A Seasonal Rainfall Performance Probability Tool for Famine Early Warning Systems over Africa

Nicholas Novella[1,2] and Wassila Thiaw[1]

[1]*Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland*
[2]*Innovim, Greenbelt, Maryland*

1. Introduction

In this extended abstract, we show the development of a new statistical tool which produces 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 Kernel Density Estimation (KDE) methods to compute smoothed, continuous density functions based on more than 30 years of historical precipitation data from the Africa Rainfall Climatology Version 2 (ARC2) dataset (Novella and Thiaw, 2013). Also presented here are various KDE parameterizations tests to determine optimality of density estimates, and thus, performance of SPP for operational monitoring. Verification results using Heidke Hit Proportion (HHP) scores from 2010-2014 suggest that SPP reliably provides probabilistic outcomes of seasonal rainfall anomaly categories by early to mid-stages of rains seasons for major monsoon regions in east, west and southern Africa. SPP has been a useful tool in operational climate monitoring at CPC International desks, where it has helped to provide early warning guidance for developing drought situations, and other related hydrometeorological climate anomalies. This is expected to promote better decision making in food security, planning and response objectives for the United States Agency for International Development / Famine Early Warning Systems Network (USAID/FEWS-NET).

2. Data and methods

This new SPP product solely uses Africa Rainfall Climatology version 2.0 (ARC2) precipitation estimate data over Africa. The features of ARC2 are suited for the development and application of SPP, since its daily resolution and 30+ year historical record allow for a sufficient number of years to quantitatively determine the probability of seasonal rainfall performances. For operational monitoring at CPC, meteorologists have designated six seasonal timeframes over three main domains in Africa. These include the East Africa domain encompassing the Mar-May, Jun-Sep, Feb-Sep, and Oct-Dec timeframes, as well as, the West Africa and Southern Africa domain, covering the May-Sep and Oct-May timeframes, respectively. These timeframes have been useful in capturing the evolution of monsoon rainfall, as they also cover pertinent agricultural calendars and cropping activities on the ground for famine early warning systems.

a. Kernel Density Estimation

The main purpose in SPP lies in determining the probability density function (PDF) of historical precipitation rates from a current point in a season to the end of season. SPP applies Kernel Density Estimation (KDE) methods on the ARC2 30+year climatology in order 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 and unknown density, $f(x)$, the KDE is defined as:

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

_____

Correspondence to: Nicholas Novella, 5830 University Research Ct., Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; E-mail: nicholas.novella@noaa.gov

where n equals the number of historical observations, x(i) are the historical observations, and h is the bandwidth parameter. The selection of the bandwidth parameter, h, and kernel type, K, both have a marked effect on the shape of the estimated density, and more discussion is included later in this section of the paper. The main advantage of KDE resides in how it evaluates point-wise contributions (*i.e.* distances between x and x(i)), and where the summation of kernels converges faster to the true underlying density for continuous random variables like precipitation. If we let x be an array of hypothetical precipitation rates (from 0 to Infinity) required to satisfy an array of percent of normal rains by the end of season, then this will not only render a smoother estimate of the density function, f(x), but the probabilities for each hypothetical precipitation rate can then be determined. Taking the integral of f(x) results in the Cumulative

Fig. 1 Example of the probability density function (upper), cumulative distribution function with SPP probabilities (lower) estimated from KDE from a sample set of historical precipitation rates for the remainder of the season.

Distribution Function (CDF), and it is here where probabilities within specified intervals along F(x) can be ascertained and plotted to render a probability value for each point in space.

