

© Copyright 2005 American Meteorological Society (AMS). Permission to use figures, tables, and *brief* excerpts from this work in scientific and educational works is hereby granted provided that the source is acknowledged. Any use of material in this work that is determined to be “fair use” under Section 107 of the U.S. Copyright Act or that satisfies the conditions specified in Section 108 of the U.S. Copyright Act (17 USC §108, as revised by P.L. 94-553) does not require the AMS’s permission. Republication, systematic reproduction, posting in electronic form on servers, or other uses of this material, except as exempted by the above statement, requires written permission or a license from the AMS. Additional details are provided in the AMS Copyright Policy, available on the AMS Web site located at (<http://www.ametsoc.org/AMS>) or from the AMS at 617-227-2425 or copyright@ametsoc.org.

Permission to place a copy of this work on this server has been provided to the author by the AMS. The AMS does not guarantee that the copy provided here is an accurate copy of the published work.

The AMS and the authors subscribe to the general scientific convention that published works not be altered without review and approval by an editor. The AMS is responsible for maintaining the definitive version of any AMS publication.

Evaluation and Comparison of MODIS and IMS Snow-Cover Estimates for the Continental United States Using Station Data

K. L. BRUBAKER

Department of Civil and Environmental Engineering, University of Maryland at College Park, College Park, Maryland

R. T. PINKER AND E. DEVIATOVA

Department of Meteorology, University of Maryland at College Park, College Park, Maryland

(Manuscript received 2 October 2004, in final form 23 March 2005)

ABSTRACT

Satellite-derived information on fractional snow cover is essential to resource monitoring, hydrologic modeling, and climate change assessment. Evaluating the accuracy of remotely sensed snow-cover products is important but difficult, largely because point-scale surface observations are spatially sparse and generally nonrepresentative of the remote sensor footprint.

In this study, two remotely sensed snow-cover products [the Interactive Multisensor Snow and Ice Mapping System (IMS) and the Moderate Resolution Imaging Spectroradiometer (MODIS) Climate Modeling Grid (CMG), v.3] are evaluated against ground observations from the Cooperative Observing Network and SNOTEL on a daily basis over the continental United States for calendar year 2000. Ground observations are treated as points in space and time; no physical modeling or statistical interpolation is applied. Hypothesis tests based on discrete and continuous distributions are developed to assess agreement between ground observations and the remotely sensed snow-cover products at 0.25° resolution. (The MODIS CMG product was degraded from 0.05° for this study, thus its potential is not fully evaluated.)

As overall snow extent increases in the course of the season, both MODIS and IMS improve in identifying snow-covered areas (fewer errors of omission), but deteriorate in identifying snow-free areas (more errors of commission). The detection of scattered areas of snow is generally better during ablation than during accumulation. Weaknesses of the statistical methods and assumptions are discussed.

This work will help to identify areas for improvement in snow-cover detection algorithms and provides a framework to assess the accuracy of remotely sensed snow cover used as model input and/or confirmation.

1. Introduction

Information on the time-varying extent of snow cover over watersheds, subareas of watersheds, and large regions is essential for addressing a range of monitoring and modeling issues in hydrology and climate. These issues include assessing climate change, input to snowmelt hydrologic models (e.g., Martinec et al. 1992), developing depletion curves that relate fractional snow cover to snow water equivalent in model parameterizations (e.g., Leavesley et al. 1983), validating distributed models that predict snow cover, and

providing boundary conditions (including albedo) to atmospheric models.

Snow-covered area (SCA) is defined as the area fraction of a region covered by snow of any depth. Roesch et al. (2001) discuss the importance of SCA and issues related to snow detection over flat terrain, mountains, and forested areas. Difficulties in obtaining representative information on SCA have been extensively discussed (Scialdone and Robock 1987; Robinson et al. 1993; Brown 2000; Tait et al. 2000). Snow-cover climatologies have been prepared, at different temporal and spatial scales, using both conventional and remotely sensed data (Robinson et al. 1993; Armstrong and Brodzik 2001, 2002; Romanov et al. 2002, 2003; Romanov and Tarpley 2003).

Remotely observed SCA is often an intermediate step toward determining variables of greater interest:

Corresponding author address: Dr. Kaye Brubaker, 1173 Glenn L. Martin Hall, University of Maryland, College Park, College Park, MD 20742.
E-mail: klbrubak@eng.umd.edu

snow depth (SD) or snow-water equivalent (SWE). Liston (1999) formalizes the mathematical relationship between SCA and the time-varying histogram of SWE. Romanov and Tarpley (2004) extend the work of Romanov et al. (2003) to infer SD based on a correlation between satellite-derived subpixel snow-cover and snowpack depth. The effects of spatial and temporal resolutions of remotely sensed SCA on modeled distributions of SWE have been explored (e.g., Cline et al. 1998). Conversely, information on SD has also been used to infer continental and global SCA using statistical interpolation of SD measurements, empirical correlation, and physical modeling (Brasnett 1999; Brown et al. 2003).

As sensors' temporal, spatial, and spectral resolutions and their classification algorithms continue to improve, users place increasing trust in remotely sensed products. In developing and applying statistical methods to make inferences about snow properties, remotely sensed data are used as input or as validation, sometimes both. Fundamental understanding of spatial patterns and scaling effects (Blöschl 1999) and quantifying confidence in inferred snow variables require accurate measurements at a range of scales. These applications underscore the need to assess satellite products' accuracy.

The goal of this study is to evaluate two remote sensing SCA approaches, considered to be state of the art: the Interactive Multisensor Snow and Ice Mapping System (IMS) of the National Oceanic and Atmospheric Administration (NOAA)/National Environmental Satellite, Data, and Information Service (NESDIS) and the Moderate Resolution Imaging Spectroradiometer (MODIS) Climate-Modeling Grid (CMG) product. This evaluation is performed on a daily basis by grid cell, over a large spatial domain. Based on sampling theory and statistical hypothesis testing, our methods are designed to evaluate spatial data using station observations as "truth," without any statistical or physical interpolation in space and time.

a. Available ground observations

Snow depth and SWE are operationally measured at weather observing stations, automated Snowpack Telemetry (SNOTEL) sites, and snow courses. Brown et al. (2003) present a comprehensive discussion. Station and SNOTEL observations are taken (generally on a daily basis) at fixed locations chosen for accessibility, relevance to aviation, or locations found to be good index indicators of seasonal accumulation and snowmelt runoff for water supply forecasting. Snow course measurements are taken over transects on the order of tens or hundreds of meters (small in comparison to

satellite footprints, kilometers or tens of kilometers); they are generally taken at most twice a month.

b. Remotely sensed snow-cover information

A number of snow-cover products based on both passive microwave and visible sensors are archived at the National Snow and Ice Data Center (NSIDC 2004), including MODIS, Special Sensor Microwave Imager (SSM/I), Scanning Multichannel Microwave Radiometer (SMMR), and Advanced Microwave Scanning Radiometer (AMSR)-E (Armstrong and Brodzik 2001). Passive microwave methods are being developed at National Aeronautic and Space Administration's (NASA's) Goddard Space Flight Center (e.g., Kelly et al. 2003) and at NSIDC. Advantages of the microwave sensors are their ability to penetrate clouds, provide data during darkness, and provide an index of SD or SWE.

The National Operational Hydrologic Remote Sensing Center (NOHRSC) provides airborne, satellite, and modeled snow data and products for the coterminous United States and Alaska, primarily to support the National Weather Service, but also available to the public (NOHRSC 2003). The operationally produced maps of SWE and SCA (snow/no snow) at NOHRSC represent a blend of in situ observations (SNOTEL, snow courses), airborne gamma, satellite information [Geostationary Operational Environmental Satellite (GOES) visible and Advanced Very High Resolution Radiometer (AVHRR)] and physically based modeling. Available (and under ongoing development) since 1990, these products are sometimes used as truth against which to evaluate other remote sensing products or model results.

Of the many products available, two were selected for this study: the NOAA/NESDIS IMS (Ramsay 1998) and MODIS CMG (Hall et al. 2002a), described below. Both produce snow cover as a primary product (rather than SD or SWE). Neither algorithm makes use of surface snow observations; therefore the ground station data can provide an independent check on their accuracy. The IMS is produced operationally and widely used. The MODIS product is itself rapidly becoming a standard of "truth," due to its global coverage and high spectral and spatial resolution. We evaluate each product individually against station data, then use the superior spatial coverage of the MODIS product (compared to station data) to test the IMS product.

1) THE IMS PRODUCT

The Interactive Multisensor Snow and Ice Mapping System (IMS) product of NOAA/NESDIS (Ramsay 1998) is available at a nominal resolution of about 24

km on a daily time scale, covering the Northern Hemisphere on a 1024 by 1024 pixel grid in a polar stereographic projection. This product involves an analyst with primary input from motion loops in the visible-wavelength images from geostationary satellites and polar orbiting satellites; it also relies heavily on passive microwave images. This product was first produced in 1997, and was compared to NOAA/NESDIS's previous weekly product until 1999, when the weekly product was discontinued. Since February 2004, the IMS snow-cover maps have been produced at 4-km resolution.

2) THE MODIS PRODUCT

The Moderate Resolution Imaging Spectroradiometer is flown aboard the *Terra* and *Aqua* satellites, launched in 1999 and 2002, respectively. MODIS snow and ice products, developed at NASA's Goddard Space Flight Center, are archived and distributed by NSIDC Distributed Active Archive Center (NSIDC 2004). MODIS snow-cover products improve and enhance previous operational products mainly because the MODIS products are global at 500-m resolution, daily (as cloud cover permits), and have the capability to better separate snow and clouds. Because the MODIS snow-mapping algorithms are automated, a consistent dataset may be generated for long-term climate studies that require snow-cover information. Extensive quality assurance information is stored with the products (Riggs et al. 2003).

