University of Reading

Advances in the stochastic modelling of satellite-derived rainfall estimates using a sparse calibration dataset

Article

Published Version

Creative Commons: Attribution 3.0 (CC-BY)

Open Access

Greatrex, H., Grimes, D. and Wheeler, T. (2014) Advances in the stochastic modelling of satellite-derived rainfall estimates using a sparse calibration dataset. Journal of Hydrometeorology, 15. pp. 1810-1831. ISSN 1525-7541 doi: 10.1175/JHM-D-13-0145.1 Available at https://centaur.reading.ac.uk/39974/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1175/JHM-D-13-0145.1

Publisher: American Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Advances in the Stochastic Modeling of Satellite-Derived Rainfall Estimates Using a Sparse Calibration Dataset

HELEN GREATREX,* DAVID GRIMES,+ AND TIM WHEELER

University of Reading, Reading, United Kingdom

(Manuscript received 3 August 2013, in final form 1 May 2014)

ABSTRACT

As satellite technology develops, satellite rainfall estimates are likely to become ever more important in the world of food security. It is therefore vital to be able to identify the uncertainty of such estimates and for end users to be able to use this information in a meaningful way. This paper presents new developments in the methodology of simulating satellite rainfall ensembles from thermal infrared satellite data. Although the basic sequential simulation methodology has been developed in previous studies, it was not suitable for use in regions with more complex terrain and limited calibration data. Developments in this work include the creation of a multithreshold, multizone calibration procedure, plus investigations into the causes of an overestimation of low rainfall amounts and the best way to take into account clustered calibration data. A case study of the Ethiopian highlands has been used as an illustration.

1. Introduction

Much of sub-Saharan Africa relies on rain-fed cropping systems for food security; thus, the timely and accurate reporting of weather observations can have sizable benefits for decision makers. Satellite-based rainfall estimates (SRFEs) are proving to be an increasingly valuable tool in modeling African rainfall. They are available in real time and have complete area coverage (Grimes 2008), allowing both local and regional estimates of rainfall. These features make them particularly attractive for end users, and there are several studies showing their use in crop yield forecasts (Thornton et al. 1997; de Wit and van Diepen 2008; Rosema et al. 2010), food security monitoring systems (Velpuri et al. 2014), or in other weather-derived products such as weather-based index insurance (Dinku et al. 2009; Hellmuth et al. 2009). In recent years, a new range of thermal infrared (TIR)-based SRFEs have also been created that extend back over 30 years, making them ideal for climate and trend analysis in a region with sparse ground-based datasets [Tropical Applications of Meteorology using Satellite Data and Ground-Based Observations (TAMSAT) African Rainfall Climatology and Time Series (TARCAT), Maidment (2014); African Rainfall Climatology, version 2 (ARC 2), Novella and Thiaw (2013)].

As technology develops, it is likely that SRFEs will gain ever more prominence in operational decision making over Africa. For example, in recent years the R4 Rural Resilience Initiative (Oxfam America 2014) and Kilimo Salama (Syngenta 2014) insurance projects have directly used SRFEs to insure many tens of thousands of poor farmers in East and West Africa. However, as mathematical algorithms are used to estimate rainfall from satellite images, there will always be some degree of inherent uncertainty in the end products. Quantifying and communicating this uncertainty is extremely important, as it is extremely difficult to assess the skill of a downstream product such as a hydrological or agricultural forecast or insurance contract without knowing the skill of the rainfall estimate that was input into the model. Consequently, it is not enough to just have a statistical description of uncertainty, it is also important for end users to be able to communicate and use the results in a meaningful way.

Denotes Open Access content.

* Current affiliation: International Research Institute for Climate and Society, Lamont-Doherty Earth Observatory, Earth Institute, Columbia University, Palisades, New York.

+ Deceased.

Corresponding author address: Helen Greatrex, International Research Institute for Climate and Society, Lamont-Doherty Earth Observatory, Earth Institute, Columbia University, 61 Route 9W, Palisades, NY 10964-1000.
E-mail: greatrex@iri.columbia.edu

DOI: 10.1175/JHM-D-13-0145.1

© 2014 American Meteorological Society

The process of characterizing the uncertainty on an SRFE is complex, particularly at a daily time scale, because it must take into account the following:

- differences in the spatial scale between the areal satellite pixel and the point-based rain gauge used to calibrate it;
- multiple sources of uncertainty, including in the calibration algorithm, the interpolation method of the gauge data, measurement errors in the gauge data themselves, and the location error of the satellite pixel (many of these errors are heteroskedastic);
- the fact that the statistics of rainfall occurrence are very different from the statistics of rainfall amount, thus making it necessary to use a joint, or mixed, distribution;
- the non-Gaussian nature of rainfall amounts, which are highly skewed and depend on any spatial and temporal averaging (these are in general modeled using a gamma or mixed distribution);
- the spatial intermittency of daily rainfall, that is, it is much more variable on a smaller spatiotemporal scale than when averaged over a larger region or time period; and
- the various spatial correlations of the rainfall and cold cloud duration (CCD) fields, which are likely to vary from location to location and month to month.

The nontrivial nature of the problem described above means that most studies designed to calculate the uncertainty on SRFEs have focused on a probabilistic approach rather than attempting to propagate the error analytically. A two-step process can be employed to achieve this; first, a full statistical description of the SRFE and its associated uncertainty is performed, before the uncertainty is sampled to create an ensemble of potential rainfall fields. An added benefit of this approach is that one might use the resulting rainfall ensemble to drive end-user applications such as crop simulation (Greatrex 2012; de Wit and van Diepen 2008) or hydrological models (Skinner 2013), providing an elegant method of assessing the impact of SRFE uncertainty on end-user needs such as crop yields and regional water availability.

Several studies have attempted to use an ensemble approach to quantify the uncertainty on an SRFE; however, all of the work reported thus far has concentrated on either relatively small homogeneous pilot studies [e.g., the Gambia in Teo and Grimes (2007)] or in a region with dense observational rainfall data [e.g., Europe in de Wit and van Diepen (2008)]. One can envision that a sizable benefit of quantifying the uncertainty on SRFEs could be if this process was operationalized for decision makers to allow them to quickly assess the robustness of the rainfall information. This is particularly the case for regions with a complex climate, sparse rainfall observations, or whose citizens are involved in activities that would be sensitive to reported rainfall amounts, for example, in areas being monitored for food security or reliant on rain-fed agriculture.

To meet this challenge, it is necessary to overcome some additional hurdles. First, the underlying skill of an SRFE can sometimes be poorer over complex topographies and climates, especially if there are sparse observations for calibration and validation (Dinku et al. 2007). It is therefore postulated that additional information could be included into the calibration from other aspects of TIR data. A second challenge is that many ensemble studies have reported a bias in low rainfall amounts, which are key for many end users. Finally, current uncertainty analyses have concentrated on areas with approximately spatially homogeneous gauge coverage, yet in most areas of Africa, the gauge data are clustered or uneven, which is not optimal for determining the spatial structure of rainfall or for ensuring that a calibration works over all microclimates (Bacchi and Kottegoda 1995). If it is going to be possible for ensemble methodology to become a viable part of SRFE uncertainty estimation across Africa, then it is important to consider a more clustered case study over a much larger study region.

This paper aims to address these challenges and presents new developments in the methodology of geostatistically simulating spatially correlated ensembles of rainfall estimates from Meteosat TIR data based on the TAMSAT algorithm. It includes a novel daily calibration algorithm based on general linear model (GLM) analysis, further investigation into the cause of excess low rainfall in the ensembles, and an investigation into how best to take into account clustered calibration data using a multizone calibration. A case study of the Ethiopian highlands has been used throughout as illustration and shows that this modified methodology can be operationalized and applied to an area over 200 times larger than the original Gambia pilot of Teo and Grimes (2007), with a climate that is much more complex. The paper starts with a discussion of the methodology behind ensemble methodology and the improvements and concludes with a validation of the approach against an independent gauge dataset.

2. Choice of methods and case studies

a. The satellite algorithm and ensemble technique

There are many satellite rainfall products and algorithms available to users, each with their own spatiotemporal scales and areas of skill, and there are many studies comparing and validating the different approaches. For example, see Thorne et al. (2001) for southern Africa; Ali et al. (2005), Jobard et al. (2011),

and Roca et al. (2010) for the Sahel; Dinku et al. (2007, 2008) for Ethiopia; Maidment et al. (2012) for Uganda; and Thiemig et al. (2012) for African river basins. Many modern SRFEs incorporate a mixture of both geostationary and low-orbit satellite data, taking advantage of the TIR sensors on the former and the microwave or radar sensors on the latter. However, geostationary estimates based purely on TIR data are of particular use across Africa, because it is possible to use them to create homogeneous 30-yr time series of rainfall, which are a valuable supplement to sparse rain gauge trend datasets (low-Earth-satellite data are only available in more recent years). In addition, TIR-based algorithms have advantages such as long-established user bases, frequent data collection, and they avoid factors such as problematic surface types for microwave sensing.

A TIR-based SRFE methodology of particular note in Africa is TAMSAT, which employs a simple, locally calibrated algorithm based on the 10–12 μm TIR channel of Meteosat (Grimes et al. 1999; Thorne et al. 2001). This has been shown to perform as well as, or better than, comparable or more sophisticated rainfall algorithms over the country, including those incorporating microwave or radar data (Dinku et al. 2007, 2008). In recent years, the dekadal, or 10-daily, TAMSAT algorithm was selected as the base for a 30-yr merged gauge–satellite product owned by the Ethiopian National Meteorology Agency (NMA; Dinku et al. 2013). This forms the base for some of the government of Ethiopia's food security programs. For example, it is widely used in the R4 Rural Resilience Initiative scheme, which directly protects over 12 000 Ethiopian farmers against drought (Dinku et al. 2009; Oxfam America 2014). Because of these factors and because TAMSAT was successfully used in the pilot study of this work, the TAMSAT approach has been used as the basis for this paper.

As discussed in the introduction, less attention has been paid to the question of quantifying the uncertainty on SRFEs, and the nontrivial nature of the problem described above has led most studies to focus on a probabilistic approach. This is often achieved in two stages: first, a full statistical description of the SRFE and its associated uncertainty is performed before the uncertainty is sampled to create an ensemble of potential rainfall fields. It is desirable for such an ensemble to have the following properties:

- each field should to have a realistic spatial correlation derived from observations;
- rainfall statistics at any pixel within a field (e.g., mean rainfall or the probability of rain) must agree with observed rainfall statistics at that location, and therefore the simulation method must take into account the non-Gaussian nature of rainfall and all other factors discussed in the introduction; and
- each ensemble member must be an equally probable and realistic estimate of rainfall over the region.

A number of techniques have been explored to achieve this aim, comprising both conditional and unconditional methods. These include **LU** decomposition (Davis 1987), which has been applied to radar precipitation in the Alps (Germann et al. 2009) and in Mexico (Bouvier et al. 2003); simulated annealing (Deutsch and Journel 1998) applied to SRFEs by Pardo-Igúzquiza (1998) and Haberlandt and Gattke (2004); turning bands (Tompson et al. 1989) applied to SRFEs over Florida by Bellerby and Sun (2005); the spectral method applied to the Global Atmospheric Research Program (GARP) Atlantic Tropical Experiment (GATE) rainfall dataset by De Michele and Bernardara (2005); and the geostatistical approach of sequential simulation (Deutsch and Journel 1998) applied to SRFEs in the Gambia (Teo and Grimes 2007) and in Europe (de Wit and van Diepen 2008). A more detailed review of the different methods available is given in Grimes and Pardo-Igúzquiza (2010). Sequential simulation was chosen for this study because of promising results through trial studies in Africa, documented by Teo and Grimes (2007), and because there has been previous experience in creating TAMSAT ensembles using this method.

b. *The case study and datasets*

Ethiopia was chosen as a case study for this work because of its complex climate and clustered available rain gauge data and because SRFEs are widely used by decision makers there. The study region was chosen as the box bounded by 3°–15°N and 33°–44°E, depicted as a thick line on Fig. 1, where any pixel falling outside Ethiopia was excluded. A resolution of 0.125° was chosen as a balance between computational efficiency and capturing local microclimates.

Two rain gauge datasets were obtained from the Ethiopian NMA, the locations of which can also be seen in Fig. 1. Both of these datasets were interpolated to 0.125° using a double kriging approach suitable for daily rainfall (Grimes and Pardo-Igúzquiza 2010; Barancourt et al. 1992) with climatological variograms (Delhomme 1978; Lebel et al. 1987; Lebel and Le Barbé 1997). The first dataset includes daily rainfall data from 276 gauges from the Oromia region of Ethiopia for the years 2002–06. This underwent significant quality control. Stations were excluded from the dataset if there was a lack of adequate or accurate location information, a statistically unrealistic time series, or if the station had less than 6 months of data.