To illustrate, let's suppose the following for a given location where: 1) the current seasonal accumulated total is 100mm at $T_{current}$, 2) the current seasonal climatological normal total is 150mm at $T_{current}$, 3) the end-of-season climatological normal total is 500mm at T_{final}, and, 4) the number of days remaining in the season equals 60. While the current seasonal percent of normal anomaly is well below-average at 66%, we would therefore find that threshold precipitation rates of 5.00 mm/day, 6.66 mm/day, and 8.33 mm/day 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 (1983-2012) from $T_{current}$ to T_{final}, as well as, an array of hypothetical precipitation rates required in the future, x, to define the PDF, plotting the Below-Average (brown) and Above-Average (green) threshold rate points along the x axis on both the PDF and CDF curves (Fig. 1), shows that the highest probability (~50%) exists for seasonal rainfall to be in the "Below-Average" category (<80% of Normal) by the end of the season. Also evident is the second highest probability (~28%) for seasonal rains to finish in the "Above-Average" category (>120% of Normal), and the lowest probability (~22%) to finish in the "Average" category (>=80% and <=120% of Normal) by the end of the season.

b. Parameterization: Kernel Type & Bandwidth Selection

In KDE literature, studies by Rajagopalan *et al.* (1997) and Rajagopalan *et al.* (1993) have referenced the implementation of the Epanechnikov kernel instead of using a Gaussian kernel when using precipitation data since it has inherent bounded support to minimize potential boundary effects. However, these studies also showed that boundary issues are ameliorated through the use of a log transformation within the kernel as it prevents any "leakage" of the probability mass extending beyond the boundary (Rajagopalan *et al.*, 1997). Regardless of kernel type selected, this log transformation was considered necessary for SPP to properly handle the fixed lower bound of precipitation, so that f(x) = 0 where x < 0, and f(x) still integrates to one. For bandwidth selection, the method that is most commonly referenced in literature is Silverman's Rule-of-Thumb (Silverman, 1986). However, some studies have suggested that this method may not be aptly suited
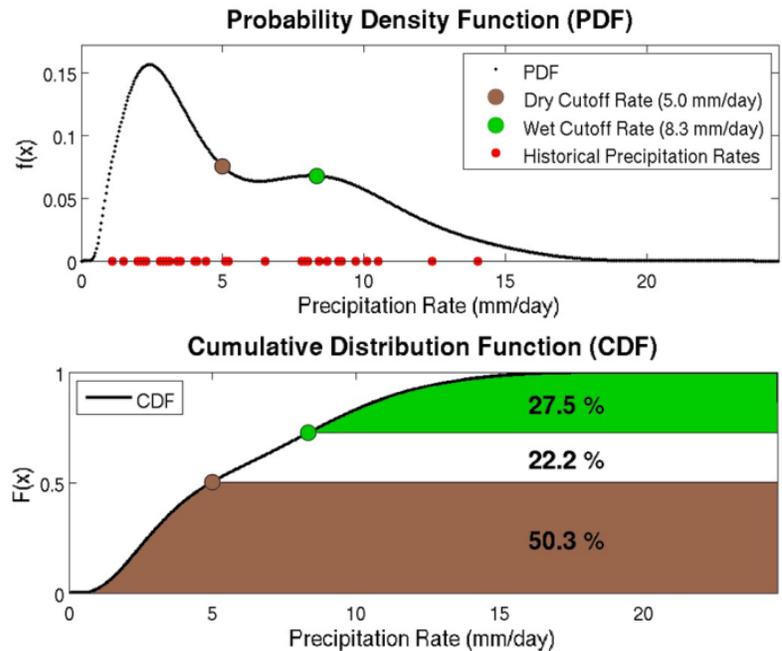
for mutli-modal distributions, and underperformance has been linked to its heavy reliance to assumptions of the underlying distribution (Rajagopalan *et al.*, 1997). As an alternative, the "plug-in" or recursive method of (Sheather and Jones, 1991) (hereafter referred to as SJ) has also been widely described in KDE associated literature. 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 analysis to determine optimality for SPP in the following section.

3. Results

a. Historical reprocessing and verification

In determining the optimal KDE parameters for SPP in operational monitoring, the SPP algorithm was reprocessed using kernels and bandwidth methods, as highlighted in the previous section, over several key monsoonal periods and regions in eastern, southern and western Africa from 2010 to 2014. No reprocessing prior to 2010 was performed, since SPP still requires a high number of years to generate densities. For this exercise, verification consisted of calculating the Heidke Hit Proportion (HHP) scores for probabilistic forecasts (IRI, 2013). This verification metric was regarded as the most straight-forward and relevant in forecasting anomaly categories corresponding to below-average, average, and above-average seasonal rainfall. HHP awards credit (hits) where the highest categorical SPP probability matches the observed category by the end of season. Hits are then summed and divided by the total number of forecasts in space.

