INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **38**: 1201–1215 (2018)
Published online 25 August 2017 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.5238

Development of high spatial resolution rainfall data for Ghana

J. N. A. Aryee,[a]* L. K. Amekudzi,[a] E. Quansah,[a] N. A. B. Klutse,[b] W. A. Atiah[a] and C. Yorke[c]

[a] *Meteorology and Climate Science Unit, Department of Physics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana*
[b] *Remote Sensing, GIS and Climate Center, Ghana Space Science and Technology Institute, Accra, Ghana*
[c] *Ghana Meteorological Agency, Accra, Ghana*

ABSTRACT: Various sectors of the country's economy – agriculture, health, energy, among others – largely depend on climate information, hence availability of quality climate data is very essential for climate-impact studies in these sectors. In this paper, a monthly rainfall database (GMet v1.0) has been developed at a $0.5^\circ \times 0.5^\circ$ spatial resolution, from 113 Ghana Meteorological Agency (GMet) gauge network distributed across the four agro-ecological zones of Ghana, and spanning a 23-year period (1990–2012). The datasets were first homogenized with quantile-matching adjustments and thereafter, gridded at a spatial resolution of $0.5^\circ \times 0.5^\circ$ using Minimum Surface Curvature with tensioning parameter, allowing for comprehensive spatial fields assessment on the developed dataset. Afterwards, point-pixel validation was performed using GMet v1.0 against gauge data from stations that were earlier excluded due to large datagaps. This proved the reliability of GMet v1.0, with high and statistically significant correlations at 99% confidence level, and relatively low biases and rmse. Furthermore, GMet v1.0 was compared with GPCC and TRMM rainfall estimates, with both products found to adequately mimick GMet v1.0, with high correlations which are significant at 99% confidence level, low biases and rmse. In addition, the ratio of 90*th* – percentile provided fairly similar capture of extremes by both TRMM and GPCC, in relation to GMet v1.0. Finally, based on annual rainfall totals and monthly variability, *k*-means cluster analysis was performed on GMet v1.0, which delineated the country into four distinct climatic zones. The developed rainfall data, when officially released, will be a useful product for climate impact and further rainfall validation studies in Ghana.

KEY WORDS rainfall climatology; GMet v1.0; homogenization; minimum surface curvature; quantile matching; validation; clustering; Ghana

Received 13 October 2016; Revised 14 June 2017; Accepted 17 July 2017

1. Introduction

Over the years, climate research studies have been carried out using various climate variables such as precipitation and surface temperature, to provide a rich evidence of climate change and variability at different spatio-temporal scales (Mengistu Tsidu, 2012). However, in developing countries such as Ghana, a key challenge to providing quality climate data for research of a sort, is the lack of dense instrumental network (Liu *et al.,* 2003; Manzanas *et al.,* 2014). It is therefore imperative to provide climate data with reasonably high spatio-temporal resolution, to improve climate-impact studies in this region.

Rainfall is highly variable spatially in the tropics, and as such employing a neighbouring station's data as replacement for a reference series, despite their proximity, may be inappropriate. Moreover, rain gauge observations usually provide longer records but insufficient spatial representation (Gruber and Levizzani, 2008; Mengistu Tsidu, 2012). This is a major limitation in the observation and spatial analysis of tropical rainfall. Data gridding alleviates this challenge by interpolating point rainfall estimates, and averaging them within grid cells. Finer resolutions provide detailed rainfall representations, whereas coarser resolutions compromise the rainfall information.

Numerous methods have been used for such spatial interpolations, which include Kriging (Price *et al.,* 2000; Jeffrey *et al.,* 2001; Dressler, 2007; Li and Heap, 2008; Yeh *et al.,* 2011; Mengistu Tsidu, 2012), minimum surface curvature (hereafter referred to as MSC) (Smith and Wessel, 1990; Dressler, 2007), inverse distance weighting (Dressler, 2007; Lu and Wong, 2008; Feng-Wen and Chen-Wuing, 2012) and regularized expectation maximization (Schneider, 2001; Mengistu Tsidu, 2012), among others. For this study, the MSC algorithm has been used due to its computational speed and suitability for large number of data points. In addition, the algorithm is suitable for smooth approximation and interpolation of climate data (Smith and Wessel, 1990; Dressler, 2007). Contrarily, for monthly data, the assumption of Gaussian distributed data will vary seasonally and spatially across the study area, thus RegEM was considered unsuitable. On the other hand, kriging has performance similar to the minimum curvature, however, they produce undesirable 'pits' and 'circular' isolines in the generated surface (Dressler, 2007).

Additionally, while a fixed climatological network is important for providing longer records which help to monitor climate variability, a limitation is that the instruments

* Correspondence to: J. N. A. Aryee, Department of Physics, Meteorology and Climate Science Unit, U.P.O-KNUST, Kumasi, Ghana.
E-mail: jeff.jay8845@gmail.com

undergo changes over time (Blair, 2012); thereby introducing non-climatic shifts in their measurements. Maintenance of these fixed station-network is mostly affected by socio-economic factors such as demographic and infrastructural changes, resulting in data inhomogeneity. Data homogenization is therefore necessary for removing these shifts in the datasets. For most operational climate datasets, information regarding data homogeneity, station operations, among others are documented in a metadata, per the World Meteorological Agency (WMO) standard. However, in Ghana, there is currently no metadata providing information of a sort. A comprehensive explanation on data homogenization and metadata documentation is presented in the WMO (2008) technical manual.

The non-existence of homogenized and highly-resolved, countrywide rainfall data is the motivation for this study. The objective of this study therefore was to develop a high, spatially-resolved rainfall dataset for Ghana, to advance climate-impact studies in the country. To this end, GMet v1.0 was developed, and has the potential to serve as quality input for hydro-meteorological and other rainfall-based modelling applications over the study region, for improved climate diagnostics, among others, over the country. Its development has generally maximized widespread country-based network of rain gauge observations – into a gridded dataset – and thus proves essential for enhanced rainfall validation studies over wider spatial domains within the country. However, the unsuitability of the MSC algorithm for extrapolation triggers an exercise of caution when using GMet v1.0 over locations along the borders. Notwithstanding this caveat, GMet v1.0 has the potential to be used as readily-available, quality dataset for climate-impact studies and subsequently, a source of climate information for stakeholders and policy makers.

The paper is structured as follows: a description of the study area and the data source is presented in Section 2, with presentation of the methodology employed in the study provided in Section 3. The results are discussed in Section 4 and finally, conclusions given in Section 5.

2. Study area and data source

2.1. Climate of the study area

The study is conducted over the West African nation of Ghana. Its climate is monsoonal, with two main seasons (dry and wet) dominating the country, due to its location in the tropics. Earlier research studies over this domain identifies rainfall to be mainly linked to development and movement of Mesoscale Convective Systems (MCS), and regulated by advection of moisture in the lower levels of the atmosphere from the Gulf of Guinea (Sultan and Janicot, 2003). This system is driven by temperature and energy contrasts between the Sahara and Gulf of Guinea. Along the inter-tropical discontinuity (ITD), maritime tropical air mass from the Atlantic Ocean converges with dry and cold tropical continental air mass (Amekudzi *et al.,* 2015 and references therein).

As enhancement to this conceptual view on tropical monsoon flow and dynamics, Nicholson (2009) further highlights three quasi-independent mechanisms that serve as controls to precipitation development over West Africa, namely ascent linked to the upper-level jet streams, convergence associated with the surface inter-tropical convergence zone (ITCZ), and a coastal circulation cell linked to sea-breeze effects. These factors induce the generation of tropical rainbelt (rainfall maxima; not similar to ITCZ) which migrates seasonally between the northern and southern halves of the continent. As a result, the south of the country experiences a bimodal rainfall pattern whereas a unimodal pattern exists over the north of the country (Owusu and Waylen, 2013; Manzanas *et al.,* 2014; Amekudzi *et al.,* 2015; Mensah *et al.,* 2016).

2.2. Description of datasets

2.2.1. GMet observed data

Monthly rainfall data measured from various meteorological stations (synoptic, agrometeorological and climatological) over Ghana were collected from the Ghana Meteorological Agency's archive. The observations made by the agency, are in conformity with the World Meteorological Organization (WMO) standards (Amekudzi *et al.,* 2015, 2016; Mensah *et al.,* 2016). However, inherent data gaps were detected within the collected time-series. A tolerance of less than 10%, non-continuous gaps was set for sorting the datasets as a check for consistency. Out of over 200 stations identified countrywide, a total of 113 stations which met this criteria were employed in this study. These stations are distributed over the four agro-ecological zones (Savannah, Transition, Forest and Coast) as shown in Figure 1, with a dense station-network in the Forest zone. Figure 1(a) shows the various types of meteorological stations and the percentage of inherent datagaps in each individual station dataset is illustrated in Figure 1(b).

2.2.2. Global Precipitation Climatology Centre data

The use of Global Precipitation Climatology Centre (GPCC) datasets in this study stems primarily from the fact that, it has been demonstrated in various studies, for example Manzanas *et al.* (2014) and Baidu *et al.* (2017), to be a reliable, gauge-based dataset over the study domain, and thus proves useful for comparison with GMet v1.0. GPCC datasets are monthly rainfall estimates covering the last century, and provided by the Earth System Research Lab (ESRL) at the National Oceanic and Atmospheric Administration (NOAA). The GPCC version 7 (v7), containing the monthly rainfall totals on a regular grid with a spatial resolution of $0.5° \times 0.5°$, was used in this study. Although the datasets spanned over a century, the period 1990–2012 was extracted for use in this study. Also, due to its similarity to GMet v1.0 on a spatial scale, no regridding was performed on GPCC v7.