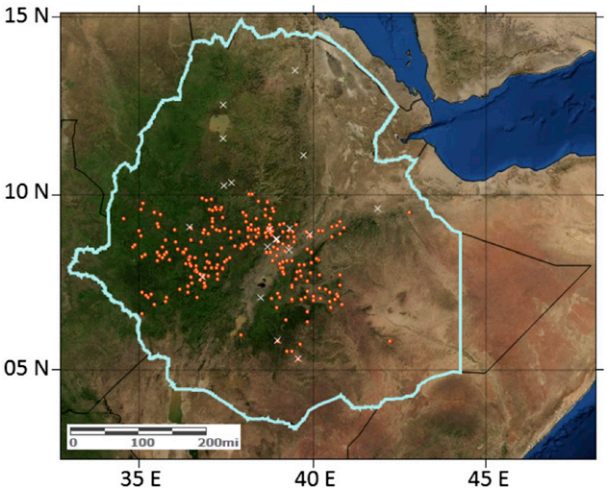

FIG. 1. The location of the calibration (orange circles) and the validation (white crosses) datasets. The thick light blue outline denotes the study area. This map was created using ArcGIS software by Esri.

It is evident from Fig. 1 that although this dataset is spatially dense, it is completely located over the Oromia region of the country, which is mountainous and relatively wet compared to many of the regions in the north and west. In addition, the data are temporally limited as only 5 years of data were available; thus, it does not sample the full Ethiopian climate and could not be used to create a climatology. This is a common occurrence in Africa, where, although some individual countries do have dense rain gauge networks, it is often logistically difficult to access data, especially if one is interested in rainfall amounts across country borders. This is one of the underlying reasons why SRFEs are so important over the continent, so an important question is whether such a clustered dataset is adequate for a full uncertainty assessment across the region.

A small amount of data were available for this study outside the Oromia region, comprising a daily rain gauge dataset from 1994 to 1999, for 20 stations spread over a larger portion of the study area (Fig. 1). These data were also quality controlled and had been used as part of a validation dataset for several publications (Diro et al. 2008, 2011a,b). As the time frame for this dataset did not overlap the Oromia dataset, the data were not dense enough to be used as calibration. It was therefore used as an independent validation dataset to assess the performance of the ensembles across the entire study region. It should be noted that there is now a significant amount of additional rain gauge data available across Ethiopia that could greatly aid further research and validation.

3. Advances in methodology

a. A multithreshold, multizone calibration

1) BACKGROUND

As discussed above, the SRFE selected for this study was the TAMSAT algorithm (Grimes et al. 1999). Similar to many other SRFE algorithms, TAMSAT methodology relies on the assumption that all rainfall falls from convective clouds and that all clouds with high cloud tops are convective. This is the reason why TAMSAT has been shown to work well in areas such as the Sahel, but less well in regions such as coastlines, where other rainfall types dominate. As the 10.8-μm TIR channel of Meteosat has been shown to represent cloud top height, a temperature threshold T can be chosen, whereby it is assumed that a pixel with a cloud top temperature colder than the threshold corresponds to a raining cloud. The length of time any pixel spends below this temperature threshold is denoted as the CCD or D.

The exact value of the threshold temperature is derived locally for manually selected climatologically homogeneous regions using historical rain gauge data. For simplicity, only specific CCD temperature thresholds are considered that correspond to the typical range of convective cloud top temperatures (−30°, −40°, −50°, and −60°C); hence, D_{30} corresponds to the CCD created using a temperature threshold of −30°C. The threshold sampling resolution of 10°C [or 1–2 km in height, depending on the humidity profile of the atmosphere (Cole and Kantor 1963; Parameswaran et al. 2000)] was chosen to maximize the possibility of identifying different cloud types or cloud properties without creating large amounts of redundant and highly correlated data. For example, because many clouds are at least 1 km thick, if one were to sample every 1°C, then many of the thresholds might correspond to the same cloud and significant model redundancy would be expected.

The dekadal rainfall amount R_{10} is then modeled using a simple linear relationship with CCD, where the parameters a_0 and a_1 are derived locally for the calibration regions using historical rain gauge data:

$$R_{10} = \begin{cases} a_0 + a_1 D_T & D_T > 0 \\ 0 & D_T = 0 \end{cases}. \quad (1)$$

It is reasonable to assume that the relationship between rainfall and CCD might change throughout the year as rainfall types shift; thus, an individual calibration is normally conducted for each calendar month.

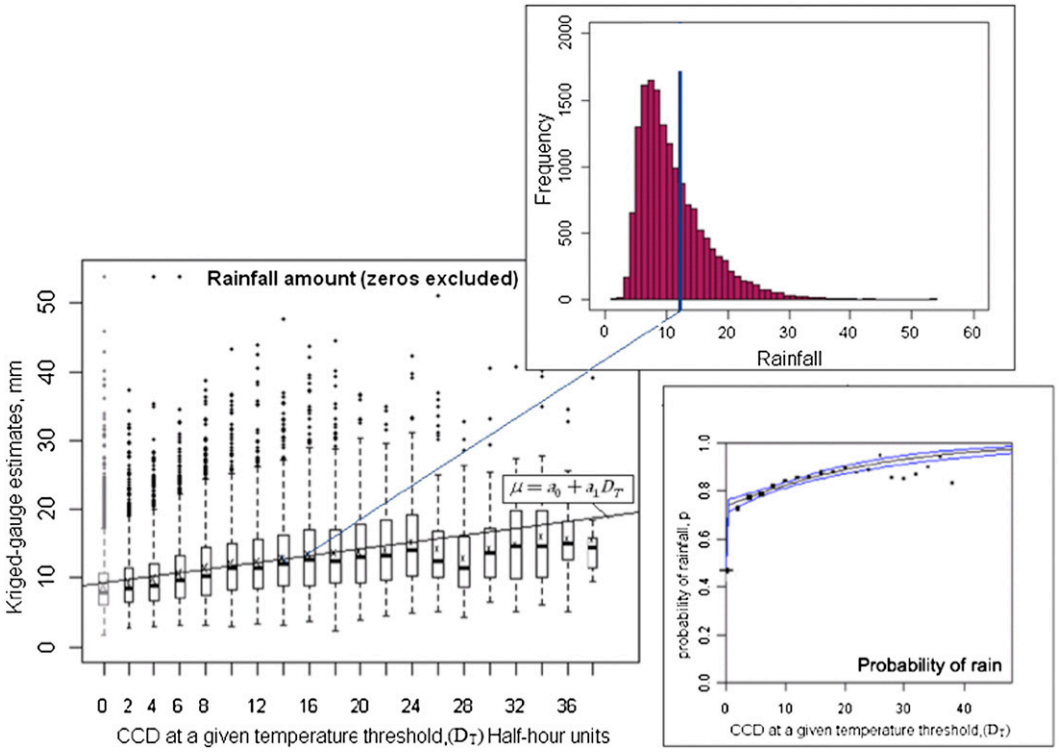

FIG. 2. Schematic of the daily TAMSIM stochastic calibration for June over the Oromia region of Ethiopia.

2) A MULTITHRESHOLD CALIBRATION METHODOLOGY

To take into account the additional complexity of modeling daily rainfall and to allow the uncertainty on the estimate to be calculated, Teo and Grimes (2007) developed an extension of the algorithm using a mixed distribution model referred to as TAMSIM, which is described as follows: for a given value of CCD, the probability p of daily rainfall Z can be modeled as a logistic regression of D_T:

$$P(Z>0 \mid D=D_T)=p=\begin{cases} e^{1/[-(b_0-b_1D_T)]} & D_T>0 \\ p_0 & D_T=0 \end{cases}. \tag{2}$$

If the CCD at that threshold is zero, then the probability of rain is fixed p_0. In pixels where it is raining, the rainfall amount if raining F for a given CCD can be modeled as a gamma distribution g, with shape and scale parameters μ and σ, respectively [Eqs. (3) and (4)], through the regression of D_T against kriged rainfall amount where D_T is measured in 0.5-h units. This work updated the original methodology to model the mean using a gamma GLM with identity link:

$$f(Z \mid Z>0)=F=g[\mu_F,\sigma_F(\mu_F)] \tag{3}$$

and

$$\mu=a_0+a_1D_T. \tag{4}$$

Figure 2 shows an example schematic using June data.

All previous works have assumed that only one temperature threshold of CCD is used in each calibration zone. This is a longstanding assumption for all previous TAMSAT and TAMSIM calibrations, although there has been some investigation into the matter. For example, Grimes et al. (2003) showed that using CCD at multiple thresholds in a dekadal calibration for the Republic of Zambia provided some additional information about whether a cloud was raining or not, but not enough to warrant the additional complexity of the method in that case.

However, the structure of daily rainfall is very different to dekadal averages. In particular, a daily infrared image is much more likely to correspond to one particular rainfall event; therefore, there is potentially more of a problem with cirrus contamination, where the cold, high, nonrainy cirrus cloud in the anvil of a storm is detected as raining. As a dekadal SRFE is generally the sum of several convective events, a smaller proportion of high cloud generally corresponds to cirrus in dekadal algorithms, and cirrus contamination is not usually an issue. Background research has shown that this is a large

problem for daily rainfall over Ethiopia, especially at the edge of the rainy season, where there can often be a large amount of nonrainy high clouds (Greatrex 2012). This work hypothesized that multiple TIR thresholds can provide additional information; for example, a cold CCD threshold might delineate cirrus, while a warmer one might delineate convective cloud. It is important to note that information from other satellite sensors (e.g., microwave or radar) was not considered because a core benefit of the TIR-based SRFEs is a temporally homogeneous calibration that stretches back over 30 years, making the estimates useful for climate studies or trend analysis over Africa.

The single threshold assumption has therefore been reinvestigated, and the linear relationship for mean rainfall given in Eq. (4) was written as

$$\mu = a_0 + a_1 D_{30} + a_2 D_{40} + a_3 D_{50} + a_4 D_{60}. \quad (5)$$

Balanced against the reasonable assumption that a multithreshold approach may provide additional information is the issue of collinearity. To investigate this, a full statistical calibration was performed to find the most skillful fit against the calibration data and to highlight any redundant terms. This investigation also considered different methods of determining calibration zones and is described fully below.

b. A multizone calibration methodology

As discussed above, there are several parameters that are locally calibrated within the TAMSIM algorithm. Each of these is assumed to be constant within regions that have a homogeneous climate, but may differ between different regions; for example, they would be expected to differ between a mountainous region and a plateau. The regional boundaries and calibration parameters are expected to vary from month to month. This was a particular issue for this work as discussed earlier, because despite there being a dense rain gauge dataset available for calibration, the gauges are clustered in a small proportion of the total area for which satellite estimates are needed. An added complication comes from the fact that the area containing the calibration gauges is exclusively within Ethiopia's highlands and so they do not cover all of Ethiopia's diverse landscapes and microclimates.

Previous TAMSAT and TAMSIM calibrations have been performed in areas that either were too small to need multiple calibration zones or where the calibration rain gauges have been more uniformly distributed. These followed a subjective calibration approach where homogeneous zones were created through expert knowledge of the area and study of contingency tables. As discussed in the introduction, this work aimed to test different methods of defining zones in order to investigate whether a clustered calibration dataset was enough to be able to calibrate TAMSIM over the larger study region. Five hypotheses were tested:

- *ONE ZONE:* The study area is left as one zone (the null hypothesis).
- *ISOHYET:* The study area is split using a monthly rainfall contour, as it is reasonable to expect that monthly rainfall amounts might give a clear indication of climatologically homogeneous regions.
- *ELEVATION:* The region is split using an elevation contour. Ethiopia's rainfall is highly dependent on its topography (Dinku et al. 2007); therefore, an elevation contour may provide a good estimate of climatologically homogeneous areas.
- *DEKADAL:* The TAMSAT dekadal calibration zones can be used for the daily Ethiopian calibration. These are a subjective fit, but created using a much wider network of dekadal rain gauges. It is reasonable to assume that daily and dekadal rainfall may have the same climatological spatial structure; thus, a dekadal calibration might be able to incorporate additional relevant information from the extra available gauges.
- *CUSTOM:* There is some other way of defining zones subjectively apparent from examination of spatial plots of dichotomous statistics, for example, a simple east–west split.

For each of these hypotheses, the zone boundaries and thresholds were determined through spatial analysis of dichotomous (rain/no rain) statistics created from the calibration dataset and CCD at different thresholds, against maps overlaid with relevant monthly rainfall and elevation contours. Thresholds of elevation and monthly rainfall and borders of other zones were then subjectively chosen that best split the region into climatologically homogeneous zones. Statistics used in this method were dichotomous bias, probability of detection, and Pierce's skill score. It is recognized that this methodology would be difficult to apply operationally without expert knowledge of Ethiopia's climate; thus, a more automated zone selection procedure is in the process of being developed in future work.

c. Fitting procedure and results

For each month, every combination of threshold and zone was modeled using Eq. (5). Multicollinearity was measured using the variance inflation factor (VIF). Any fits with VIF > 5 were deemed unacceptable and violating predictors were removed. Initial fits were also compared using Akaike's information criterion (AIC) skill score, analysis of variance (ANOVA) tables, and

F tests. Finally, the resulting fits were examined manually using knowledge of Ethiopia's climate as a reality check. To avoid over complication, the dominant temperature threshold in the fit for Eq. (5) was used as the input for the probability of rain described in Eq. (2). This method was then repeated using a leave-one-year-out process to test the sensitivity of the calibration to the input data.