The MODIS 500-m-resolution, 2330-km-swath snow-cover map is gridded to an integerized sinusoidal grid to produce daily and 8-day composite tile products. Both daily and 8-day composites are aggregated to a climate-modeling grid (CMG) product at 0.05°; these products include estimates of grid cell fractional snow cover.

c. Approaches and issues in validating remotely sensed snow cover

Validation against "ground truth" is always a goal for a remotely sensed product. Issues surrounding ground truth for SCA are 1) the point dimension of most ground measurements versus the spatial nature of SCA, 2) the intermittency of ground observations in space and time (remotely sensed products also suffer from space/time intermittency, in particular, clouds for the visible-spectrum instruments), and 3) the lack of ground observations at high elevations in remote regions, where snow data are particularly vital. Many validation studies of remotely sensed SCA have focused on small areas or on cloud-free time periods. The SNOTEL network provides the only available U.S. ground measurements of snow at high elevations and in rugged terrain. However, for purposes of validating re-

mote observations, the SNOTEL network is quite sparse.

The NOHRSC snow-cover maps have been used as ground truth for evaluating the accuracy of satellite-derived SCA. Bitner et al. (2002) compared the NOHRSC snow-cover maps with NESDIS and MODIS observations for the period of November 2000–February 2001. The two products showed an average agreement of 96%. In the MODIS comparison the agreement was 94% in one study area and 95% in a second study area. (Confirmation of the NOHRSC product itself is challenging because the operational mission dictates the use of all available observations, leaving no independent ground truth for comparison.)

Anticipating the launch of MODIS aboard the *Terra* and *Aqua* spacecraft, Hall et al. (2000) used 30-m Landsat Thematic Mapper (TM) images as the baseline for comparison of a prototype MODIS algorithm with 1-km NOHRSC and 0.5° SSM/I snow maps of four locations in North America. The study highlighted the difficulty of mapping snow-covered fraction in areas of patchy snow. Hall et al. (2001) presented a preliminary calculation of global-scale snow mapping errors in the MODIS 500-m data, based on the ability of MODIS to correctly detect complete snow cover under cloud-free conditions and minimal forest cover. An early validation case study of the MODIS snow product (Hall et al. 2002a), involving field and aircraft measurements, reported low errors of commission between 0.02% and 0.10%. An intercomparison of the MODIS snow-cover product with NOHRSC snow cover and SSM/I passive-microwave snow maps attributed differences to the coarse footprint of the SSM/I sensor and the MODIS instrument's ability to capture transient snow cover due to its temporal frequency (Hall et al. 2002b).

Klein and Barnett (2003) compared the MODIS MOD10A1 daily snow-cover product to SNOTEL station data and NOHRSC operational snow maps for the Upper Rio Grande basin of Colorado and New Mexico. They compared maps using methods for categorical data (pixels were classified as "snow," "snow free," etc.). Compared to point SNOTEL sites, MODIS was more likely to map too much snow (errors of commission) and NOHRSC to miss snow (errors of omission). Klein and Barnett point out the difficulty in comparing any remotely sensed product to point SNOTEL observations, in particular the fact that the SNOTEL is not necessarily representative of the remote sensor footprint area.

Simpson and McIntire (2001) used California SNOTEL data to validate their neural network snow-cover classifier for GOES and AVHRR scenes. In addition to scale incompatibility, they raised the issue of positional un-

certainty in the two datasets. They constructed a 7×7 pixel kernel centered on the nominal location of each SNOTEL site; the kernel was compared to SNOTEL only if all nonexcluded (cloud free) pixels in the kernel shared the same label (all snow or all clear land).

Romanov et al. (2000) used in situ observations from the U.S. Cooperative Observing Network (Coop) stations to evaluate the IMS snow-cover maps. Station data were compared with snow classification results within a 3×3 pixel snow map cell centered at the location of the station. The result was considered a “snow–snow” hit if the ground observer reported non-zero SD and snow cover was detected in at least one of the nine pixels; if no snow was detected around such a station, the result was labeled a “snow miss” (omission error). A “no snow–no snow” hit was assigned when all pixels around a station with zero snow depth were classified as snow-free. If more than one pixel in the nine-pixel neighborhood was labeled as snow and snow was not reported at the ground station, the result was labeled a “false alarm” (commission error). Cases when no snow was reported at the station and only one of nine pixels was classified as snow were considered too uncertain to include.

Simic et al. (2004) used SD observations from almost 2000 stations to evaluate daily satellite snow-cover products over Canada: MODIS at 500-m resolution, a NOAA GOES–SSM/I combination at 4-km resolution, and SPOT-VEGETATION (VGT) at 1-km resolution. Their findings showed that the VGT had a bias toward mapping snow-free conditions (errors of omission). The MODIS and NOAA products had similar levels of accuracy, agreeing with ground data at about 80%–100% on a monthly basis. The authors estimated that nonrepresentativeness of the station observations accounted for about 10% disagreement.

As this brief overview suggests, there is no single widely accepted method for evaluating the accuracy of remotely sensed SCA. Building on and complementing the aforementioned work, the methods presented here (a) use only station observations as ground truth; (b) are based on quantitative (rather than categorical) data; (c) allow a quantitative statement of confidence in the conclusions, by means of a dictated level of significance in statistical hypothesis tests; (d) are applied over the entire continental United States, not limited to a specific region; and (e) apply to daily (not monthly) data.

2. Methods

We have performed comparisons between and among: the IMS product, the MODIS Climate Modeling Grid daily snow-cover product (MOD10C1, v.3),

and a ground station dataset consisting of the U.S. Coop stations (NCDC 2004) and SNOTEL sites (NRCS 2004). All analyses were performed on a 0.25° latitude \times 0.25° longitude grid, consistent with the IMS product’s resolution. An IMS determination of “snow covered” is assumed to indicate that the land surface represented by a grid cell is at least 50% covered by snow of any depth.

The IMS maps of snow/no snow in a polar stereographic projection were reprojected to the 0.25° study grid using routines in Grid Analysis and Display System (GrADS; Doty 2002), with a nearest-neighbor resampling procedure to preserve the binary nature of the images. At the latitude range of the continental United States, the 0.25° resolution differs by a few kilometers from the nominal 24-km resolution of the IMS map; therefore, the reprojection does not impose aggregation or disaggregation; however, by introducing up to half a pixel (12 km) of translation, the process has the potential to misidentify grid cells at the snow/no-snow boundary or in areas of scattered snow cover.

Each comparison consists of a statistical hypothesis test performed by grid cell over the entire continental United States on a daily basis for 6 months (January–March and October–December) of the year 2000. The hypothesis tests are now described.

a. Test 1: Using ground station data to test the IMS snow-cover product

The ground stations (SNOTEL and Coop) within each grid cell were identified by their latitude–longitude coordinates from the respective metadata (NRCS 2004; NCDC 2004). Station measurements within the grid cell are treated as Bernoulli trials in the probabilistic sense; namely, the stations are treated as random samples of a binary population, snow or no-snow, where the Bernoulli probability of “success,” p_S , is the true snow-covered fraction in the grid cell. The population consists of infinitely many points in the grid cell, and success is defined as snow of any depth at the sampled point. The n ground station observations within a cell are “truth” in the sense of being a *random sample* of the population rather than *representative*. This treatment addresses the dual problem of point measurement versus an areally defined quantity and nonrepresentativeness of ground truth data. A random, not a representative, sample is the foundation of the sampling theory underlying statistical hypothesis testing.

For each of n point observations within a grid cell, the Coop stations’ SD reports and the SNOTEL sites’ SWE reports are examined to see whether they report snow of any depth. Bernoulli success is defined as SD or

SWE > 0 and “failure” as SD or SWE = 0. The number of successes is recorded as y . Grid cells containing no ground station reports for a given date are skipped.

The hypothesis test determines whether the observed proportion of successes in the grid cell, y/n , is consistent with IMS identification of the pixel as snow covered or a snow free. To test whether the cell is snow-free, the research hypothesis is (following the assumed IMS threshold)

$$H_A: p_S < 0.5, \quad (1)$$

and the corresponding null hypothesis is

$$H_0: p_S \geq 0.5. \quad (2)$$

The cumulative binomial formula is used to calculate the probability of observing y or fewer successes in n trials if p_S equals the threshold value, 0.5. If this probability is small (less than a predetermined level of significance, α), then H_0 is rejected and H_A accepted; the station data are consistent with true snow cover less than 50% in the cell with confidence level $1 - \alpha$. The probability of observing y or fewer successes in n trials decreases as p_S is increased beyond 0.5; therefore contradicting H_0 requires calculations only for the threshold value. Failing to reject H_0 does not constitute proof that p_S exceeds the threshold value (see, e.g., Hayter 2002).

To test whether a cell is snow covered, the research and null hypotheses are, respectively,

$$\begin{aligned} H_A: p_S > 0.5 \\ H_0: p_S \leq 0.5 \end{aligned} \quad (3)$$

The cumulative binomial formula is used to calculate the probability of observing y or more successes in n trials given that p_S equals the threshold value, 0.5. If this probability is less than α , H_0 is rejected; the station data are consistent with true snow cover greater than 50% in the cell, with $1 - \alpha$ confidence. [The strict inequality used for H_A in Eq. (3) may appear inconsistent with our assumption that IMS “snow covered” grid cells have fractional coverage greater than or equal to 0.5. In statistical hypothesis testing, equality is included in the null hypothesis, in order to calculate the critical value of a test statistic. The probability of a continuous random variable, such as the unknown population parameter,

p_S , assuming any single exact value is zero, and thus, $\Pr(p_S \geq 0.5) = \Pr(p_S > 0.5)$.]