2.2.3. Tropical rainfall monitoring mission data

For the study region, a major-performing, satellite-based dataset is the tropical rainfall monitoring mission

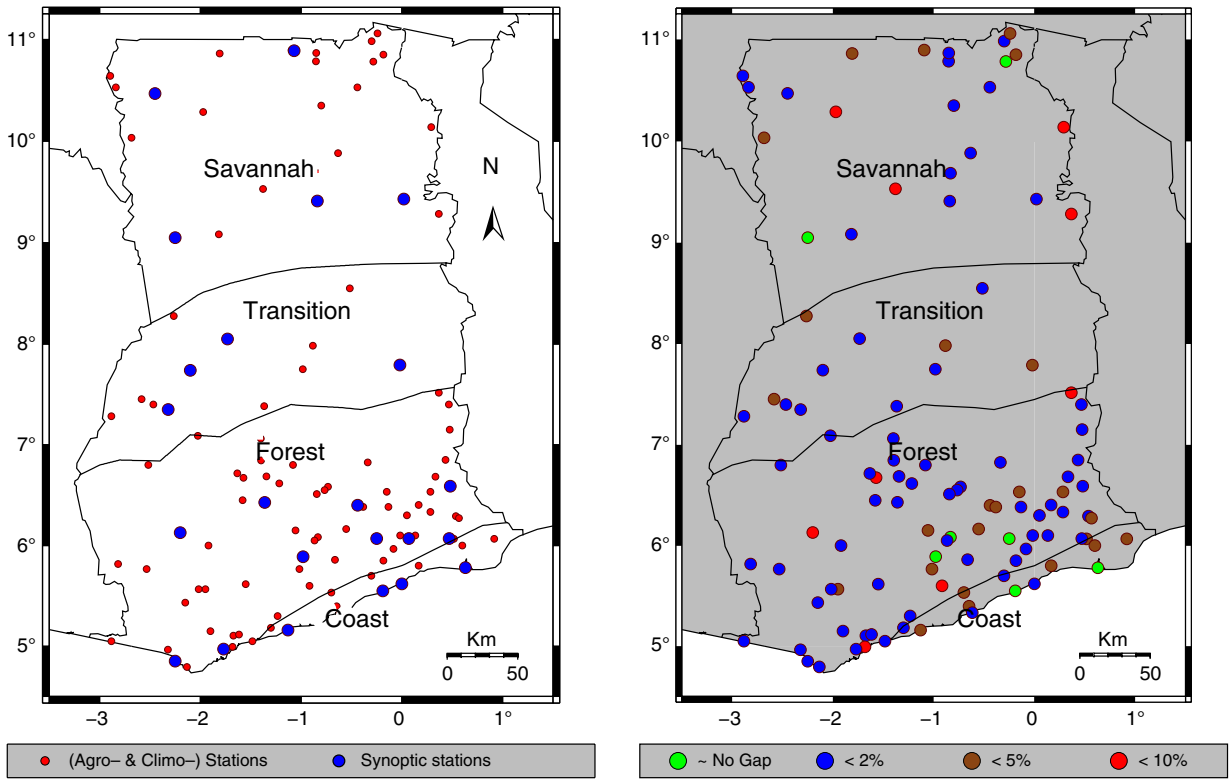

Figure 1. Spatial distribution of 113 meteorological stations over the four agro-ecological zones, comprising the types of stations (left) and percentage of gaps inherent in each of the 113 station data (right). [Colour figure can be viewed at wileyonlinelibrary.com].

(TRMM) data. It has the ability to rightly capture rainfall events and other mesoscale convective activities (Amekudzi *et al.,* 2016). Thus, statistical assessment of TRMM – GMet v1.0 has been performed and presented further in this study. TRMM 3B43 product of the TRMM multi-satellite precipitation analysis (TMPA) version 7 spans a shorter period of 1998–2012. Its algorithm combines 3-hourly merged high-quality/IR estimates (3B42), SSM/I estimates (3A46) and the monthly accumulated GPCC rain gauge analysis (3A45) to produce best estimates of precipitation rate and root mean square precipitation error (Huffman *et al.,* 2007, 2010). The product is gridded on $0.25° \times 0.25°$ spatial, and monthly temporal resolutions. For comparison with GMet v1.0, TRMM was re-gridded on a $0.5° \times 0.5°$ spatial scale.

3. Methodology

3.1. Homogenization and gridding

Monthly rainfall data from 113 meteorological stations were subjected to homogenization by quantile-matching adjustments (QMadj), to adjust all de-trended data points. Generally, two types of homogenization (absolute and relative) procedures can be performed on any time-series. However, due to the sparse datasets from neighbouring stations, absolute homogenization was preferred. This point-by-point homogenization was carried out using the RHtestsV4 package described in Wang *et al.* (2010), which includes provision of QMadj. The homogenization employed a linear regression analysis of the reference series to identify changepoints. Based on the magnitude of changepoints, the reference series was augmented. Thereafter, the homogenized data was gridded on $0.5° \times 0.5°$ spatial scale, using minimum surface curvature (MSC) with tensioning parameter. MSC is a surface interpolation method, similar to a slender and linearly-elastic plate moved through data values with infinitesimal bending. The method employs a surface of relatively small total squared curvature, as well as, continuous second derivatives. According to Smith and Wessel (1990), MSC generates finest possible surface but is an inexact interpolator, hence, it is better resolved by introducing a tension parameter (T) to the algorithm. Generally, the algorithm is a numerical solution given in Equation (1) as

$$(1-T)\,\nabla^4 f(x,y) - (T)\,\nabla^2 f(x,y) = 0 \tag{1}$$

with three boundary conditions given in Equations (2)–(4).

$$(1-T)\,\frac{\partial^2 f}{\partial n^2} + T\frac{\partial f}{\partial n} = 0 \tag{2}$$

$$\frac{\partial\left(\nabla^2 f\right)}{\partial n} = 0 \quad \text{on the edges} \tag{3}$$

$$\frac{\partial^2 f}{\partial x \partial y} = 0 \quad \text{at the corners} \tag{4}$$

where T is the tensioning parameter for the boundary ($T \in 0, 1$), ∇^2 is the Laplacian operator, ∇^4 is the biharmonic operator and n is the normal.

Procedurally, the 113 point data were first pre-processed with 'blockmean', per every $0.5° \times 0.5°$ spatial grid, to avoid spatial aliasing and also remove redundant data. This allowed for averaging of datasets per each grid for all points that were bounded. Thereafter, a tension factor of 1, based on recommendations of Smith and Wessel (1990), was used in gridding. This parameter is found to (1) be good for potential field data such as the rainfall data for this study, (2) suppress undesired oscillations and false local maxima or minima that are sometimes generated by MSC and also (3) generate a harmonic surface. Afterwards, the surface algorithm, together with the set tension factor, was iterated till convergence (default used here; which is e^{-4} of the root mean square deviation of the data from a best-fit plane) was met. Despite the suitability of the algorithm to interpolating incomplete datasets, it has complexities dealing with extrapolation (Dressler, 2007). In most instances, inflections were observed at the boundaries, which are likely attributable to the method's unsuitability for extrapolation.

Afterwards, a point-pixel validation was performed using the gridded data (GMet v1.0) and data from stations that were earlier excluded from the gridding due to the presence of large and continuous gaps. Assessment of spatial fields of GMet v1.0 rainfall data was performed alongside GPCC to provide further information on monthly, seasonal and yearly rainfall variability from both products. Finally, TRMM as a better-performing merged gauge-satellite product over the study domain, together with GPCC was statistically assessed, with respect to GMet v1.0.

3.2. Validation

In order to test the reliability of GMet v1.0 in capturing rain that fell on ground within the time domain, a point-pixel validation was carried out using some of the stations that were previously rejected due to their inadequate data. Three statistical metrics, described in Amekudzi *et al.* (2008), Quansah *et al.* (2014) and Mensah *et al.* (2016) were used to assess the performance of GMet v1.0. These statistical metrics include Pearson's correlation coefficient, bias and root mean squared error (hereafter termed as rmse), given in Equations 5–7 as:

$$r_{xy} = \frac{\sum_{i=1}^{N} \left(x_i - \mu_x\right) \left(y_i - \mu_y\right)}{(N-1) \; S_x S_y} \tag{5}$$

$$\text{bias} = \frac{1}{N} \sum \left(x_i - y_i\right) \tag{6}$$

$$\text{rmse} = 1 - \frac{\sum_{i=1}^{N} \left(y_i - x_i\right)^2}{\sum_{i=1}^{N} \left(x_i - \mu_x\right)^2} \tag{7}$$

where x is the observed rainfall data, y is the GMet v1.0, μ_x and μ_y are the arithmetic mean of x and y, respectively, S_x and S_y are the standard deviation of x and y, respectively, N is the sample size, r_{xy} is the Pearson correlation coefficient ($r \in -1,1$) and rmse is the root mean squared error. Moreover, statistical significance of the correlation was tested at a 99% confidence level.

4. Results

4.1. Homogenization

For this step, a pointwise homogeneity check was performed for each station using the RHtestV4 package, before gridding the homogenized datasets. Subjecting the various station data to homogenization by QMadj, revealed that most of the station data had undergone varying degrees of change. For example, one changepoint was identified in Sefwi-Bekwai, as shown in Figure 2(a). However, there is no metadata providing information on the station history and change to determine the likely cause of the changepoint. This changepoint is likely attributable to migration of the station, instrumental rearrangements within the enclosure, other demographical changes, or possibly even a climate-induced change. The de-trended datasets were augmented to provide a homogenized series (see Figure 2(c)), based on the magnitude of the identified change points (see Figure 2(b)), which ranged up to about 65 mm. On the contrary, due to the quality manning of the synoptic stations and the large dependence on their datasets, data from many of them were found to be already homogeneous. Included in the Appendix are the changepoints as detected in each of the stations that were employed for the gridding procedure.

4.2. Point-pixel validation of GMet v1.0

To evaluate reproducibility of GMet v1.0, relative to gauge, a point-pixel validation was carried out. Stations that were earlier discarded, due to the presence of a continuous and long gap in data, were used in validating GMet v1.0. Data used for validation were sampled from stations that, although had continuous gaps greater than 10% and as such were not included in the gridding, still had high rainfall data to gap ratio. In all, a total of 78 stations meeting this criteria were used for the point-pixel validation. For each of the 78 stations validated, we select data from the nearest gridbox in GMet v1.0 that each station is bounded by and validate with gauge data.

