JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 111, D16113, doi:10.1029/2005JD006276, 2006

Evaluation of model-derived and remotely sensed precipitation products for continental South America

L. Gustavo Goncalves de Goncalves,[1,2] W. James Shuttleworth,[1] Bart Nijssen,[1] Eleanor J. Burke,[3] Jose A. Marengo,[4] Sin Chan Chou,[4] Paul Houser,[5] and David L. Toll[6]

Received 26 May 2005; revised 9 January 2006; accepted 8 May 2006; published 30 August 2006.

[1] This paper investigates the reliability of some of the more important remotely sensed daily precipitation products available for South America as a precursor to the possible implementation of a South America Land Data Assimilation System. Precipitation data fields calculated as 6 hour predictions by the CPTEC Eta model and three different satellite-derived estimates of precipitation (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN), National Environmental Satellite, Data and Information Service (NESDIS), and Tropical Rainfall Measuring Mission (TRMM)) are compared with the available observations of daily total rainfall across South America. To make this comparison, the threat score, fractional-covered area, and relative volumetric bias of the model-calculated and remotely sensed estimates are computed for the year 2000. The results show that the Eta model-calculated data and the NESDIS product capture the area without precipitation within the domain reasonably well, while the TRMM and PERSIANN products tend to underestimate the area without precipitation and to heavily overestimate the area with a small amount of precipitation. In terms of precipitation amount the NESDIS product significantly overestimates and the TRMM product significantly underestimates precipitation, while the Eta model-calculated data and PERSIANN product broadly match the domain average observations. However, both tend to bias the zonal location of precipitation more heavily toward the equator than the observations. In general, the Eta model-calculated data outperform the several remotely sensed data products currently available and evaluated in the present study.

Citation: de Goncalves, L. G. G., W. J. Shuttleworth, B. Nijssen, E. J. Burke, J. A. Marengo, S. C. Chou, P. Houser, and D. L. Toll (2006), Evaluation of model-derived and remotely sensed precipitation products for continental South America, *J. Geophys. Res.*, *111*, D16113, doi:10.1029/2005JD006276.

1. Introduction

[2] Weather and climate predictions are known to be sensitive to surface storage of water and energy, both at regional and global scales [e.g., *Koster and Suarez*, 1999; *Beljaars et al.*, 1993; *Betts et al.*, 1996; *Fast and McCorkle*, 1991; *Fennessey and Shukla*, 1999]. In the last three decades, land surface models (LSMs) have been used extensively in coupled land surface-atmospheric models to provide a description of the feedback from the underlying soil and vegetation to the atmosphere during model integration [e.g., *Sellers et al.*, 1986; *Xue and Shukla*, 1993]. Recently, so-called land data assimilation schemes (LDAS) [*Mitchell et al.*, 2000, 2004; *Rodell et al.*, 2004; *Koster et al.*, 2004] have been successfully employed to provide improved initial surface fields of soil moisture for use in predictive meteorological models (in near real time) and to address land surface management issues. LDAS comprise two-dimensional arrays of LSMs arranged to match the grid squares used in the predictive model, which are forced by model-derived near-surface fields supplemented, to the maximum extent possible, with surface observations of meteorological variables.

[3] An important challenge when using LDAS or when assessing the performance of coupled (or uncoupled) land-atmosphere parameterizations is the scarcity of comprehensive land surface data at the spatial and temporal resolutions at which the models operate [*Maurer et al.*, 2002]. Providing adequate observations of precipitation is particularly problematic because precipitation is so spatially variable,

[1]Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA.
[2]Now at Hydrological Sciences Branch, Code 614.3, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
[3]Hadley Centre for Climate Prediction and Research, Met Office, Exeter, UK.
[4]Centro de Previsão do Tempo e Estudos Climáticos, Instituto Nacional de Pesquisas Espaciais, Cachoeira Paulista, Sao Paulo, Brazil.
[5]Center for Research on Environment and Water, George Mason University, Calverton, Maryland, USA.
[6]Hydrological Sciences Branch, Code 614.3, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

0148-0227/06/2005JD006276$09.00

and often only point sample data from well-separated rain gauges are available. Some regions of the globe (e.g., North America, Europe, Japan, and parts of the former Soviet Union and China) have a reasonably dense coverage of observations. However, in the context of the present paper, it is significant that South America has very sparse temporal and spatial data coverage, and that this coverage is biased toward populated centers near the edge of the continent and along the main river course in the Amazon region. Consequently, simple interpolation of daily total precipitation is fraught with difficulty. Remotely sensed estimates of precipitation inferred, for example, from infrared cloud top temperatures may provide a means of filling the gaps between surface observations in remote regions. Unfortunately, the calibration and validation of such remotely sensed estimates is also difficult because ground-based observations are so sparse. Nonetheless, in the context of LDAS and despite these validation difficulties, remotely sensed estimates (perhaps merged with surface observations) remain the best hope for providing a spatially comprehensive set of precipitation observations in regions like South America, where surface observations are limited [*Peterson et al.*, 1998; *New et al.*, 2001].

