NOTES AND CORRESPONDENCE

Evaluating Detection Skills of Satellite Rainfall Estimates over Desert Locust Recession Regions

TUFÁ DINKU AND PIETRO CECCATO
International Research Institute for Climate and Society, The Earth Institute at Columbia University, Palisades, New York

KEITH CRESSMAN
Desert Locust Information Service, United Nations Food and Agriculture Organization, Rome, Italy

STEPHEN J. CONNOR
International Research Institute for Climate and Society, The Earth Institute at Columbia University, Palisades, New York

(Manuscript received 11 May 2009, in final form 6 February 2010)

ABSTRACT

This paper evaluates rainfall detection capabilities of seven satellite rainfall estimates over the desert locust recession regions of the world. The region of interest covers the arid and semiarid region from northwestern Africa to northwestern India. The evaluated satellite rainfall products are the African rainfall climatology (ARC), rainfall estimation algorithm (RFE), Tropical Rainfall Measuring Mission 3B42 and its real-time version (3B42RT), NOAA/Climate Prediction Center morphing technique (CMORPH), and two versions of the Global Satellite Mapping of Precipitation moving vector with Kalman filter (GSMaP-MVK and GSMaP-MVK+). The reference data were obtained from the Desert Locust Information Service of the United Nations Food and Agriculture Organization (FAO). The FAO data are qualitative information collated by desert locust survey teams from different countries during field campaigns. Such data can only be used to assess the rainfall detection capabilities of the satellite products. The validation region is divided into four subregions and validations statistics are computed for each region. The probability of detection varies from 70% for the relatively wet part of the validation region to less than 20% for the driest part. The main weakness of the satellite products is overestimation of rainfall occurrences. The false-alarm ratio was as high as 84% for the driest part and as high as 57% for the wetter region. The satellite products still exhibit positive detection skill for all of the subregions. A comparison of the different products shows that no single product stands out as having the best or the worst overall performance.

1. Introduction

The desert locust affects a vast area of about 28 million km² that extends from the Atlantic coast of Africa to eastern India and from northern Turkey to Tanzania in the south (FAO 2007). The breeding areas, also known as desert locust recession regions (FAO 2007), are limited to an area of about 16 million km², which covers arid and semiarid regions. From time to time, locusts form swarms that fly or are carried by wind to greater distances and wipe out crops and pasture located hundreds of kilometers away from the recession region. Early detection of locusts over the recession area is very critical to contain them from spreading to the surrounding regions. This requires constant monitoring of desert locust activities, weather, and vegetation conditions (FAO 2007; van Huis et al. 2007). The Desert Locust Information Service (DLIS; http://www.fao.org/ag/locusts/en/activ/DLIS/index.html) of the United Nations Food and Agriculture Organization (FAO) provides a decision support system required for monitoring desert locusts (Ceccato et al. 2007). The

Corresponding author address: Tufa Dinku, IRI, The Earth Institute at Columbia University, 61 Route 9W, Monell Bldg., Palisades, NY 10964.
E-mail: tufa@iri.columbia.edu

DOI: 10.1175/2010JAMC2281.1

© 2010 American Meteorological Society
information produced by DLIS is used by about 30 countries in the affected region to plan survey-and-control operations. This information may also be used by the international donor community to target assistance, especially during emergencies.

Rainfall data are one of the inputs into the DLIS decision support system. Rainfall is very important in determining the extent and intensity of desert locust breeding areas. Rainfall is required to produce sufficient green vegetation that could sustain hopper development following hatching. Thus, rainfall information could be very useful in pinpointing potential breeding areas and in helping to avoid random searches. This makes rainfall information a critical part of the desert locust monitoring system. However, the number of rain gauge stations over the potential breeding region is very limited because the area spreads across a sparsely populated and difficult-to-access region. As a result, satellite rainfall estimates are the only source of rainfall information over these vast, remote areas. FAO has been using satellite information to monitor desert locusts for a very long time (e.g., Hielkema and Sanders 1994). DLIS currently uses rainfall maps specifically prepared for this purpose by the International Research Institute for Climate and Society at Colombia University (http://iridl.ldeo.columbia.edu/maproom/.Food_Security/.Locusts/.Regional/). These maps are based on the National Oceanic and Atmospheric Administration (NOAA)/Climate Prediction Center (CPC) morphing technique (CMORPH; Joyce et al. 2004) rainfall estimates.