Point-pixel validation for two (Akaa in the Transition zone and Nungua in the Coast) of the 78 stations have been shown in Figure 3. These stations have been selected from zones within which rainfall patterns are bimodal and thus, depict high variabilities. These were purposely sampled in order to identify how well GMet v1.0 captures the rainfall amounts and monthly variability. Generally, the gridding technique proved useful in reproducing rainfall patterns and amounts over the selected locations. For example, as shown in Figure 3(a), the monthly variability of rainfall has been effectively mimicked by GMet v1.0 over Akaa, with

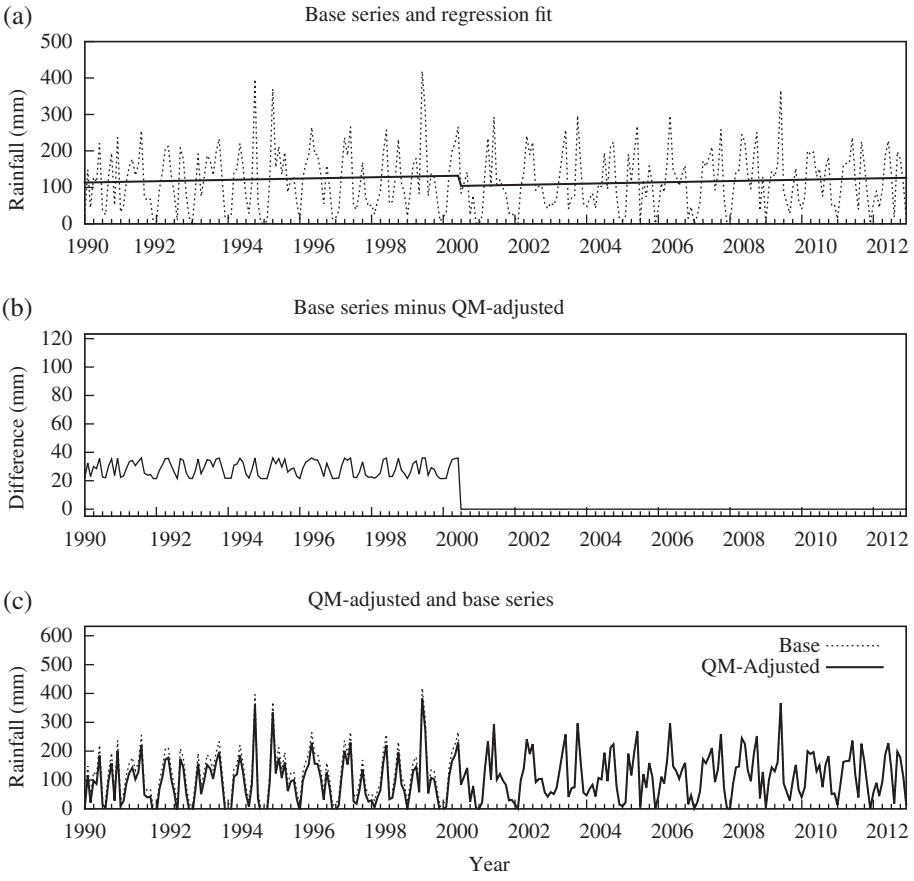

Figure 2. Homogenization by QMadj performed on one of the stations (Sefwi-Bekwai).

a correlation coefficient of 0.84 found statistically significant at 99% confidence level, as shown in Figure 3(b). In addition, a wet bias of 0.69% was observed between GMet v1.0 and the point data as shown in Figure 3(b), with rmse of 53.26 mm month^{-1}.

Again, in Figure 3(c), the monthly rainfall variability for Nungua has been effectively mimicked by GMet v1.0, with a correlation coefficient of 0.89, statistically significant at 99% confidence level. Rainfall estimates of GMet v1.0 were found to averagely overestimate that of the gauge by approximately 2.72% (see Figure 3(d)), and with rmse of 31.70 mm month^{-1}.

Additionally, countrywide performance assessment of GMet v1.0, by way of correlation, bias and rmse analyses was performed over all four agro-ecological zones. As shown in Figure 4(a), point – pixel correlation coefficients were – on average – greater than 0.7, and statistically significant at 99% confidence level. This indicates at least, 70% probability that rainfall over the country have been adequately reproduced by GMet v1.0. Moreover, four out of the 78 sampled stations were found to have correlations above 0.9, with two of them lying in the upper Savannah, one in the south-west Forest zone and the other in the Coast. Furthermore, a pointwise bias was computed for matching grids on GMet v1.0 to the 78 select stations for validation, as shown in Figure 4(b). These computations provided the relative dry and wet biases of GMet v1.0, relative to the select gauge data. Generally, countrywide wet and dry biases ranged between ±30%, with one in the lower Savannah having wet bias >30%, and two stations in the north-western Savannah recording dry biases of about 50%. Additionally, the rmse values were found to be generally about 70 mm per month.

4.3. Measure of spatial representativeness of GMet v1.0

As a measure of spatial representativeness of GMet v1.0, goodness-of-fit, based on station density per grid was performed. This step was primarily carried out because the accuracy of gridded precipitation datasets (such as GMet v1.0), based on gauge measurements, mainly depends on the spatial density of stations being used.

As shown in Figures 5(a) and (b), a dense gauge network is evenly distributed over the Coast and Forest zones. These dense network further highly correlates with GMet v1.0, with statistically significant coefficients of approximately 0.9. On the other hand, the Savannah and Transition zones were found to house a sparse gauge network. Notwithstanding, correlation coefficients in these zones were observed to be also approximately 0.9 and statistically significant.

The World Meteorological Organization (WMO) recommends the distribution of stations in the Regional

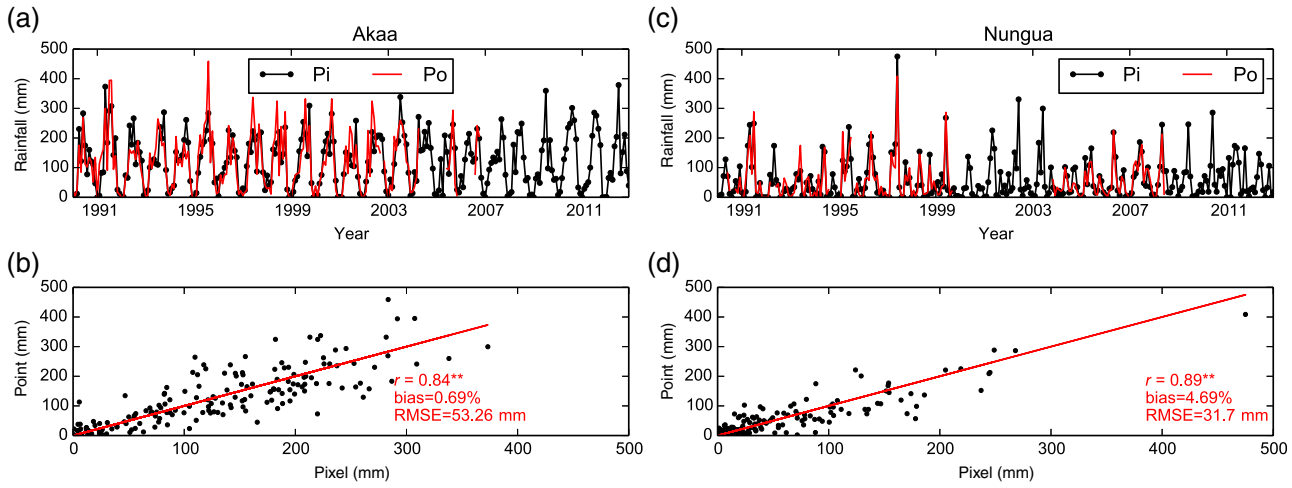

Figure 3. Point-pixel rainfall validation over Akaa (7.427°N, 0.407°E) (a and c) in the Transition zone and Nungua [5.600°N, 0.067°W] (b and d) in the Coast, with point observation shown in red and pixel data in black. Goodness-of-fit is as shown in (c) and (d). [Colour figure can be viewed at wileyonlinelibrary.com].

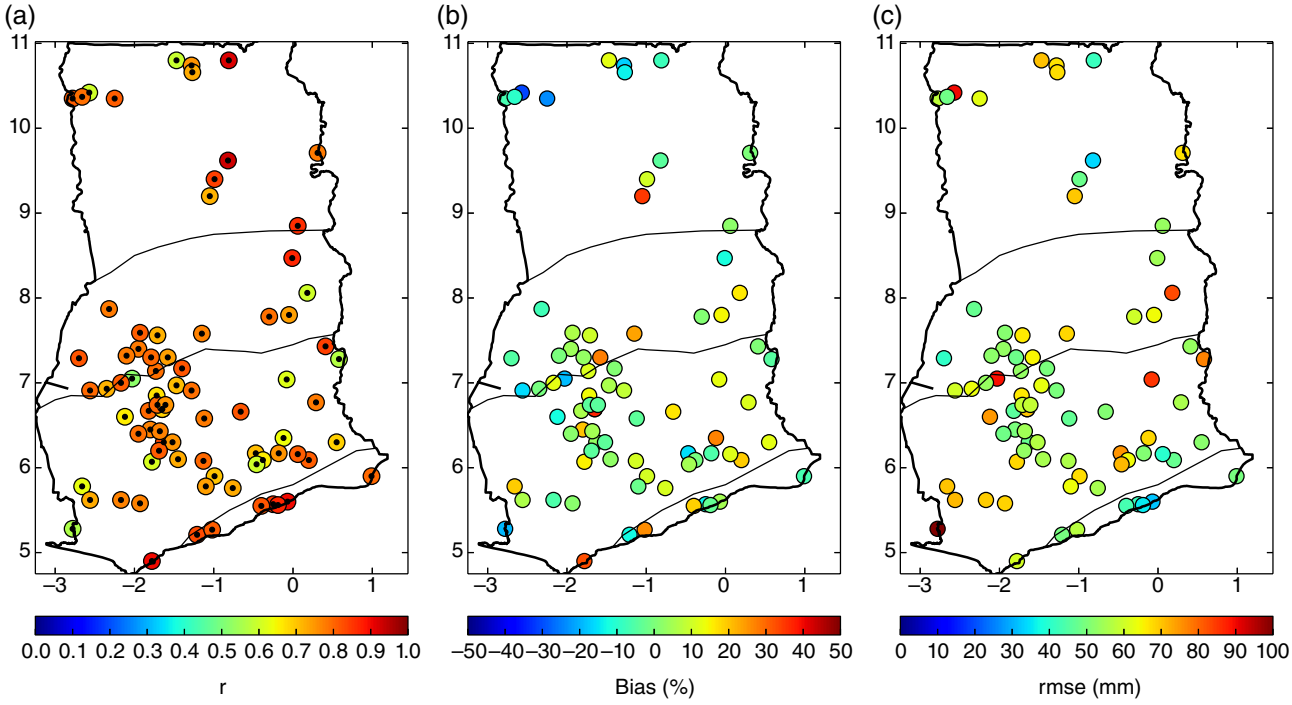

Figure 4. Point-pixel rainfall validation showing correlation coefficients (a) with statistical significance at 99% confidence level shown as black dots, bias (b) and rmse (c) over all four agro-ecological zones. [Colour figure can be viewed at wileyonlinelibrary.com].