The process described above was applied from March to December and the full results are included in Greatrex (2012). Note that January and February fall at the heart of Ethiopia's dry season and are not part of the cropping calendar; thus, it was decided to exclude these months from the study.

d. Analysis of the calibration results

1) Analysis of the zone choice

The best fit for zone choice during each month shows a progression throughout the year linked to the seasonal rainfall cycle as depicted in Fig. 3. At the height of the main rainy season in July and August, the simple ONE ZONE hypothesis outperformed the others. This is likely to be due to the limited distribution of the rain gauges, as preliminary research indicated that the monthly rainfall amount in these two months over the calibration region was relatively stable and heavy ($>200\,\text{mm month}^{-1}$); thus, all of the calibration gauges fell into one homogeneous climate zone, making it difficult to determine any others. It is important to note that this result is almost certainly due to the data artifact of limited calibration data rather than being "realistic," especially as most areas of Ethiopia are subject to significantly different climate drivers during these months (Gissila et al. 2004; Diro et al. 2011a). It also means that one might expect the product to overestimate rainfall in areas without such heavy rainfall (e.g., southern Ethiopia), although we have no observations available to test this hypothesis. However, this outcome is mitigated by the fact that the calibration zones are simply used to fine tune the TAMSAT calibration using locally available data—most other SRFEs work using a global calibration.

The theory that the ONE ZONE choice was found to be statistically most skillful in July and August because of the placement of calibration gauges is confirmed by the fact that in the other main rainy season months (May, June, and September), the chosen zone hypothesis was the monthly rainfall ISOHYET, which picked out these regions of higher rainfall. The picture was slightly different during the Belg rains (April) and toward the end of the main rainy season (October). Here, the more complex DEKADAL hypothesis had the best fit, suggesting that the situation was complicated enough to warrant the extra information included in the DEKADAL zone choice. Finally, in the drier months (March, November, and December), there was less consistency in the zone choice apart from a general wet-west/dry-east divide, which again follows the main rainfall progression, but is perhaps indicative of a lack of information.

2) Analysis of the multiple threshold choice

Seasonal patterns were also observed in the calibration parameters and in the most skillful combinations of temperature thresholds. The variation in the parameters described in Eq. (5) is shown in Fig. 4 for each month and zone. A leave-one-year-out cross-validation process was used to create a box plot in each case. In each month and zone, the final fit chosen had the lowest AIC value and was significantly the most skillful at a significance level of $p = 0.05$.

A noticeable seasonal cycle can be seen in the intercept parameter, with both Ethiopia's Belg (spring) and Kiremt (summer) rains noticeable. The higher intercept in wetter months could be either assigned to heavy rain from short-duration convective events or rain from warm clouds. There is very little literature into rainfall types in Ethiopia, so it would be difficult to comment on the second hypothesis. However, Birhanu and Alamirew (2008) suggest that Ethiopian rainfall has a higher intensity during short convective events, which perhaps adds to the argument that the high intercept is due to these storms.

The results from this plot show that a multiple calibration was significantly more skillful ($p < 0.05$) for every month outside the main rainy season (July and August). Specifically, Fig. 4 indicates that CCD_{30} is important in nearly every month in the calibration, but that apart from in the wettest months (July and August), a colder temperature threshold is needed to supplement the CCD_{30} fit. This result agrees well with the assumptions stated earlier in the paper, as the single threshold assumption (and other TAMSAT products) have been shown to work well if rainfall comes from purely convective systems, which are dominant during July and August. In other months, it was postulated that there might be more cirrus contamination, which was identified through a multiple threshold approach. As expected, the cross-validation approach showed that the results were less stable in cases with less calibration data. This was either toward the dry season (November and December) or in months with multiple zones and thresholds.

Figure 5 shows the monthly variation in the parameters of Eq. (2) (probability of rain). The main monthly

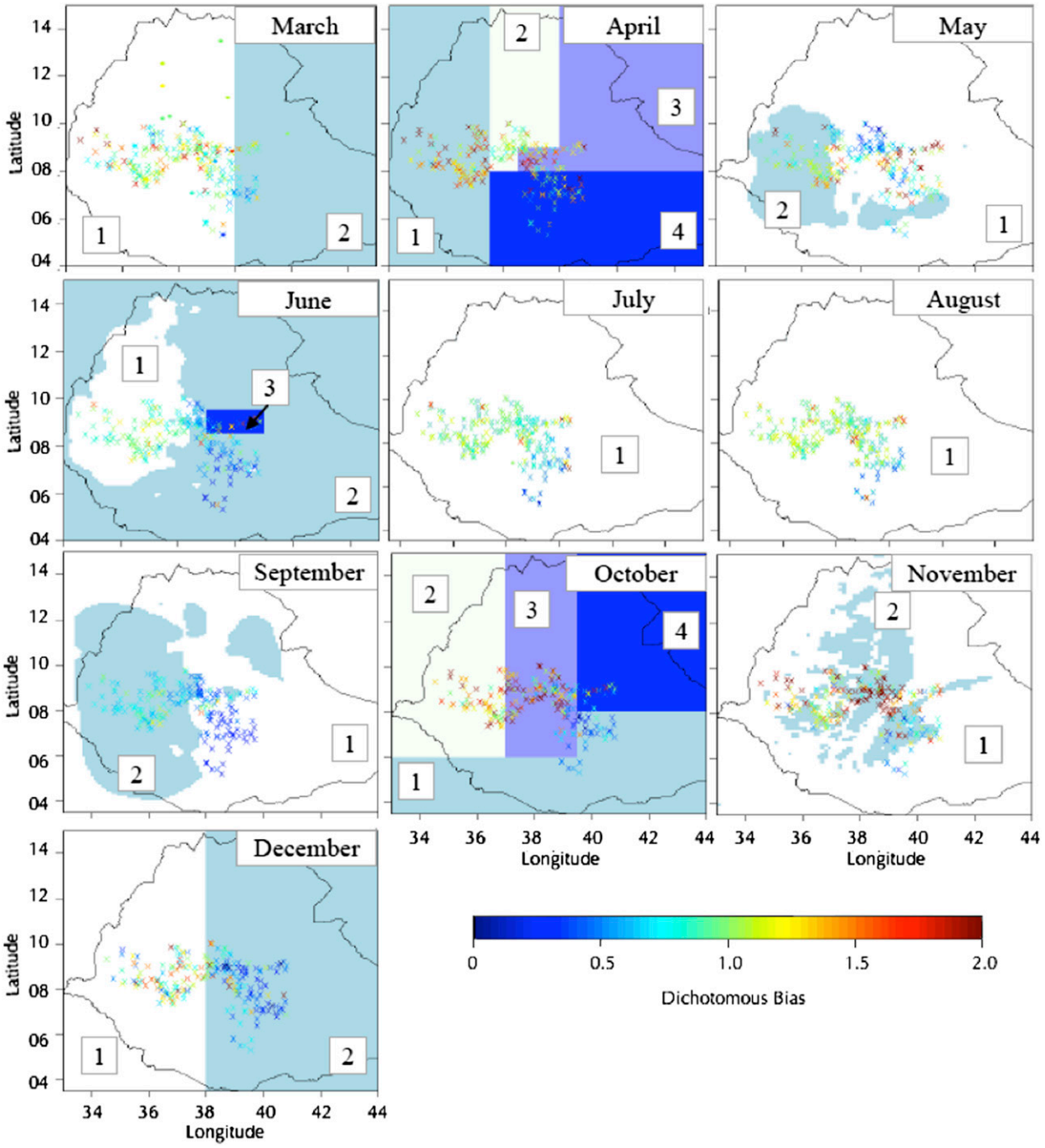

FIG. 3. The zonal divisions that provided the most skillful fit in each month. Superimposed on each subplot are the rain/no-rain bias for CCD threshold 30 for each rain gauge during that month.

driver is shown to be the probability of rain at zero CCD p_0 and in the intercept of the logistic regression b_0. This again could be due to an increase in short-duration convective events in wetter months or because of rainfall from warm clouds. The importance of using multiple calibration zones is particularly apparent in the plot for October, where the southernmost zone 1 has a markedly different p_0 from the other October zones, but similar to the September p_0, because it is the only area in Ethiopia where the rains have not finished. This is a further indication that rainfall isohyets might be a sensible automatic zone choice.

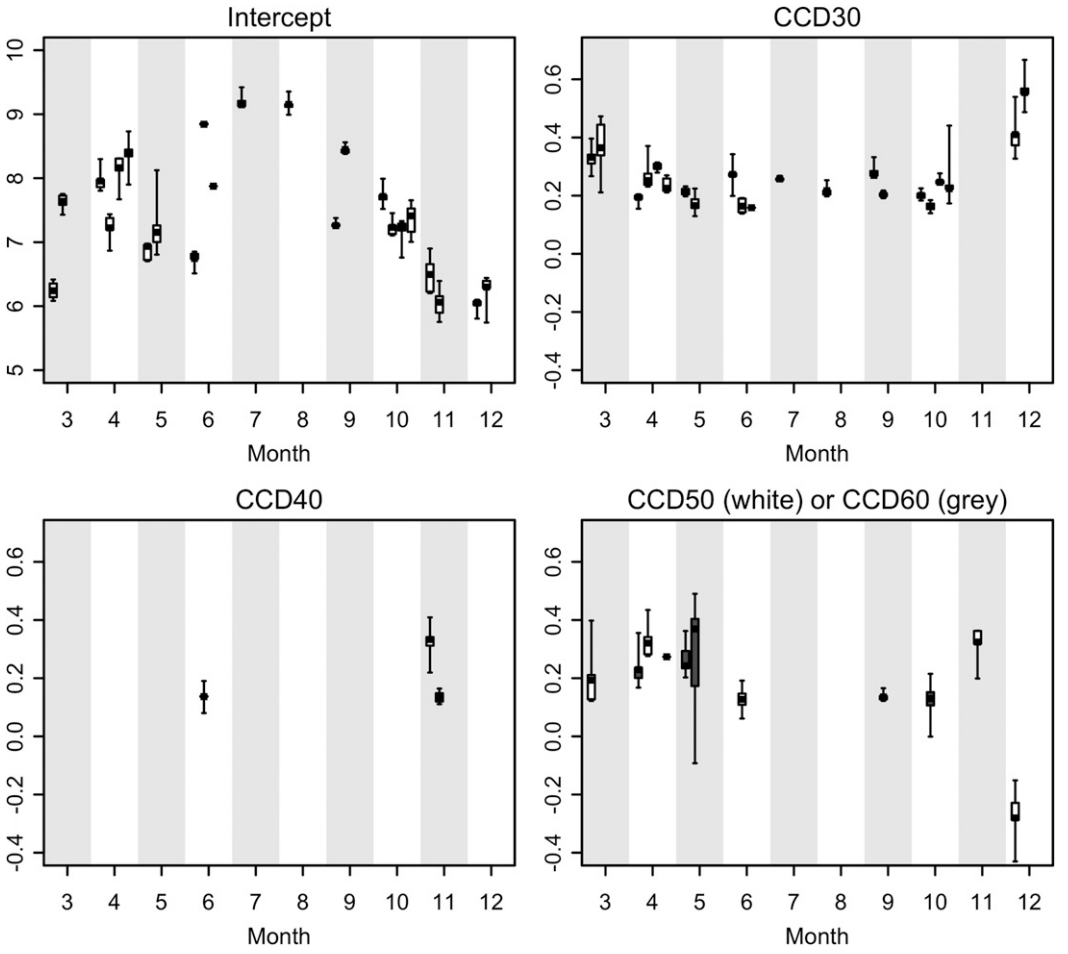

FIG. 4. Parameter a_0 (intercept), a_1 (CCD = 30), a_2 (CCD = 40), a_3 (CCD = 50), and a_4 (CCD = 60) of the multiple linear regression of rainfall amount for each zone and month described in Eq. (5). Within each month, multiple zones are plotted from left to right as numbered in Fig. 3. Each box plot is derived from a leave-one-year-out cross-validation process.