There are many satellite rainfall products available over the region of interest. However, the accuracy of satellite rainfall estimates has not been assessed over most of this region. Thus, validation of satellite rainfall products over this region would be an important contribution as a feedback to the developers of the different satellite products. Seven satellite rainfall products are evaluated here. The different satellite products are described in section 2c.

The reference data used here were provided by FAO DLIS. These data consist of qualitative information collected during field surveys. The data provided correspond only to observations reported when rainfall events occurred (i.e., dry conditions are not reported). Dry cases are extracted from other information provided in the field reports. Only detection capabilities of the satellite products are evaluated because of the nature of the data. Because satellite estimates are the only data available over most of the desert locust recession regions, information on the detection skills of the satellite products would be very useful to FAO DLIS. To that effect, FAO DLIS has indicated that they are less concerned with the exact quantity of estimated rainfall and are more interested in whether the rainfall events that occur could be detected by the satellite products.

2. Validation region and data

a. Validation region

The validation region extends from the Atlantic coast of northwestern Africa to western India (Fig. 1). It covers the driest parts of the world, including the Sahara and Arabian Deserts. Rainfall is very small except over the southern parts of the region (Fig. 2). The validation region is vast, with different climate and environmental characteristics; thus, it has been divided into four subregions (Fig. 1) for better understanding of the performance of the satellite products over the different subregions. The delineation of the four regions is intuitive rather than objective; it is mainly based on the rainfall patterns over the different regions (Figs. 2, 3). There is
a very strong north–south rainfall gradient, and the subdivisions follow this to some extent. Region 1 (R1) is the driest of all regions, and region 2 (R2) is relatively wet (Fig. 3). The main rainy season for R1 is during northern winter, whereas the other regions have rainfall peaks during the summer months. The summer rainfall is mainly associated with the north–south movements of the intertropical convergence zone, and the winter rains over R1 are associated with temperate frontal-type disturbances or elongated troughs in the midlatitude westerlies (e.g., Nicholson 2000).

b. Reference data

The reference data were provided by DLIS FAO. The data consist of over 30 000 reports from 1992 to 2006. Only data from 2003 to 2006 were used for this investigation. These data are reports collected during different field campaigns by desert locust survey teams in the different countries. These are not rain gauge measurements; rather, they are qualitative information collected during the field campaigns. Rainfall is either observed by the survey team itself (rare) or is reported to them by the locals (farmers or nomads) that they encounter in the field during the survey. Sometimes the dates are precise, but often the timing could be vague. Here the data were used only when the time and location of observations were reported to be exact. The locations from where the reports were collected are given in Fig. 1. The total number of reports between 2003 and 2006 was over 22 000. However, most of the data were removed owing to uncertainties in time and location of the observations. A total of about 9000 observations were used for the study.

The field reports represent only cases in which it rained. To get information for the dry cases, it was assumed that there was no rain on days on which the survey teams were in the field but did not report any rainfall. This is possible because the survey teams report the coordinates and duration of each of their visits. However, the observations are still biased toward wet events because the survey teams go to the field mainly during the wet seasons. To overcome this problem, the data were resampled based on dry/wet climatological data obtained from synoptic stations within each validation region. The data for the synoptic stations were obtained online (http://www.ncdc.noaa.gov/cgi-bin/res40.pl?page=climvisgsod.html). According to this climatological data, the proportions of rainy days to total observations are 5%, 25%, 10%, and 5% for regions R1, R2, R3, and R4, respectively. Because the synoptic stations available over this region are very few, these figures are just rough.
estimates. The number of nonzero values for R1 and R4 would be very small at 5% resampling. Thus, 10% is used instead.

