

The Global Impact of Clouds on the Production of MODIS Bidirectional Reflectance Model-Based Composites for Terrestrial Monitoring

D. P. Roy, P. Lewis, C. B. Schaaf, S. Devadiga, and L. Boschetti

Abstract—A global data set of cloud occurrence probability derived from Moderate Resolution Imaging Spectroradiometer (MODIS) Terra and Aqua gridded daily data is analyzed to investigate the probability of obtaining at least a minimum number of cloud-free observations within various compositing periods. The probabilities derived from Terra and Aqua, with morning and afternoon overpass times, respectively, are similar and increase with compositing period. Compositing both Terra and Aqua observations results in considerably higher probabilities of obtaining a sufficient number of observations for bidirectional reflectance model-based compositing. Given that the only alternative to obtaining sufficient samples is to extend the observation period, which can cause significant problems when the surface state changes, it is concluded that using data from the two MODIS sensors provides the most effective way of generating composited products. Findings with respect to the availability of cloud-free composites when n -day composites are generated on a temporally overlapping daily rolling basis, i.e., every day, rather than every n -days, are also discussed for regional and global applications.

Index Terms—Bidirectional reflectance distribution function (BRDF), cloud, compositing, direct broadcast, Moderate Resolution Imaging Spectroradiometer (MODIS), remote sensing.

I. INTRODUCTION

POLAR-ORBITING wide-field-of-view sensors provide nearly daily global coverage observations at resolutions appropriate for regional- and global-scale terrestrial monitoring. Monitoring is conventionally undertaken using temporally composited data rather than instantaneous orbital swath data. This is because composites have less data volume and because the compositing process reduces remotely sensed variations

that are often considered as noise. Compositing processes conventionally select from colocated pixels in different orbits of geometrically registered data a pixel that best satisfies some compositing criteria [1]. The maximum normalized difference vegetation index (NDVI) criteria was the first to be applied in this context, being developed for application to Advanced Very High Resolution Radiometer (AVHRR) NDVI time series to preferentially select pixels with reduced cloud and atmospheric contamination [2]. Other compositing criteria with similar goals have since been developed, e.g., [3]. Bidirectional reflectance effects have been identified as a major source of residual noise in such composites, although they may be accounted for by use of appropriate models [4]–[7]. Recently, Schaaf *et al.* [8] designed and implemented such an approach for Moderate Resolution Imaging Spectroradiometer (MODIS) data that estimate nadir reflectance at a consistent solar zenith angle through inversion of a linear model of the surface bidirectional reflectance distribution function (BRDF), providing consistent composited time series suitable for global terrestrial monitoring.

With all compositing approaches, the availability of sufficient cloud-free observations is important to the generation of high quality products. The impact of clouds is not easy to predict, however, as it is dependent on the spatiotemporal dynamics of clouds and satellite observations, and on the seasonality and duration of the compositing period. Solutions to reducing cloud impacts include extending the period [9] and using the last n cloud-free observations regardless of the period [7]. As compositing procedures implicitly assume that the surface state remains static, increasing the period increases the incidence of temporal biases associated with violating this assumption. For example, the maximum NDVI procedure preferentially selects the most vegetated pixel [1], and so composites generated during periods of vegetation growth (increasing NDVI) are biased temporally to the selection of later observations. Other approaches that may reduce cloud impacts include the generation of composites on a temporally overlapping “rolling” basis and the generation of composites from observations sensed at different overpass times that have different cloud characteristics [10], [11]. These approaches are potentially advantageous as they may reduce the need for longer compositing periods.

This letter examines the global impact of cloud on the production of MODIS BRDF model-based composites for a year of MODIS observations. Cloud impacts on the generation of n -day products produced every n -days, and on a temporally overlapping daily rolling basis, using MODIS Terra (nominal

Manuscript received January 27, 2006; revised March 14, 2006. This work was supported by the National Aeronautics and Space Administration’s Earth System Science Research using Data and Products from TERRA, AQUA, and ACRIM Satellites program (NNG04GP09).

D. P. Roy is with the Geographic Information Science Center of Excellence, South Dakota State University, Brookings, SD 57007 USA (e-mail: david.roy@sdstate.edu).