Averaged HHP scores (from 2010-2014) using various parameterizations in SPP for all seasons and regions in shown in Table 1. The most salient observation is that there doesn't appear to be any distinct advantage in using a particular kernel, or a particular bandwidth method in terms of improved HHP verification scores, since differences in HHP scores between kernel types and bandwidth methods appear to be quite negligible at seasonal stages. By the end of the first month and through mid-point 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. While not perfect, these scores suggest a level of confidence for operational monitoring, where we can provide reasonable guidance of a seasonal rainfall outcome to users before halfway through the season. Based on these results, the Gaussian kernel and Silverman's bandwidth method was selected for operational SPP implementation purely due their efficiency in daily processing.

b. SPP case studies

Perhaps the most well-known drought case study in recent years was the severe drought that devastated East Africa from 2010-2011. In our monitoring of precipitation, ARC2 accurately depicted the onset of the drought during the Oct-Dec 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). Figs. 2a-d show the reprocessed SPP for the Oct-Dec, 2010 rainfall season in East Africa. After one month (1/3) into the season,

Table 1 Averaged HHP scores from 2010-2014 using various KDE parameterizations in SPP for all seasons and regions (h1=Silverman, h2=SJ).

East Africa (March-May)

Kernel	Bandwidth	15-Mar	1-Apr	15-Apr	1-May	15-May
Gaussian	h1	0.5332	0.6142	0.7153	0.8149	0.8955
Gaussian	h2	0.5345	0.6168	0.7156	0.8133	0.8945
Epanechnikov	h1	0.5334	0.6148	0.7157	0.8146	0.8955
Epanechnikov	h2	0.5345	0.6167	0.7164	0.8136	0.8943

West Africa (July - September)

Kernel	Bandwidth	15-Jul	1-Aug	15-Aug	1-Sep	15-Sep
Gaussian	h1	0.6169	0.6542	0.7354	0.8141	0.8752
Gaussian	h2	0.6167	0.6539	0.7348	0.8138	0.8752
Epanechnikov	h1	0.6169	0.6541	0.7355	0.8140	0.8753
Epanechnikov	h2	0.6159	0.6537	0.7345	0.8137	0.8754

East Africa (October - December)

Kernel	Bandwidth	15-Oct	1-Nov	15-Nov	1-Dec	15-Dec
Gaussian	h1	0.5855	0.6296	0.7141	0.7829	0.8651
Gaussian	h2	0.5830	0.6270	0.7131	0.7808	0.8652
Epanechnikov	h1	0.5860	0.6293	0.7141	0.7824	0.8654
Epanechnikov	h2	0.5834	0.6268	0.7126	0.7814	0.8649

Southern Africa (December - February)

Kernel	Bandwidth	15-Dec	1-Jan	15-Jan	1-Feb	15-Feb
Gaussian	h1	0.5811	0.6323	0.7081	0.7817	0.8866
Gaussian	h2	0.5813	0.6318	0.7099	0.7831	0.8873
Epanechnikov	h1	0.5815	0.6322	0.7085	0.7819	0.8867
Epanechnikov	h2	0.5813	0.6326	0.7109	0.7834	0.8876

the percent of normal ARC2 rainfall on November 1st, 2010 (Fig. 2a) begins to depict developing dryness throughout much 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. 2c), with 80% to 90% probabilities over areas where climatologically, lesser amounts of rainfall are expected for remainder of season, thus reflecting the increased likelihood of drought development and persistence before the end of the season. Analysis of the final season percent of normal rainfall (Fig. 2b) and HHP verification score map (Fig. 2d) on November 1st, 2010 indicate that nearly 70% of the seasonably active areas in east Africa had SPP probabilities that correctly verified in the respective anomaly categories.

