Seasonal Forecasting of Onset of Summer Rains over South Africa

J. V. RATNAM AND TAKESHI DOI
Application Laboratory, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

WILLEM A. LANDMAN
Department of Geography, Geoinformatics and Meteorology, University of Pretoria, South Africa

SWADHIN K. BEHERA
Application Laboratory, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

(Manuscript received 14 March 2018, in final form 12 October 2018)

ABSTRACT

In this study, we attempted to forecast the onset of summer rains over South Africa using seasonal precipitation forecasts generated by the Scale Interaction Experiment–Frontier Research Center for Global Change, version 2 (SINTEX-F2), seasonal forecasting system. The precipitation forecasts of the 12-member SINTEX-F2 system, initialized on 1 August and covering the period 1998–2015, were used for the study. The SINTEX-F2 forecast precipitation was also downscaled using dynamical and statistical techniques to improve the spatial and temporal representation of the forecasts. The Weather Research and Forecasting (WRF) Model with two cumulus parameterization schemes was used to dynamically downscale the SINTEX-F2 forecasts. The WRF and SINTEX-F2 precipitation forecasts were corrected for biases using a linear scaling method with a 31-day moving window. The results indicate the onset dates derived from the raw and bias-corrected model precipitation forecasts to have realistic spatial distribution over South Africa. However, the forecast onset dates have root-mean-square errors of more than 30 days over most parts of South Africa except over the northeastern province of Limpopo and over the Highveld region of Mpumalanga province, where the root-mean-square errors are about 10–15 days. The WRF Model with Kain–Fritsch cumulus scheme (bias-corrected SINTEX-F2) has better performance in forecasting the onset dates over Limpopo (the Highveld region) compared to other models, thereby indicating the forecast of onset dates over different regions of South Africa to be model dependent. The results of this study are important for improving the forecast of onset dates over South Africa.

1. Introduction

Most of the subsistence farmers across South Africa depend on the onset of summer rains for planting maize, the staple food of the country (Tadross et al. 2005; Moeletsi et al. 2011). Successful long-lead forecasting of the dates of onset would be beneficial to the farmers in planning their farming activities. There have been some studies, for example, Reason et al. (2005), Tadross et al. (2005), and Moeletsi et al. (2011), to understand the processes associated with the interannual variation of onset of rains over South Africa. The dates of onset over South Africa have variability from interannual to decadal time scales (Tadross et al. 2005). Using self-organizing maps of the 500-hPa geopotential anomalies, Tadross et al. (2005) found the timing of the onset of rains over South Africa in the summer season to be associated with the location of the midtropospheric height anomalies in the vicinity of South Africa. Their study indicates that an early onset over South Africa is generally favored when positive midtropospheric height anomalies are observed on the southeast of the continent. Similar results were reported in the study of Reason et al. (2005). The studies of Reason et al. (2005) and Moeletsi et al. (2011) indicate the leading modes of global climate variability, such as El Niño–Southern Oscillation (ENSO), have no significant influence on the interannual variability of the dates of onset over South Africa during the austral summer season, with early (delayed) onset dates observed during

Corresponding author: J. V. Ratnam, jvratnam@jamstec.go.jp

DOI: 10.1175/JAMC-D-18-0067.1
© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).
both phases of ENSO. Though these studies made an effort to understand the physical mechanisms for the onset of rains over South Africa and its variability, there have been few efforts to forecast the onset of the summer rains over South Africa at seasonal time scales. We try to fill this gap by using seasonal precipitation forecasts from the Scale Interaction Experiment–Frontier Research Center for Global Change, version 2 (SINTEX-F2; Doi et al. 2016, 2017), seasonal forecasting system to forecast the dates of onset of austral summer rains over South Africa. The onset of rains over South Africa is during the early summer months, from September to November. The seasonal forecast skill over South Africa during the early summer months is relatively low compared to December–February (Landman and Mason 1999; Landman et al. 2012) because of the absence of tropical influences. This relatively low skill during spring and the lack of significant relation of onset of summer rains to ENSO is indicative of the challenges faced when trying to predict dates of onset over South Africa.

Similar attempts to forecast onset of rains over Africa and other parts of the world have been reported, for example, Vellinga et al. (2013), Pradhan et al. (2017), MacLeod (2018), Phakula et al. (2018). Vellinga et al. (2013) found the skill in forecasting the regional onset of West Africa monsoon to be modest, with relative operating characteristic (ROC) scores of 0.6–0.8, at a 2–3-month lead times in the hindcasts of the Met Office seasonal forecasting system Global Seasonal forecast system, version 4 (GloSea4), and six forecasting systems from the ENSEMBLES project. The skill in forecasting the onset over West Africa comes from the models’ ability to realistically reproduce the teleconnections of the atmospheric response to tropical oceans (Vellinga et al. 2013). MacLeod (2018), evaluating the ECMWF seasonal hindcasts, found the skill in forecasting the onset of short rains over East Africa to be higher than that of forecasting the long rains. The higher skill in forecasting the short rains is mainly due to the influence of the tropical oceans on the short rains. Pradhan et al. (2017) found the onset of monsoon over India to be predictable 3 months in advance by using the Indian Institute of Tropical Meteorology (IITM) CFSv2 T382 model. In a recent study Phakula et al. (2018), attempted to forecast the onset months over South Africa using global model outputs. The study found the forecast of the precipitation totals for the onset months of the regions of South Africa to have ROC score of >0.5 with modest reliability and limited sharpness in the forecasts. The onset of rains over South Africa, unlike West Africa, East Africa, and India, is less influenced by the tropical oceans and more by internal dynamics, which limits the skill of predicting onset of rains over South Africa.