Both one-sided tests are performed for each study grid cell. If neither results in a rejection of its null hypothesis, the cell is labeled “nonconclusive”; the observed y/n is reasonably probable in either case, snow cover less than 50% or snow cover greater than 50%. Otherwise, the results are counted into the appropriate category (Table 1): IMS “snow” cells where the station data agree by accepting H_A in Eq. (1); IMS “snow” cells where the stations disagree by accepting H_A in Eq. (3) (a false positive or error of commission); IMS “no snow” cells where stations disagree (a nondetect or error of omission); and IMS “no snow” cells where station data agree.

If a grid cell is in reality exactly 50% snow covered, $p_S = 0.5$, then the probability of observing exactly 1 or exactly 0 successes in one trial ($n = 1$) is 0.5. Typical values of α used in hypothesis testing are 0.05 or 0.10. Clearly, a grid cell containing only one station could never give results contradicting the null hypothesis in either hypothesis test, and is automatically inconclusive. In other words, a single station reporting snow is not sufficient to confirm a remote sensing classification of “snow covered,” even though 1/1 represents 100% snow reports, nor is a single station reporting “no snow” sufficient to confirm a classification of “snow free.” If in reality $p_S = 0.5$, the probability of observing exactly 2 or exactly 0 successes in two trials ($n = 2$) is 0.25. To produce more than a handful of results, this study selected $\alpha = 0.26$ as the level of significance. This decision allows grid cells containing two observations to produce statistically significant results, but only if both stations agree (if $p_S = 0.5$, the probability of observing exactly 1 success in 2 trials is 0.5). This value means that all conclusions are made with 74% confidence. A smaller α would give more confidence, but fewer results.

Sampling theory assumes only a random sample, and allows for the possibility that a sample is nonrepresentative. The rigorous probabilistic definition of a random sample of size m is that any set of m individuals is equally likely to be chosen from the population. It could be argued that using fixed observing stations means that not every m points in the pixel have an

TABLE 1. Table of outcomes, testing IMS snow-cover product with station data.

IMS	Hypothesis test on station data		
	$H_A: p > 0.5$ accepted	$H_A: p < 0.5$ accepted	Neither research hypothesis accepted
Snow	Snow: stations agree	Snow: stations disagree (false positive)	Nonconclusive
No snow	No snow: stations disagree (NonDetect)	No snow: stations agree	Nonconclusive

“equal probability of being chosen”—that, in fact, the probability of selecting points other than the stations is zero. On the other hand, if storm systems produce random patterns of snow cover within the cell, then the fixed locations of the points are random with respect to that population. Whether the assumption of random patterns of snow coverage at the 0.25° scale is reasonable is discussed later; it is questionable, at best, in regions of pronounced orography and preferred storm tracks. Nonetheless, it allows a starting point for the application of statistical hypothesis testing to the “ground truth” problem using uninterpolated station observations.

The results are mapped and tabulated for each day. Example input and results are shown in Fig. 1 for 22 December 2000. Figure 1a shows the IMS Snow/No Snow product for that date; Fig. 1b is the MODIS snow-cover estimate at the resolution of the study grid (discussed below); Fig. 1c gives the percentage of ground stations in each pixel reporting nonzero SWE or SD; and Fig. 1d gives the number of stations per pixel reporting on that date. Results of test 1 are plotted in Fig. 1e. Confirmed IMS snow or no snow, false positives, and nondetects are shown in different colors; cells with no station data or with inconclusive results are not shown. A large number of cells contain only one or two stations (Fig. 1d), resulting in sparse coverage of conclusive results. The remaining parts of Fig. 1 are discussed in the corresponding sections below.

b. Test 2: Using ground station data to test the MODIS product

The MODIS CMG daily product MOD10C1, Version 3, provides an estimate of percent snow cover, either equal to 0 or between 40 and 100, on a 0.05° latitude \times 0.05° longitude geographic grid. (Version 4 estimates snow cover over the full range, 0 through 100.) A binomial distribution technique similar to test 1 is applied in testing the MOD10C1 product, except that, instead of imposing an assumed threshold, the hypothesized grid cell SCA or Bernoulli success rate, p_s , can vary.

We first aggregate the CMG snow product to the 0.25° study grid. The original MODIS swath images have a nominal resolution of 500 m; the MODIS algorithm labels each 500-m pixel as clear land, snow, cloud, water, etc. The parameters provided for each CMG pixel are based on n original MODIS observations, where

$$n = n_{\text{snow}} + n_{\text{clear-land}} + n_{\text{cloud}} + n_{\text{unknown-not-water}} \quad (4)$$

The number of pixels used in analyzing a CMG pixel, n , varies with latitude.

MOD10C1 v.3, provides a percent snow cover, representing the fraction of original MODIS pixels making up the CMG pixel that were identified as “snow”:

$$\% \text{snow} = 100 \frac{n_{\text{snow}}}{n}. \quad (5)$$

The algorithm imposes no assumptions about snow under cloud cover. The fraction of the CMG pixel covered by cloud is reported as

$$\% \text{cloud} = 100 \frac{n_{\text{cloud}}}{n}. \quad (6)$$

A Confidence Index (CI), equal to the fraction of pixels determined to be cloud-free land, is provided:

$$CI = 100 \frac{n_{\text{snow}} + n_{\text{clear-land}}}{n}. \quad (7)$$

The total number of original MODIS pixels used to create the CMG information, n , is not reported (NASA 2004).

We infer a CMG pixel snow cover, $p_{\text{CMG_pixel}}$, consistent with the MODIS data and accounting for the uncertainty introduced by partial cloud cover in the pixel. Rather than a single number, $p_{\text{CMG_pixel}}$ is described by a probability density function (pdf) expressing our knowledge, belief, and uncertainty about its value. A reasonable first estimate, based on the visible portion of the pixel, assumes that the snow-covered fraction of the entire pixel equals the fraction of cloud-free area covered by snow. We represent our level of belief by making this estimate the mode, or most likely value. The areas identified as cloud covered could be completely snow-free; thus, the minimum possible value of $p_{\text{CMG_pixel}}$ is $\% \text{snow}$. At the other extreme, the entire area blocked by cloud may be snow covered, making the maximum possible value of $p_{\text{CMG_pixel}}$ equal to the sum of $\% \text{snow}$ and $\% \text{cloud}$. The triangular pdf depicts the possibilities and belief expressed in these statements. An illustrative range of hypothetical triangular pdfs is graphed in Fig. 2 for cases in which $\% \text{snow}$ equals 20, and $\% \text{cloud}$ varies from 80 to 0. The triangles along the snow-cover axis indicate the mean (expected value) of the distribution and two standard deviations above and below it. The expected value of is given by

$$p_{\text{CMG_pixel}} = \frac{1}{3} \left(\frac{\% \text{snow}}{100} + \frac{\% \text{snow}}{CI} + \frac{\% \text{snow} + \% \text{cloud}}{100} \right) \quad (8)$$