Also observed was the suppression of the probability of rain at high CCD during the drier months of the year (particularly March, November, and December), as shown in Fig. 5 for March. This was attributed to the relatively short time series of calibration data (5 years) and the complex climate found at the edges of the rainy season. Climatically, this can be explained by the presence of a weather type with a blanket of high nonrainy cloud, mixed with the start of the large convective storms in the rainy season. It should be noted that there are very few days recording high CCD in the dry season; thus, there are very little data available to achieve a robust fit or to perform a weather typing experiment. This can be seen in the b_1 plot, which shows a much higher variation in dry months from the leave-one-year-out analysis. Overall, this suggests that care must be taken in using the ensembles at the very edges of the rainy season with a fixed monthly calibration.

e. *Variance modeling*

Thus far, there has been little mention of the variance term in Eq. (3). As summarized in the introduction and described in detail in Grimes et al. (1999), the final uncertainty on an SRFE must take into account more than simply the regression variance of the calibration plot σ_C^2. Instead, this can be considered to be made up of the underlying "true" rainfall variance σ_S^2 plus several other errors, including the block kriging error σ_G^2 and other uncertainties assumed to be negligible ε. Thus, the final uncertainty can be modeled as follows:

$$\sigma_S^2 = \sigma_C^2 - \sigma_G^2 + \varepsilon. \tag{6}$$

Both the regression error and the block kriging error are heteroskedastic and vary with mean rainfall. This was taken into account using the gamma GLM methodology

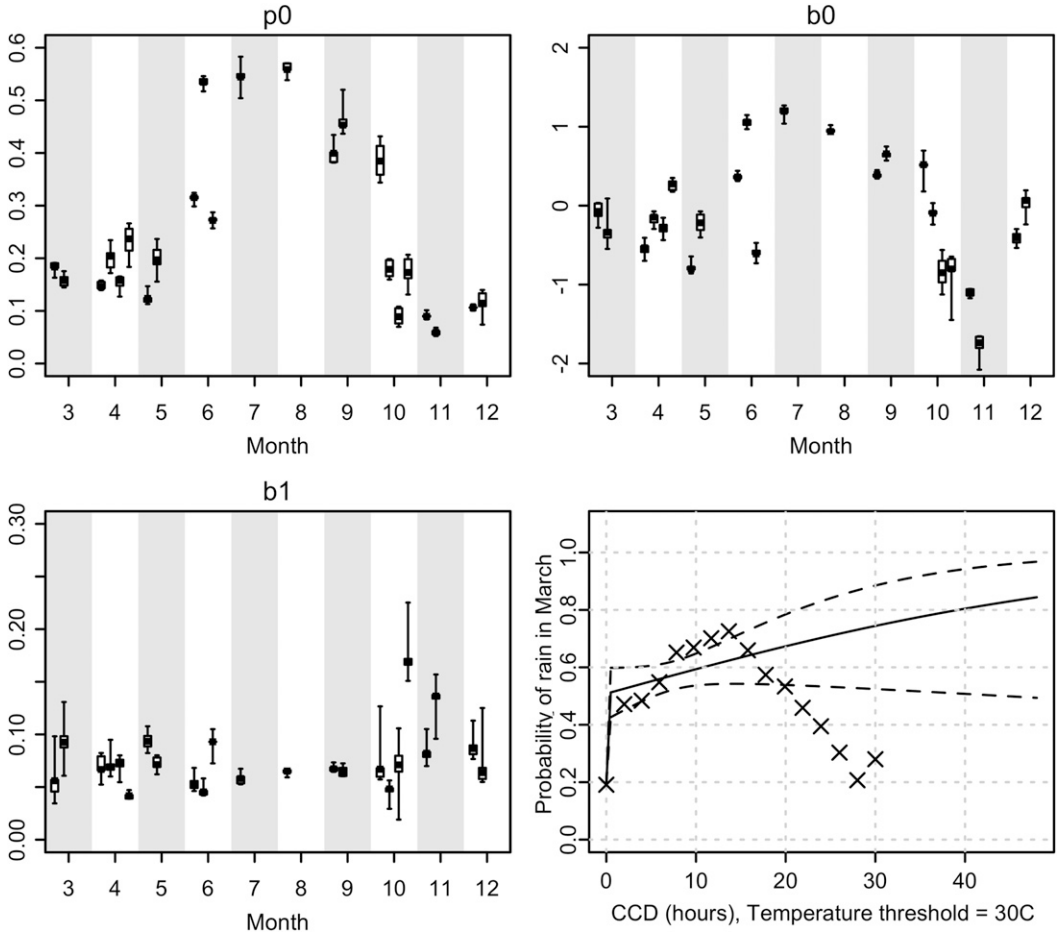

FIG. 5. Parameter (top left) p_0 (probability of rain at zero CCD), (top right) b_0 (logistic regression intercept), and (bottom left) b_1 (logistic regression gradient) as described in Eq. (2). Within each month, multiple zones are plotted from left to right as numbered in Fig. 3. Each box plot is derived from a leave-one-year-out cross-validation process. (bottom right) The full logistic regression for March (solid line); the dashed lines indicate 90% confidence intervals.

for the regression error and as an empirical power-law function of μ for the block kriging error:

$$\sigma_G^2 = \kappa\mu^{\theta}, \tag{7}$$

where κ and θ are empirical constants determined using an iterative least squares approach (Caroll and Ruppert 1988). Other studies employing sequential simulation on smaller datasets have used methods including maximum likelihood estimators (Skinner 2013) and the method of moments (Teo and Grimes 2007) to fit the gamma distribution, which could be employed in future studies if needed.

There is evidence to suggest that ε is not negligible at low values of rainfall, where a combination of low regression error and high kriging error can often lead to rainfall with negative errors. This is clearly unrealistic; thus, the missing error has been attributed in the previous studies to a collocation error σ_l^2, which can be split into accuracy with which the geographical location of any pixel is known and measurement error from the TIR sensor on board the satellite. This is extremely difficult to measure analytically and so was modeled as a constant equivalent to an uncertainty in pixel location to plus or minus one pixel. Thus, the final error analysis was presented as

$$\sigma_F^2 = \max[\sigma_l^2, (\sigma_C^2 - \sigma_G^2 + \sigma_l^2)]. \tag{8}$$

The results from the validation of this work suggest that this assumption does still not fully characterize the uncertainty at low CCD as discussed later in the paper.

f. *Ensemble methodology*

It should be noted that the TAMSAT calibration described by Eqs. (2), (3), (5), and (8) does not lead directly to a rainfall estimate. Instead, the equations describe the statistical properties of the relationship

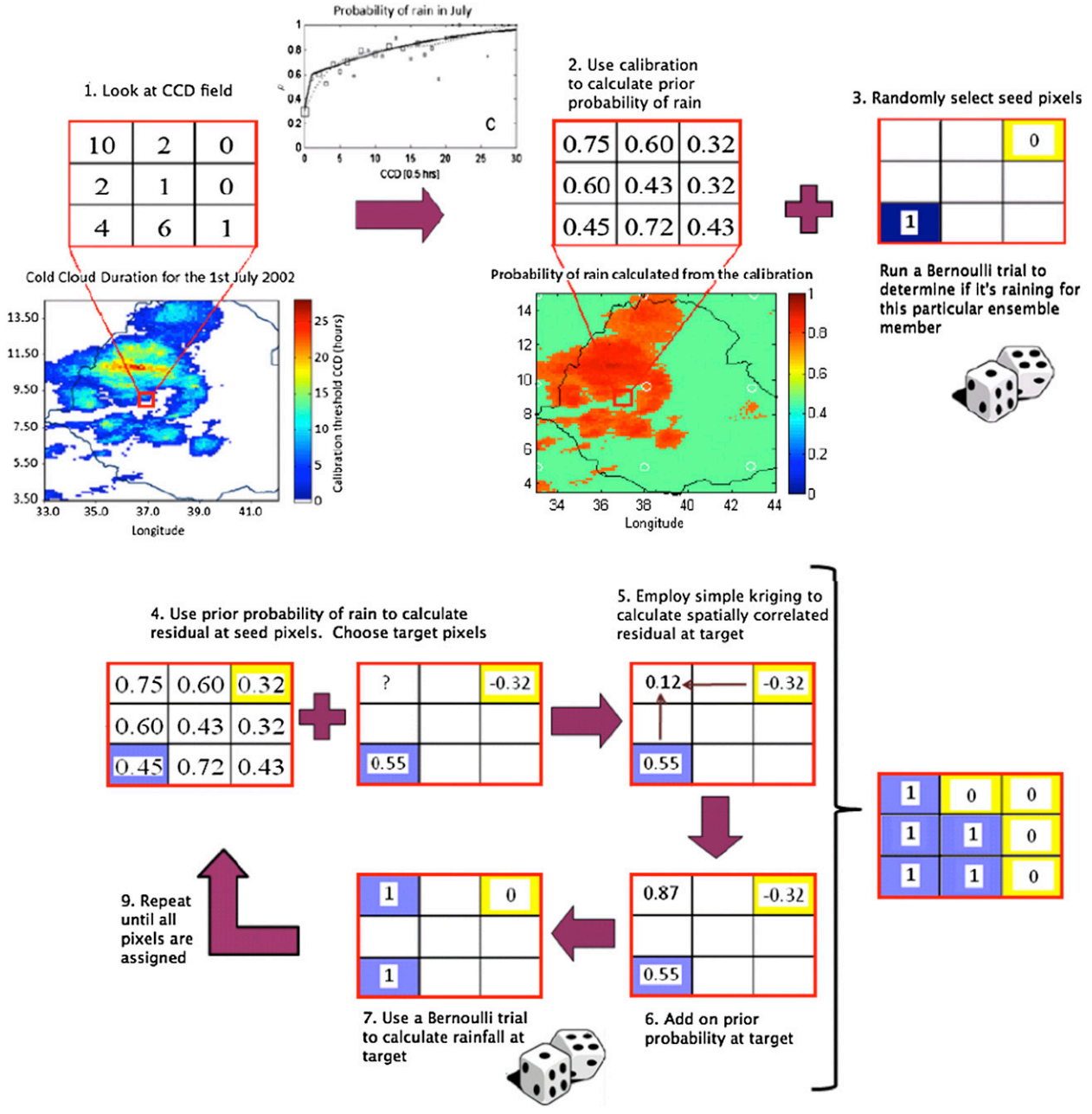

FIG. 6. Sequential indicator simulation.

between CCD and rainfall, for example, the probability of rain or the shape of the rainfall amount distribution. A probabilistic approach can then be taken where a large ensemble of potential rainfall values is generated for a given CCD using the relationship

$$Z_j = I_j F_j . \tag{9}$$

Here I_j is a binary variable representing whether it is raining or not at pixel j, and F_j represents rainfall amount at the pixel. Specifically in this case, the geostatistical approach of sequential indicator simulation was used to create the resulting ensemble of rainfall occurrence and sequential Gaussian simulation was used to create the rainfall amount if raining, including a reversible normal scores transformation to include the gamma distribution of the data. This process is described mathematically in full in Teo and Grimes (2007), de Wit et al. (2008), and Greatrex (2012) and is summarized in Figs. 6 and 7. This approach allows the generation of an

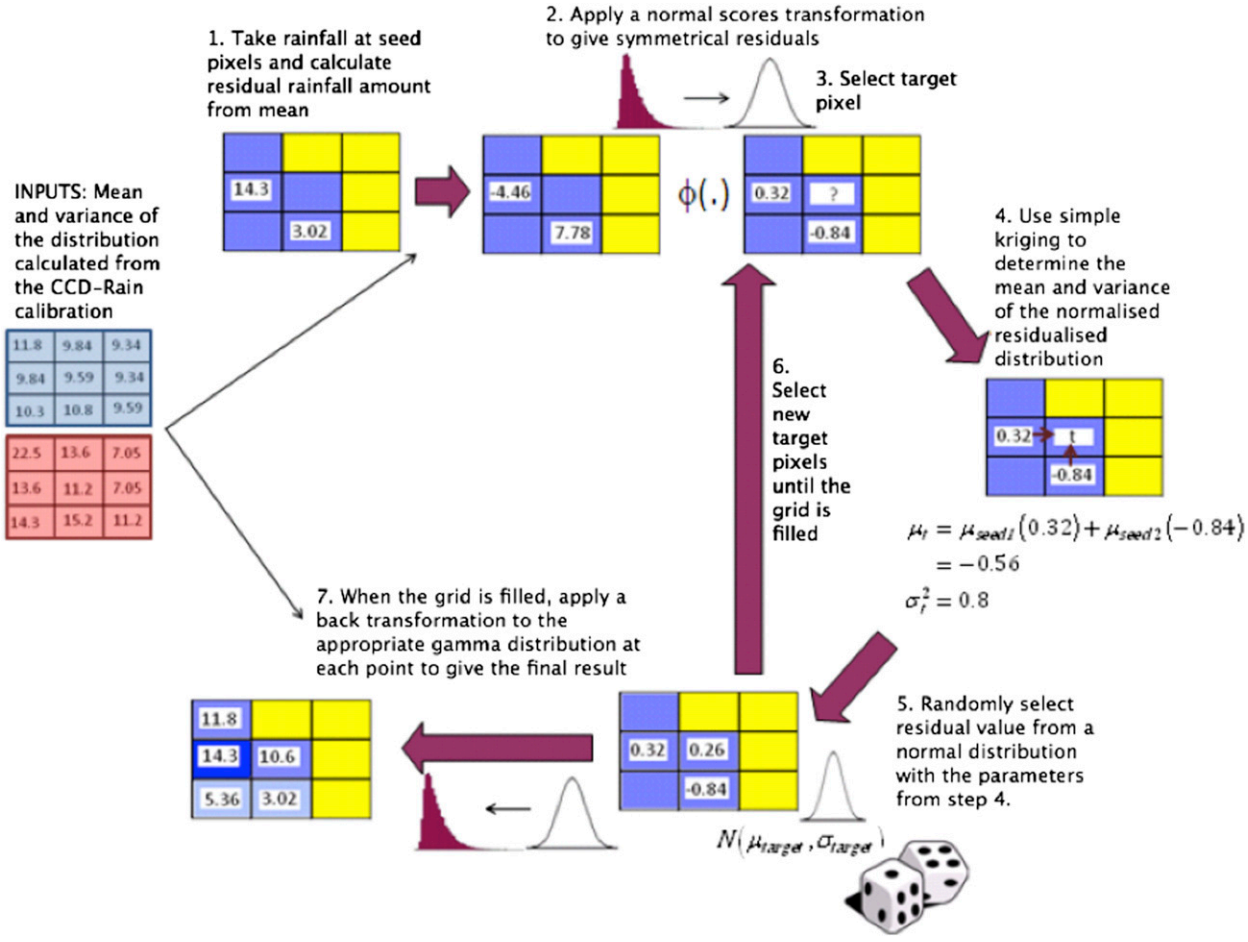

FIG. 7. Sequential Gaussian simulation including a normal scores transformation.

ensemble of rainfall fields based on the stochastic calibration at each pixel, but with a spatial correlation structure derived from observations within each ensemble member.

g. *Ensemble creation*

A TAMSAT daily ensemble was created using the methodology described above at a resolution of 0.125° for the years 1994–99, then compared to the validation gauge dataset (also kriged to a resolution of 0.125°). An ensemble size of 200 was selected through preliminary research.