Basic Synoptic Network from which monthly surface climatological data are collected to be such that every 250 000 km^2 is represented by at least one station and by up to 10 evenly spread stations, if possible (WMO, 2011). Per these recommended densities, a 2.5° × 2.5° longitude–latitude goodness-of-fit was performed, as shown in Figure 5(b). Here, the country was divided into six zones. The average rainfall from all point stations was compared to the grid average. Despite the relative low station-network in the upper zones, correlations between the averaged precipitation of the stations within a grid, to GMet v1.0, were relatively high. These zones have a unimodal rainfall pattern and thus, there is relatively low variability in their rainfall patterns, hence accounting for these relatively high correlation coefficients. Generally, for all six zones, coefficients were above 0.9 and statistically significant at 99% confidence level, which is indication of good spatial performance of GMet v1.0 over the entire country, irrespective of station density.

4.4. Assessment of spatial fields for monthly, seasonal and annual rainfall

Presently, GPCC provides a gridded gauge-based dataset that has been found to perform well over the study domain (Manzanas *et al.,* 2014), thus, spatial fields of monthly,

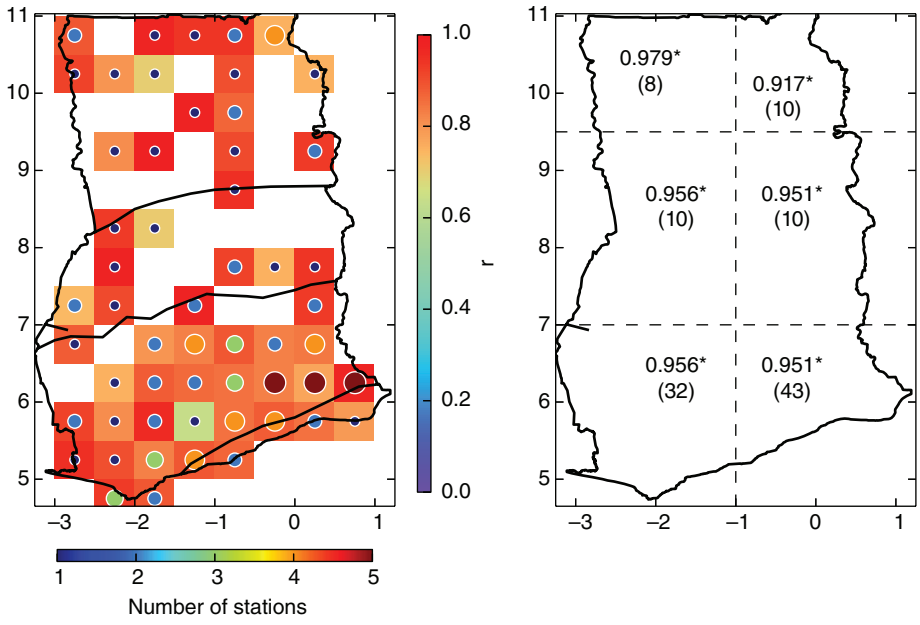

Figure 5. Measure of goodness-of- fit on (a) 0.5° × 0.5° and (b) 2.5° × 2.5° spatial resolution. Raster shown in (a) indicate correlation coefficients for total number of stations per grid (circles). Number of stations per grid has been classified per the horizontal legend. Further shown in (b) are correlation coefficients (above) statistically significant at 99% confidence level (*), and total stations per grid (below and in parenthesis). [Colour figure can be viewed at wileyonlinelibrary.com].

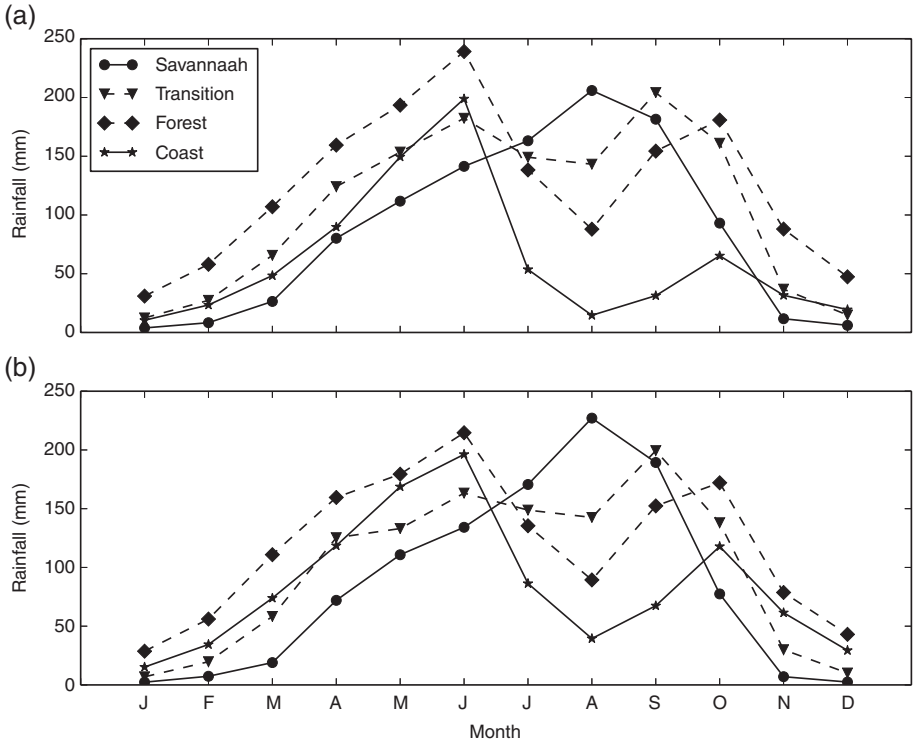

Figure 6. Mean monthly total rainfall over all four agro-ecological zones from GMet v1.0 (a) and GPCC (b).

seasonal and annual rainfall have been assessed for both GPCC and GMet v1.0.

First, monthly rainfall patterns of the four agro-ecological zones are illustrated in Figure 6(a) (for GMet v1.0) and Figures 6(b) (for GPCC), with each point on the graph representing the mean monthly rainfall amount, over the 23-year period, for each zone. A unimodal rainfall pattern, associated with the northward migration of the tropical rainbelt (Amekudzi *et al.*, 2015), was observed in the Savannah zone, with peaks in August (Nicholson, 2009), whereas the Transition, Forest and Coastal zones were characterized by bi-modal rainfall patterns, separated by a temporary cessation (referred to as little dry season, due to its reduced

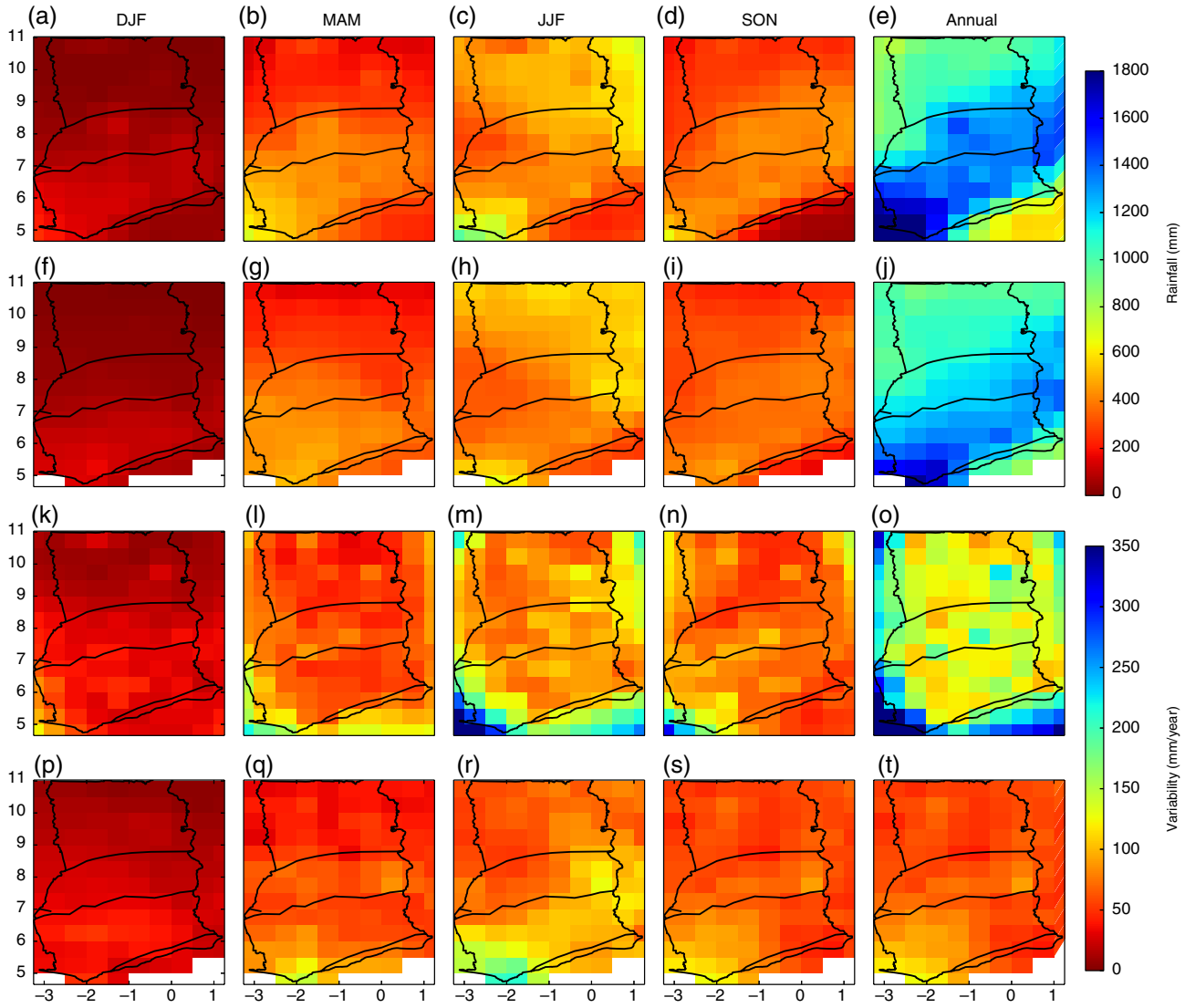

Figure 7. Mean seasonal and annual rainfall total from GMet v1.0 (a–e) and GPCC (f–j), as well as, rainfall variability from GMet v1.0 (k–o) and GPCC (p–t) for the study domain. [Colour figure can be viewed at wileyonlinelibrary.com].

rainfall amount and/or frequency; see Figures 6, 8(a) and (b)) that is linked to the break between the northward ascent of the tropical rainbelt and its sharp return (Mensah *et al.,* 2016).