and the variance is

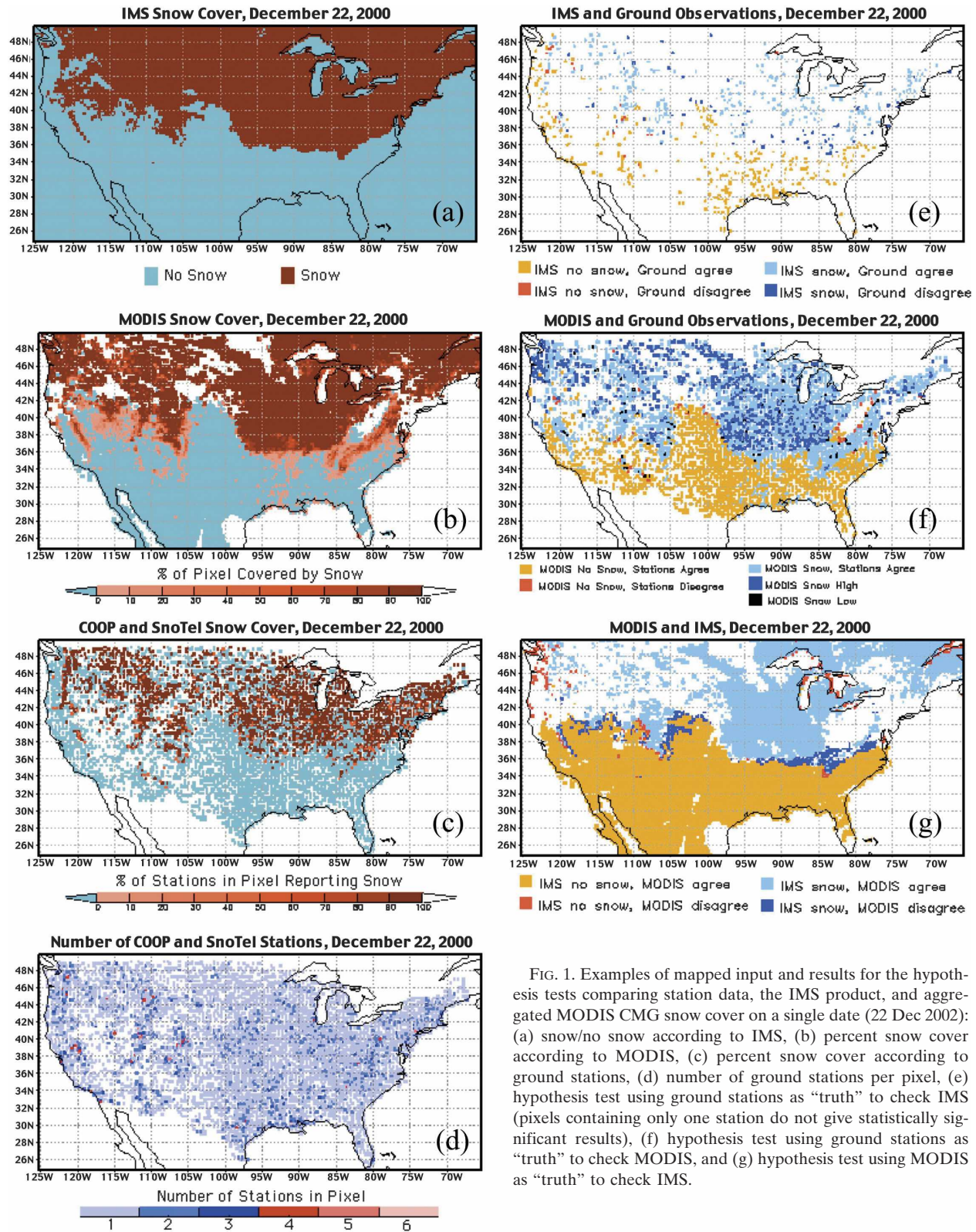


FIG. 1. Examples of mapped input and results for the hypothesis tests comparing station data, the IMS product, and aggregated MODIS CMG snow cover on a single date (22 Dec 2002): (a) snow/no snow according to IMS, (b) percent snow cover according to MODIS, (c) percent snow cover according to ground stations, (d) number of ground stations per pixel, (e) hypothesis test using ground stations as “truth” to check IMS (pixels containing only one station do not give statistically significant results), (f) hypothesis test using ground stations as “truth” to check MODIS, and (g) hypothesis test using MODIS as “truth” to check IMS.

$$\text{Var}[p_{\text{CMG_pixel}}] = \frac{1}{18} \left[\left(\frac{\% \text{snow}}{100} \right)^2 + \left(\frac{\% \text{snow}}{\text{CI}} \right)^2 + \left(\frac{\% \text{snow} + \% \text{cloud}}{100} \right)^2 - \left(\frac{\% \text{snow}}{100} \right) \left(\frac{\% \text{snow}}{\text{CI}} \right) - \left(\frac{\% \text{snow}}{\text{CI}} \right) \left(\frac{\% \text{snow} + \% \text{cloud}}{100} \right) - \left(\frac{\% \text{snow} + \% \text{cloud}}{100} \right) \left(\frac{\% \text{snow}}{100} \right) \right]. \quad (9)$$

In the first case (Fig. 2a), 20% of the CMG pixel is identified as snow and the remaining 80% is cloud covered (%snow = 20; CI = 20; %cloud = 80); the range of the triangular distribution reflects the possibility that snow cover could lie anywhere between 20 and 100. The apex, the most likely value of pixel snow cover, lies at 100, because 100% of the cloud-free land is identified as snow. In the second example (Fig. 2b), only 70% of the pixel is identified as cloud and 10% as snow-free land; the pixel snow cover might lie anywhere between 20 and 90%, with the most likely value 20/30 or 67%. As the cloud-covered portion of this hypothetical pixel decreases, the mean and mode of the distribution both approach 20, and the variance decreases. Thus, the distribution and its statistics reflect increased confidence in CMG pixel snow-cover estimates as cloud cover decreases. Entirely cloud-covered pixels are excluded.

The snow-cover estimates for all the 0.05° CMG pixels contained in each study grid cell must be further aggregated to obtain a MODIS SCA estimate at 0.25° resolution. The study grid snow cover is estimated as a simple average of the expected values in all CMG pixels that contribute data:

$$E[p_{\text{MODIS}}] = \frac{1}{n_{\text{valid}}} \sum_{i=1}^{n_{\text{valid}}} E[p_{\text{CMG_pixel}}(i)], \quad (10)$$

$$\text{Observed Level Of Significance} = p = \begin{cases} \Pr(y \leq y_{\text{obs}}) = \sum_{y=0}^{y_{\text{obs}}} \binom{n}{y} y^{p_{\text{MODIS}}} (n-y)^{1-p_{\text{MODIS}}} & \text{if } \frac{y}{n} < p_{\text{MODIS}} \\ \Pr(y \geq y_{\text{obs}}) = \sum_{y=y_{\text{obs}}}^n \binom{n}{y} y^{p_{\text{MODIS}}} (n-y)^{1-p_{\text{MODIS}}} & \text{if } \frac{y}{n} > p_{\text{MODIS}} \end{cases}. \quad (12)$$

If the probability is small (less than $\alpha/2$ for the two-tailed test), then H_0 is rejected. If the null is not rejected, the station data confirm the MODIS estimate and the pixel is counted as “MODIS Snow – Agree.”

A one-sided test determines whether MODIS over or underestimates the snow cover with respect to the stations. In this one-sided test, the research hypothesis is either

$$H_A: p_S < p_{\text{MODIS}} \quad (13)$$

where n_{valid} is the count of valid CMG pixels—those for which data are given and are not completely cloud-covered—contained in the study grid cell (maximum 25). Cells containing no valid MOD10C1 pixels are excluded.

The station reports (SD and SWE) are counted for each grid cell as in test 1, only now the hypothesis test is based on the MODIS snow-cover estimate, Eq. (10). If both the MODIS and station estimates are 0, the pixel is counted as “MODIS No Snow – Stations Agree.” If the MODIS estimate is 0, and the station estimate nonzero, the pixel is counted as “MODIS No Snow – Stations Disagree.” In other words, one station reporting snow is necessary and sufficient to disprove the hypothesis that p_S equals 0. If the MODIS estimate is nonzero, a two-sided hypothesis test for agreement is performed based on the binomial distribution, where the research and null hypothesis are, respectively,

$$\begin{aligned} H_A: p_S &\neq p_{\text{MODIS}} \\ H_0: p_S &= p_{\text{MODIS}} \end{aligned}. \quad (11)$$

The binomial probability of observing y/n at least as contradictory to the null hypothesis as that observed is computed, *assuming* $p_S = p_{\text{MODIS}}$. This probability, also known as the “observed level of significance,” is computed for this two-tailed test as

or

$$H_A: p_S > p_{\text{MODIS}} \quad (14)$$

as appropriate. Rejection of H_0 in the two-sided test ($p < \alpha/2$) necessarily implies rejection in the one-sided test ($p < \alpha$). Therefore, when p_{MODIS} is less (greater) than y/n and H_0 is rejected in Eq. (11), MODIS is said to under(over)estimate snow cover with respect to station data (“MODIS Snow Low” or “MODIS Snow

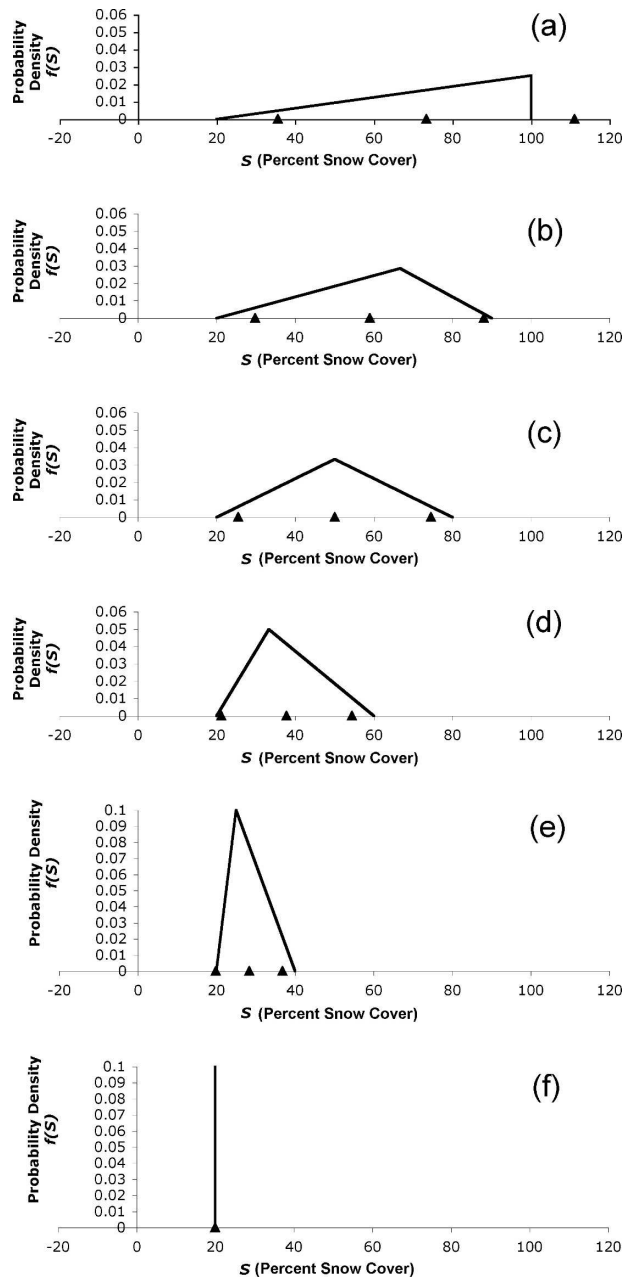


FIG. 2. Example probability density functions (pdfs) for Pixel Snow Cover based on MOD10C1. In all cases [(a)–(f)], 20% of the $0.05^\circ \times 0.05^\circ$ climate model grid is identified as “snow covered”: (a) 80% cloud, (b) 70% cloud, (c) 60% cloud, (d) 40% cloud, (e) 20% cloud, (f) 0% cloud. The small triangles on the horizontal axis indicate the mean and ± 2 standard deviations.