P. Lewis is with the Remote Sensing Unit and NERC Centre for Terrestrial Carbon Dynamics, Department of Geography, University College London, London WC1H 0AP, U.K. (e-mail: plewis@geog.ucl.ac.uk).

C. B. Schaaf is with the Center for Remote Sensing, Department of Geography, Boston University, Boston, MA 02215 USA (e-mail: schaaf@stanislaus.bu.edu).

S. Devadiga is with Science Systems and Applications, Inc., Lanham, MD 20706 USA and also with the NASA Goddard Space Flight Center, Greenbelt, MD 20071 USA (e-mail: devadiga@ltpmail.gsfc.nasa.gov).

L. Boschetti is with the Department of Geography, University of Maryland, College Park, MD 20742 USA (e-mail: luigi.boschetti@hermes.geog.umd.edu).

Digital Object Identifier 10.1109/LGRS.2006.875433

10:30 A.M. descending equatorial crossing time), MODIS Aqua (nominal 1:30 P.M. ascending equatorial crossing time), and both Terra and Aqua observations, are considered for different compositing periods.

II. DATA

Cloud probability statistics are derived from cloud fraction metadata that describe the fraction of land pixels in MODIS land tile products flagged as “cloudy” or “probably cloudy” by the MODIS cloud mask algorithm [12]. Globally, there are 460 nonoverlapping MODIS land tiles defined in the equal area sinusoidal projection, each covering approximately 1200×1200 km ($10^\circ \times 10^\circ$ at the equator) [13]. Daily tile cloud fraction metadata for tiles that contain 25% or more land pixels, not including Antarctica, for the 366 days of 2004 are considered. An average of 3% and 7% of the 2004 daily cloud fraction data are missing for the Terra and Aqua tiles used in this study, respectively, due to infrequent MODIS production or sensor failures [14]. This data set provides a typical year of cloudy and missing observations sensed by the Terra and Aqua MODIS sensors.

III. METHOD

The following analysis is undertaken independently for each MODIS land tile. The probability of a pixel in a tile being cloudy is assumed to be equal to the tile cloud fraction for that day. For mathematical tractability, it is assumed that there is no temporal correlation of occurrence probability.

Given the probability of a MODIS observation being cloudy $P_{\text{cloud}}(i)$, then the probability of a particular combination j of k different noncloudy observations being selected from n observations is defined from the product of n probabilities as

$$P_C(j) = \prod_{i \in A} (1 - P_{\text{cloud}}(i)) \prod_{i \in B} P_{\text{cloud}}(i) \quad (1)$$

where A is the set of k noncloudy observations and B is the set of $n-k$ cloudy observations (with $A \not\subset B$). For composites generated using only Terra or only Aqua daily gridded products, n is not greater than the number of days defining the compositing period. For composites generated using both Terra and Aqua data, n may be up to twice the number of days defining the compositing period.

The probability of there being k different noncloudy observations selected from n observations is

$$P_k^n = \sum_{j=1}^{nC_k} P_C(j) \quad (2)$$

where $nC_k = n! / [(n-k)!k!]$ is the total number of combinations that k different observations can be selected from n observations. The probability of there being m or more noncloudy observations selected from n observations is

$$\sum_{k=m}^n P_k^n \quad (3)$$

Probability (3) is computed for n -day compositing periods moved every n -days, and also on a rolling basis in daily steps, starting January 1 and moving through the year of daily $P_{\text{cloud}}(i)$ data. Only compositing periods that fall entirely within the year are considered. As cloud cover may have seasonal variability, the percentage of compositing periods over the year where probability (3) is greater or equal to 0.9 is also derived.

The BRDF model-based compositing approach requires a minimum number of cloud-free observations for model inversion. In this study, we consider $m = 7$, as the semi-empirical RossThick–LiSparse model used to generate the MODIS BRDF/albedo product requires seven or more cloud-free observations for reliable inversion [8].