The most widely used definition of onset of rains over South Africa is based on the amount of precipitation required for the successful germination of maize (Reason et al. 2005; Tadross et al. 2005). Though there are other definitions of onset of rains over Africa, for example, Boyard-Micheau et al. (2013) and Dunning et al. (2016), we use the widely accepted definition of Reason et al. (2005) and Tadross et al. (2005) in our study. To improve the precipitation forecasts over South Africa, we downscale the SINTEX-F2 forecasts using both the dynamical and statistical downscaling methods. The technique of dynamical downscaling using a regional climate model has been widely used to improve the simulation of precipitation over South Africa (Joubert et al. 1999; Tadross et al. 2006; Hansingo and Reason 2008; Kgatuke et al. 2008; MacKellar et al. 2009; Landman et al. 2009; Créta et al. 2011; Ratnam et al. 2012, 2013, 2015, 2016; Ratna et al. 2014; Diallo et al. 2015; Moalafhi et al. 2017) In this study, we use a regional model, the Advanced Research version of the Weather Research and Forecasting (WRF) Model (ARW; Skamarock et al. 2008) to dynamically downscale the SINTEX-F2 forecasts. Two sets of WRF Model forecasts were generated, one with the Kain–Fritsch (KF; Kain 2004) and the other with the Betts–Miller–Janjić (BMJ; Betts and Miller 1986; Janjić 1994) cumulus parameterization scheme, as the precipitation forecasts over South Africa are sensitive to the cumulus parameterization scheme used in the WRF Model (Ratna et al. 2014). The statistical downscaling technique we adopted in this study was to first linearly interpolate the SINTEX-F2 forecast precipitation to the validation data grid and then correct the biases in the interpolated precipitation using the linear scaling method with a 31-day moving window. The moving window approach is often used to reduce the inhomogeneity to changing months and seasons (Theoßl et al. 2011; Wilcke et al. 2013). We used the same technique to correct the biases in the WRF forecast precipitation.

2. Model and methodology

The ensemble-mean precipitation forecast by a 12-member ensemble of the SINTEX-F2 seasonal prediction system initialized on 1 August and was used for deriving the onset of rains over South Africa for the period 1998–2015. The SINTEX-F2 seasonal prediction system is a coupled global model with ECHAM5 (Roeckner et al. 2003) at T106 resolution (horizontal resolution of about 125 km at the equator) as the atmospheric component and the Nucleus for European
Modelling of the Ocean (NEMO; Madec 2008) model at $0.5^\circ \times 0.5^\circ$ horizontal resolution as the oceanic component. The 12 ensemble members of the seasonal prediction system were generated in three steps, using 1) two kinds of observational SST datasets [the weekly OISSTv2 data with 1.0° latitude $\times$ 1.0° longitude global grid (Reynolds et al. 2002) and the high-resolution daily NOAA OISST analysis with 0.25° latitude $\times$ 0.25° longitude global grid (Reynolds et al. 2007)]; 2) three large negative feedback values ($-2400$, $-1200$, and $-800$ W m$^{-2}$ K$^{-1}$) for the SST-nudging scheme in the initialization phase; and 3) two different modeling ways for the ocean vertical mixing induced by small-vertical-scale structures within and above the equatorial thermocline (Sasaki et al. 2012). Therefore, the 12-member ensemble prediction system considers to some extent uncertainties in both initial conditions and model physics. 3DVAR correction is conducted every first day of each month using subsurface ocean temperature and salinity observation. An ensemble of 12 members of SINTEX-F2 forecasts were used to drive the WRF Model to generate an ensemble of 12-member WRF downscaled forecasts for the years 1998–2015 (18 years) covering a 6-month period from August to January.

The WRF Model, version 3.8.1, was used for dynamically downsampling the SINTEX-F2 seasonal forecasts. The WRF Model was configured with two-way interacting domains at horizontal resolutions of 27 and 9 km with 30 vertical levels extending from surface to 50 hPa. The first domain covered most parts of southern Africa (8.4°–44.6°S, 1°–60°E) and the second domain covered South Africa and the neighboring countries (19.4°–36.5°S, 10°–38°E) (Fig. 1a). Using two-way nested domains is useful for improving the regional climate forecasts (Giorgi and Mearns 1999). We used two cumulus schemes, namely, Kain–Fritsch and Betts–Miller–Janjic, to downscale the SINTEX-F2 forecasts as the simulation of precipitation over southern Africa is sensitive to cumulus parameterization schemes in the WRF Model (Ratna et al. 2014). The other physical parameterization schemes used in the WRF Model are 1) the Rapid Radiative Transfer Model (RRTM) for the longwave radiation (Mlawer et al. 1997), 2) a simple cloud-interactive shortwave radiation scheme (Dudhia 1989), 3) the Yonsei University (YSU) planetary boundary layer scheme (Hong et al. 2006), and 4) the WRF single-moment 3-class (WSM3) microphysics scheme (Hong et al. 2004). A time step of 90 s was used for the model runs. The radiation schemes were called every 1800 s, and the Kain–Fritsch cumulus was called every 300 s (Ratnam et al. 2017); all the other physical parameterization schemes were called every time step. The WRF Model was initialized using 1 August initial conditions; however, the dates of onset are calculated with respect to 1 September of each year.

