology using satellite and gauge data. *14th Conf. on Applied Climatology*, Seattle, WA, Amer. Meteor. Soc., P5.4. [Available online at <https://ams.confex.com/ams/pdfpapers/67484.pdf>.]
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- Novella, N. S., and W. M. Thiaw, 2013: African Rainfall Climatology version 2 for famine early warning systems. *J. Appl. Meteor. Climatol.*, **52**, 588–606, doi:10.1175/JAMC-D-11-0238.1.
- Rajagopalan, B., U. Lall, and D. G. Tarboton, 1993: Simulation of daily precipitation from a nonparametric renewal model. Utah State University Rep. WP-93-HWR-UL/003, 24 pp. [Available online at [http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1145&context=water\\_rep](http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1145&context=water_rep).]
- , —, and —, 1997: Evaluation of kernel density estimation methods for daily precipitation resampling. *Stoch. Hydrol. Hydraul.*, **11**, 523–547, doi:10.1007/BF02428432.
- Sheather, S. J., 2004: Density estimation. *Stat. Sci.*, **19**, 588–597, doi:10.1214/088342304000000297.
- , and M. C. Jones, 1991: A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Stat. Soc.*, **53B**, 683–690.
- Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis*. *Monogr. Stat. Appl. Probab.*, No. 26, Chapman and Hall, 176 pp.
- Verdin, J., C. Funk, G. Senay, and R. Choularton, 2005: Climate science and famine early warning. *Philos. Trans. Roy. Soc.*, **360B**, 2155–2168, doi:10.1098/rstb.2005.1754.
- Wand, M. P., and M. C. Jones, 1995: *Kernel Smoothing*. *Monogr. Stat. Appl. Probab.*, No. 60, Chapman and Hall, 224 pp.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 676 pp.
- Williams, A. P., and C. Funk, 2011: A westward extension of the warm pool leads to a westward extension of the Walker circulation, drying eastern Africa. *Climate Dyn.*, **37**, 2417–2435, doi:10.1007/s00382-010-0984-y.