Our assumption that the probability of cloud occurrence is constant within each tile and that there is no temporal correlation of occurrence probability may result in the overestimation of probability (3). Overestimation is particularly likely where there are persistent moderate to high cloud probabilities, for example, in regions with stationary weather systems or orographic clouds.

IV. RESULTS

Fig. 1 shows summary global tile-level results computed for a 16-day compositing period moved on a rolling basis in daily steps through the 366 days of 2004. The mean annual probability of obtaining seven or more noncloudy observations (left column) and the percentage of windows over the year where the probability of obtaining seven or more noncloudy observations is ≥ 0.9 (right column) are shown for compositing Terra, Aqua, and both Terra and Aqua observations. Regional variations are clearly evident in Fig. 1, with greatest reductions in the availability of cloud-free composites occurring in equatorial West Africa, equatorial South America, Southern and Southeast Asia, and in boreal regions of Eurasia and Canada.

Summarizing the results in Fig. 1, the global mean annual probability of obtaining seven or more cloud-free observations is higher for Terra (0.636) than for Aqua (0.595), implying that globally at this scale there are fewer clouds over land at the 10:30 A.M. Terra overpass time than at the 1:30 P.M. Aqua overpass time. The global standard deviations of these probabilities are relatively large and similar however (0.263). Remote sensing artifacts, such as different MODIS cloud mask detection sensitivities to cloud type and properties that vary diurnally, may bias these results. These biases are complicated to understand and may be compounded by the varying sensing time across the MODIS swath and due to the descending and ascending Terra and Aqua orbits [15].

Compositing both Terra and Aqua observations provides a significantly higher likelihood of obtaining cloud-free observations (mean global probability 0.894, standard deviation 0.142) than compositing Terra or Aqua observations alone. The percentage of windows over the year where the probability of obtaining seven or more noncloudy observations is ≥ 0.9 shows similar patterns (Terra 43%, Aqua 37%, Terra and Aqua 80%). This is consistent with Moody *et al.* [16], who observed that annually half the global land surface is