The Climate Prediction Center (CPC)–developed precipitation dataset, Africa Rainfall Climatology, version 2 (ARC2; Novella and Thiaw 2013), is used as the verification dataset. The ARC2 data use inputs from the 3-hourly geostationary infrared (IR) data centered over Africa and the quality-controlled Global Teleconnection System (GTS) gauge observation reporting 2-h rainfall accumulations over Africa for deriving the precipitation over Africa. The ARC2 data are at horizontal grid resolution of 0.1° $\times$ 0.1° and is available at daily temporal resolution. The ensemble average of the SINTEX-F2 forecast precipitation and the ensemble mean precipitation of the second domain of WRF forecasts, at 9-km horizontal resolution, are linearly interpolated to the ARC2 grid (at $\sim$10-km resolution) for calculating the dates of onset and for other analysis. The linearly interpolated SINTEX-F2 and WRF forecasts are hereafter referred to as SINTEXF2, WRFkf, and WRFbmkj forecasts (Table 1).

Because of the uncertainties in the observed precipitation over southern Africa (Nikulin et al. 2012; Sylla et al. 2013), we have used the daily precipitation estimates from Tropical Rainfall Measuring Mission (TRMM; Huffman et al. 2007) (3B42V7 product) in addition to the ARC2 precipitation estimates over southern Africa to derive the dates of onset. The TRMM data are at a horizontal resolution of 0.25° $\times$ 0.25°. The TRMM precipitation is estimated from multiple satellites as well as rain gauges (Huffman et al. 2007). The TRMM data were linearly interpolated to the ARC2 grid for comparing the results at same grid resolution. The TRMM data are available from 1998 to 2015. The analysis of the onset of rains over South Africa is carried out over the common period of both the observed precipitation datasets.

We used the linear scaling (LS; Lenderink et al. 2007) technique to correct the biases in the interpolated SINTEX-F2 and WRF precipitation forecasts. The temporal jackknife cross-validation technique, in which the data are repeatedly divided into calibration (17 years) and verification (1 year) datasets to reduce overfitting of data, is used for applying the statistical technique for correcting the biases in the precipitation forecasts. However, we found that the conclusions do not change even if the data are divided into two halves and the first half of the data (9 years) is used for training and the second half of the data (9 years) for verification, and vice versa. A moving window approach is often employed for applying the bias-correction methods to avoid inhomogeneity in the calibration data to changing months and seasons. Themeßl et al. (2011) employed a moving window of 61 days centered on the
FIG. 1. (a) Topography of the WRF domains. (b) Map showing the provinces of South Africa. The rectangular box represents the Highveld region of the west Mpumalanga province. (c) Mean September–December ARC2-estimated precipitation. (d)–(i) As in (c), but forecast by SINTEX-F2, WRFkf, WRFbmj, SINTEXF2corr, WRFkfcorr, and WRFbmjcorr, respectively. (j)–(l) Significant bias in the precipitation forecast by SINTEX-F2, WRFkf, and WRFbmj, respectively. Significance is at 90% using the Student’s two-tailed 𝑡 test.
day of interest; Wilcke et al. (2013) used a moving window of 31 days for bias correcting the regional climate model simulated precipitation. In our study, we used a moving window of 31 days for bias correcting the forecast precipitation using the LS method.

In the LS technique, we correct the daily precipitation in the verification period by multiplying it with the ratio of mean of observed daily precipitation and mean of daily model precipitation in the calibration period. We applied this technique to the daily precipitation forecast by the SINTEX-F2 and the WRF Model using a 31-day moving window centered on the day of interest covering the period from 1 September to 30 December. The LS technique can be represented as

$$P_{\text{verf}}^* = P_{\text{verf}} \times \left[ \frac{\mu(P_{\text{obs}})}{\mu(P_{\text{cali}})} \right],$$  

(1)

where $P_{\text{verf}}$ and $P_{\text{verf}}^*$ are the uncorrected and corrected daily precipitation in the verification period and $\mu(P_{\text{obs}})/\mu(P_{\text{cali}})$ is the ratio of the mean over all the years in the calibration period of the 31-day-averaged daily observed precipitation and the 31-day-averaged daily model forecast precipitation. The bias-corrected SINTEX-F2 and the WRF precipitation forecasts are hereafter referred to as SINTEXF2corr, WRFFkcorr, and WRFbmjcorr (Table 1). SINTEXF2corr is therefore the statistically downscaled SINTEX-F2 precipitation, and WRFFkcorr and WRFbmjcorr are the bias-corrected WRF forecast precipitation.