The reference data have not been interpolated because they are just binary (rain/no rain) data collected during ad hoc field campaigns. Thus, the field reports are compared with the closest satellite pixels. This means comparing a point datum with area-averaged satellite estimates, which would affect the validation results. This may also affect the comparison among the different products at different resolutions (see section 4a).

c. Satellite data

The evaluated satellite rainfall products are the African rainfall estimation algorithm (RFE; Herman et al. 1997; Xie et al. 2002), African rainfall climatology (ARC; Love et al. 2004), CMORPH (Joyce et al. 2004), the National Aeronautics and Space Administration (NASA) Tropical Rainfall Measuring Mission (TRMM) 3B42 and its real-time version 3B42RT (Huffman et al. 2007), and the Global Satellite Mapping of Precipitation moving vector with Kalman filter (GSMaP-MVK, hereinafter GSMaP) and its latest version GSMaP-MVK+ (hereinafter GSMaP+: Okamoto et al. 2007). The NOAA/CPC and NASA products are selected for validation here because they have been available for some time and are widely used. The GSMaP product is included here because it uses a relatively new algorithm and is available at high spatial (0.1°) and temporal (hourly) resolutions. The main characteristics of the different products are summarized in Table 1, and a brief description of each product is given below.

The algorithms used for the ARC and RFE products are very similar. They use a thermal infrared (TIR) brightness temperature threshold of 235°K for discriminating raining clouds from nonraining ones. This temperature threshold is used to compute cold cloud duration (CCD) from TIR images. Then a simple linear relationship is used to convert CCD into rainfall. Both ARC and the current version of RFE produce daily rainfall at 0.1° spatial resolution. ARC is designed specifically to produce consistent high-resolution daily rainfall time series for climatological applications over Africa, whereas RFE is an operational product. At this time, ARC data are available starting from 1995. CPC is currently working to extend the time series back to 1982. The main differences between ARC and RFE are that 1) RFE uses half-hourly TIR observations whereas ARC uses 3-hourly observations and that 2) RFE incorporates passive microwave (PM) rainfall estimates (Xie et al. 2002).

The CMORPH algorithm combines the better accuracy of PM rainfall estimates with better sampling frequency of TIR observations. The PM rainfall estimates are interpolated (morphed) using motion vectors derived from half-hourly TIR observations (Joyce et al. 2004). The final product is a spatially and temporally complete PM rainfall estimate that is independent of the TIR rainfall estimates. The TRMM Multisatellite Precipitation Analysis (TMPA; Huffman et al. 2007) algorithm is used to produce 3B42 and 3B42RT. This algorithm combines TIR data from geostationary satellites and PM retrievals from different sources in four steps: 1) the different PM estimates are adjusted and combined, 2) TIR precipitation estimates are created using the PM estimates for calibration, 3) PM and TIR estimates are combined, and 4) the final product is rescaled to monthly totals whereby gauge observations are used to adjust the satellite product. The 3B42 product is available 2 days after the end of each month. The real-time version (3B42RT) is a product from the third step above and thus does not use gauge adjustment. It is made available with a lag time of a few hours after the TIR and PM inputs are obtained.

The GSMaP algorithm starts with the CMORPH approach of using TIR-derived motion vectors for propagating PM estimates in time and space, but it also uses the TIR rainfall estimates at times for which the PM estimates are not present, along with the propagated PM estimates using a Kalman filter framework (Okamoto et al. 2007). Thus, in this case the TIR observations provide more than just the evolution of the PM rain rates. This is an improvement over the CMORPH algorithm. NOAA/CPC is also currently working to implement the Kalman filter approach for the next version of CMORPH. The main difference between GSMaP and GSMaP+ is that the latter uses PM estimates from PM sounders, mainly the Advanced Microwave Sounding Unit, that were not used in the former.

3. Results

The FAO data are mainly qualitative and may not reliably be used for evaluating the skill of the satellite products in estimating rainfall amounts. Thus, only error
statistics describing the rainfall detection capabilities of the satellite products are presented. Probability of detection (POD), false-alarm ratio (FAR), probability of false detection (POFD), and Heidke skill score (HSS) are used here. The expressions for these statistics are based on a contingency table (Table 2), where $A$, $B$, $C$, and $D$ represent hits, false alarms, misses, and correct negatives, respectively, and are given as

\[
POD = \frac{A}{A+C},
\]

\[
FAR = \frac{B}{A+B},
\]

\[
POFD = \frac{B}{B+D},
\]

\[
HSS = \frac{2(AD - BC)}{(A+C)(C+D) + (A+B)(B+D)}.
\]