High”). Figure 3 shows this procedure graphically for two examples where the binomial distribution is asymmetric ($p_{\text{MODIS}} \neq 0.5$). The test is performed by grid cell for each day, over the same time period as test 1. The five possible outcomes are shown as a flowchart in Fig. 4.

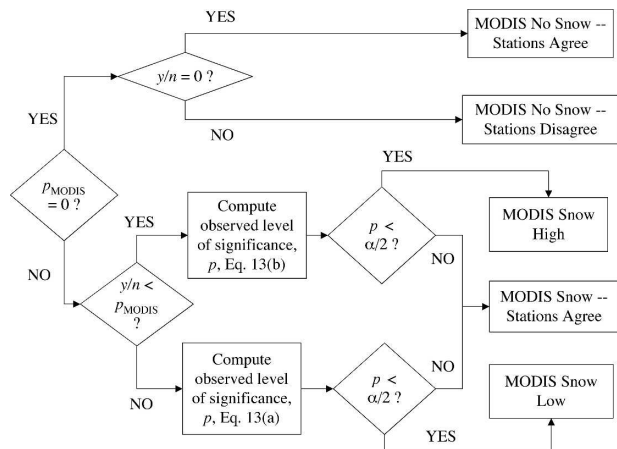


FIG. 3. Flowchart for hypothesis test 2: using station data to test 0.25° by 0.25° pixel snow cover derived from MODIS MOD10C1 (p_{MODIS}). In this hypothesis test, n represents the number of stations in the pixel and y the number of stations reporting snow of any depth on the ground. The observed level of significance, p , is the binomial probability of observing y or fewer (or greater) stations with snow if the true snow-covered fraction of the pixel equals p_{MODIS} .

Percent snow cover for the study grid as determined from MOD10C1 is shown in Fig. 1b and results of the hypothesis test in Fig. 1f for 22 December 2000. At the same level of significance as test 1 ($\alpha = 0.26$), this hypothesis test is conclusive for a larger number of cells because cells containing only one station can give statistically significant results when $E[p_{\text{MODIS}}]$, Eq. (10), is either small or large. An undesirable corollary is the “confirmation” of a number of MODIS “snow” cells in Louisiana and neighboring states, where physical reasoning and additional information (e.g., temperature) indicate that no snow is present. These results were not edited out, because we wish to explore all the properties of the proposed testing method. This problem is discussed in section 4.

c. Test 3: Using MODIS to test IMS product

One shortcoming of the use of ground stations to test a remotely sensed snow-cover product such as IMS was noted in test 1: sparse station coverage results in numerous grid cells that contain either no stations or too few stations to give statistically significant results. Having established the level to which ground station data confirm MODIS estimates of snow cover on the study grid, we use the MODIS product’s superior coverage to evaluate the IMS product. The final set of hypothesis tests determines whether the grid cell fraction of snow cover inferred from the MOD10C1 product confirms the IMS characterization as snow covered or snow free.

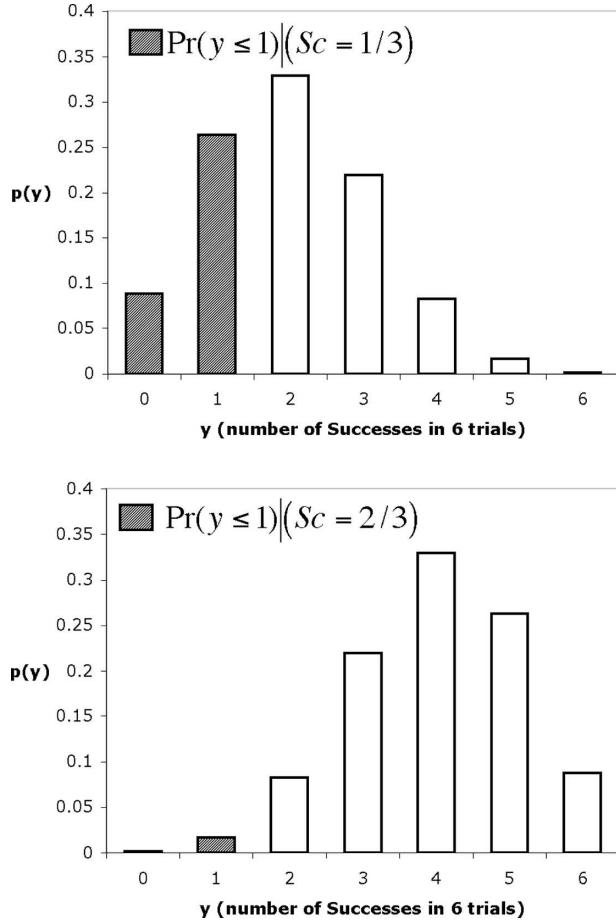


FIG. 4. Examples of hypothesis test outcomes when hypothesized snow cover is (a) 1/3 and (b) 2/3. In each case, the pixel contains six station observations and only one station reports non-zero snow depth or snow water equivalent. In (a) the observed level of significance is 0.30 and the null hypothesis of equality cannot be rejected. In (b) the observed level of significance is 0.017 and the null hypothesis of equality is rejected for any $\alpha/2$ (two-sided test) or α (one-sided test) greater than that value.

Again, we assume that IMS “snow” means a cell is at least 50% covered by snow of any depth. In this test, the MODIS10C1 product is treated as “truth” and used to test the hypothesis posed by the IMS classification. The estimated “truth,” however, is subject to variance introduced by partial cloud cover. Using the triangular distribution assumption described under test 2, the expected value of the aggregated MODIS estimate of SCA at the study grid resolution is given by Eq. (10) and the variance of the study grid cell estimate is

$$\text{Var}[p_{\text{MODIS}}] = \frac{1}{(n_{\text{valid}})^2} \sum_{i=1}^{n_{\text{valid}}} \text{Var}[p_{\text{CMG_pixel}(i)}] \quad (15)$$

(assuming that the estimate of p_{MODIS} for a given cell is statistically independent of all other cells). The variance given by Eq. (15) establishes a confidence interval around the expected value of p_{MODIS} given by Eq. (10). The test statistic is the standard normal random variable, z , based on the hypothesized p_S :

$$z = \frac{E[p_{\text{MODIS}}] - 0.5}{(\text{VAR}[p_{\text{MODIS}}])^{1/2}}. \quad (16)$$

(Because p_{MODIS} is an average of up to 25 random variables drawn from triangular distributions, the sampling distribution would be expected to be at least roughly mound shaped, and thus hypothesis testing based on a normal distribution assumption is appropriate.)

To confirm a pixel as snow free, according to MODIS “truth,” the research and null hypotheses are as in Eqs. (1) and (2). The null hypothesis is rejected if z [Eq. (16)] is less than the critical value $-z_\alpha$, where $-z_\alpha$ corresponds to the lower tail of the normal distribution,

$$P[z < -z_\alpha] = \alpha. \quad (17)$$

To confirm a cell as snow covered, the research and null hypotheses are as in Eq. (3). The null hypothesis is rejected if z [Eq. (16)] exceeds the critical value z_α , where z_α corresponds to the upper tail of the normal distribution

$$P[z > z_\alpha] = \alpha. \quad (18)$$

As in test 1, both tests are performed for each cell of the study grid. If neither null is rejected, the pixel is labeled “nonconclusive,” meaning that the uncertainty in the MODIS estimate is so great that snow cover less than or greater than 50% cannot be confirmed. Otherwise, the cell is categorized, and the results mapped and summarized, as for test 1: IMS Snow – MODIS agrees, IMS Snow – MODIS disagrees (false positive), IMS No snow – MODIS disagrees (nondetect), and IMS No Snow – MODIS agrees. Example mapped results, with $\alpha = 0.26$, are shown in Fig. 1g. Disagreements appear mostly along the snow/no-snow boundary, IMS having a general tendency to map more snow than MODIS on this particular date, but missing some scattered patches, such as in the southern Appalachians. Gaps in the northwest United States in Fig. 1g, despite apparent agreement in that region between Figs. 1a and 1b, are due to variance in the MODIS grid cell estimate intro-