Moreover, seasonal and annual rainfall patterns of GMet v1.0 (Figures 7(a)–(e)) were analogous to GPCC (Figures 7(f)–(j)), with both rightly illustrating the migration of the tropical rainbelt over the country (see Figure 7). Generally, the Forest zone recorded the most rainfall, followed by the Transition, Savannah and Coastal zones (see Table 1). Seasonal analysis of both GMet v1.0 and GPCC captured DJF (December–January–February) as the driest period over the entire country, with a rainfall total below 300 mm (see Figures 7(a)–(f)). Similar to findings in Owusu and Waylen (2013), Amekudzi *et al.* (2015) and Mensah *et al.* (2016), the MAM (March–April–May) season was observed to be the rainfall onset period for all four agro-ecological zones (see Figures 7(b)–(g)), which is attributable to the northward migration of the tropical rainbelt (N'Tchayi Mbourou *et al.,* 1997; Sultan and Janicot, 2003; Flamant *et al.,* 2007; Nicholson, 2009). Within this season, suspected heightening in solar activity (Chakraborty and Bandyopadhaya, 1987) over the zones leads to the generation of more convective systems. Amekudzi *et al.* (2015) (and references therein), highlighted that the nature of vegetation and terrain also influence the development of these local convective systems, which tend to induce the onset of rains. Also, JJA (June–July–August) was identified as the wettest season, with likely occurrence of most probable extreme rainfall events (see Figures 7(c) and (h)). The SON (September–October–November) season was identified as the minor rainy season for the Transition, Forest and Coastal zones (see Figures 7(d) and (i)). Rainfall in this season is mostly attributed to the sharp return of the tropical rainbelt southwards, which introduces a near-dry condition in the upper Savannah zone.

Also, countrywide rainfall variabilities have been equally provided in Figures 7(k)–(t), showing each grid's observed seasonal change in rainfall amount per

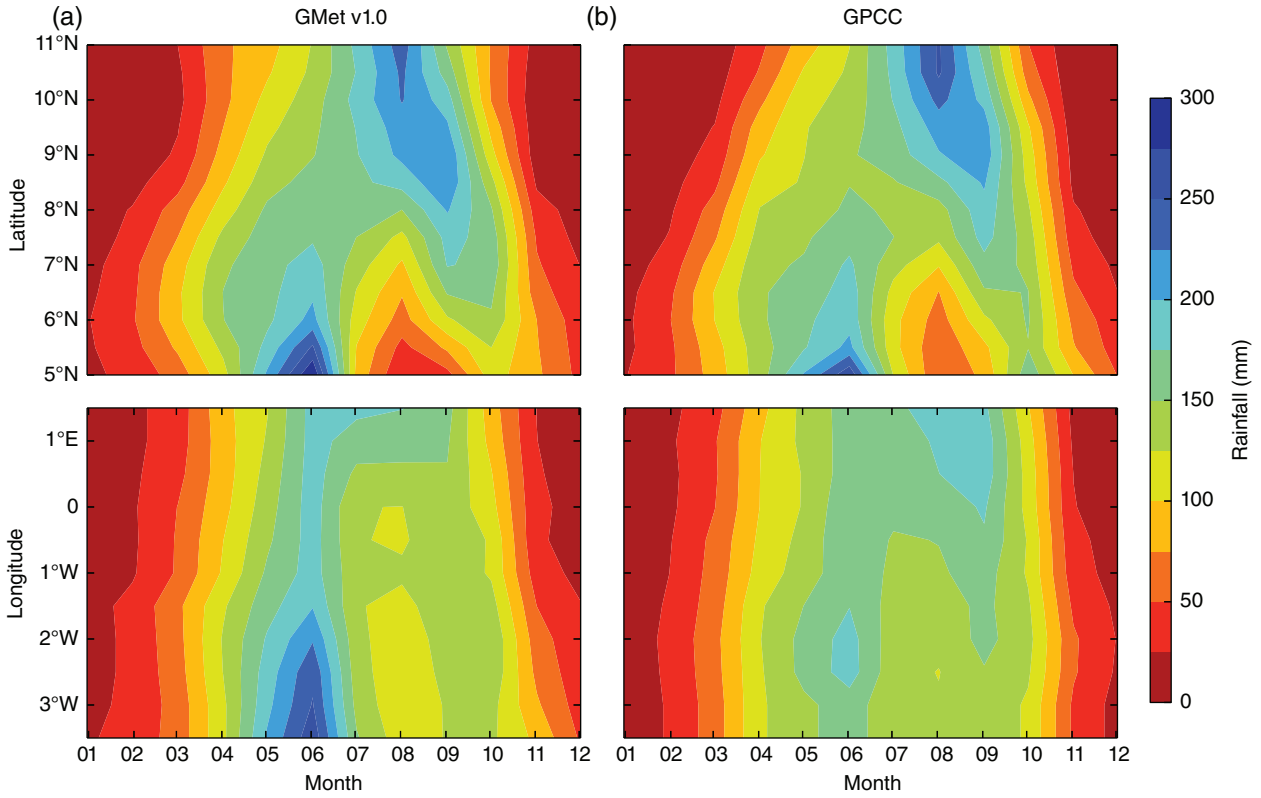

Figure 8. Zonal (top) and Meridional (bottom) Hovmoller diagrams from GMet v1.0 (a) and GPCC (b), showing month-to-month latitudinal and longitudinal migration of the tropical rainbelt over the entire country. [Colour figure can be viewed at wileyonlinelibrary.com].

year, for the study period, as observed by both GMet v1.0 and GPCC. The dry season (DJF) was found to have low variabilities, which was generally less than 50 mm year^{-1}. On the other hand, rainfall variability during the MAM, JJA and SON seasons were generally about 50–150 mm year^{-1}, and also widespread. However, from GMet v1.0, the south-west Forest zone was found to have variability of about 250 mm year $^{-1}$ for the JJA season. Annually, rainfall variability as detected by GPCC was generally lower than 150 mm year^{-1}, whereas GMet v1.0 ranged between 100 and 300 mm year^{-1}, with large variabilities over the south-west Forest zone.

Furthermore, annual assessment of both GMet v1.0 and GPCC was performed to quantify on average the rainfall distribution per year over the entire country. Generally, the Savannah and Coastal zones were found to record the least amount of rainfall per year, whereas higher rainfall amounts were observed for the Transition and Forest zones (see Figures 7(e) and (j)), with rainfall peaks in the south-west Forest zone.

In addition, a zonal assessment was performed for both GMet v1.0 and GPCC via Hovmoller diagram, as shown in Figures 8(a) and (b), respectively. Various features were derived from the plot, spanning rainfall onsets, monsoon flows and transitions, wet and dry periods and monthly rainfall variabilities. To begin with, as opposed to local knowledge that rainfall onsets start from the lower latitudes, it was found to be between latitude 6 and 7 N (given threshold of 50–100 mm per month). After the early onsets in latitude 6 –7 N, onsets were spread over the low latitudes, before its ascent to the high latitudes by March. Also, a bimodal rainfall pattern was captured in the low latitudes with rainfall peaks in June, and a unimodal rainfall pattern in the high latitudes, with peaks in August. Again, the smooth seasonal migration of the tropical rainbelt is observed from the low latitudes up, followed by its return. Moreover, the little dry spell, as discussed in Section 4.4, was also identified. Finally, smooth transitions in rainfall cessation were also captured from the north of the country (by early November) to the Coasts (from mid to late November).

Furthermore, a meridional rainfall variability was assessed, as shown in Figures 8(c) and (d) for GMet v1.0 and GPCC. Earlier rainfall onsets were captured between 3° and 1.5° W. Also, the monsoon flow was seen to be propagating eastwards during the first half of the year. Rainfall maxima were observed in June for the westward longitudes, whereas the eastward longitudes had a widespread meridional rainfall maxima between May and September. These maxima however were relatively lower than the westward longitudes.

Consequently, superimposing the zonal Hovmoller diagram on the meridional, it was observed that earliest onsets were captured around 6°N, 3°W and 7°N, 1.5°W. Also, the tropical rainbelt was generally found to be migrating from south-west to north-east of the country during its ascent, and then later retreats generally southwards (see Figures 8(a) and (b)), and widespread across all latitudes (see Figures 8(c) and (d)).