We identified the date of onset of rains over South Africa using the widely used definition based on the water required for the successful germination of the maize plants (Reason et al. 2005; Tadross et al. 2005; Moellets et al. 2011). In our study, the onset of rainy season over South Africa is defined as the last day starting from 1 September of each year in which 25 mm or more precipitation has been accumulated over the previous 10 days and at least 20 mm accumulated in the subsequent 20 days. We applied these criteria to extract the onset dates during 1 September–30 December of every year, based on both the ARC2- and TRMM-estimated precipitation as well as the SINTEX-F2 and WRF precipitation forecasts. The chosen period is a common period after considering the 31-day moving window for the LS bias correction technique. The biases and the root-mean-square error (RMSE) in the forecast dates of onset precipitation using both the ARC2 and TRMM estimates are calculated. The anomaly correlation of the onset dates is calculated using the bootstrapping method with a sample size of 500. The median and the 95% confidence values of the correlation coefficients are estimated for both the SINTEX-F2 and WRF forecasts.

3. Results

a. Mean precipitation and biases in the precipitation forecasts

The provinces of South Africa are shown in Fig. 1b for the ease of explanation of the results. The spatial distribution of the mean ARC2 precipitation, averaged over 1998–2015 from September to December, has magnitudes of 1.5–3.5 mm day$^{-1}$ over South Africa to the east of 25°E, with a peak of about 3–3.5 mm day$^{-1}$ along the coastal regions of KwaZulu-Natal province (Fig. 1c). To the west of 25°E, the precipitation is less than 1.5 mm day$^{-1}$. The small magnitude of the precipitation indicates that during the early summer months, from September to December, most parts of South Africa receive very little precipitation. The mean of the TRMM-estimated precipitation (Fig. 2a) has similar spatial distribution to the ARC2 precipitation. However, the peak in precipitation in the TRMM precipitation dataset is over Mpumalanga province and northern parts of Eastern Cape province with a magnitude of about 3.5–4.0 mm day$^{-1}$ (Fig. 2a). The magnitudes of precipitation in the TRMM dataset are about 0.5 mm day$^{-1}$ higher than the magnitudes of the ARC2 precipitation data over the eastern parts of South Africa. The SINTEX-F2 forecast precipitation has a realistic spatial distribution over South Africa, with high precipitation over the eastern parts of South Africa and less precipitation over the western parts of South Africa (Figs. 1d and 2b). However, the forecast SINTEX-F2 precipitation has a significant positive mean bias, exceeding 2 mm day$^{-1}$, over KwaZulu-Natal and

---

**Table 1. List of model experiments.**

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SINTEXF2</td>
<td>SINTEX-F2 model precipitation linearly interpolated to ARC2 precipitation grid</td>
</tr>
<tr>
<td>SINTEXF2corr</td>
<td>SINTEX-F2 model precipitation linearly interpolated to ARC2 precipitation grid and corrected for biases</td>
</tr>
<tr>
<td>WRFkf</td>
<td>WRF Model (with KF scheme) precipitation linearly interpolated to ARC2 precipitation grid</td>
</tr>
<tr>
<td>WRFkfcorr</td>
<td>WRF Model (with KF scheme) precipitation linearly interpolated to ARC2 precipitation grid and corrected for biases</td>
</tr>
<tr>
<td>WRFbmj</td>
<td>WRF Model (with BMJ scheme) precipitation linearly interpolated to ARC2 precipitation grid</td>
</tr>
<tr>
<td>WRFbmjcorr</td>
<td>WRF Model (with BMJ scheme) precipitation linearly interpolated to ARC2 precipitation grid and corrected for biases</td>
</tr>
</tbody>
</table>
northern parts of Eastern Cape province with respect to both ARC2 (Fig. 1) and TRMM (Fig. 2h) precipitation. The dynamical downscaling of the SINTEX-F2 forecasts using WRFkf enhanced the magnitude of the forecast precipitation over the southeastern parts of South Africa (Figs. 1e and 2c) compared to the SINTEX-F2 forecast precipitation (Figs. 1d and 2b). However, the WRFbmj cumulus scheme reduced the precipitation over the southeastern parts of South Africa (Figs. 1f and 2d) with respect to the SINTEX-F2 forecast precipitation, indicating the Kain–Fritsch cumulus scheme generated higher precipitation over South Africa compared to the Betts–Miller–Janjic cumulus scheme. The significant biases in the WRFkf forecast precipitation are higher than 2 mm day$^{-1}$ over most parts of South Africa to the east of 25°E (Figs. 1k and 2i). The biases in the WRFbmj forecast precipitation are on the order of −0.5–1.5 mm day$^{-1}$ over parts of South Africa to the east of 25°E (Figs. 1l and 2j).