POD assesses what fraction of the actual rainfall events were detected by the satellite products, and FAR gives the fraction of false alarms to rainfall occurrences detected by the satellite. On the other hand, POFD is used to assess what fraction of the no-rain events were identified as rain by the satellite products. The POFD statistics are included here to give some perspective to large FAR values observed for the current validation region. HSS measures the overall detection skill accounting for matches due to random chance. As of the time of writing, there is a more detailed description of these statistics online (http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html).

Some of these statics are sensitive to the climatic behavior of the validation region. This sensitivity is more pronounced for the FAR statistic and for regions where rainfall is a rare event. The FAR values could vary significantly even within one region, depending on the data sample used. Figures 4 and 5 compare POD, FAR, and POFD over R1 and R2. The different statistics are computed using 10% of the data (subscript 1) and using all available data (subscript 2). The POD and POFD do not show big differences for the sampled and the whole data, but the FAR exhibits big differences—in particular for R1. These effects need to be taken into account when comparing these statistics across the different validation regions.

The validation statistics for the different regions are presented in Figs. 6–9. Figure 6 shows that POD is very low (<30%) for all of the satellite products. The lowest value is observed for 3B42. On the other hand, the FAR values are extremely high (about 80%) for all of the products. These statistics show that satellite products

<table>
<thead>
<tr>
<th>Table 2. Contingency table comparing gauge area averages and satellite rainfall estimates. The rainfall threshold used is 0.5 mm.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauge $\geq$ threshold</td>
</tr>
<tr>
<td>Satellite $\geq$ threshold</td>
</tr>
<tr>
<td>Satellite $&lt;$ threshold</td>
</tr>
</tbody>
</table>

![Figure 4](image-url)  
*Fig. 4. Comparison of POD, FAR, and POFD over validation region 1 (R1). POD1, FAR1, and PFD1 are computed using 10% of the data, whereas POD2, FAR2, and POFD2 are computed using all of the available data.*
have serious problems detecting rainfall when it occurs and that they detect rain that does not reach the ground. On the other hand, the POFD shows that the ratio of false alarms to the total no-rain events is small (<20%). For instance, although 84% of the ARC-estimated rainfall occurrences are false alarms they represent just 19% of the total no-rain events. Thus, the high FAR values may partly be ascribed to the fact that rainfall is a rare event over this region. That could also be why the HSS statistics show positive skill despite the low/high POD/FAR values. Results for R2 are presented in Fig. 7. The POD is higher (close to 70%) while the FAR is much lower (30%–57%). The FAR is still high for some products, although R2 is relatively wetter with more frequent rainfall events. The POFD are higher (up to 30%) relative to R1, but vary from 4% for 3B42 to 30% for RFE. The TRMM products have the lowest FAR and POFD values, but POD values are also low. Again, all the products show positive skills despite the high FAR values. In particular, 3B42, CMORPH, and GSMaP have relatively higher skills. Region 3 is similar to R1 in that it has high FAR values (Fig. 8), but its POD values are higher than those for R1. The POFD is low (<23%), and HSS shows positive skills. The NOAA products, RFE and
ARC, are not available for R4. The statistics for the other products (Fig. 9) show high FAR values (over 67%), POD values of less than 50%, and POFD values that range between 10% and 14%. The HSS values again show that the skill of the satellite products in detecting rainfall occurrences is much better than random chance.

4. Discussion

a. Comparison of different products

The satellite products compared here have diverse attributes. Algorithms range from the simplest one for ARC to the more involved algorithms such as those for CMORPH and GSMap. The spatial resolution is 0.1° for RFE, ARC, and GSMap and is 0.25° for the other products. Some of the products use gauge adjustments, and others do not. However, there is no single product that stands out as consistently having the best or worst performance. In most cases, the ARC product is as good as products that use more involved algorithms and PM inputs. This could partly be because the bulk of the information comes from TIR data for all of the products. However, one would have still expected a somewhat different result for CMORPH because it does not use
the TIR rainfall estimates. All products are equally bad over region 1. 3B42RT stands out as having the worst performance for R2 while 3B42, CMORPH, and GSMaP are relatively better. The performance of all of the products is again poor over R3. The 3B42RT product has the worst performance (lower POD and HSS) while RFE and GSMaP+ have slightly higher POD and HSS values. Among the products evaluated over R4, the GSMaP products have slightly better POD and HSS values.