[4] This paper evaluates three remotely sensed daily precipitation products currently available for South America. This is done as a precursor to the possible implementation of a South America LDAS for the Eta regional model [*Mesinger et al.*, 1988; *Black*, 1994] coupled with the Simplified Simple Biosphere (SSiB) LSM [*Xue et al.*, 1991], which is used for weather forecasting at CPTEC (Centro de Previsão de Tempo e Estudos Climáticos), the Brazilian Center for Weather Forecasts and Climate Studies. Consequently, the primary concern is how well the remotely sensed precipitation products reflect the daily precipitation inputs to the land surface. Because the goal of any LDAS would be to improve the initial energy and moisture state of the land surface model for a given moment in time, the precipitation products should reflect as accurately as possible the spatially distributed precipitation amount on any given day. Note that this is a stricter and more challenging requirement than matching the observed precipitation in some average or climatological sense. In addition, to be useful in an LDAS context, the precipitation products should better match the actual precipitation than the model derived product.

[5] In this study, precipitation data fields calculated as 24 hour accumulated predictions by the CPTEC Eta model and several different satellite-derived estimates of precipitation are compared with the available observations of daily total rainfall across South America, and the threat score, fractional covered area, relative volumetric bias, root-mean-square error and spatial cross correlation coefficients of the model-calculated and remotely sensed estimates are computed for the year 2000. In turn, these estimates are compared with gauge observations, which, although not without problem, remain the standard measure of surface precipitation input.

[6] The paper is divided into seven sections. The general climatology of South America is described briefly first, followed by a description of the Eta model, the remotely sensed data products, and the surface observations (section 3). Section 4 presents the analysis methods and performance measures. Results, discussion, and conclusions are presented in sections 5, 6, and 7, respectively.

7. Summary and Conclusions

[40] The fractional covered area (FCA) determines whether the product overestimates or underestimates the area of precipitation. This is often used in conjunction with the threat score (TS), which measures the model's ability to forecast the location of events. The relative volumetric bias (RVB), measures the total volume of predicted precipitation relative to that observed. The RMSE gives a measure of the absolute value of the departure from the observation, while the distribution of cross correlation coefficients with distance measures the model ability to reproduce the spatial structure of precipitation. These measures were used in conjunction with available observations of daily total rainfall across South America to evaluate four precipitation products: the 6 hour predictions from the CPTEC Eta model, and three different satellite-derived estimates of precipitation (PERSIANN, NESDIS, and TRMM).

[41] In a LDAS context, precipitation generated by the atmospheric model is used, along with other forcing information, by the LSM to create a background field of land surface states. Remote sensed information along with sparse surface data are then assimilated to correct inherent systematic errors in the land surface states errors in the atmospheric model. Thus in this study, the above mentioned characteristics of precipitation are analyzed and compared between the Eta model and three precipitation products to determine whether weaknesses of the former could be addressed by the latter. It is important to remember that the analysis presented here is for only for 1 year (2000) and the number of stations with sufficient observations is limited. However, to be useful in an LDAS context, the precipitation products should reflect as accurately as possible spatially distributed precipitation on any given day. This is a stricter and more challenging requirement than matching the observed precipitation in some average or climatological sense, and this has been the focus of the present study.

[42] The Eta model-calculated data and NESDIS product are broadly comparable in terms of success in defining the fractional covered area of the domain with precipitation, the Eta model data being better for low precipitation thresholds and the NESDIS product for high precipitation threshold. However, the comparative success of the NESDIS product in terms of fractional coverage area for high precipitation thresholds may just be a consequence of its tendency to heavily overestimate precipitation in general. The NESDIS product is also the most successful at identifying the location of areas with very heavy precipitation ($P > 50$ mm), although it is still relatively poor. The Eta model-calculated precipitation and PERSIANN product broadly match the zonal average observations. However, both tend to bias the zonal location of precipitation toward the Equator more than the observations. The PERSIANN product also tends to underestimate the area without precipitation, to overestimate the area with a small amount of precipitation (less than 5 mm), and has difficulty locating areas with and without precipitation. The TRMM product underestimates overall precipitation and does not capture the location of precipitation and the fractional area with precipitation as efficiently as the other products.

[43] In general, the Eta model-calculated precipitation outperforms the remotely sensed products evaluated in the present study in that it provides the best agreement with both the overall observed amount of precipitation and the observed location of precipitation. However, some specific features merit mention.