Interestingly, statistical downscaling of SINTEX-F2 forecast precipitation reduced the precipitation over the southeastern parts of South Africa in the SINTEX-F2 forecast (Figs. 1g and 2e) and the spatial pattern of the SINTEXF2corr forecast looks similar to the ARC2- and TRMM-estimated precipitation. Similarly, the LS technique reduces the precipitation over southeastern parts of South Africa in the WRFkf (Figs. 1h and 2f) and WRFbmj forecasts (Figs. 1i and 2g). The biases in the LS-corrected SINTEXF2corr and WRF forecasts (WRFkfccorr and WRFbmjcorr) are small and are not significant (figure not shown). The above analysis indicates that statistically downscaling the SINTEX-F2 forecast precipitation by applying the LS technique with a 31-day moving window has the same effect on the mean forecast precipitation as dynamically downscaling the SINTEX-F2 forecasts and then bias correcting them.

b. Mean onset dates and biases in the onset dates forecast

Spatial distribution of mean dates of onset of summer rains over South Africa has a gradient across the country, with early onset over the southern and eastern parts...
of the country and late onset over central and northern parts of South Africa (Figs. 3a and 4a). The southern and eastern parts of South Africa experience onset of rains from the end of September to the middle of October (Figs. 3a and 4a), and the central and northern parts of South Africa have onset of rains from the end of October to early November (Figs. 3a and 4a). The mean dates of onset derived from both ARC2 (Fig. 3a) and TRMM (Fig. 4a) precipitation estimates broadly agree with each other, though with slight changes in the spatial distribution of the dates of onset. The periods of onset of rains identified using both the ARC2 and the TRMM precipitation data agree well with the onset periods reported earlier (Reason et al. 2005; Tadross et al. 2005). The dates of onset calculated over the South Africa using the SINTEX-F2 forecast precipitation has early onset over the KwaZulu-Natal, Eastern Cape, Free State, and Mpumalanga provinces (Figs. 3b and 4b) compared to both the ARC2 and TRMM precipitation-derived onsets (Figs. 3a and 4a). The predicted onset is early by more than 20 days over KwaZulu-Natal, Eastern Cape, and Free State (Figs. 3h and 4h). The bias in the forecast of onset date is between 8 and 12 days over the Highveld region of west Mpumalanga (Figs. 3h and 4h) province. The Highveld region west of Mpumalanga was considered in the study as a rectangular box in Fig. 1b. The bias in the dates of onset over these regions is as expected, as the dates of onset are derived from the precipitation forecasts and the SINTEX-F2 has large positive bias over the KwaZulu-Natal, Eastern Cape, Free State, and Mpumalanga provinces, leading to a forecast of early onset of rains over the regions. The Limpopo province has late onset of rains in the SINTEX-F2 forecast by about 8–16 days (Figs. 3h and 4h) compared to the onset dates derived from both the ARC2 and the TRMM precipitation. The RMSE values of the onset dates derived from the SINTEX-F2 forecast precipitation are higher than 30 days over KwaZulu-Natal, Free State, Eastern Cape, Northern Cape, and North West provinces (Figs. 3n and 4n). The RMSE values are about 22–30 days over Limpopo (Figs. 3n and 4n). However, the RMSE values over the west Mpumalanga and Gauteng provinces are less compared to other provinces and the values are on the order of 14–18 days (Figs. 3n and 4n).

The onset derived from the WRFkf forecast precipitation (Figs. 3e and 4e) is early over most parts of South Africa to the east of 25°E (Figs. 3e and 4e) because of large positive biases in the precipitation forecast by the WRFkf (Figs. 1e and 2e) model. However, the dates of onset are improved (Figs. 3i and 4i) over the Limpopo province compared to the onset dates from the SINTEX-F2 forecasts (Figs. 3b and 4b) because of smaller biases in the precipitation forecast over Limpopo in the WRFkf forecast compared to the SINTEXF2 forecast. The RMSE values are higher over most parts of South Africa except over the Limpopo province in the onset dates derived from the WRFkf precipitation (Figs. 3o and 4o) compared to the RMSE values in the onset dates from SINTEX-F2 precipitation (Figs. 3n and 4n). The RMSE in the dates of onset over the Limpopo province in the WRFkf forecast varies from 16 to 24 days. The dates of onset are improved in the WRFbmj forecasts (Figs. 3d and 4d) over most parts of South Africa compared to both SINTEXF2 (Figs. 3b and 4b) and WRFkf forecasts (Figs. 3c and 4c) because of smaller biases in the WRFbmj forecast precipitation. However, the biases in the dates of onset over Limpopo are significantly positive in the WRFbmj forecast (Figs. 3j and 4j) because of significant negative biases in the precipitation over the region in the WRFbmj forecast (Figs. 11 and 2), and they are higher in magnitude compared to those forecast by WRFkf.

The statistical downscaling of the SINTEX-F2 and bias correcting the WRFkf and WRFbmj precipitation using 31-day moving-window LS technique improved the onset date forecasts over South Africa. The spatial pattern of the dates of onset in the SINTEXF2corr (Figs. 3e and 4e), WRFkfcorr (Figs. 3f and 4f), and WRFbmjcorr (Figs. 3g and 4g) is similar to the onset dates estimated by ARC2 (Fig. 3a) and TRMM (Fig. 4a) datasets. The biases in the onset dates are reduced in the SINTEXF2corr, WRFkfcorr, and WRFbmjcorr forecasts and are significant only over scattered regions of South Africa with respect to onset dates derived from AR2C and TRMM estimates (Figs. 3k–m and 4k–m). The RMSE in the dates of onset is also reduced because of bias correction of the forecast precipitation. The RMSE in the SINTEXF2corr (Figs. 3q and 4q), WRFkfcorr (Figs. 3r and 4r), and WRFbmjcorr (Figs. 3s and 4s) is reduced compared to the SINTEXF2 (Figs. 3n and 4n), WRFkf (Figs. 3o and 4o), and WRFbmj (Figs. 3p and 4p) forecasts, respectively. Large improvement in the forecast of date of onset is evident over the Mpumalanga province of South Africa in the WRFkfcorr (Figs. 3r and 4r) forecast compared to WRFkf (Figs. 3o and 4o) forecast.