A variety of different checks were performed to ensure that the ensembles were internally consistent and were faithfully reproducing model parameters (Greatrex 2012). This analysis uncovered the feature that the ensembles overestimate the occurrence of low rainfall. Hypotheses investigated to explain this include the following.

- As discussed earlier in the paper, the error distribution for low rainfall amount contains a simple approximation at low CCD, where the statistical description of the error analysis mathematically breaks down. Further investigation into these errors shows that, unlike original studies in the Gambia, where they were rarely employed, in Ethiopia the collocation error played a role in the low rainfall variance in every month. This is a significant finding and suggests that much more research needs to be employed in understanding the full error analysis from satellites.
- Teo and Grimes (2007) suggested that excess low rainfall was caused by a tendency to overestimate the probability of rain for CCD = 0 in those cases for which zero CCD covers most of the domain—a situation that occurs more frequently at the beginning and end of the rainy season. No evidence of this was seen in this study, suggesting that the result is more complex than originally expected.
- The excess rainfall might be caused by an incorrect parameterization of the relationship between the probability of rain and CCD. In particular, it is reasonable to expect that the probability of rain for a given

value of CCD would be lower during a dry spell than a day falling in a rainy period. Consequently, a Markov analysis was conducted that showed that although the situation is complicated and varies from month to month, CCD on a given day does not depend on whether previous days were rainy. This was especially the case during dry months.

The exact reason behind the overestimation of low rainfall amounts is likely to be a combination of these and other factors and in order to address it properly, a significant amount of time would have to be spent reinvestigating the calibration procedure. This was beyond the scope of this work; thus, an empirical correction was applied to the existing dataset. This involved, at a pixel scale, randomly setting 75% of ensemble members to zero for any pixel day with an ensemble mean of less than 2.5 mm. Importantly, this correction was not found to significantly affect the spatial scale of individual ensemble members. It is suggested, however, that his correction should be reevaluated if the ensembles are to be applied to a new region. Finally, it should be noted that this bias could also be corrected through postprocessing, before it was input into end-user models such as crop simulation model (Ines and Hansen 2006).

4. Validation

This section first considers a validation of the ensemble mean before moving on to describe a validation of the full ensemble dataset. Only pixels containing validation gauges (Fig. 1) were used in the validation. The validation dataset is completely independent of the calibration dataset, covering different years and including areas in very different climates in the calibration. This allowed us to explicitly test whether it is possible to extend the ensemble methodology outside the temporal and spatial extent of a clustered calibration dataset. A leave-one-year-out cross-validation process was also performed using the calibration dataset to assess instabilities in the calibration parameters.

a. Ensemble mean comparison with other SRFEs

The mean of the TAMSAT ensembles corresponds to the ''deterministic best guess'' of the method; thus, it is interesting to compare it to observations and other SRFEs. A comparison between the ensemble mean and kriged gauge rainfall (at a daily 0.125° scale) is shown in Fig. 8 and the parameters of the fit for each case are recorded in Table 1. For ease of comparison, the statistics used for evaluation match those described in another satellite comparison for Ethiopia (Dinku et al. 2008). These comprise the adjusted coefficient of determination R^2_{adj}, root-mean-square error (RMSE), relative RMSE (RMS), and the multiplicative bias (BIAS). Each month exhibits a relatively poor fit on a daily basis, with adjusted R^2 values ranging from 0.09 to 0.27; however, these are still within the same range as those reported for other daily satellite estimates in Dinku et al. (2008), including Tropical Rainfall Measuring Mission (TRMM) 3B42 (Huffman et al. 2007), the National Oceanic and Atmospheric Administration (NOAA)/Climate Prediction Center (CPC) African Rainfall Estimation Algorithm (RFE; Herman et al. 1997), the CPC morphing technique (CMORPH; Joyce et al. 2004), and Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Network (PERSIANN; Sorooshian et al. 2000). This is particularly promising because the ensemble mean is being compared to gauge data at a resolution of 0.125° rather than 0.25°. In addition, with the exemption of December (BIAS = 3.03), the results show a low bias in most months, ranging from 0.84 to 1.39, which outperforms many other estimates in Dinku et al. (2008). Figure 9 shows the results of applying the leave-one-year-out calibration process to these parameters of fit.

As mentioned above, Table 1 and Fig. 8 show that the ensemble mean has a worse fit in the two dry months, November and December. Figure 9 also shows that the fit is much more sensitive to the amount of calibration data in these months, with a much wider variation in skill scores. This strongly suggests that the poor performance is due to a lack of calibration data, which is expected because of fewer rain days. Therefore, there is evidence in future work to suggest that a single ''dry season'' calibration might perform better than individual months. However, this result is less important than earlier months when considering that the end use of the product is for crop simulation modeling and that November and December fall at the end of the growing season, when crops are less sensitive to rainfall.

No dichotomous statistics were included in the ensemble mean validation because, by its nature, an ensemble mean of 200 rainfall pixels will almost always be rainy. Consequently, the ensemble mean by itself can never be considered as a deterministic estimate of daily rainfall if one is interested in rainfall occurrence. However, its promise and comparative skill in predicting rainfall amounts suggest that it may be worth future studies modifying the approach to create such a product.

b. Validation of the probabilistic ensemble

1) CUMULATIVE FREQUENCY ANALYSIS

The distribution of daily rainfall from the ensemble was assessed using a reliability plot approach, where

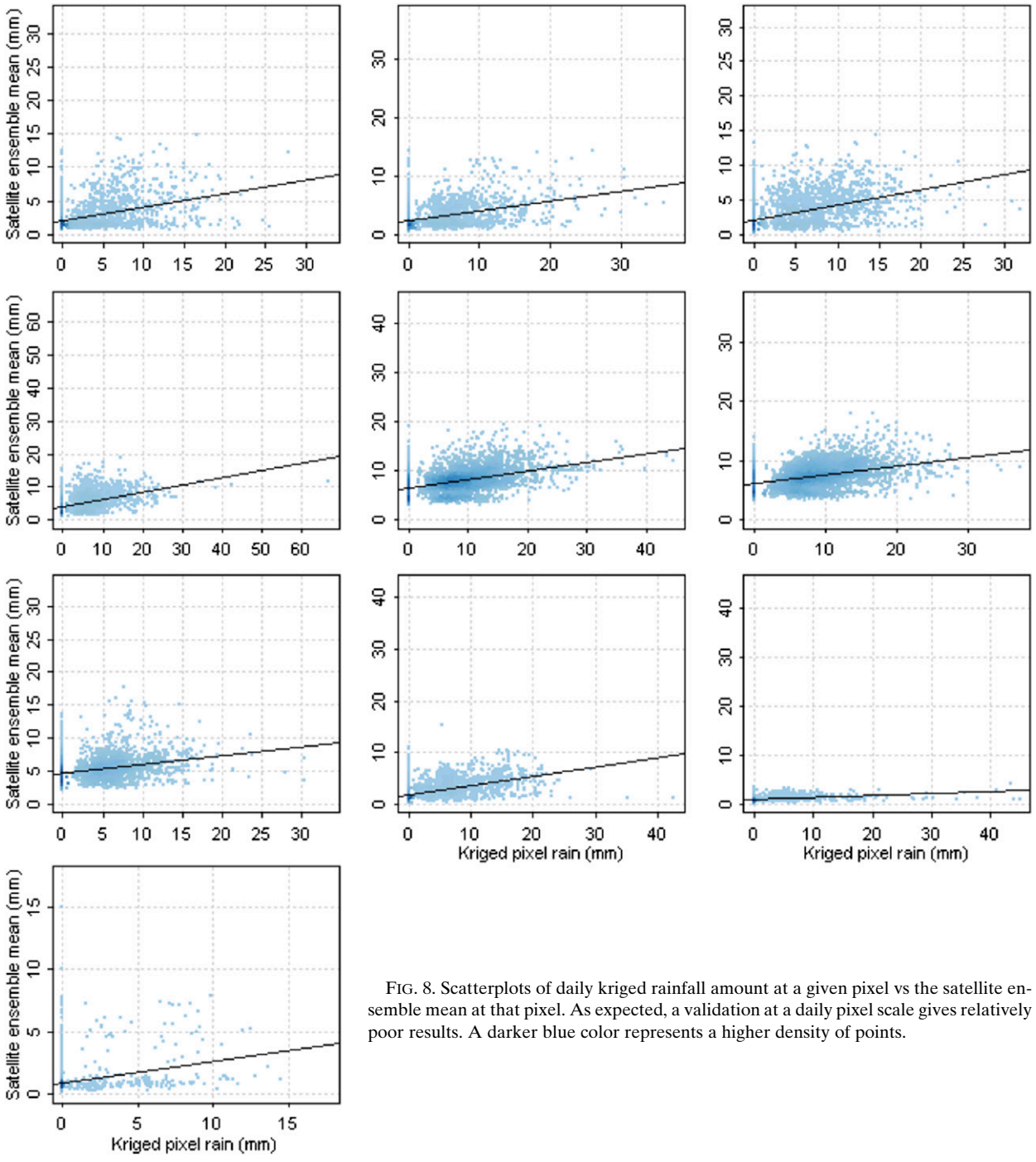

FIG. 8. Scatterplots of daily kriged rainfall amount at a given pixel vs the satellite ensemble mean at that pixel. As expected, a validation at a daily pixel scale gives relatively poor results. A darker blue color represents a higher density of points.

coincident simulated and observed probabilities of rainfall exceedance were compared for different rainfall thresholds, as shown in Fig. 10. Probability bins with less than five observed gauge estimates have been omitted and less than 25 have been marked in gray. Note that the top four plots show the probability of having less than or equal to the rainfall threshold, while the bottom two plots show the probability of receiving greater rainfall than the threshold. The plot shows that distribution of rainfall from the kriged gauge values is extremely well replicated by the ensemble. Particularly encouraging is the $P(Z \leq 0)$ plot on the top left, because it shows that the ensemble is correctly replicating the occurrence of dry days. The fit appears to be slightly less sound at high

TABLE 1. Statistics of the linear fits shown in Fig. 8. These have also been compared with the same statistics taken from a range of products in Dinku et al. (2008). The statistics in the Dinku et al. validation were for July–September; thus, these months have been highlighted in boldface in the table. The TAMSAT ensemble mean shows a low bias in comparison to the other estimates. For intercept and gradient, the standard error (s.e.) is given in parentheses.

Month	Intercept (s.e.)	Gradient (s.e.)	p value	R^2_{adj}	RMSE	RMS	BIAS
Range of statistic	—	—	—	0–1	0–∞	0–∞	−∞–∞
"Perfect" value	—	—	—	1	0	0	1
Mar	2.06 (0.03)	0.20 (0.01)	<0.001	0.17	2.00	1.48	1.23
Apr	2.37 (0.03)	0.16 (0.01)	<0.001	0.16	2.13	1.58	1.07
May	2.13 (0.04)	0.22 (0.01)	<0.001	0.23	1.98	1.61	0.93
Jun	4.19 (0.06)	0.22 (0.01)	<0.001	0.18	2.56	1.08	1.37
Jul	**6.56 (0.06)**	**0.17 (0.01)**	**<0.001**	**0.19**	**2.72**	**0.74**	**0.98**
Aug	**6.21 (0.05)**	**0.15 (0.01)**	**<0.001**	**0.17**	**2.33**	**0.76**	**0.96**
Sep	**4.75 (0.05)**	**0.13 (0.01)**	**<0.001**	**0.09**	**2.39**	**0.85**	**1.39**
Oct	1.85 (0.03)	0.18 (0.01)	<0.001	0.27	2.00	1.96	0.84
Nov	0.85 (0.01)	0.04 (<0.01)	<0.001	0.09	0.50	3.86	0.92
Dec	0.86 (0.01)	0.16 (0.01)	<0.001	0.10	0.90	1.79	3.03
Range from satellite products discussed in Dinku et al. (Jul–Sep)			—	**0.1–0.2**	—	**1.5–2.6**	**0.7–1.6**

rainfall thresholds, where there are more points with a high simulated rainfall probability compared to observed (suggesting that the ensemble distribution has a longer tail). However, very little data went into the majority of these points because high rainfall is a relatively rare event. When a weighted linear fit is applied, the results appear to be unbiased, as shown by the red line on the graph.