The effect of spatial resolution needs to be explored in comparing the different products. Table 3 compares GSMaP+ over R1 and R2 at its original resolution (0.1°) and averaged over 0.25°. The two regions are selected because R1 is the driest region and R2 is the wettest region. The spatial variability of the mean rainfall is also higher over R2 relative to R1 or the other validation regions (Fig. 2). Table 3 shows that the effect of spatial resolution is very different over the two regions. Whereas there are no visible effects for R1, there are appreciable differences among the statistics for R2. Averaging to coarser resolution increased the difference between GSMaP+ estimates and the field reports for R2. This means that the effect from the point-to-area comparison may also vary from one region to another. This also means that part of the differences between the performance of GSMaP and those of CMORPH and the TRMM products could be ascribed to differences in the spatial resolutions.

b. Factors affecting satellite rainfall estimation over the current validation region

The major weakness of satellite rainfall estimates over the desert locust recession regions is overestimation of the occurrence rainfall (high FAR values). This overestimation is severe over the drier parts. Subcloud evaporation, rainfall suppression by desert aerosols, and surface effects are among the possible factors. The atmosphere is normally very dry over most of this region. Figure 10 compares mean relative humidity (RH) at the 1000- and 500-hPa levels over the current validation region and surrounding areas. These data are obtained from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis (Kalnay et al. 1996). The atmosphere over the current validation region is drier relative to the surrounding regions, and R1 is the driest. Note also that the atmosphere over R1 is drier at the lower level (1000 hPa) than at the upper level (500 hPa). This dry atmosphere may result in high lifting condensation levels. This means that raindrops have to pass through a thick dry and hot atmosphere before reaching the ground. This leads to evaporation of raindrops in the dry atmosphere beneath the cloud base (Rosenfeld and Mintz 1988; Takemi 1999; McCollum et al. 2000). As a result, even

![Fig. 9. As in Fig. 6 but for R4.](image)

<table>
<thead>
<tr>
<th>Resolution (°)</th>
<th>POD</th>
<th>FAR</th>
<th>POFD</th>
<th>HSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.10</td>
<td>0.27</td>
<td>0.80</td>
<td>0.12</td>
</tr>
<tr>
<td>R2</td>
<td>0.10</td>
<td>0.71</td>
<td>0.48</td>
<td>0.22</td>
</tr>
<tr>
<td>Change (%)</td>
<td>-4</td>
<td>+10</td>
<td>+18</td>
<td>-14</td>
</tr>
</tbody>
</table>
though the satellite sensors detect rainfall aloft, it may evaporate before reaching the surface. Rosenfeld and Mintz (1988) investigated this effect over a semiarid region of central South Africa. They report that, for a rain rate of 1 mm h\(^{-1}\) at the cloud base, 50% of the rain evaporated 1 km below the cloud base and all rain evaporated at 1.6 km below the cloud base. The current study region is drier than central South Africa, implying even more severe subcloud evaporation. This may partly explain the extremely high FAR values over R1.

Suppression of precipitation by desert dust could be another factor. Dust and air pollution can suppress rainfall by inhibiting droplet coalescence and precipitation formation (e.g., Rosenfeld 2000, 2006; Rosenfeld et al. 2001; Mahowald and Kiehl 2003). Desert dust acts as cloud condensation nuclei (CCN). The availability of a large number of CCN may result in a larger number of small water droplets. These small droplets may not be able to achieve velocities high enough to overcome updrafts. Thus, although these clouds may not produce rain at the surface, the satellite sensors may detect rain, leading to the observed large FAR values. McCollum et al. (2000) made a comparison of rainfall efficiency over the Amazon region and equatorial Africa. They found that the efficiency over the Amazon is 2 times that over equatorial Africa. One of the factors they ascribed to