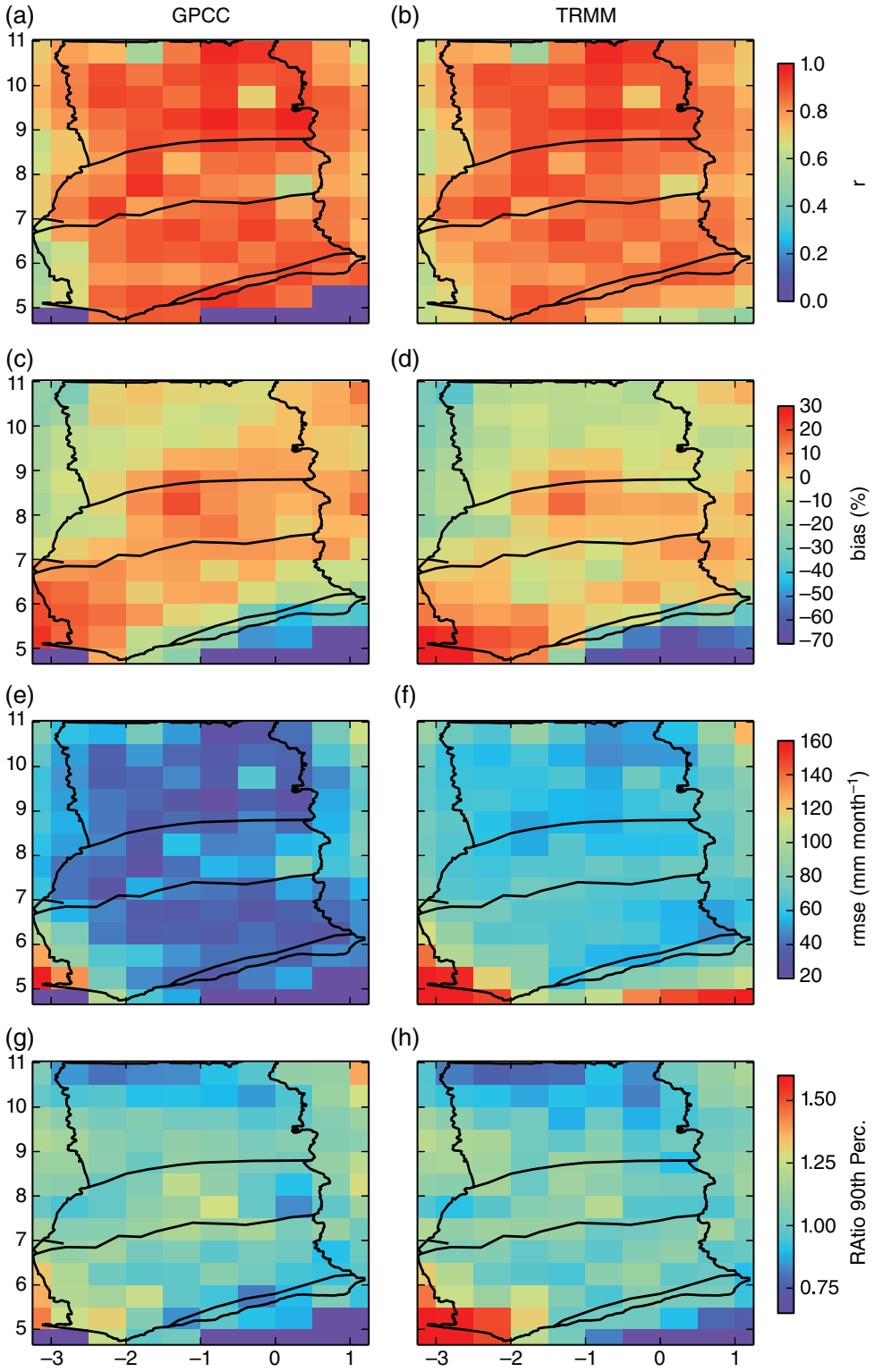

Figure 9. Countrywide GMet v1.0 – GPCC (a) and GMet v1.0 – TRMM (b) comparison, showing correlations (a and b) statistically significant at 99% confidence level, bias (c and d), rmse (e and f) and ratio of 90th percentile (g and h). [Colour figure can be viewed at wileyonlinelibrary.com].

4.4.1. Statistical comparison of GMet v1.0 data with other products

Afterwards, a comparison of the performance of GMet v1.0 against TRMM and GPCC was carried out using Pearson's correlation coefficient at 99% confidence level, bias, rmse and ratio of 90th percentiles (see Figure 9). These metrics generally showed a good agreement of GMet v1.0 with the other products over the entire country, however with higher correlation coefficients over the Savannah zone and reduced coefficients towards the south-west Forest zone.

Table 1. Mean seasonal and annual total rainfall in the four agro-ecological zones.

Period	Rainfall (mm)			
	Savannah	Transition	Forest	Coast
December–January–February (DJF)	76.8	138.5	190.1	100.9
March–April–May (MAM)	322.7	432.9	544.4	398.1
June–July–August (JJA)	529.0	376.4	380.4	175.1
September–October–November (SON)	135.2	213.5	289.9	183.5
Annual	1063.7	1161.3	1404.8	857.6

Table 2. Description of gauge and satellite products used.

Product	Spatial res.	Temporal res.	Duration	Product type
Default				
GMet v1.0	$0.5^\circ \times 0.5^\circ$	Monthly	1990–2012	Gauge
TRMM	$0.25^\circ \times 0.25^\circ$	Daily	1998–2015	Satellite + gauge
GPCC	$0.25^\circ \times 0.25^\circ$	Monthly	1901–2015	Gauge
For comparison				
GMet v1.0	$0.5^\circ \times 0.5^\circ$	Monthly	1990–2012	Gauge
TRMM	$0.5^\circ \times 0.5^\circ$	Monthly	1998–2012	Satellite + gauge
GPCC	$0.5^\circ \times 0.5^\circ$	Monthly	1990–2012	Gauge

Over the Savannah zone, both products seem to rightly mimick rainfall estimates of GMet v1.0. High correlations and low biases, on average, greater than 0.8 and ±10% respectively buttress these findings. However, in the north-western Savannah zones, relatively higher dry biases are detected for both GPCC and TRMM, with TRMM fairly widespread. For the Transition, Forest and Coast, both products fairly correlated that of GMet v1.0 over most grids, with correlation coefficients ranging between 0.6 and 0.9. The Transition and Forest zones were however dominated by wet biases of about 20%, whereas the Coast was also dominated by dry bias of about 30%. Rainfall estimates over the south-western Forest zone were strongly overestimated in TRMM than GPCC on order of about 30%.

Subsequent computation of root mean square error also provided further information on deviation of both TRMM and GPCC, relative to GMet v1.0. Generally, rmse were on the order of 30–70 mm per month countrywide for GMet v1.0 – GPCC, and 50–100 mm per month for GMet v1.0 – TRMM. Again, the south-west Forest zones recorded higher rmse values in both. Furthermore, a measure of extremes was analysed via the ratio of 90th percentile, as shown in Figures 9(g) and (h). Here a value of 1 shows similar capture of extremes between GMet v1.0 and either GPCC or TRMM, whereas values less than one suggest that extremes detected by GMet v1.0 are lesser than that of either product, and values greater than one suggest otherwise. It is observed countrywide that extremes as detected by both GMet v1.0 and either TRMM or GPCC greatly match, however with strong overestimates by GMet v1.0 in the south-west Forest zone, especially against TRMM.

In addition, it is worth noting that variations in the statistical comparisons of GMet v1.0 with TRMM and GPCC could be likely due to the different time domains over which these datasets have been compared as shown in Table 2 (GPCC from 1990–2012; TRMM from 1998–2012) and/or their sampling techniques.

4.5. Climatic delineation of Ghana based on GMet v1.0 analysis

As a final step, a new climatic delineation was done by means of clustering. First, the datasets were sorted by both monthly variability (for seasonal pattern detection) and annual totals. For each grid, the mean annual rainfall total was deduced and then collectively, the means were all standardized on a countrywide-scale using a common mean and standard deviation. Thereafter, they were clustered using the *k*-means cluster (precisely, Lloyd's algorithm), described in Kanungo *et al.* (2002), which is based on the simple observation that optimal placements of a data centre is at the centroid of the associated cluster. The method assigns similar labels to regions bounding the same centre, and with homogeneous rainfall patterns and standardized annual means. Consequently, four homogeneous climatic zones were delineated from clustering of GMet v1.0, as shown in Figure 10(a), with their respective monthly rainfall variability provided in Figure 10(b).

To begin with, Zone A was found to have a unimodal rainfall pattern, with majority of rains captured between April and October, as shown in Figure 10(b). Also, Zone B was observed to have a bimodal rainfall pattern with rainfall amounts greater than A. Zone C has a similar bimodal pattern as B, however with higher rainfall amounts. Finally, Zone D was observed to have low rainfall amounts, almost similar to Zone A, however with a bimodal rainfall pattern.

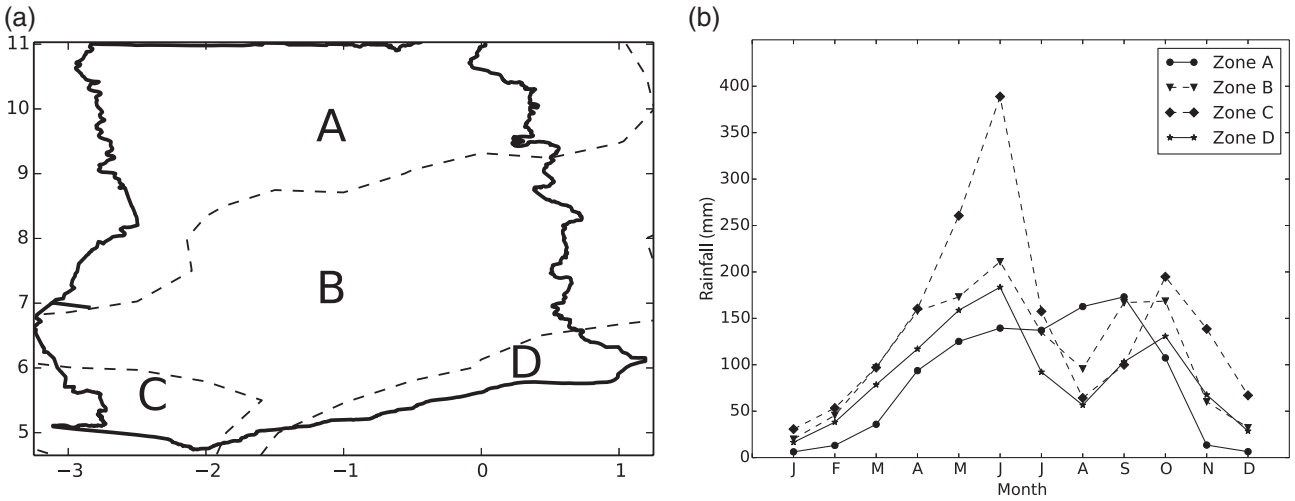

Figure 10. Climatic zonation of Ghana (a) and the monthly rainfall variability (b) for each of the delineated zones.

5. Conclusions

In this study, a rainfall climatological database has been developed by first homogenizing rainfall data from 113 meteorological stations distributed within the four agro-ecological zones of Ghana. First, absolute homogenization was performed to detect the various change points within the observed data, which were then augmented based on the magnitude of the changepoints, using QMadj. Afterwards, the homogenized rainfall data was gridded at a $0.5° \times 0.5°$ spatial resolution using the MSC with tensioning parameter algorithm, allowing for comprehensive spatial fields assessment on monthly, seasonal and annual time-scales.

Furthermore, point-pixel validation was performed using GMet v1.0 data against gauge data from stations that were earlier excluded from the gridding due to large datagaps. This validation proved the reliability of GMet v1.0, since it was able to rightly mimick the rainfall for these stations by at least 70%, which was statistically significant at 99% confidence level, and with dry and wet biases of at maximum ±30%. Also, rmse was found to be generally about 70 mm per month.