Finally, there were more pixel days where observations suggest that there was a small chance of less than 10 mm of rainfall compared to the ensemble. This can be explained when comparing the monthly distributions of observed and forecast rainfall (Greatrex 2012). Most months showed an overestimation of low rainfall (especially during drier months) and an underestimation of high rainfall (especially during wetter months). They also show that, in general, the peak of the gamma distribution is approximately 10 mm; thus, values of rainfall of approximately 10 mm are much more likely to be randomly selected by the stochastic process.

2) SPATIAL DISTRIBUTION

Experimental indicator and positive rainfall amount variograms were calculated from the Oromiya dataset and from a selection of ensemble members for the same time and pixels. Unfortunately, the validation dataset was deemed to be too small to use in the variogram analysis; however, because the raw Oromiya gauge data were only used indirectly in the calibration process through the derivation of calibration parameters and residual variograms (created using the "residual" between the rainfall amount and mean monthly value, or

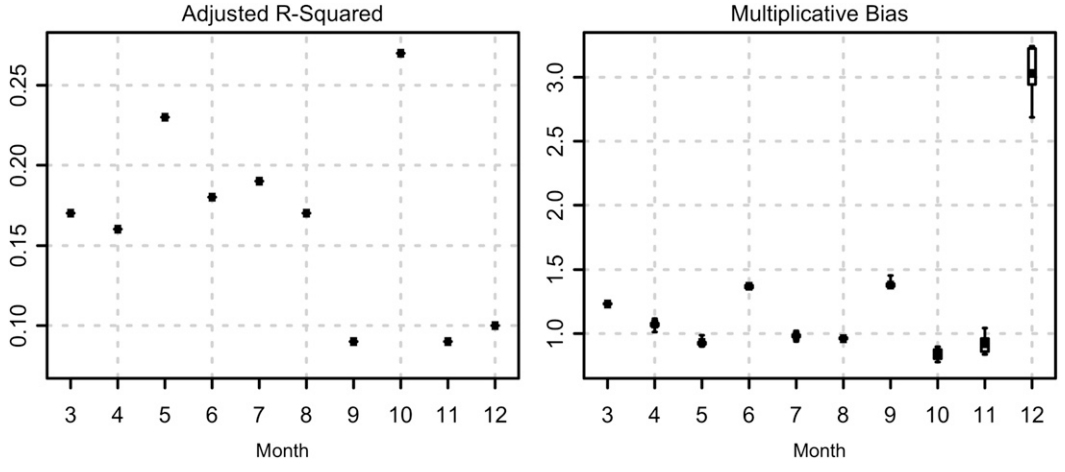

FIG. 9. Box plots of the monthly variation in the (left) R^2_{adj} and (right) multiplicative bias skill scores. Each box plot is made up of data from a leave-one-year-out cross-validation approach applied to the calibration data. As expected, more variation is seen in months when fewer data were available for calibration, especially in December.

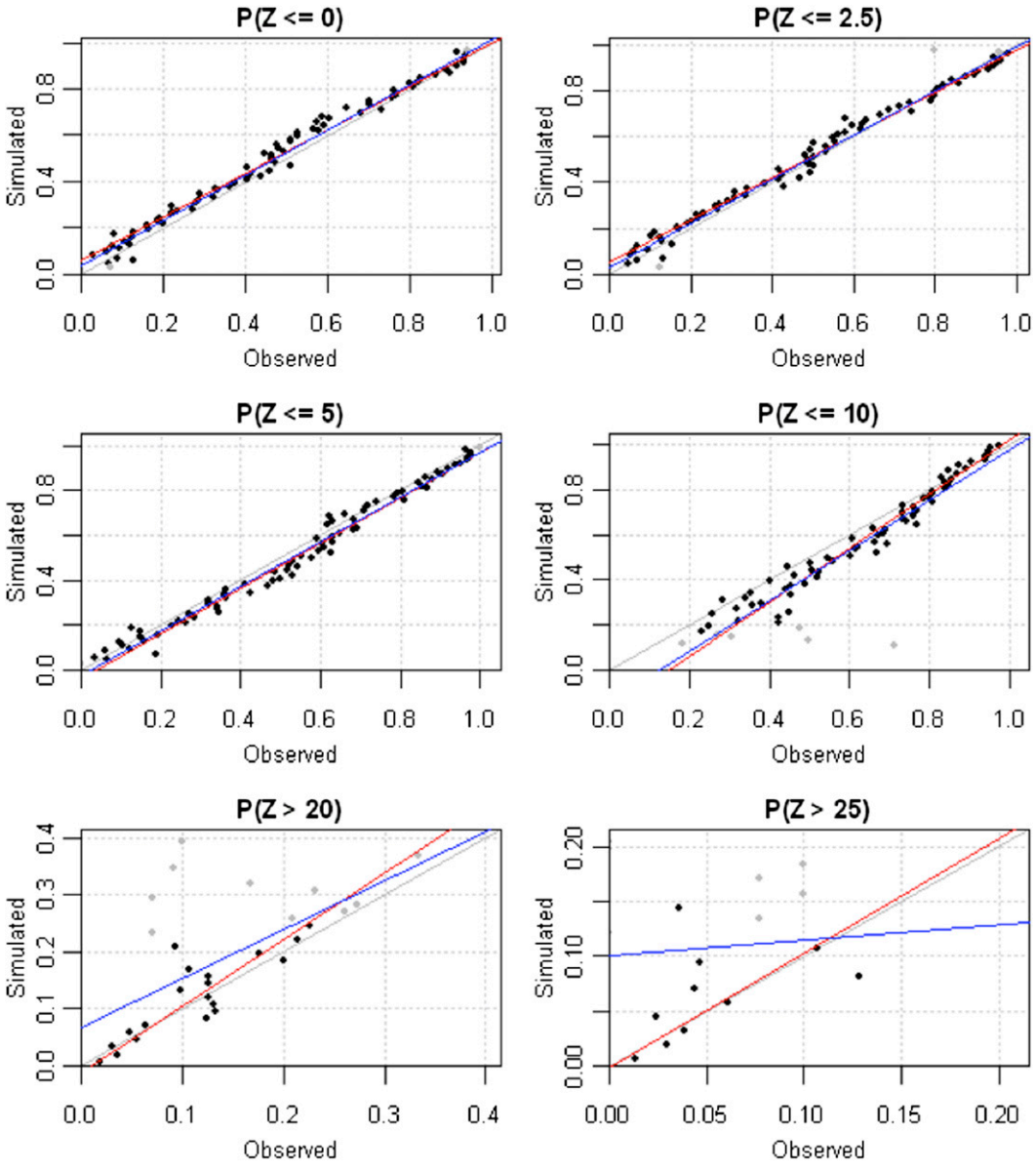

FIG. 10. Reliability plot: the solid gray line is the one-to-one line, the blue line is the linear fit of the data, and the red line is the linear fit weighted using the number of values that went into each bin.

between the binary occurrence and monthly probability of rain), it was felt to be suitable for use in the validation. This approach has also been taken in previous comparable validation studies. Indicator variograms can be seen in Fig. 11 (which model the spatial correlation of rainfall occurrence), and positive rainfall amount variograms can be seen in Fig. 12 (which model the spatial correlation of rainfall on rainy days). In general, the ensembles show a good fit to the data, although the effect of an early software choice forcing the nugget (or semivariance at zero distance) to zero for June, July, and August is reflected in the appropriate plots. Choosing a zero nugget is often a reasonable choice when modeling rainfall, as one might expect that rainfall from two gauges in the same location would be perfectly correlated. However, this assumption relies on limited measurement errors, especially for gauge location, and in this case, modeling the nugget clearly shows a more realistic result. Months that have an observed nested distribution (i.e., one of more clear scales of spatial correlation) were also modeled less well, although contrary to Teo and Grimes (2007), the variogram sills and long-range behavior are captured. It is postulated that some of the inability to capture some of the nested

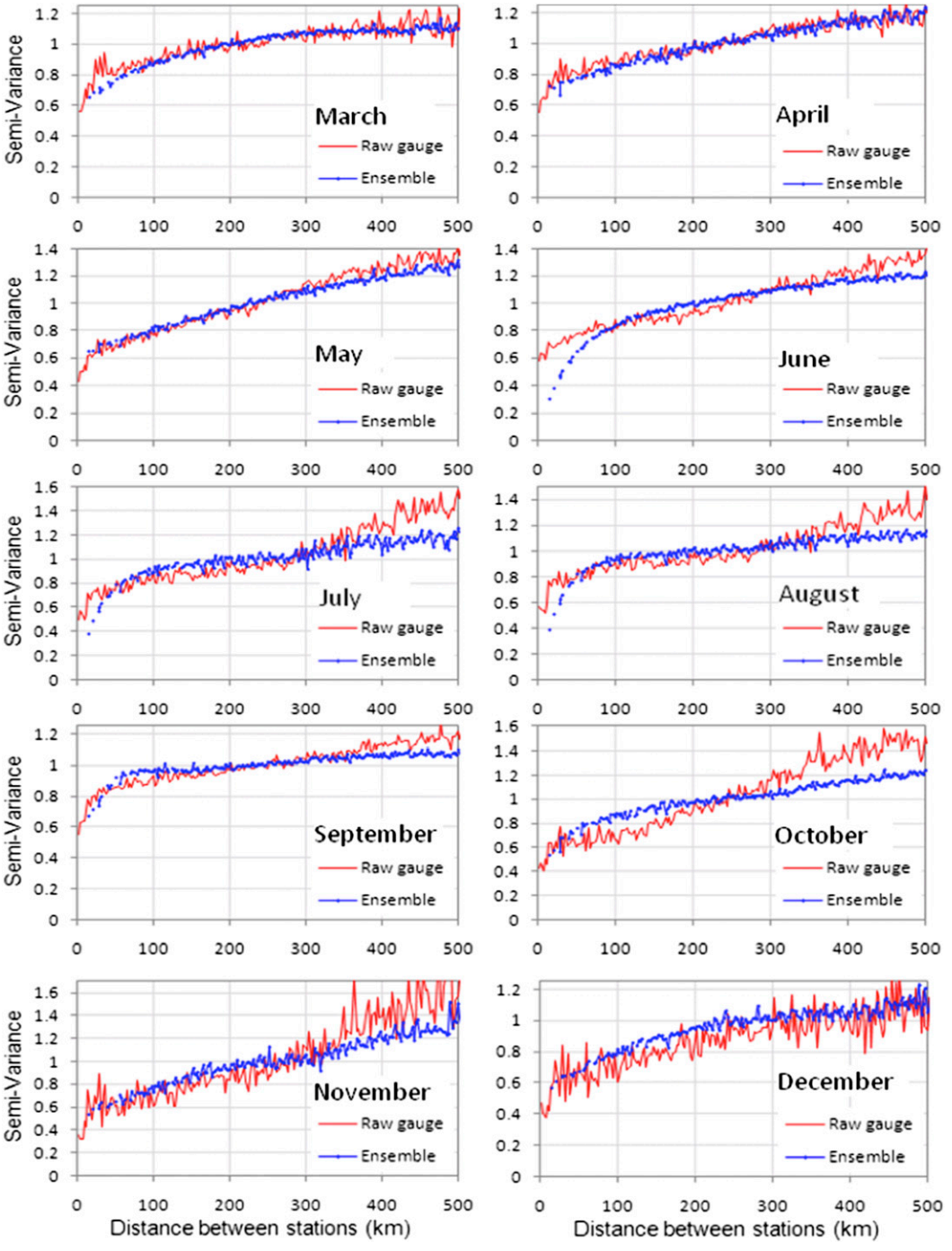

FIG. 11. Indicator variograms of rainfall occurrence for each month. The red line corresponds to the variogram generated by the raw gauge dataset and the blue line from the ensembles. A good comparison can be seen for most months.

behavior at small spatial scales is due to the fact that the range of the nested component of the variogram is approximately equivalent or smaller than 0.125°, the resolution of the ensemble product. In addition, there were very few gauges situated less than 20 km apart, leaving limited data available for calibration.