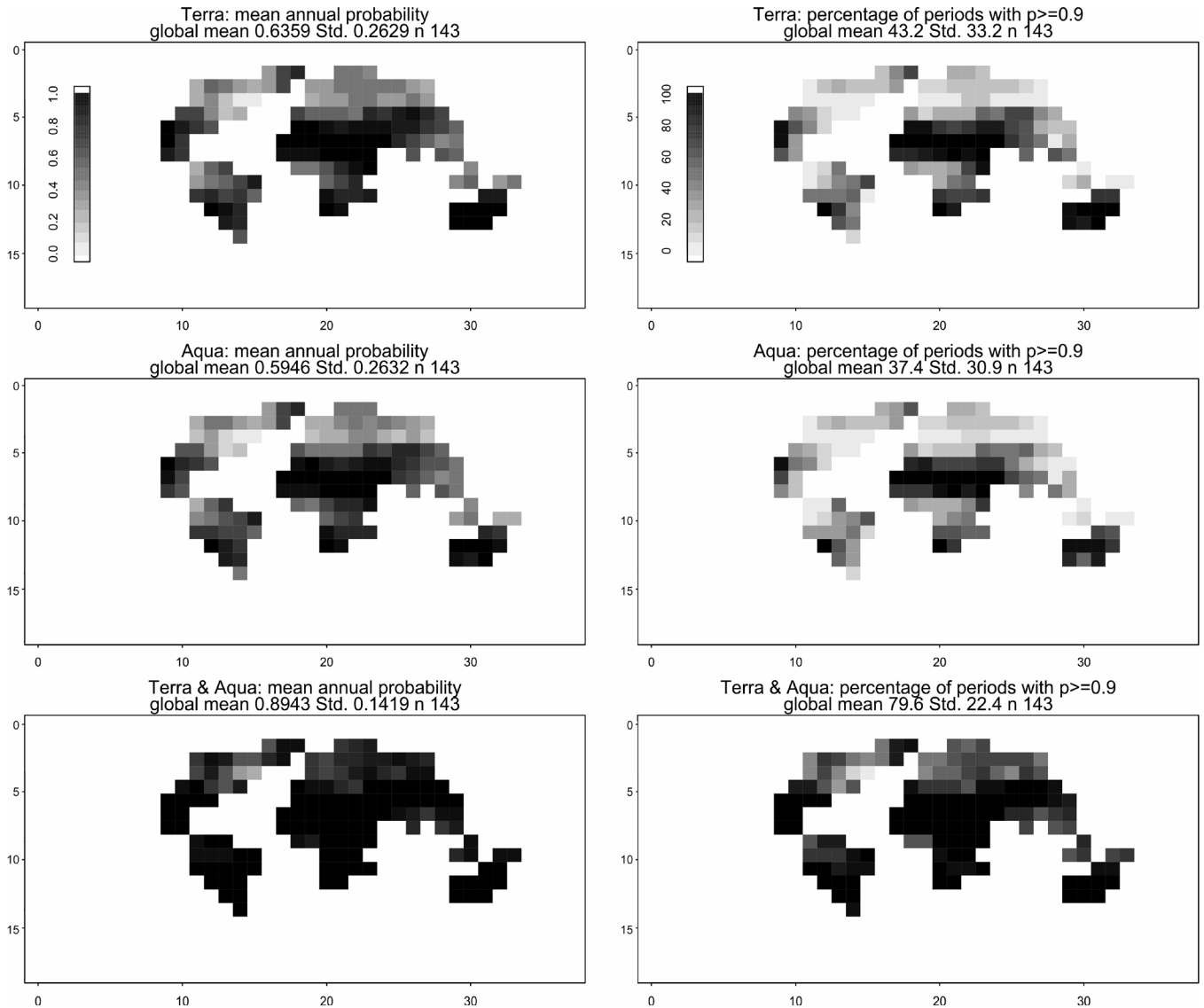


Fig. 1. Global analysis of the impact of clouds on the availability of seven or more noncloudy MODIS observations for a 16-day compositing period moved on a rolling basis in daily steps through the 366 days of 2004. (Left column) Mean annual probability of obtaining seven or more noncloudy observations. (Right column) Percentage of compositing periods over the year where the probability of obtaining seven or more noncloudy observations is greater than 0.9. Results shown for Terra only (top row), Aqua only (middle row), and both Terra and Aqua (bottom row) observations. Results illustrated for 143 MODIS land tiles that cover $\geq 25\%$ land pixels (not including Antarctica). Each tile has fixed Earth locations covering an area of approximately 1200×1200 km ($10^\circ \times 10^\circ$ at the equator). The figure axes show the 36×18 MODIS land tile coordinate system [13].

obsured in MODIS Terra data due to cloud cover as well as snow effects.

Figs. 2 and 3 summarize the 16-day results in Fig. 1 and the results derived for the other (10-, 20-, 24-, and 32-day) compositing periods. Increasing the compositing period increases the global availability of composites with seven or more cloud-free observations.

In Fig. 2, the global mean annual probability of obtaining seven or more noncloudy Terra observations increases with compositing period duration from 0.35 ($n = 10$ days), 0.64 ($n = 16$ days), 0.75 ($n = 20$ days), 0.83 ($n = 24$ days), to 0.91 ($n = 32$ days). The standard deviation of these probabilities decreases with increasing compositing period, reflecting reduced regional variation when longer compositing periods are used. Similarly, in Fig. 3, the global mean percentage of windows over the year where the probability of obtaining seven

or more noncloudy Terra observations is ≥ 0.9 increases from 18% ($n = 10$ days), 43% ($n = 16$ days), 57% ($n = 20$ days), 68% ($n = 24$ days), to 82.5% ($n = 32$ days). The availability of cloud-free composites is consistently slightly greater when Terra rather than Aqua observations are composited, and greater still when both Terra and Aqua observations are used. Differences between these reduce as the compositing period increases.