The spatial pattern of the anomaly correlation coefficients of the onset dates derived from SINTEX-F2 and WRF precipitation forecasts with 95% significance values (2.5% lower confidence bound and 97.5% upper confidence bound) are shown in Fig. 5. The median of the bootstrapped estimates of the correlation coefficients between SINTEX-F2 precipitation-derived onset anomalies and the ARC2-estimated onset anomalies shows positive values over parts of the Eastern Cape, Free State, Mpumalanga, Limpopo, KwaZulu-Natal, North West, and Gauteng provinces (Fig. 5b). The upper confidence bound (97.5%) estimate (Fig. 5c)
FIG. 3. (a) Onset dates over South Africa derived from ARC2 precipitation estimates. (b)–(g) As in (a), but derived from SINTEX-F2, WRFkf, WRFbmj, SINTEXF2corr, WRFkfcorr, and WRFbmjcorr precipitation forecasts, respectively. (h)–(m) Significant biases in the onset dates forecast by SINTEX-F2, WRFkf, WRFbmj, SINTEXF2corr, WRFkfcorr, and WRFbmjcorr, respectively, with respect to onset dates derived from ARC2 precipitation estimates. (n)–(s) As in (h)–(m), but RMSE in the forecast onset dates. Significance is at 90% using the Student's two-tailed $t$ test.
FIG. 4. (a)–(s) As in Fig. 3, but with respect to TRMM estimates.
has positive values over most parts of South Africa; however, the lower confidence bound (2.5%) estimate (Fig. 5a) has positive values over scattered regions of South Africa, indicating the anomaly correlation coefficients to be not significant. Similar results are obtained by calculating the anomaly correlation coefficients between WRFkf (Figs. 5d–f), WRFbmj (Figs. 5g–i), SINTEXF2corr (Figs. 5j–l), WRFkfcorr (Figs. 5m–o), and WRFbmjcorr (Figs. 5p–r). The correlation coefficients with respect to the onset dates derived from TRMM estimates are also not significant (not shown). The results indicate that the models have limited skill in forecasting the anomalies of the dates of onset over South Africa because of large biases and RMSE values in the forecast onset dates. The anomalies are not improved by correcting for the mean biases in the precipitation forecast by the models, indicating the global as well as the regional models are to be improved to boost the skill in forecasting the anomalies of the dates of onset over South Africa.

c. Interannual variability of onset dates over Limpopo and the Highveld region of Mpumalanga

The mean onset dates are area averaged over the Limpopo province and the Highveld region of Mpumalanga (27°–25.5°S, 29°–30.5°E; rectangular box in Fig. 1b) for the years 1998–2015 (Fig. 6) to get a quantitative estimate of the interannual variability in the forecasted onset dates. We chose these regions for the analysis as these have less bias and RMSE values in the onset dates compared to other regions of South Africa. Over Limpopo province, the onset dates have large interannual variation. The onset of rains over Limpopo was early in 2007 and 2012 and delayed in 2004 (Figs. 6a,b). Comparing Fig. 6a and Fig. 6b, it is evident that there are differences in the dates of onset derived from the two observation precipitation datasets ARC2 and TRMM, though both of them agree on early and delayed onset years over the Limpopo province. All the model forecasts failed to capture the extremes in the onset of rains over the Limpopo region (Figs. 6a,b).

However, as seen in Table 2, the onset dates derived from the WRFkf forecast precipitation has smaller errors over the region compared to onset derived from other models. The onset date has a mean bias of 0.37 days, mean absolute error (MAE) of 8.13 days, and RMSE of 11.23 days in the onset dates derived from WRFkf precipitation. However, the biases are enhanced in the bias-corrected WRFkfcorr forecasts WRFkfcorr. The bias correction of WRFkf precipitation delayed the onset of rains over the Limpopo (Figs. 6a,b) compared to WRFkf. The increases in biases in the WRFkfcorr is partly due to the cross-validation technique applied for bias correction in which the weights from the training period are applied to the precipitation in the validation period. The onset date from SINTEXF2 forecasts precipitation has a bias of 5.95 days, MAE of 9.84 days, and RMSE of 13.01 days (Table 2). Similar to WRFkfcorr, SINTEXF2corr biases are enhanced compared to SINTEXF2 forecast onset dates. The mean bias in the SINTEXF2corr forecast onset dates is 9.13 days, with MAE and RMSE of 11.22 and 14.72 days.
respectively (Table 2). The dates of onset forecast by WRFbmj have a mean bias of 8.72 days, MAE of 11.98 days, and RMSE of 14.74 days (Table 2). These biases are slightly reduced in the bias-corrected forecast WRFbmjcorr (Table 2). The dates of onset are the same for both WRFkfcorr and WRFbmjcorr (Fig. 6a) over the Limpopo province. The biases with respect to onset dates derived from TRMM estimates also demonstrates the WRFkf to perform better than other models in forecasting onset dates over the Limpopo region (Table 2).