---

**FIG. 10.** NCEP–NCAR reanalysis mean (1968–96) RH (%) at (top) 1000 and (bottom) 500 hPa for July. Note that the color scales are different. The rectangular boxes are the different validation regions in Fig. 1. (These images are provided by the Physical Sciences Division of the NOAA/Earth System Research Laboratory on their Internet site: http://www.esrl.noaa.gov/psd/.)
this is the proximity of equatorial Africa to desert dust. The two factors, subcloud evaporation and CCN effect, may act together in that the CCN effect produces raindrops of small sizes that, even if they fall out of the cloud, will be evaporated in the dry atmosphere beneath the cloud. A third factor is that desert surfaces could be confused with rain signatures in PM retrievals. For instance, Wang et al. (2009) and Seto et al. (2009) report significant false rain detection over the Sahara Desert owing to surface misclassification.

Overestimation is not the only problem. The satellite products also have problems detecting rainfall when it occurs (low POD). Again the problem is worse over the drier regions. One possible cause could be the combination of coarse spatial resolution of the satellite products and the very hot background surface. Satellite pixels represent the average of whatever is in the sensor’s footprint. If part of the footprint is rain and the other part is background surface, the average value may not be detected as rain.

c. Improving satellite rainfall estimation over desert locust recession regions

Satellite rainfall retrieval algorithms are improving continuously. However, comparison of ARC with the rest of the products has shown that the improvements in the different algorithms have not resulted in any significant improvement for the current validation region. This could partly be due to the unique characteristics of this region, and also because all of the products compared here rely, in one way or another, on TIR inputs. One approach that may help to alleviate the problem is calibrating a selected algorithm using locally available rain gauges. Estimates from TRMM precipitation radar (PR) could also be used where gauges are not available. This local calibration does not need to be complicated. Dinku et al. (2007) show that a simple algorithm calibrated using locally available gauges could be as good as the more sophisticated algorithms. For instance, ARC and RFE use a single brightness temperature threshold (235°K) for the whole of Africa to discriminate between raining and nonraining clouds. This threshold could be too cold for most of the current validation region. Thus, just adjusting the temperature threshold may help to reduce the large FAR values. Another approach to improving the accuracy of the satellite products could be incorporating the available rain gauge observations into the satellite products. Many current satellite products do include rain gauge observations. The problem is that only very few observations are available from the current validation region. Though the gauge network over this region is extremely sparse, it is believed that the meteorological services in the different countries have access to many more data than are available to the developers of the satellite products. These data could be used to improve the quality of satellite products.

5. Summary

Rainfall detection skills of seven satellite rainfall estimates were evaluated over desert locust recession regions. The validation region covers the arid and semiarid region from the Atlantic coast of northwestern Africa to northwestern India. The validation region has been divided into four subregions. Although the satellite estimates exhibited positive skills in rainfall detection, the false-alarm ratios were very high while the probabilities of detection were low. The performances were much better over the relatively wet subregion, whereas they were the poorest over the driest part. The POD ranges from 0.21 for 3B42 over R1 (the driest region) to 0.71 for GSMaP over R2 (the wettest region). The FAR ranges from 0.30 for 3B42 over R2 to 0.84 for ARC over R1. There is no single product that stood out as having the best or the worst performance consistently across subregions. Possible factors for the poor performances of the satellite products could be subcloud evaporation, rainfall suppression by desert aerosols, surface identification problems by PM algorithms, coarse spatial resolution of the products, and hot background surfaces. Local calibration, using locally available rain gauge data or PR estimates, and blending satellite products with locally available rain gauge measurements may help to alleviate the poor performance of satellite estimates.

Acknowledgments. This work was funded by a grant/cooperative agreement from the National Oceanic and Atmospheric Administration, NA05AR4311004, and NASA HQ Science Mission Directorate Cooperative Agreement Notice NN-H-04-Z-YO-010-C through Cooperative Agreement 06CRAG0028 with the U.S. Geological Survey. However, the views expressed herein are those of the author(s) and do not necessarily reflect the views of NOAA or NASA, or any of their subagencies. We are very grateful to FAO for providing the reference data used for this work.

REFERENCES