Globally, the availability of cloud-free composites is slightly, but not significantly, greater for rolling daily composites rather than for those produced in a conventional nonrolling manner. The global mean statistics illustrated in Figs. 2 and 3 are greater than those found for conventional processing, but only in the third and second decimal places of the global mean probability and global mean percentage statistics, respectively. Differences depend on the temporal frequency of clouds relative to the

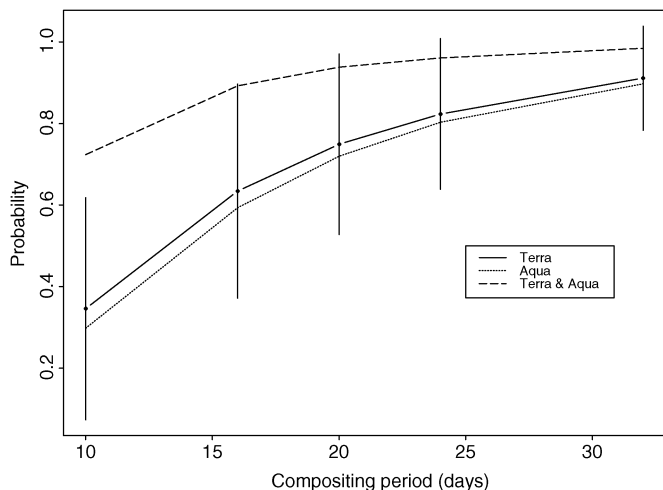


Fig. 2. Global mean annual probability of obtaining seven or more noncloudy observations computed over the 143 MODIS land tiles for compositing MODIS Terra, Aqua, and both Terra and Aqua observations. Results shown for different compositing periods moved on a rolling basis in daily steps through the 366 days of 2004. The error bars show the global mean Terra probability plus or minus the standard deviation of the Terra probabilities.

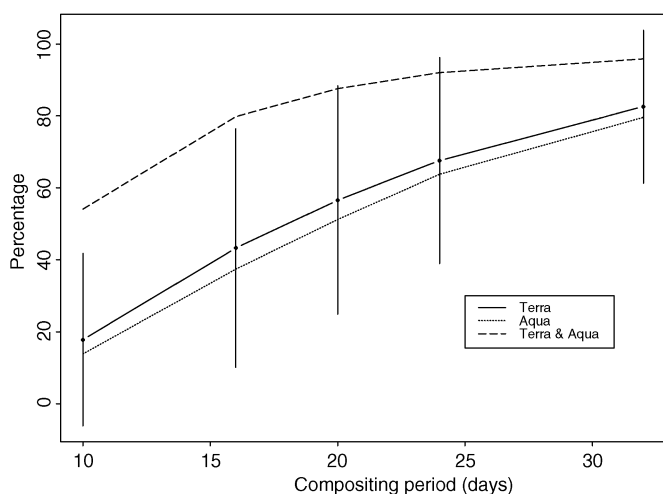


Fig. 3. Global mean percentage of compositing periods over the year where the probability of obtaining seven or more noncloudy observations is ≥ 0.9 computed over the 143 MODIS land tiles for compositing MODIS Terra, Aqua, and both Terra and Aqua observations. Results shown for different compositing periods moved on a rolling basis in daily steps through the 366 days of 2004. The error bars show the global mean Terra percentage plus or minus the standard deviation of the Terra percentages.

timing of the compositing periods and may be more apparent at the local and weekly scale. This is illustrated in Fig. 4, which shows a year of tile cloud fraction data (top) and the probability of obtaining seven or more noncloudy observations for a 16-day composite moved on a daily rolling basis (bottom). The wet summer (December to March), dry winter (May to October), and intermediary periods in this tropical savanna region of northern Australia [17] are evident in the temporal cloud fraction variation. Fig. 4 illustrates that when persistent cloud occurs locally within an n -day period (e.g., around day 148), the conventional generation of composites every n -days is less likely to produce a cloud-free time series suitable for terrestrial monitoring than the generation of composites on a daily or nearly daily rolling basis.