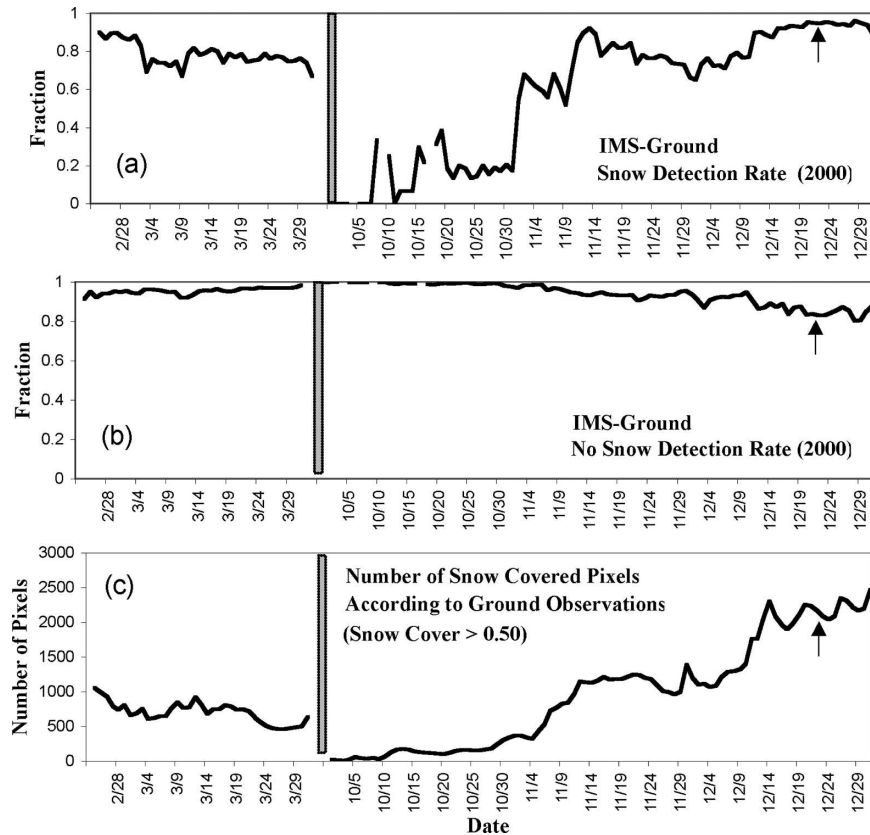


FIG. 5. Summary time series of results for test 1: using ground stations to evaluate IMS snow/no snow product: (a) snow detection rate: fraction of pixels which IMS correctly identifies as snow-covered; (b) no-snow detection rate: fraction of pixels which IMS correctly identifies as snow-free; (c) number of snow-covered pixels according to ground observations (snow covered is defined as more than 50% coverage at $\alpha = 0.26$). The arrow indicates 22 Dec, the date for which results are shown in Fig. 1.

duced by cloud fraction reported in MOD10C1; the variance was sufficient to render the hypothesis test inconclusive in many cells on that date.

3. Results

The testing procedures were conducted for 126 days in calendar year 2000, producing maps corresponding to Fig. 1 for each day. These voluminous results are summarized as time series in Figs. 5, 6, and 7. Each point plotted measures agreement in the respective grid cell comparisons on a daily basis over the entire spatial domain. For comparison with Fig. 1, 22 December 2000 is indicated on these figures by a small arrow.

Figure 5 summarizes the evaluation of IMS snow/no-snow data against ground stations as truth—that is, the maps analogous to Fig. 1e. In Fig. 5a, Snow Detection Rate is defined as the fraction of pixels at least 50% snow covered according to station data (at 26% level of

significance) that were correctly identified by IMS as snow covered—that is,

Snow Detection Rate =

$$\frac{\text{Pixels with "IMS Snow - Stations Agree"}}{\text{Pixels with "IMS Snow - Stations Agree"} + \text{Pixels with "IMS No Snow - Stations Disagree"}} \quad (19)$$

In Fig. 5b, “No Snow Detection Rate” is similarly defined as the fraction of pixels less than 50% snow covered, according to station data at $\alpha = 26\%$, correctly identified by IMS as snow free—that is,

No Snow Detection Rate =

$$\frac{\text{Pixels with "IMS No Snow - Stations Agree"}}{\text{Pixels with "IMS No Snow - Stations Agree"} + \text{Pixels with "IMS Snow - Stations Disagree"}} \quad (20)$$

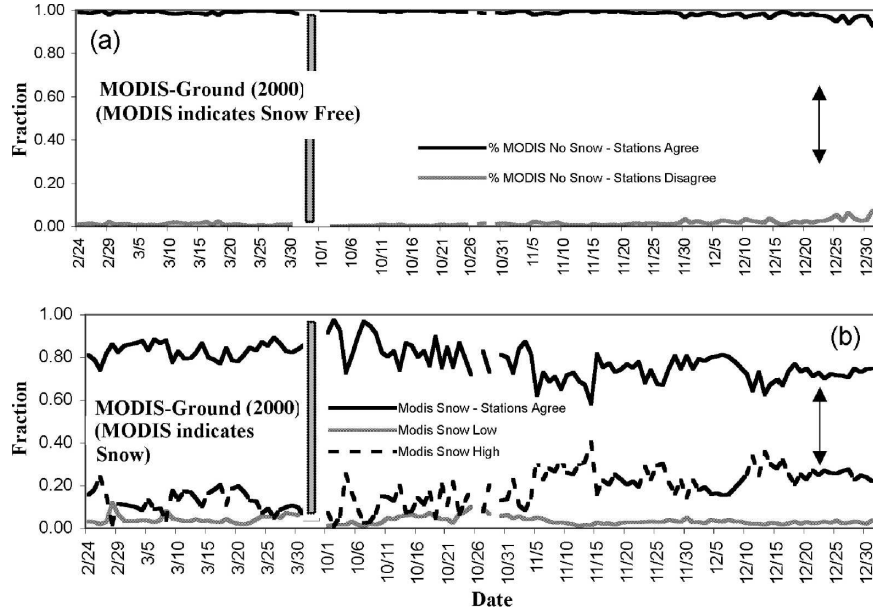


FIG. 6. Summary time series of results for test 2: using ground station as “truth” to test MODIS percent snow cover over $0.25^\circ \times 0.25^\circ$ pixels. (a) Pixels for which MODIS indicates 0 snow cover: percent of pixels where stations agree and disagree; (b) pixels for which MODIS indicates snow cover greater than 0; percent of pixels where stations do not contradict the MODIS estimate (at $\alpha = 0.26$), and where stations indicate MODIS estimate is low and high. The arrow indicates 22 Dec, the date for which results are mapped in Fig. 1.

Figure 5c shows the number of study grid pixels identified as snow-covered ($\geq 50\%$ snow) according to the station data.

Figure 6 summarizes the evaluation of MODIS data, aggregated to the 0.25° study grid, against ground stations as truth, analogous to Fig. 1f. Figure 6a plots hypothesis test results for cases when the MODIS estimates indicate zero snow cover (p_{MODIS} equals 0): cells where neither MODIS nor the stations indicated snow, and cells where the MODIS estimate of zero snow cover was contradicted by at least one station. The graphed fraction equals the number of cells in each situation divided by the total number of cells identified as snow-free by MODIS. Figure 6b summarizes the results when MODIS indicates any nonzero SCA: all cells where the null hypothesis $p_s = p_{\text{MODIS}}$ was not rejected, and cells where MODIS snow cover was greater or less than that supported by the station data (at 26% significance). Each graphed fraction equals the number of cells in that category, divided by the total number of cells where MODIS indicated nonzero snow cover.

Figure 7 summarizes the evaluation of IMS snow/no-snow product using MODIS snow cover (aggregated to 0.25°) as truth, analogous to Fig. 1g. In Figs. 7a and 7b, “Snow Detection Rate” and “No-Snow Detection Rate” are as defined as in Eqs. (19) and (20), except that the denominator refers to cells with, respectively,

50% snow cover or greater, and less than 50%, according to MODIS ($\alpha = 26\%$). Figure 7c gives the value of that denominator for each day.

Compared to station data, IMS snow detection improves as snow cover increases (Fig. 5). During the accumulation season (October), IMS is not very successful at detecting the few snow-covered pixels. Later in the season (November) when snow cover increases to about half its ultimate value, the snow detection rate improves to about 80%. The IMS snow detection rate is best (about 90%) in December when snow cover reaches its maximum. The IMS snow detection also deteriorates somewhat (from 90%–80%) during the snowmelt season in late February–March; it is interesting to note that in March 2000 the snow detection is better than in November 2000, even though the snow cover is less during March, according to the station data. The IMS algorithm is successful (nearly 100%) in detecting snow-free pixels early in the snow accumulation season (October), but the no-snow detection rate deteriorates slightly (to 80% in late December) as snow cover increases. During the melt season (February–March), no-snow detection is between 95% and 100%.

The aggregated MODIS CMG product correctly identifies snow-free pixels at a nearly 100% rate, with only a small loss of accuracy late in the accumulation season (December; Fig. 6a). At most 3% or 4% of pix-