In a more recent case study, the core of the southern Africa rainy season, Dec, 2014 – Feb, 2015 has been characterized as

Fig. 2 East Africa spatial maps of (a) percent of normal seasonal rainfall anomaly on Nov 1st, 2010, (b) the final percent of normal seasonal rainfall anomaly captured Dec 31st, 2010, (c) SPP reprocessed on Nov 1st, 2010, (d) Heidke Hit Proportion (HHP) of verified hits (green) and misses (red) on Nov 1st, 2010.

being poor and highly erratic. This had presented a greater challenge to SPP during operational monitoring due to unusual reversals in the monsoon circulation that had been observed throughout the course of the season. By the end of February, a dipole anomaly pattern emerged with the southeastern portion of the Africa continent having experienced well above-average rainfall, and below-average moisture conditions prevailing throughout much of southwestern Africa (Fig. 3a). However, the evolution of this dipole was not straight-forward nor gradual as one might expect. In the middle of December, 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 Dec-Feb 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 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 transpired during January in the region did SPP point to a high probability for below-average rains by the end of the season. In Figs. 3c-d, we see the SPP probabilities and HHP hit map illustrating nearly 70% of the seasonably active areas in southern Africa had SPP probabilities that correctly verified in the respective anomaly categories by January 15th.

c. SPP real-time operational output

Consistent with the real-time, daily maps and time series products updated at CPC, the SPP algorithm consists of generating probabilistic output for every gridded pixel, every day over Africa. In an effort to
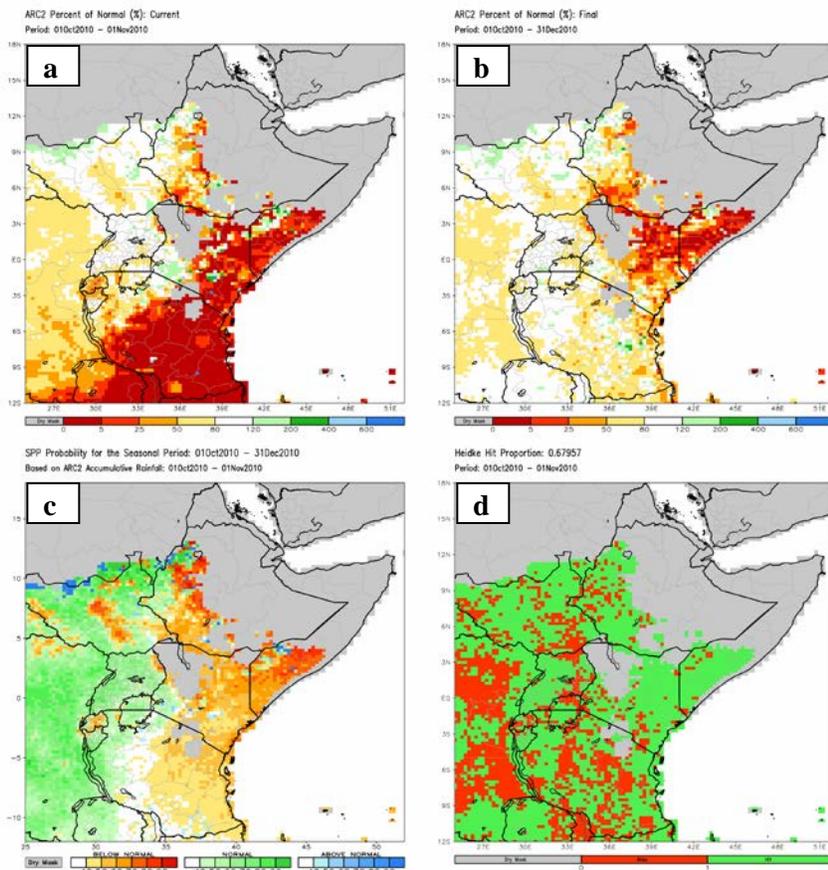
further relax an intense computational environment, the resulting SPP spatial fields are aggregated from a 0.1° to 0.25° resolution. To best cover the seasonality of precipitation over Africa, users will be able to choose any base period spanning 1 to 4 months of ARC2 accumulated rainfall, and a probabilistic outlook period ranging from the end of the current month out to 3 months. SPP output consists of single map depicting all probabilities corresponding to Below-Average (<80% of Normal), Average (>=80% and <=120% of Normal), and Above-Average (>120% of Normal) rainfall for the end of every projection period.