The onset dates over the Highveld region (Figs. 6c,d) have less interannual variation compared to the onset dates over the Limpopo region (Figs. 6a,b). The performance of the models is different over the two regions. The onset dates from the two observations datasets show differences in the dates of onset though they agree on the early and late onset dates over the region (Figs. 6c,d). The dates of onset in the WRFkf forecast are early compared to the onset dates derived from both ARC2 (Fig. 6c) and TRMM (Fig. 6d) precipitation in all the years because of large positive biases in the precipitation in WRFkf. The WRFkf forecast onset dates over the Highveld region have a mean bias of −17.18 days, MAE of 18.68 days, and RMSE of 22.79 days (Table 2) with respect to onset dates from the ARC2 data. The errors are also large in the onset dates from SINTEXF2 and WRFbmj though smaller compared to WRFkf (Table 2). The bias in the onset dates over the Highveld region in the SINTEX-F2 forecasts is −10.90 days, with an MAE of 13.76 days and an RMSE of 16.09 days with respect to ARC2-derived onset dates. The bias, MAE, and RMSE in the onset dates forecast by WRFbmj are −10, 14.85, and 18.12 days, respectively (Table 2), with respect to ARC2-derived onset dates. The onset dates are improved by correcting for mean biases in the SINTEXF2, WRFkf, and WRFbmj precipitation. The errors in the SINTEXF2corr, WRFkfcorr, and WRFbmjcorr are smaller compared to SINTEXF2, WRFkf, and WRFbmj (Table 2). The onset dates from SINTEXF2corr have the smallest MAE of 9.9 days and RMSE of 12.62 days with respect to ARC2 data compared to other models (Table 2). The errors compared to onset dates derived from TRMM estimates (Table 2) are similar to the errors compared to ARC2 (Table 2), with WRFkf having the largest errors and SINTEXF2corr having the smallest errors among the models.
The above results indicate that the performance of the models in forecasting the dates of onset over Limpopo and over the Highveld region of the Mpumalanga province is model dependent. Over Limpopo, the WRFkf performs better compared to the other models, whereas over the Highveld region, the global model SINTEX-F2 performs better. Because of the high horizontal resolution of 9 km, the regional model has enhanced precipitation over the mountainous region of the Highveld, leading to large positive biases in precipitation and early onset of rains over the region. The bias-correction technique applied to the WRFkf and WRFbmj forecast precipitation corrects the large bias in precipitation, and the onset dates are much improved.

### 4. Conclusions

In this study, we attempted to predict the dates of onset of summer rains over South Africa using seasonal precipitation forecasts from the SINTEX-F2 model. The SINTEX-F2 forecasts were dynamically and statistically downscaled to ~10-km horizontal resolution to improve the precipitation forecasts. We compared the dates of onset derived from 1) linearly interpolated SINTEX-F2 precipitation (SINTEXF2), 2) bias-corrected SINTEX-F2 precipitation (SINTEXF2corr), 3) WRF Model forecast precipitation (WRFkf and WRFbmj), and 4) bias-corrected WRF Model precipitation (WRFkfcorr and WRFbmjcorr) (Table 1).

Comparison of the onset dates forecast by the models with respect to the onset dates derived from the validating datasets ARC2 and TRMM demonstrates the forecasts have a realistic spatial distribution of onset over South Africa, with early onset over KwaZulu-Natal province and late onset over the interior regions of South Africa. However, the SINTEXF2 and WRFkf forecasts have early onset over the eastern provinces of South Africa because of positive biases in the forecast precipitation over the region. The biases in the onset from the WRFbmj precipitation forecasts are less compared to onset dates derived from both SINTEX-F2 and WRFkf precipitation forecasts.

Correcting the mean biases in the forecast precipitation of SINTEXF2, WRFkf, and WRFbmj resulted in reducing the errors in the forecast of dates of onset over South Africa. The average dates of onset derived

### Table 2. Bias, MAE, and RMSE in forecast of onset dates over Limpopo with respect to ARC2 data and TRMM data and also over the Highveld region of Mpumalanga with respect to ARC2 data and TRMM data.