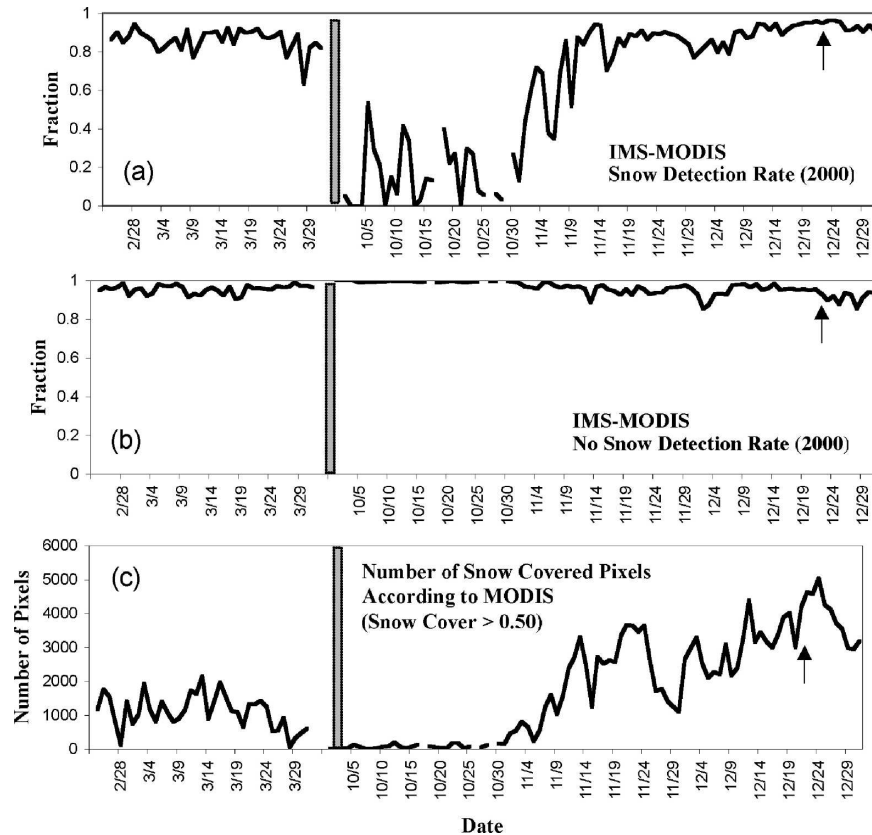


FIG. 7. Summary time series of results for test 3: using MODIS as “truth” to evaluate IMS snow/no snow product at $0.25^\circ \times 0.25^\circ$ resolution. (a) Fraction of pixels that IMS correctly identifies as snow covered ($>50\%$ snow); (b) fraction of pixels that IMS correctly identifies as snow-free ($<50\%$ snow); (c) total number of snow-covered pixels according to MODIS (snow cover >0.5 at $\alpha = 0.26$). The arrow indicates 22 Dec, the date for which results are mapped in Fig. 1.

els identified by MODIS as snow free are contradicted by station data. Of pixels identified by MODIS as having any fraction of snow cover, 80% are confirmed by station data during the months of February–March, between 70% and 95% early in the accumulation season (October) and about 75% later in the accumulation season (December; Fig. 6b). In cells where MODIS and ground stations disagree, MODIS is more likely to overestimate snow cover than to underestimate it, especially early in the accumulation season, when cells with detectable SCA are few and scattered. As snow cover increases, MODIS shows an increasing rate of overestimation, but this rate is lower during the melt season. The tendency to underestimate snow cover appears to be fairly consistent throughout the year. It is reasonable to infer that overestimation is due to cloud contamination; the seasonal pattern of “MODIS Snow – High” is consistent with greater cloud cover during periods of snow accumulation.

Using MODIS as “truth,” IMS is quite successful in

detecting snow (90%–95%) in the months of February, March, November, and December, but less successful in October. During October, IMS matches MODIS very well (near 100%) in terms of identifying snow-free cells (Fig. 6b). During early spring, the IMS no-snow detection rate is about 95%, similar to results in the IMS-Ground Station test (Fig. 4b); however, the no-snow detection rate during the accumulation season is somewhat lower, generally between 85% and 95%.

4. Discussion

As remote sensing science continues to advance, and applications of remote sensing products increase, users need to have confidence in the accuracy of the products, as well as methods to quantify their inevitable uncertainty. Most validation studies have focused on small regions, using a variety of approaches to grapple with the many challenges, including scale incompatibility, sparse and nonrepresentative ground truth, cloud

interference, and space–time intermittency in satellite products. We have designed and implemented statistical methods to evaluate two remote sensing products, the IMS snow/no-snow maps and MODIS fractional snow cover, at 0.25° resolution using ground measurements of SD or SWE.

We treat the SNOTEL and Cooperative Observing System point observations as Bernoulli trials, the fraction of stations reporting snow as a binomial random variable, and the remote sensing snow-cover estimate as the hypothesized Bernoulli success probability. This treatment avoids a major challenge of ground truth in remote sensing data: the fact that point measurements are not generally representative of the remote sensing pixel. We have applied this method to a binary remotely sensed product (IMS snow/no snow) as well as a continuously valued product (MODIS snow cover). For the IMS product, it was necessary to assume a threshold snow cover above which the algorithm should designate a pixel “snow”; we have assumed that this is 50%. For the MODIS product, a triangular probability density function was applied to reflect the level of belief and uncertainty about 0.05° × 0.05° CMG pixel snow cover when the pixel is partially cloud covered. The hypothesized snow cover in the 0.25° × 0.25° pixel was computed using the mathematical expected value of the triangular distribution.

Treating ground stations as random samples of a binary population is a major assumption in our approach. Although it represents an advance over assuming that the point measurement is representative of the pixel,

this assumption is questionable in regions where the horizontal pattern of snow cover is consistent in time. In mountain regions orography drives consistent patterns in precipitation supply, precipitation phase, and melt energy; neither a valley Coop station nor a high-elevation SNOTEL could be considered a random sample of this population. For plains and prairies, where the pattern of snow cover is driven more by storm patterns, the assumption is expected to be more valid.

The denominators in the computed percentages for Fig. 6 are based upon MODIS identification of “snow free” or “snow present,” rather than a percentage of pixels that are snow free according to station data. Therefore, the rates graphed in Fig. 6 do not represent detection rates as in Fig. 5, where the denominator is based on “truth.” The reason for this approach is that, in test 2, there is not a single threshold for identifying a cell as “snow free” or “snow covered.” A cell in which no stations report snow is not necessarily inconsistent with nonzero Bernoulli p_s . As a result, a number of pixels for which MODIS reported some degree of snow cover, yet no Coop or SNOTEL stations reported any snow, were still counted as “MODIS Snow – Stations Agree.” Table 2 enumerates observed levels of significance for a range of such situations. To reject the null hypothesis, the observed probability must be less than $\alpha/2$. When the hypothesized success probability (p_{MODIS}) is small, the probability of observing zero successes (stations with snow) is quite high, even when the sample size is six stations in a cell (the maximum in this

TABLE 2. Probability of observing zero Bernoulli successes (stations reporting snow) in n trials (station reports) if true success probability (p_s) equals SC_{MODIS} . Italics indicate probabilities less than $\alpha/2$ when $\alpha = 0.26$, implying rejection of the null hypothesis (equality).

SC_{MODIS}	$n = \text{no. stations in pixel}$									
	1	2	3	4	5	6	7	8	9	10
0.03	0.970	0.941	0.913	0.885	0.859	0.833	0.808	0.784	0.760	0.737
0.06	0.940	0.884	0.831	0.781	0.734	0.690	0.648	0.610	0.573	0.539
0.09	0.910	0.828	0.754	0.686	0.624	0.568	0.517	0.470	0.428	0.389
0.12	0.880	0.774	0.681	0.600	0.528	0.464	0.409	0.360	0.316	0.279
0.15	0.850	0.723	0.614	0.522	0.444	0.377	0.321	0.272	0.232	0.197
0.18	0.820	0.672	0.551	0.452	0.371	0.304	0.249	0.204	0.168	0.137
0.21	0.790	0.624	0.493	0.390	0.308	0.243	0.192	0.152	<i>0.120</i>	<i>0.095</i>
0.24	0.760	0.578	0.439	0.334	0.254	0.193	0.146	<i>0.111</i>	<i>0.085</i>	<i>0.064</i>
0.27	0.730	0.533	0.389	0.284	0.207	0.151	<i>0.110</i>	<i>0.081</i>	<i>0.059</i>	<i>0.043</i>
0.3	0.700	0.490	0.343	0.240	0.168	<i>0.118</i>	<i>0.082</i>	<i>0.058</i>	<i>0.040</i>	<i>0.028</i>
0.33	0.670	0.449	0.301	0.202	0.135	<i>0.090</i>	<i>0.061</i>	<i>0.041</i>	<i>0.027</i>	<i>0.018</i>
0.36	0.640	0.410	0.262	0.168	<i>0.107</i>	<i>0.069</i>	<i>0.044</i>	<i>0.028</i>	<i>0.018</i>	<i>0.012</i>
0.39	0.610	0.372	0.227	0.138	<i>0.084</i>	<i>0.052</i>	<i>0.031</i>	<i>0.019</i>	<i>0.012</i>	<i>0.007</i>
0.42	0.580	0.336	0.195	<i>0.113</i>	<i>0.066</i>	<i>0.038</i>	<i>0.022</i>	<i>0.013</i>	<i>0.007</i>	<i>0.004</i>
0.45	0.550	0.303	0.166	<i>0.092</i>	<i>0.050</i>	<i>0.028</i>	<i>0.015</i>	<i>0.008</i>	<i>0.005</i>	<i>0.003</i>
0.48	0.520	0.270	0.141	<i>0.073</i>	<i>0.038</i>	<i>0.020</i>	<i>0.010</i>	<i>0.005</i>	<i>0.003</i>	<i>0.001</i>
0.51	0.490	0.240	<i>0.118</i>	<i>0.058</i>	<i>0.028</i>	<i>0.014</i>	<i>0.007</i>	<i>0.003</i>	<i>0.002</i>	<i>0.001</i>

study). Even the probability of zero successes in ten trials is not inconsistent with MODIS snow cover as high as 12%. As a result, a number of pixels in the Gulf Coast states, for example, were identified as “MODIS Snow – Stations Agree” because the number of stations reporting was simply insufficient to contradict the null hypothesis. In other words, nonzero station observations can disprove the MODIS identification of “snow free” but it is difficult to disprove MODIS “snow” with point SD/SWE observations alone at this level of coverage.

Measurement sites tend to be in open areas, where ablation is likely to occur earlier than elsewhere in the grid cell. Assuming that satellite sensors are able to detect more sheltered snow-covered areas, this would lead to systematic overestimation of SCA by the remote sensors during melt events. However, snow detection during spring 2000 was quite good for both products.