3) TIME SERIES

Daily time series were plotted over the entire validation time period (1994–99) and the ensemble values were compared with gauge data in Figs. 13 and 14. The two statistics chosen for the time series were the daily

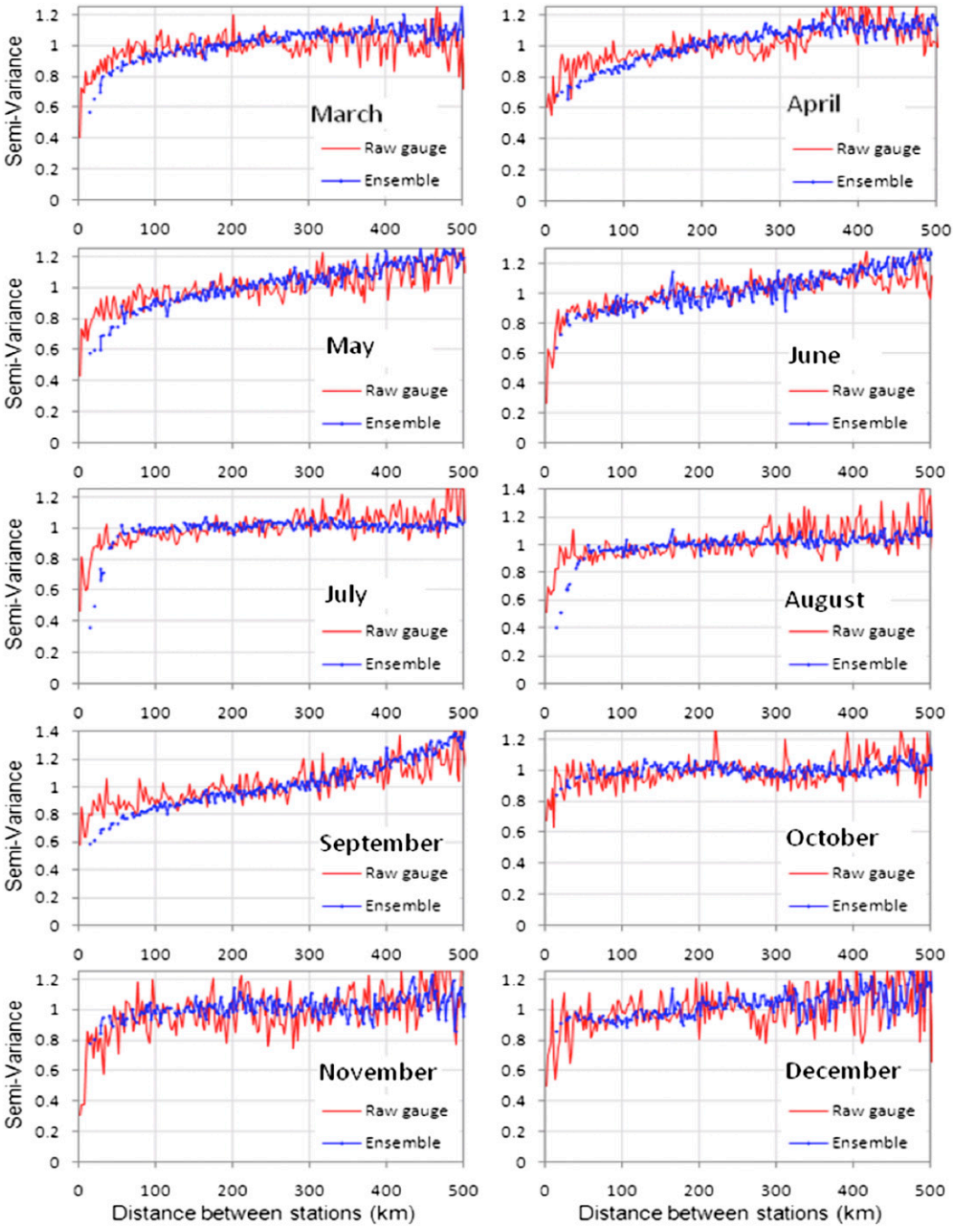

FIG. 12. Variograms of positive rainfall occurrence for each month. The red line corresponds to the variogram generated by the raw gauge dataset and the blue line from the ensembles. A good comparison can be seen for most months.

mean over the validation pixels and the proportion of rainy validation pixels each day. The results suggest the ensemble is capturing the seasonal cycle and that, in general, the ensemble is suggesting a reasonable range of values on a given day. There are very few days when at least one ensemble member does not capture the observed rainfall—one ensemble member captured observed rainfall on over 98% of days at locations within the calibration region and at over 95% of days for the other validation locations.

The analysis was repeated at a 10-day scale in Fig. 15, allowing the ensembles to be compared against the

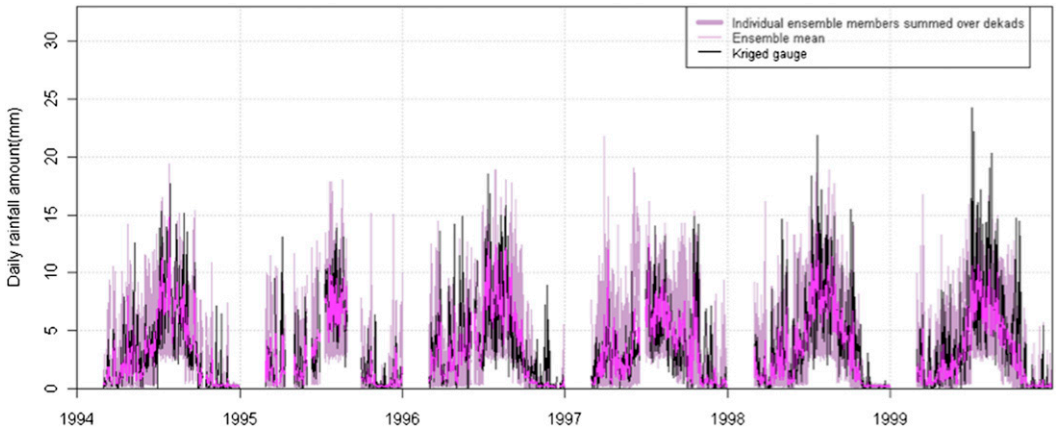

FIG. 13. Time series of the mean rainfall from gauge (black), the modified ensemble members (purple cloud), and the modified ensemble mean (pink).

TAMSAT 30-yr rainfall climatology (TARCAT), which has been calibrated on an extensive gauge dataset of 700 stations throughout Ethiopia spanning 30 years and has been comprehensively validated (Maidment 2014). It is encouraging to see that the time series from the ensembles closely resemble the one produced by TARCAT, although both struggle to capture the Belg rainfall season accurately. The ensemble cloud is also capturing the observed rainfall in nearly every dekad.

5. Conclusions

As satellite technology develops, SRFEs are likely to become ever more important in the world of food security, particularly in developing countries where large numberss of people rely on rain-fed agriculture, yet are not able to access adequate ground-based rainfall monitoring systems. Determining the uncertainty on these estimates can be extremely complex, especially over regions such as Africa where there are few rain gauges available for calibration and validation. The distribution of these stations is also often uneven with most stations located in lowlands, cities, and towns. As a result, an assessment of SRFE uncertainty is often not carried out over regions where the information is most needed. In addition, it is important that any uncertainty estimates take into account factors such as the skewed and mixed distribution of rainfall, its spatial correlations, differences in spatial scale between the calibration gauges, and the areal satellite estimates, plus all sources of uncertainty from the satellite instrument to the rain gauge data used to calibrate it.

Probabilistic ensembles of satellite rainfall are an effective way of achieving this goal, not least because it would be extremely complex to calculate the error analytically, but also because they allow the ensemble to be fed through an impact model to see how any uncertainty in rainfall propagates. Several studies have

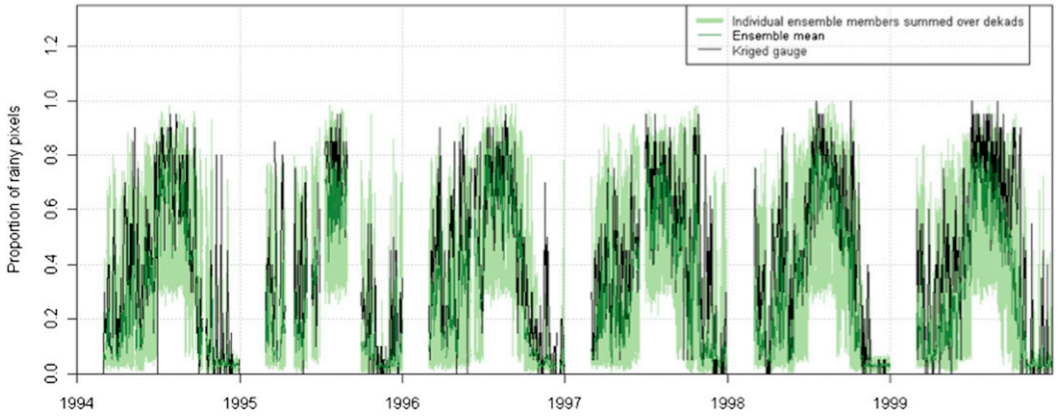

FIG. 14. Time series of the proportion of rainy pixels from gauge (black), the modified ensemble members (green cloud), and the modified ensemble mean (green).

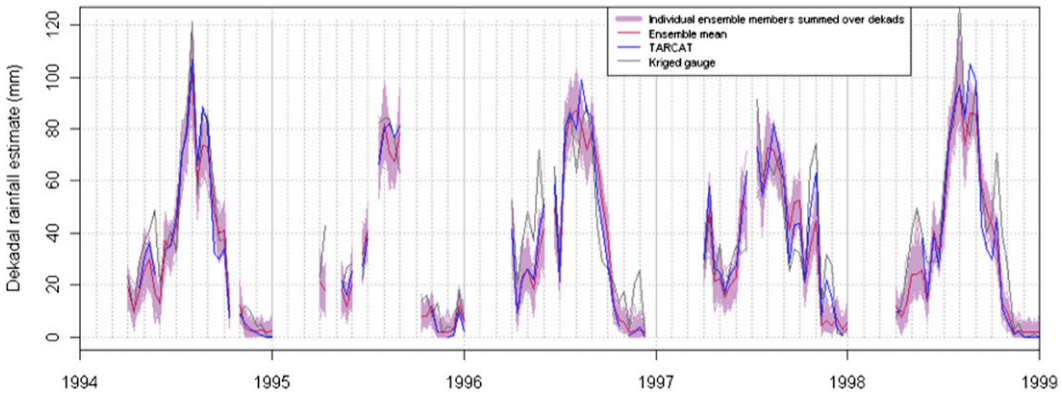

FIG. 15. Time series of the mean dekadal rainfall from the average of the 20 validation stations from kriged gauge data (black), the TARCAT rainfall climatology (blue), the ensemble mean (red), and the 200 individual ensemble members (purple cloud). Values were only plotted for dekads where all the data were present.

attempted such an ensemble approach, which has shown promise in areas with dense gauge networks such as Europe or in pilot studies over Africa. There are, however, still several challenges that need to be overcome before this approach could become a routine product alongside SRFEs.

This paper attempts to address two of these hurdles, through showing how the ensemble methodology could be applied to a much larger study region with a much more complex climate and a clustered calibration dataset. To do this, a new multithreshold, multizone calibration was proposed and an investigation was performed into the source of excess low rainfall reported in several rainfall sequential simulation studies (Teo and Grimes 2007; Skinner 2013). A case study of the Ethiopian highlands was used as an illustration.

The results were encouraging and showed that even with calibration data spanning only a small part of the region for which estimates were needed and for only 5 years, it was possible to create realistic rainfall fields that captured the rainfall recorded at the validation gauges. Particularly promising was the fact even though there were no calibration data at all in the drier north and west of the country, the ensembles still managed to capture observed rainfall in over 95% of cases in these locations (rising up to 98% in regions containing calibration data). The validation in section 4 showed that the uncertainty was well captured in most cases.

Several hypotheses were investigated with regard to the overestimation at low rainfall amounts, including the possibility of a Markov effect, overestimation on days with zero CCD [as postulated by Teo and Grimes (2007)], and a discussion of the breakdown of the calibration error parameters. The situation in this case was shown to be extremely complex, with all factors playing some role. This adds evidence to the suggestion that the solution would be to postprocess the final ensembles to ensure that the low rainfall bias is removed; in this case, a basic stochastic filter was successfully applied. Alternatively, significant additional research needs to be performed into qualifying the sources of SRFE error, noting that these will be different for each satellite and algorithm.

This technique still relies on a relatively spatially dense rain gauge dataset, although as this work found, the time series of these data do not have to be particularly long (>5 years); thus, the method might not be suitable for expansion to all locations. This does not, however, stop it from being extended to many other regions in Africa, as there is often a large amount of historical rain gauge data available from national meteorological agencies. In addition, a less dense calibration dataset might not be an insurmountable issue; the climate of Ethiopia is one of the most variable in sub-Saharan Africa, so a less dense dataset might be suitable for other regions where the climate varies less rapidly (as long as the data contains at least one area with more closely spaced gauges in order to create the variogram). In conclusion, the approach shows strong promise, and there is good potential for using such a method to routinely quantify the uncertainty on satellite rainfall estimates.