A comparison of GMet v1.0 with GPCC and TRMM precipitation estimates was carried out, and both products were found to rightly mimick GMet v1.0 with correlations greater than 0.8 at 99% confidence level, and biases of about ±30% countrywide. Moreover, rmse values countrywide were found to range from 20–70 mm and 40–100 mm per month for GMet v1.0 – GPCC and GMet v1.0 – TRMM, respectively. Generally, the south-west Forest zone was found to exhibit relatively lower correlations, and high bias and rmse values. Additionally, ratio of 90*th* percentiles captured almost similar percentile values countrywide for both GPCC and TRMM, however, the south-west Forest zone was strongly overestimated, especially by GMet v1.0 against TRMM.

In addition, based on spatial assessments of patterns and rainfall amounts of GMet v1.0, the country was further re-delineated into four climatic zones using *k*-means cluster analysis. Zones A and D were observed to have almost similar rainfall amounts, but unimodal and bimodal rainfall patterns respectively. Zones B and C also had bimodal patterns but recorded relatively higher rainfall amounts than A and D, however zone C had the maximum rainfall amounts.

To conclude, the developed GMet v1.0 data will be potentially useful for validation studies and other rainfall-based applications such as disease-modelling, hydrological modelling, among others. This is a preliminary study, with the intent of updating the database for years beyond the current study scope. Also, the data is currently available only to weather and climate data users in Ghana, till approved for upload to a wider user community, e.g. via the GTS by Ghana Meteorological Agency.

Acknowledgements

The authors wish to acknowledge the Ghana Meteorological Agency for providing monthly rainfall data for this study. The TRMM 3B43 data were provided by the NASA/Goddard Space Flight Center and PPS, which develop and compute the TRMM 3B43 as a contribution to TRMM, and archived at the NASA GES DISC. The authors also appreciate the Earth System Research Lab for provision of GPCC data. Extra appreciation to the Climate Research Division of Environment Canada for the provision of the RHtestsV4 package. Finally, the authors are also thankful to the anonymous reviewers for their extensive reviews that helped to shape this paper in the best possible way.

Conflict of interest

The authors declare no conflict of interest.

Appendix:

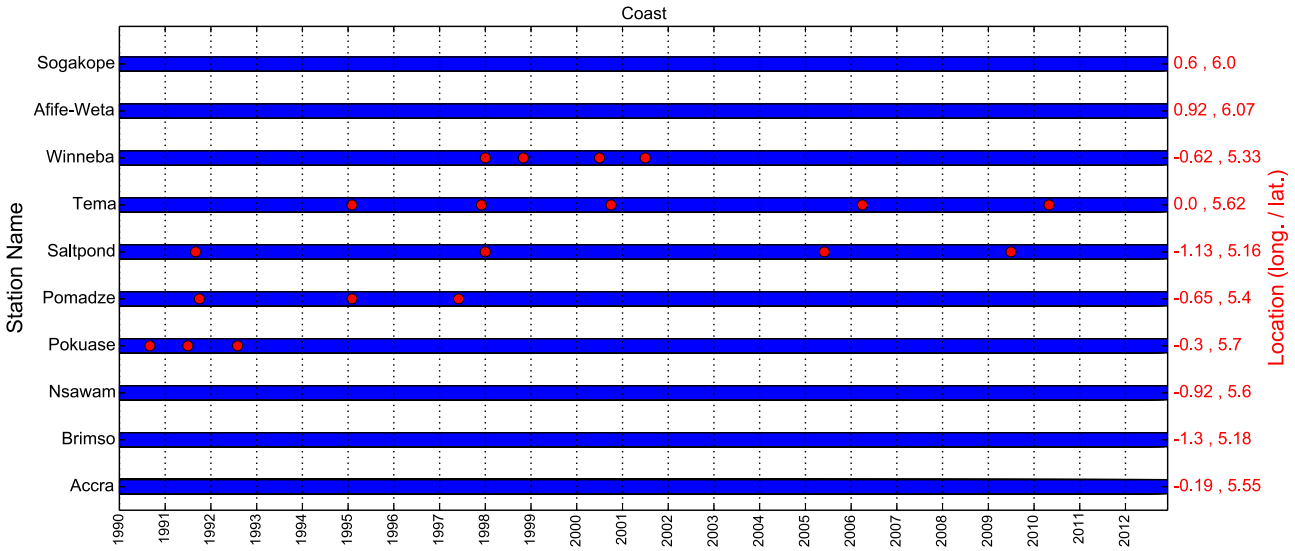

Figure A1. Changepoints identified in stations within coastal zone. [Colour figure can be viewed at wileyonlinelibrary.com].

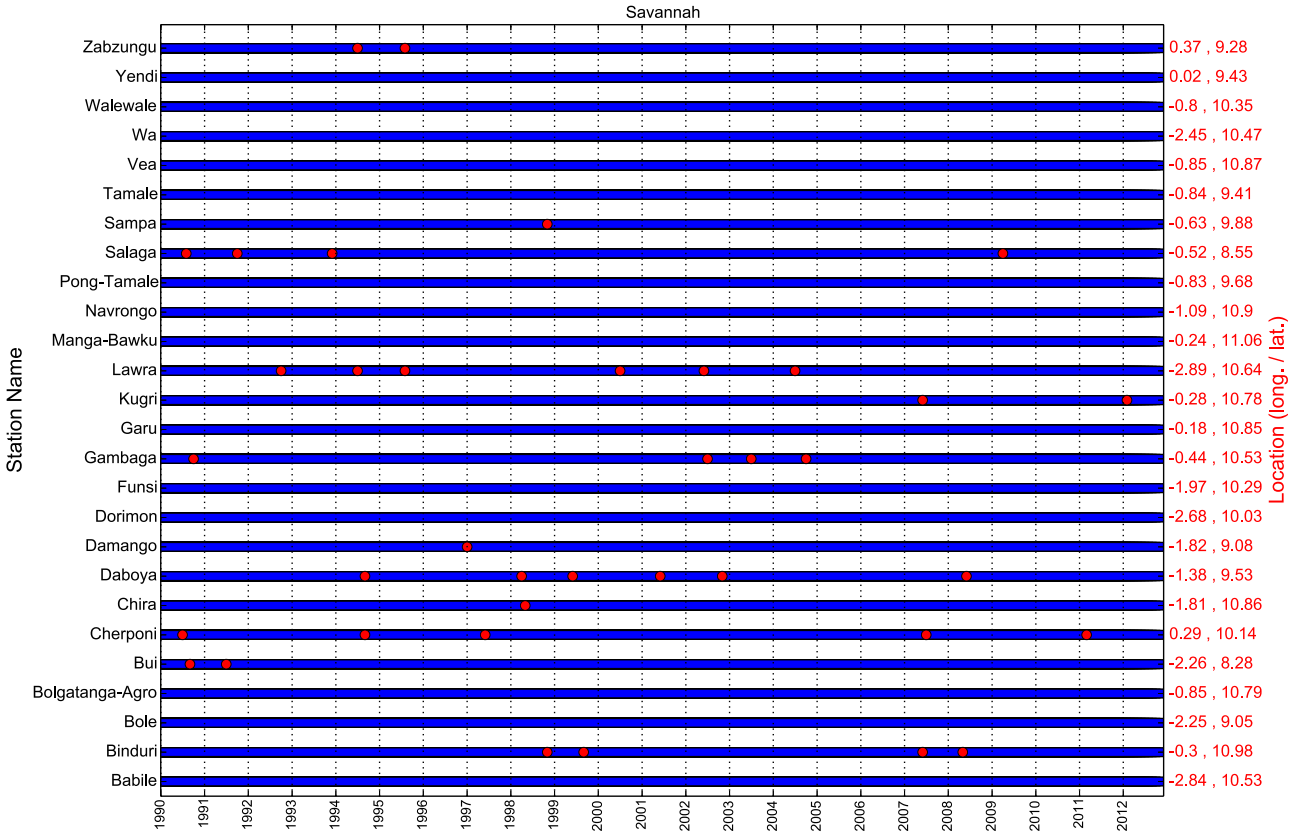

Figure A2. Changepoints identified in stations within the Savannah zone. [Colour figure can be viewed at wileyonlinelibrary.com].

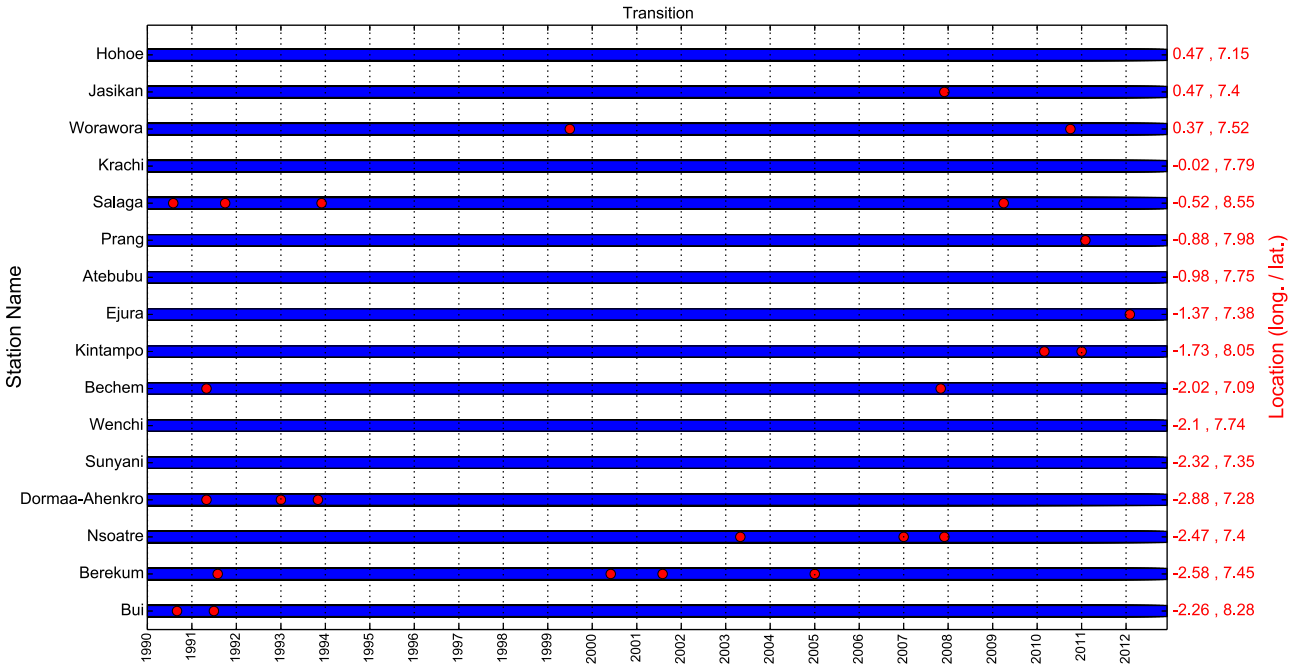

Figure A3. Changepoints identified in stations within the transition zone. [Colour figure can be viewed at wileyonlinelibrary.com].