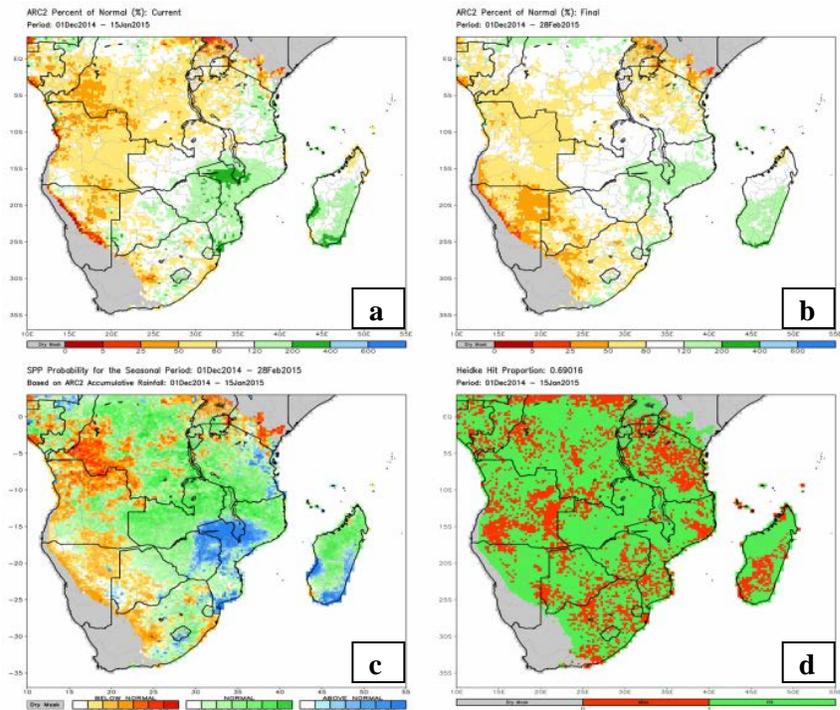## 4. Conclusions

This paper describes a new statistical tool, called SPP, which computes spatial probability maps for seasonal precipitation to finish at rainfall anomaly categories

Fig. 3 Southern Africa spatial maps of (a) percent of normal seasonal rainfall anomaly on Jan 15th, 2015, (b) the final percent of normal seasonal rainfall anomaly captured Feb 28th, 2015, (c) SPP reprocessed on Jan 15th, 2015, (d) Heidke Hit Proportion (HHP) of verified hits (green) and misses (red) on Jan 15th, 2015.

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 (KDE) methods which yield probability density functions (PDF's) and cumulative density functions (CDF's) 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.

Reprocessing and verification results indicate that, on average, at least 60% of the SPP probability fields had correctly verified in their respective anomaly category. This suggests 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.

References

Novella, N., W. Thiaw, 2013: Africa rainfall climatology version 2 for Famine Early Warning Systems. *Journal for Applied Meteorology and Climatology*, **52**, 588-606.

Rajagopalan, B., U. Lall, D.G. Tartoton, 1993: Simulation of daily precipitation from a nonparametric renewal model. *Utah State University Digital Commons Reports Paper 146*.

Rajagopalan, B., U. Lall, D.G. Tartoton, 1997: Evaluation of kernel density estimation methods for daily precipitation resampling. *Stochastic Hydrology and Hydraulics*, **11**, 523-547.

Sheather, S.J., M.C. Jones, 1991: A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B*. **53**, 683-690.

Silverman, B.W., 1986: Density estimation for statistics and data analysis. *Chapman and Hall, New York*.