Acknowledgments. This research was funded through a National Environmental Research Council Ph.D. grant, with support from the SAIN project. In addition, this paper was supported by the CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), which is a strategic partnership of CGIAR and Future Earth, and by the European Union (EU) with technical support from the International Fund for Agricultural Development (IFAD). The views expressed in

this publication cannot be taken to reflect the official opinions of CGIAR or Future Earth. The authors are grateful to the National Meteorological Agency of Ethiopia for their expertise and datasets, plus to Emily Black, Chris Skinner, Fiona Underwood, Allard de Wit, Elena Tarnavsky, Teo Chee Kiat, and Ross Maidment for their support. Thank you also to the reviewers for their detailed and helpful comments; the paper is much improved as a result.

We dedicate this paper to the expertise, commitment, and passion for African rainfall and its importance to society by our coauthor and friend, Dr. David Grimes, who sadly passed away in December 2011 and who is deeply missed by a great many people.

REFERENCES

Ali, A., A. Amani, A. Diedhiou, and T. Lebel, 2005: Rainfall estimation in the Sahel. Part II: Evaluation of rain gauge networks in the CILSS countries and objective intercomparison of rainfall products. *J. Appl. Meteor.,* **44,** 1707–1722, doi:10.1175/JAM2305.1.

Bacchi, B., and N. T. Kottegoda, 1995: Identification and calibration of spatial correlation patterns of rainfall. *J. Hydrol.,* **165,** 311–348, doi:10.1016/0022-1694(94)02590-8.

Barancourt, C., J. D. Creutin, and J. Rivoirard, 1992: A method for delineating and estimating rainfall fields. *Water Resour. Res.,* **28,** 1133–1144, doi:10.1029/91WR02896.

Bellerby, T. J., and J. Sun, 2005: Probabilistic and ensemble representations of the uncertainty in an IR/microwave satellite precipitation product. *J. Hydrometeor.,* **6,** 1032–1044, doi:10.1175/JHM454.1.

Birhanu, S., and T. Alamirew, 2008: Rainfall intensity–duration–frequency analysis for selected parts of eastern Oromia. *Proc. Workshop on Hydrology and Ecology of the Nile River Basin under Extreme Conditions,* Addis Ababa, Ethiopia, National Science Foundation, 290. [Available online at http://www2.fiu.edu/~melessea/Nile%20Workshop%20Proceedings_%20June%2016-19,%202008.pdf.]

Bouvier, C., L. Cisneros, R. Dominguez, J.-P. Laborde, and T. Lebel, 2003: Generating rainfall fields using principal components (PC) decomposition of the covariance matrix: a case study in Mexico City. *J. Hydrol.,* **278,** 107–120, doi:10.1016/S0022-1694(03)00122-7.

Carroll, R. J., and D. Ruppert, 1988: *Transformation and Weighting Regression.* Chapman and Hall, 249 pp.

Cole, A. E., and A. J. Kantor, 1963: Tropical and sub-tropical atmospheres. *J. Appl. Meteor.,* **2,** 90–98, doi:10.1175/1520-0450(1963)002<0090:TASA>2.0.CO;2.

Davis, M., 1987: Production of conditional simulations via the **LU** triangular decomposition of the covariance matrix. *Math. Geol.,* **19,** 91–98, doi:10.1007/BF00898189.

Delhomme, J. P., 1978: Kriging in hydrosciences. *Water Resour.,* **1,** 251–266, doi:10.1016/0309-1708(78)90039-8.

De Michele, C., and P. Bernardara, 2005: Spectral analysis and modeling of space–time rainfall fields. *Atmos. Res.,* **77,** 124–136, doi:10.1016/j.atmosres.2004.10.031.

Deutsch, C. V., and A. G. Journel, 1998: *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, 384 pp.

de Wit, A. J. W., and C. A. van Diepen, 2008: Crop growth modelling and crop yield forecasting using satellite-derived meteorological inputs. *Int. J. Appl. Earth Obs. Geoinf.,* **10,** 414–425, doi:10.1016/j.jag.2007.10.004.

——, S. de Bruin, and P. J. J. F. Torfs, 2008: Representing uncertainty in continental-scale gridded precipitation fields for agrometeorological modeling. *J. Hydrometeor.,* **9,** 1172–1190, doi:10.1175/2008JHM899.1.

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.

——, S. Chidzambwa, P. Ceccato, S. J. Connor, and C. F. Ropelewski, 2008: Validation of high-resolution satellite rainfall products over complex terrain. *Int. J. Remote Sens.,* **29,** 4097–4110, doi:10.1080/01431160701772526.

——, and Coauthors, 2009: Designing index-based weather insurance for farmers in Adi-Ha, Ethiopia. IRI Tech. Rep. 09-04, 82 pp. [Available online at http://iri.columbia.edu/resources/publications/pub_id/908/.]

——, K. Hailemariam, R. Maidment, E. Tarnavsky, and S. Connor, 2013: Combined use of satellite estimates and rain gauge observations to generate high-quality historical rainfall time series over Ethiopia. *Int. J. Climatol.,* **34,** 2489–2504, doi:10.1002/joc.3855.

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

——, D. I. F. Grimes, and E. Black, 2011a: Teleconnections between Ethiopian summer rainfall and sea surface temperature: Part I—Observation and modelling. *Climate Dyn.,* **37,** 103–119, doi:10.1007/s00382-010-0837-8.

——, ——, and ——, 2011b: Teleconnections between Ethiopian summer rainfall and sea surface temperature: Part II. Seasonal forecasting. *Climate Dyn.,* **37,** 121–131, doi:10.1007/s00382-010-0896-x.

Germann, U., M. Berenguer, D. Sempere-Torres, and M. Zappa, 2009: REAL—Ensemble radar precipitation estimation for hydrology in a mountainous region. *Quart. J. Roy. Meteor. Soc.,* **135,** 445–456, doi:10.1002/qj.375.

Gissila, T., E. Black, D. I. F. Grimes, and J. M. Slingo, 2004: Seasonal forecasting of the Ethiopian summer rains. *Int. J. Climatol.,* **24,** 1345–1358, doi:10.1002/joc.1078.

Greatrex, H., 2012: The application of seasonal rainfall forecasts and satellite rainfall estimates to seasonal crop yield forecasting for Africa. Ph.D. thesis, University of Reading, 320 pp. [Available online at http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.578012.]

Grimes, D. I. F., 2008: An ensemble approach to uncertainty estimation for satellite-based rainfall estimates. *Hydrological Modelling and the Water Cycle,* S. Sorooshian et al., Eds., Water Science and Technology Library, Vol. 63, Springer, 145–162, doi:10.1007/978-3-540-77843-1_7.

——, and E. Pardo-Igúzquiza, 2010: Geostatistical analysis of rainfall. *Geogr. Anal.,* **42,** 136–160, doi:10.1111/j.1538-4632.2010.00787.x.

——, ——, and R. Bonifacio, 1999: Optimal areal rainfall estimation using raingauges and satellite data. *J. Hydrol.,* **222,** 93–108, doi:10.1016/S0022-1694(99)00092-X.

——, E. Coppola, M. Verdecchia, and G. Visconti, 2003: A neural network approach to real-time rainfall estimation for Africa using satellite data. *J. Hydrometeor.,* **4,** 1119–1133, doi:10.1175/1525-7541(2003)004<1119:ANNATR>2.0.CO;2.

Haberlandt, U., and C. Gattke, 2004: Spatial interpolation vs. simulation of precipitation for rainfall–runoff modelling—

A case study in the Lippe River basin. *Hydrology Science and Practice for the 21st Century,* Vol. 1, B. Webb et al., Eds., British Hydrological Society, 120–127.

Hellmuth, M. E., D. E. Osgood, U. Hess, A. Moorhead, and H. Bhojwani, 2009: Index insurance and climate risk: Prospects for development and disaster management. Climate and Society No. 2, IRI Rep., Columbia University, 122 pp.

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. *Int. J. Remote Sens.,* **18,** 2147–2159, doi:10.1080/014311697217800.

Huffman, G. J., and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeor.,* **8,** 38–55, doi:10.1175/JHM560.1.

Ines, A. V. M., and J. W. Hansen, 2006: Bias correction of daily GCM rainfall for crop simulation studies. *Agric. For. Meteor.,* **138,** 44–53, doi:10.1016/j.agrformet.2006.03.009.

Jobard, I., F. Chopin, J. C. Berges, and R. Roca, 2011: An intercomparison of 10-day satellite precipitation products during West African monsoon. *Int. J. Remote Sens.,* **32,** 2353–2376, doi:10.1080/01431161003698286.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeor.,* **5,** 487–503, doi:10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.

Lebel, T., and L. Le Barbé, 1997: Rainfall monitoring during HAPEXSahel: 2. Point and areal estimation at the event and seasonal scales. *J. Hydrol.,* **188–189,** 97–122, doi:10.1016/S0022-1694(96)03325-2.

——, G. Bastin, C. Obled, and J. D. Creutin, 1987: On the accuracy of areal rainfall estimation: A case study. *Water Resour. Res.,* **23,** 2123–2134, doi:10.1029/WR023i011p02123.

Maidment, R. I., 2014: Development and exploitation of a 30-year African rainfall climatology and time-series. Ph.D. thesis, University of Reading, 234 pp.

——, D. I. F. Grimes, R. P. Allan, H. Greatrex, O. Rojas, and O. Leo, 2012: Evaluation of satellite-based and model reanalysis rainfall estimates for Uganda. *Meteor. Appl.,* **20,** 308–317, doi:10.1002/met.1283.

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.

Oxfam America, 2014: R4 Rural Resilience Initiative: Quarterly report, January–March 2014. Oxfam Rep., 24 pp. [Available online at http://www.oxfamamerica.org/static/media/files/R4_Rural_Resilience_Initative_January_to_March_2014_WEB.pdf.]

Parameswaran, K., and Coauthors, 2000: Altitude profiles of temperature from 4 to 80 km over the tropics from MST radar and lidar. *J. Atmos. Sol.-Terr. Phys.,* **62,** 1327–1337, doi:10.1016/S1364-6826(00)00124-3.

Pardo-Igúzquiza, E., 1998: Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. *J. Hydrol.,* **210,** 206–220, doi:10.1016/S0022-1694(98)00188-7.

Roca, R., P. Chambon, I. Jobard, P.-E. Kirstetter, M. Gosset, and J. C. Bergès, 2010: Comparing satellite and surface rainfall products over West Africa at meteorologically relevant scales during the AMMA campaign using error estimates. *J. Appl. Meteor. Climatol.,* **49,** 715–731, doi:10.1175/2009JAMC2318.1.

Rosema, A., M. De Weirdt, and S. Foppes, 2010: METEOSAT based agrometeorological monitoring and crop yield forecasting using the energy and water balance monitoring system. *Workshop Proceedings: Remote Sensing to Support Crop Yield Forecasts and Area Estimates,* ISPRS, 117–122.

Skinner, C. J., 2013: Ensemble-characterisation of satellite rainfall uncertainty and its impacts on the hydrological modelling of a sparsely gauged basin in western Africa. Ph.D. thesis, University of Hull, 338 pp. [Available online at https://hydra.hull.ac.uk/assets/hull:7199a/content.]

Sorooshian, S., K.-L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc.,* **81,** 2035–2046, doi:10.1175/1520-0477(2000)081<2035:EOPSSE>2.3.CO;2.

Syngenta, cited 2014: The Kalimo Salama Project. [Available online at www.syngentafoundation.org/index.cfm?pageID=562.]

Teo, C., and D. I. F. Grimes, 2007: Stochastic modelling of rainfall from satellite data. *J. Hydrol.,* **346,** 33–50, doi:10.1016/j.jhydrol.2007.08.014.

Thiemig, V., R. Rojas, M. Zambrano-Bigiarini, V. Levizzani, and A. De Roo, 2012: Validation of satellite-based precipitation products over sparsely gauged African river basins. *J. Hydrometeor.,* **13,** 1760–1783, doi:10.1175/JHM-D-12-032.1.

Thorne, V., P. Coakeley, D. Grimes, and G. Dugdale, 2001: Comparison of TAMSAT and CPC rainfall estimates with raingauges, for southern Africa. *Int. J. Remote Sens.,* **22,** 1951–1974, doi:10.1080/01431160118816.

Thornton, P. K., W. T. Bowen, A. C. Ravelo, P. W. Wilkens, G. Farmer, J. Brock, and J. E. Brink, 1997: Estimating millet production for famine early warning: An application of crop simulation modelling using satellite and ground-based data in Burkina Faso. *Agric. For. Meteor.,* **83,** 95–112, doi:10.1016/S0168-1923(96)02348-9.

Tompson, A. F. B., R. Ababou, and L. W. Gelhar, 1989: Implementation of the three-dimensional turning bands random field generator. *Water Resour. Res.,* **25,** 2227–2243, doi:10.1029/WR025i010p02227.

Velpuri, N. M., G. B. Senay, J. Rowland, J. P. Verdin, and H. Alemu, 2014: Africa-wide monitoring of small surface water bodies using multisource satellite data: A monitoring system for FEWS NET. *Nile River Basin,* A. M. Melesse, W. Abtew, and S. G. Setegn, Eds., Springer, 69–95.