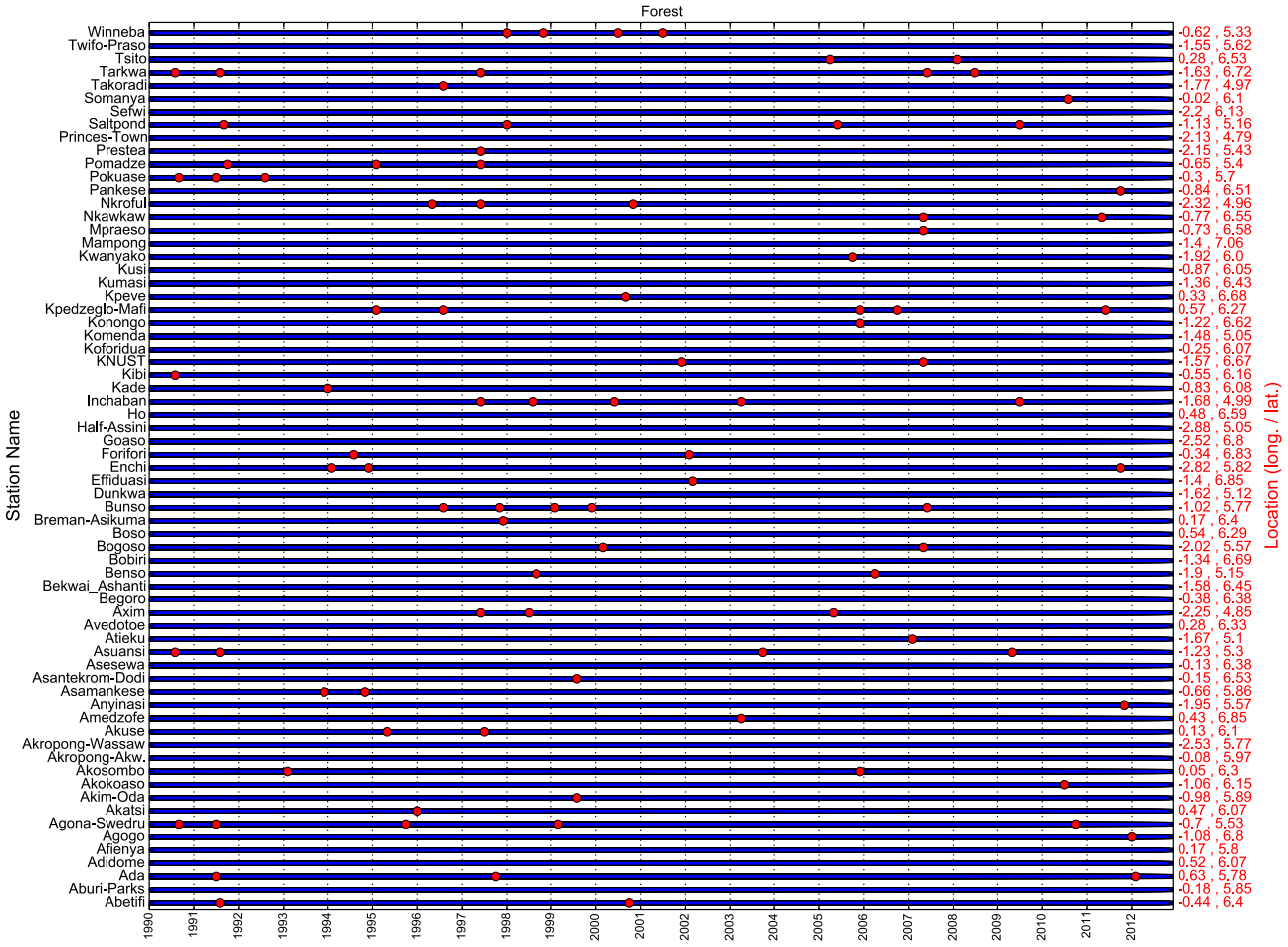

Figure A4. Changepoints identified in stations within the forest zone. [Colour figure can be viewed at wileyonlinelibrary.com].

References

Amekudzi LK, Bracher A, Bramstedt K, Rozanov A, Bovensmann H, Burrows JP. 2008. Towards validation of SCIAMACHY lunar occultation NO_2 vertical profiles. *Adv. Space Res.* **41**(11): 1921–1932.

Amekudzi LK, Osei MA, Atiah WA, Aryee JNA, Ahiataku MA, Quansah E, Preko K, Danuor SK, Fink AH. 2016. Validation of TRMM and FEWS satellite rainfall estimates with rain gauge measurement over Ashanti region, Ghana. *Atmos. Clim. Sci.* **6**(04): 500.

Amekudzi LK, Yamba E, Preko K, Asare EO, Aryee J, Baidu M, Codjoe NAS. 2015. Variabilities in rainfall onset, cessation and length of rainy season for the various agro-ecological zones of Ghana. *Climate* **3**: 416–434.

Baidu M, Amekudzi LK, Aryee JNA, Annor T. 2017. Assessment of long-term spatio-temporal rainfall variability over Ghana using wavelet analysis. *Climate* **5**(2): 30.

Blair T. 2012. Techniques involved in developing the Australian climate observations reference network - surface air temperature (ACORN-SAT) dataset. Technical Report 049, Centre for Australian Weather and Climate Research.

Chakraborty PK, Bandyopadhaya R. 1987. Variation of the onset of Indian south-west monsoon over Kerala with solar activity. *Proc. Indian Nat. Sci. Acad.* **53**(2): 303–307.

Dressler M. 2007. Interpolation methods for construction of surfaces. Technical University of Liberec [links].

Feng-Wen C, Chen-Wuing L. 2012. Estimation of spatial rainfall distribution using inverse-distance weighting (idw) in the middle Taiwan. *Paddy Water Environ.* **10**: 209–222.

Flamant C, Chaboureau JP, Parker DJ, Taylor CM, Cammas JP, Bock O, Timouk F, Pelon J. 2007. Airborne observations of the impact of a convective system on the planetary boundary layer thermodynamics and aerosol distribution in the inter-tropical discontinuity region of the west African monsoon. *Q. J. R. Meteorol. Soc.* **133**: 1175–1189.

Gruber A, Levizzani V. 2008. Assessment of Global Precipitation Products: A Project of the World Climate Research Programme Global Energy and Water Cycle Experiment (GEWEX) Radiation Panel. Technical report, WMO/TD.

Huffman GJ, Adler RF, Bolvin DT, Nelkin EJ. 2010. The TRMM multi-satellite precipitation analysis (TMPA). In *Satellite Rainfall Applications for Surface Hydrology*. Springer: Dordrecht, The Netherlands, 3–22.

Huffman GJ, Bolvin DT, Nelkin EJ, Wolff DB, Adler RF, Gu G, Hong Y, Bowman KP, Stocker EF. 2007. The trmm multisatellite precipitation analysis (tmpa): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* **8**(1): 38–55.

Jeffrey SJ, Carter JO, Moodie KB, Beswick AR. 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environ Model Softw* **16**(4): 309–330.

Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY. 2002. An efficient k-means clustering algorithm: analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7): 881–892.

Li J, Heap AD. 2008. Spatial interpolation methods: a review of environmental scientists. *Geosci. Australia*: Record 2008/23, 1–137.

Liu Z, Riu H, Teng W, Chiu L. 2003. Online intercomparison of trmm and global ridde precipitation products. In: Proc. 17th Conf. Hydrology, 2003 AMS Annual Meeting, Long Beach, CA.

Lu GY, Wong DW. 2008. An adaptive inverse-distance weighting spatial interpolation technique. *Comput. Geosci.* **34**(9): 1044–1055.

Manzanas R, Amekudzi LK, Preko K, Herrera S, Gutierrez JM. 2014. Precipitation variability and trends in Ghana: an intercomparison of observational and reanalysis products. *Clim. Change* **124**: 805–819.

Mengistu Tsidu G. 2012. High resolution monthly rainfall database for Ethiopia: homogenization, reconstruction and gridding. *J. Clim.* **25**(24): 8422–8443.

Mensah C, Amekudzi LK, Klutse NAB, Aryee JNA, Asare K. 2016. Comparison of rainy season onset, cessation and duration for Ghana from regcm4 and gmet datasets. *Atmos. Clim. Sci.* **6**(02): 300.

Nicholson SE. 2009. A revised picture of the structure of the monsoon and land itcz over west Africa. *Clim. Dyn.* **32**(7–8): 1155–1171.

N'Tchayi Mbourou G, Bertrand JJ, Nicholson SE. 1997. The diurnal and seasonal cycles of wind-borne dust over Africa north of the equator. *J. Appl. Meteorol.* **36**(7): 868–882.

Owusu K, Waylen PR. 2013. The changing rainy season climatology of mid-Ghana. *Theor. Appl. Climatol.* **112**(3): 419–430.

Price DT, McKenney DW, Nalder IA, Hutchison MF, Kesteven JL. 2000. A comparison of two statistical methods for interpolation of Canadian monthly mean climate. *Agric Meteorol.* **101**: 81–94.

Quansah E, Amekudzi LK, Preko K, Aryee J, Boakye RO, Boli D, Salifu MR. 2014. Empirical models for estimating global solar radiation over the Ashanti region of Ghana. *J. Solar Energy* **2014**: 897970.

Schneider T. 2001. Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. *J. Clim.* **14**: 853–871.

Smith WHF, Wessel P. 1990. Gridding with continuous curvature splines in tension. *Geophysics* **55**(3): 293–305.

Sultan B, Janicot S. 2003. The west African monsoon dynamics part ii: preonset and onset of the summer monsoon. *J. Clim.* **16**: 3407–3427.

Wang X, Chen H, Wu Y, Feng Y, Pu Q. 2010. New techniques for detection and adjustment of shifts in daily precipitation data series. *J. Appl. Meteorol. Climatol.* **49**: 2416–2436.

WMO. 2008. Guide to meteorological instruments and methods of observation.

WMO. 2011. Manual on the Global Observing System - Volume II.

Yeh HC, Chen YC, Wei C, Chen RH. 2011. Entropy and kriging approach to rainfall network design. *Paddy Water Environ.* **9**: 343–355.