<table>
<thead>
<tr>
<th></th>
<th>Bias (days)</th>
<th>MAE (days)</th>
<th>RMSE (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Limpopo (with respect to ARC2)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SINTEXF2</td>
<td>5.95</td>
<td>9.84</td>
<td>13.01</td>
</tr>
<tr>
<td>WRFkf</td>
<td>0.37</td>
<td>8.13</td>
<td>11.23</td>
</tr>
<tr>
<td>WRFbmj</td>
<td>8.72</td>
<td>11.98</td>
<td>14.74</td>
</tr>
<tr>
<td>SINTEXF2corr</td>
<td>9.13</td>
<td>11.22</td>
<td>14.72</td>
</tr>
<tr>
<td>WRFkfcorr</td>
<td>8.18</td>
<td>10.76</td>
<td>13.44</td>
</tr>
<tr>
<td>WRFbmjcorr</td>
<td>8.18</td>
<td>10.76</td>
<td>13.44</td>
</tr>
<tr>
<td><strong>Limpopo (with respect to TRMM)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SINTEXF2</td>
<td>7.04</td>
<td>11.64</td>
<td>15.19</td>
</tr>
<tr>
<td>WRFkf</td>
<td>1.46</td>
<td>9.77</td>
<td>13.18</td>
</tr>
<tr>
<td>WRFbmj</td>
<td>9.80</td>
<td>13.23</td>
<td>16.89</td>
</tr>
<tr>
<td>SINTEXF2corr</td>
<td>7.73</td>
<td>11.69</td>
<td>15.67</td>
</tr>
<tr>
<td>WRFkfcorr</td>
<td>8.01</td>
<td>11.75</td>
<td>15.27</td>
</tr>
<tr>
<td>WRFbmjcorr</td>
<td>6.78</td>
<td>10.66</td>
<td>15.55</td>
</tr>
<tr>
<td><strong>Highveld (with respect to ARC2)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SINTEXF2</td>
<td>−10.90</td>
<td>13.76</td>
<td>16.09</td>
</tr>
<tr>
<td>WRFkf</td>
<td>−17.18</td>
<td>18.68</td>
<td>22.79</td>
</tr>
<tr>
<td>WRFbmj</td>
<td>−10.0</td>
<td>14.85</td>
<td>18.12</td>
</tr>
<tr>
<td>SINTEXF2corr</td>
<td>2.25</td>
<td>9.90</td>
<td>12.62</td>
</tr>
<tr>
<td>WRFkfcorr</td>
<td>1.69</td>
<td>12.69</td>
<td>14.84</td>
</tr>
<tr>
<td>WRFbmjcorr</td>
<td>1.69</td>
<td>12.69</td>
<td>14.84</td>
</tr>
<tr>
<td><strong>Highveld (with respect to TRMM)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SINTEXF2</td>
<td>−7.27</td>
<td>11.39</td>
<td>13.33</td>
</tr>
<tr>
<td>WRFkf</td>
<td>−13.56</td>
<td>16.04</td>
<td>18.81</td>
</tr>
<tr>
<td>WRFbmj</td>
<td>−6.37</td>
<td>12.33</td>
<td>14.79</td>
</tr>
<tr>
<td>SINTEXF2corr</td>
<td>0.84</td>
<td>8.83</td>
<td>11.10</td>
</tr>
<tr>
<td>WRFkfcorr</td>
<td>1.02</td>
<td>10.53</td>
<td>12.72</td>
</tr>
<tr>
<td>WRFbmjcorr</td>
<td>1.38</td>
<td>9.66</td>
<td>12.24</td>
</tr>
</tbody>
</table>
from SINTEXF2corr, WRFkfcorr, and WRFbmjcorr precipitation over the study period have a spatial pattern with magnitudes similar to the onset dates derived from the ARC2 and TRMM precipitation datasets. However, the RMSE values of the onset dates from all six precipitation forecasts exceed 30 days over most parts of South Africa, indicating the forecasts to be less useful over those regions. Interestingly, the onset dates over the Limpopo province and the Highveld region of Mpumalanga have smaller errors compared to other regions of South Africa in all the model forecasts. The WRFkf model has better performance in forecasting the dates of onset over Limpopo, with a mean bias of 0.37 days, MAE of 8.13 days, and RMSE of 11.23 days with respect to ARC2-derived onset dates compared to the onset dates forecast by other models. Over the Highveld region, the bias-corrected SINTEXF2 forecast SINTEXF2corr has a better performance in forecasting the dates of onset compared to other model forecasts. The smaller errors in WRFkf (SINTEXF2corr) over Limpopo (Highveld region) are due to smaller errors in the precipitation forecast over the regions by the respective models.

The computation of the spatial distribution of the anomaly correlation coefficients between anomalies of the forecast onset dates and the verifying datasets reveals no significant correlation between the two, thereby indicating the skill of the models to be less in forecasting the dates of onset over the region. The bias correction of the forecast precipitation also does not improve the correlation coefficients as the bias correction techniques do not correct the timing errors produced by the models (Boulard et al. 2017). These results indicate that the global model SINTEX-F2 and the regional model WRF are to be improved to better represent the atmospheric processes responsible for rains over South Africa, which would lead to an improvement in the spatial distribution of precipitation over South Africa and hence the forecast of onset dates of rains over South Africa during the summer season. These conclusions are consistent with that of Phakula et al. (2018).

Our study has brought out some important aspects of forecasting the onset of the summer rains over South Africa. First, the dates of onset over most parts of South Africa are less predictable and have large RMSE values exceeding 30 days. Second, the forecast onset dates over the Limpopo province, where the forecast skill is generally higher because of the influence of the tropical processes (Landman et al. 2012), and the Highveld region of Mpumalanga have fewer errors and are more predictable compared to other regions of South Africa and can be predicted with an error of about 10 days. Third, the forecast of onset dates over South Africa is model dependent and is also sensitive to the cumulus parameterization schemes used in a regional model. The global and the regional models are to be improved to better represent the timing of the onset of rains over South Africa, which would improve the anomaly correlation coefficients and hence the skill of the forecast of dates of onset over South Africa.

Acknowledgments. Authors thank the three reviewers for insightful comments that substantially improved the manuscript. This research is supported by The Japan Agency for Medical Research and Development (AMED; Grant JP18jm0110007) and Japan International Cooperation Agency (JICA) through Science and Technology Research Partnership for Sustainable Development (SATREPS) project for iDEWS South Africa.

REFERENCES