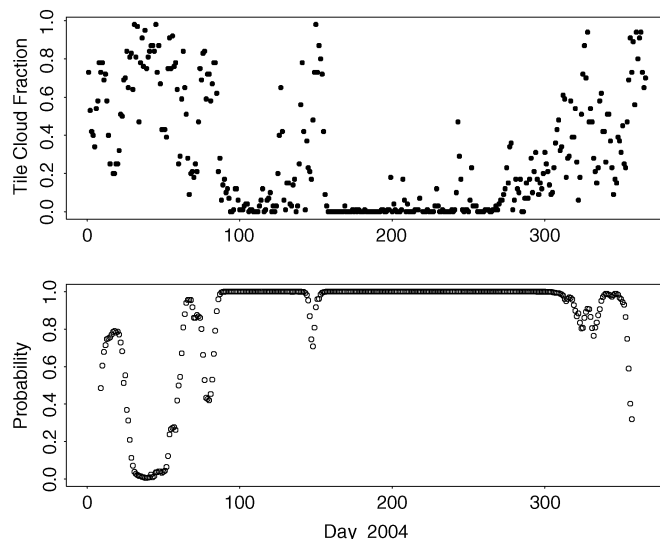


Fig. 4. (Top) Year of daily cloud fraction data for a 1200×1200 km MODIS tile, northern Australia (latitude -10.0° to -20.0° , longitude 121.8512° to 138.3520°). (Bottom) Probabilities of obtaining seven or more noncloudy observations computed for 16-day compositing periods moved on a rolling basis in daily steps (probabilities plotted at the middle of each compositing period).

V. CONCLUSION

Wide-field-of-view sensors such as MODIS have the capacity to provide nearly daily observations of the Earth’s surface; but globally, around half of the observations in gridded daily MODIS land products are unusable due to cloud. Compositing methods combine satellite observations over a given time period, but a minimum number of observations are required to reduce cloud impacts reliably [18], particularly if BRDF normalization is undertaken [8]. If insufficient observations are available, then more observations must be obtained for reliable compositing. Our analysis demonstrates that this can be achieved either by increasing the duration of the compositing period (but with a subsequent decrease in temporal resolution and a higher probability of violating assumptions of static surface conditions) or by adding observations from another sensor with a different overpass time.

The global mean annual probability of obtaining seven or more noncloudy MODIS observations increases with compositing period. For a 16-day compositing period, this probability is about 0.6 for Terra and Aqua, with only a small difference observed due to varying overpass time. If cloud occurrences were too highly correlated between the Terra and Aqua overpass times, then there would be little value in compositing observations from both sensors. However, such a combination is found to increase the probability of obtaining the minimum number of samples to around 0.9, suggesting that this is an effective way of obtaining more reliable composited data. This pattern may not be found at finer spatial scales, especially in regions where diurnal and daily cloud occurrences are temporally correlated.

To date, computational and data distribution constraints have restricted the global production of temporally overlapping “rolling” composites. The rolling approach is particularly appropriate for BRDF model-based compositing as, unlike conventional compositing approaches, every cloud-free

observation in each compositing period is used to estimate nadir reflectance. Consequently, rolling composites allow for terrestrial monitoring at higher temporal frequencies than composites generated in a conventional manner every n -days.

Daily or nearly daily rolling composites may provide a better coherence in time series of products than those derived from composites generated every n -days. We illustrate that in isolated periods of persistent cloud, the rolling approach may generate a more cloud-free time series suitable for regional terrestrial monitoring than conventional n -day composites. Such rolling composites are ideal for MODIS direct broadcast sites with rapid response requirements, where the regional data coverage reduces computational and distribution constraints. Globally, however, the mean availability of cloud-free compositing periods is not significantly higher when rolling daily processing is undertaken rather than conventional processing. Therefore, to maximize cloud-free retrievals while minimizing global data processing constraints, the Collection 5 MODIS BRDF/Albedo product [8] will adopt a quasi-rolling strategy where the compositing period will remain at 16 days but the product will be produced every eight days.

REFERENCES

- [1] D. P. Roy, "The impact of misregistration upon composited wide field of view satellite data and implications for change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 4, pp. 2017–2032, Jul. 2000.
- [2] B. N. Holben, "Characteristics of maximum-value composite images from temporal AVHRR data," *Int. J. Remote Sens.*, vol. 7, no. 11, pp. 1417–1434, Nov. 1986.
- [3] J. Cihlar, D. Manak, and M. D'Iorio, "Evaluation of compositing algorithms for AVHRR data over land," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 2, pp. 427–437, Mar. 1994.
- [4] J. L. Roujean, M. Leroy, A. Podaire, and P. Y. Deschamps, "Evidence of surface bi-directional effects from NOAA/AVHRR multi-temporal data set," *Int. J. Remote Sens.*, vol. 13, no. 4, pp. 685–698, 1992.
- [5] W. J. D. van Leeuwen, A. R. Huete, and T. W. Laing, "MODIS vegetation index compositing approach—A prototype with AVHRR data," *Remote Sens. Environ.*, vol. 69, no. 3, pp. 264–280, Sep. 1999.
- [6] F. Gao, Y. Jin, L. Xiaowen, C. B. Schaaf, and A. H. Strahler, "Bi-directional NDVI and atmospherically resistant BRDF inversion for vegetation canopy," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1269–1278, Jun. 2002.
- [7] P. Maisongrande, B. Duchemin, and G. Dedieu, "VEGETATION/SPOT: An operational mission for the Earth monitoring: Presentation of new standard products," *Int. J. Remote Sens.*, vol. 25, no. 1, pp. 9–14, Jan. 2004.
- [8] C. Schaaf, F. Gao, A. Strahler, W. Lucht, X. Li, T. Tsang, N. Strugnell, X. Zhang, Y. Jin, J.-P. Muller, P. Lewis, M. Barnsley, P. Hobson, M. Disney, G. Roberts, M. Dunderdale, R. d'Entremont, B. Hu, S. Liang, J. Privette, and D. Roy, "First operational BRDF, albedo and nadir reflectance products from MODIS," *Remote Sens. Environ.*, vol. 83, no. 1/2, pp. 135–148, Nov. 2002.
- [9] J. Cihlar, "Detection and removal of cloud contamination from AVHRR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 583–589, May 1994.
- [10] D. P. Roy, P. E. Lewis, and C. O. Justice, "Burned area mapping using multi-temporal moderate spatial resolution data—A bi-directional reflectance model-based expectation approach," *Remote Sens. Environ.*, vol. 83, no. 1/2, pp. 263–286, Nov. 2002.
- [11] D. P. Roy, Y. Jin, P. E. Lewis, and C. O. Justice, "Prototyping a global algorithm for systematic fire-affected area mapping using MODIS time series data," *Remote Sens. Environ.*, vol. 97, no. 2, pp. 137–162, Jul. 2005.
- [12] S. Platnick, M. D. King, S. A. Ackerman, W. P. Menzel, B. A. Baum, J. C. Riédi, and R. A. Frey, "The MODIS cloud products: Algorithms and examples from Terra," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 459–473, Feb. 2003.
- [13] R. E. Wolfe, D. P. Roy, and E. Vermote, "The MODIS land data storage, gridding and compositing methodology: L2 grid," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 4, pp. 1324–1338, Jul. 1998.
- [14] D. P. Roy, J. Borak, S. Devadiga, R. E. Wolfe, M. Zheng, and J. Descloitres, "The MODIS land product quality assessment approach," *Remote Sens. Environ.*, vol. 83, no. 1/2, pp. 62–76, Nov. 2002.
- [15] Y. J. Kaufman, L. A. Remer, D. Tanré, R.-R. Li, R. Kleidman, S. Mattoo, R. Levy, T. Eck, B. N. Holben, C. Ichoku, V. Martins, and I. Koren, "A critical examination of the residual cloud contamination and diurnal sampling effects on MODIS estimates of aerosol over ocean," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 12, pp. 2886–2897, Dec. 2005.
- [16] E. G. Moody, M. D. King, S. Platnick, C. B. Schaaf, and F. Gao, "Spatially complete global spectral surface albedos: Value added datasets derived from Terra MODIS land products," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 144–158, Jan. 2005.
- [17] J. Gentilli, *Australian Climate Patterns*. Melbourne, Australia: Nelson, 1972.
- [18] Y. J. Kaufman, "The effect of subpixel clouds on remote sensing," *Int. J. Remote Sens.*, vol. 8, no. 6, pp. 839–857, 1987.