In this study, the MODIS CMG product was degraded to a latitude/longitude grid consistent with the coarser resolution of the IMS product. This allowed comparison with the IMS product and ground measurements. However, the potential of the 500-m MODIS snow-cover detection algorithm was not fully evaluated.

5. Conclusions

This paper has proposed and applied a method to maximize the use of available ground measurements in evaluating remote sensing snow-cover products over the entire continental United States on a daily basis, without invoking statistical interpolation between measurements. The method was applied to the IMS product and aggregated MODIS CMG product at 0.25° resolution for 126 days in winter, spring, and fall of calendar year 2000.

Detection rates were tested separately for snow-covered and snow-free areas. Testing against station measurements, IMS snow detection rates were low (<20%) early in the accumulation season (October), rising to over 60% in November, and as high as 95% in December; during March, snow detection rates were about 70%. IMS detected snow-free cells at a rate between 95% and 100% in March and October, but the no-snow detection rate declined to around 80% in December. Of grid cells that the aggregated MODIS product identified as having nonzero snow cover, station measurements confirmed from 75 to nearly 100% of in March and October, with that rate declining to 60%–80% in December; MODIS was more likely to overestimate than to underestimate or miss snow cover in

a grid cell. Cells that MODIS identified as snow free were confirmed by station data at nearly 100% throughout the year, with a small decline to 90%–95% in December. Using the aggregated MODIS estimates as “truth,” IMS showed better snow detection rates than when tested against station measurements: mostly 80%–90% in March, up to 40% in October, 75%–80% in November, and around 95% in December. All tests used a 26% level of significance.

Further research will focus on identifying times and locations where the satellite algorithms perform well and poorly; it is hoped that such analysis will be of value in improving the snow detection algorithms. Satellite products from past years are increasingly being used in hydrologic and atmospheric modeling; the global coverage and fine resolution of the MODIS snow-cover products make them ideal for evaluating other products and models. Data users need to know when and where the data are most reliable, and account for uncertainty when ingesting satellite information into models. Further analysis of our results and continued validation against independent measurements will help to quantify time- and space-varying error bars in archived and new remotely sensed snow-cover data.

Acknowledgments. The authors thank the Research Internships in Science and Engineering (RISE) summer 2002 team: scholars (M. Conlon, A. Johnson, M. Nguyen, and C. Shea) and Graduate Research Fellow (T. D. Kasabova) for their valuable assistance in framing the research questions and launching this study. We thank D. Hall, NASA Goddard Space Flight Center, for generous advice and assistance with the MODIS data, and B. Ramsay, NOAA/NESDIS, for providing the IMS data. We are grateful to R. Armstrong and two anonymous reviewers for their helpful critique. The National Science Foundation supported the authors and the RISE team under Grant DUE-0120786 to Women in Engineering at the University of Maryland. E. Deviatova received support under NASA Grant NAG513259.

REFERENCES

- Armstrong, R. L., and M. J. Brodzik, 2001: Recent Northern Hemisphere snow extent: A comparison of data derived from visible and microwave satellite sensors. *Geophys. Res. Lett.*, **28**, 3673–3676.
- , and —, 2002: Hemispheric-scale comparison and evaluation of passive-microwave snow algorithms. *Ann. Glaciol.*, **34**, 38–44.
- Bitner, D., T. Carroll, D. Cline, and P. Romanov, 2002: An assessment of the differences between three satellite snow cover mapping techniques. *Hydrol. Proc.*, **16**, 3723–3733.

- Blöschl, G., 1999: Scaling issues in snow hydrology. *Hydrol. Proc.*, **13**, 2149–2175.
- Brasnett, B., 1999: A global analysis of snow depth for numerical weather prediction. *J. Appl. Meteor.*, **38**, 726–740.
- Brown, R. D., 2000: Northern Hemisphere snow cover variability and change, 1915–1997. *J. Climate*, **13**, 2339–2355.
- , B. Brasnett, and D. Robinson, 2003: Gridded North American monthly snow depth and snow water equivalent for GCM evaluation. *Atmos.–Ocean*, **41**, 1–14.
- Cline, D., K. Elder, and R. Bales, 1998: Scale effects in a distributed snow water equivalence and snowmelt model for mountain basins. *Hydrol. Proc.*, **12** (10–11), 1527–1536.
- Doty, B., 2002: GrADS (Grid Analysis and Display System). Center for Ocean-Land-Atmosphere Studies (COLA), Institute of Global Environment and Society. [Available online at www.grads.iges.org/grads/.]
- Hall, D. K., and Coauthors, 2000: Intercomparison of satellite derived snow cover maps. *Ann. Glaciol.*, **31**, 369–376.
- , and Coauthors, 2001: Development of technique to assess snow cover mapping errors from space. *IEEE Trans. Geosci. Remote Sens.*, **39**, 432–438.
- , and Coauthors, 2002a: MODIS snow cover products. *Remote Sens. Environ.*, **83** (1–2), 181–194.
- , R. E. J. Kelly, G. A. Riggs, A. T. C. Chang, and J. L. Foster, 2002b: Assessment of the relative accuracy of hemispheric-scale snow-cover maps. *Ann. Glaciol.*, **34**, 24–30.
- Hayter, A. J., 2002: *Probability and Statistics for Engineers and Scientists*. Duxbury/Thomson Learning, 916 pp.
- Kelly, R. E., A. T. Chang, L. Tsang, and J. L. Foster, 2003: A prototype AMSR-E global snow area and snow depth algorithm. *IEEE Trans. Geosci. Remote Sens.*, **41**, 230–242.
- Klein, A. G., and A. C. Barnett, 2003: Validation of daily MODIS snow cover maps of the Upper Rio Grande River Basin for the 2000–2001 snow year. *Remote Sens. Environ.*, **86**, 162–176.
- Leavesley, G. H., R. W. Lichty, B. M. Troutman, and L. G. Saindon, 1983: Precipitation–runoff modeling system. User’s manual, U.S. Geological Survey Water Resources Investigation Rep. 83-4238, 207 pp.
- Liston, G. E., 1999: Interrelationships among snow distribution, snowmelt, and snow cover depletion: Implications for atmospheric, hydrologic, and ecologic modeling. *J. Appl. Meteor.*, **38**, 1474–1487.
- Martinez, J., A. Rango, and R. T. Roberts, 1992: User’s manual for the Snowmelt Runoff Model (SRM): Updated ed. 1992, version 3.2. USDA-ARS Hydrology Laboratory Tech. Rep. HL-17, 70 pp.
- NASA, cited 2004: MODIS Home Page. [Available online at www.modis.gsfc.nasa.gov/.]
- NCDC, cited 2004: NCDC: Get/View Online Climate Data. [Available online at www.ncdc.noaa.gov/oa/climate/climatedata.html.]
- NOHRSC, 2003: NOHRSC Technology. [Available online at www.nohrsc.nws.gov/technology/.]
- NRCS, cited 2004: NRCS National Water and Climate Center: SNOTEL Data Network—SNOTEL Data—All Sensors. [Available online at www.wcc.nrcs.usda.gov/snotel/.]
- NSIDC, cited 2004: National Snow and Ice Data Center (NSIDC). [Available online at www.nsidc.org.]
- Ramsay, B., 1998: The interactive multisensor snow and ice mapping system. *Hydrol. Proc.*, **12**, 1537–1546.
- Riggs, G. A., D. K. Hall, and V. V. Salomonson, 2003: MODIS Snow Products User Guide. [Available online at www.modis-snow-ice.gsfc.nasa.gov/sugkc2.html.]
- Robinson, D. A., K. F. Dewey, and R. R. Heim, 1993: Global snow cover monitoring: an update. *Bull. Amer. Meteor. Soc.*, **74**, 1689–1696.
- Roesch, A., M. Wild, H. Gilgen, and A. Ohmura, 2001: A new snow cover fraction parameterization for the ECHAM4 GCM. *Climate Dyn.*, **17**, 933–946.
- Romanov, P., and J. D. Tarpley, 2003: Automated monitoring of snow cover over South America using GOES Imager data. *Int. J. Remote Sens.*, **24**, 1119–1125.
- , and —, 2004: Estimation of snow depth over open prairie environments using GOES imager observations. *Hydrol. Proc.*, **18**, 1073–1087.
- , G. Gutman, and I. Csizsar, 2000: Automated Monitoring of Snow Cover over North America with Multispectral Satellite Data. *J. Appl. Meteor.*, **39**, 1866–1879.
- , —, and —, 2002: Satellite-derived snow cover maps for North America: Accuracy assessment. *Earth’s Atmos. Ocean Surface Studies*, **30** (11), 2455–2460.
- , D. Tarpley, G. Gutman, and T. Carroll, 2003: Mapping and monitoring of the snow cover fraction over North America. *J. Geophys. Res.*, **108**, 8619, doi:10.1029/2002JD0031428619.
- Scialdone, J., and A. Robock, 1987: Comparison of Northern Hemisphere snow cover data sets. *J. Climate Appl. Meteor.*, **26**, 53–68.
- Simic, A., R. Fernandes, R. Brown, P. Romanov, and W. Park, 2004: Validation of VEGETATION, MODIS, and GOES plus SSM/I snow-cover products over Canada based on surface snow depth observations. *Hydrol. Proc.*, **18** (6), 1089–1104.
- Simpson, J. J., and T. J. McIntire, 2001: A recurrent neural network classifier for improved retrievals of areal extent of snow cover. *IEEE Trans. Geosci. Remote Sens.*, **39**, 2135–2147.
- Tait, A. B., D. K. Hall, J. L. Foster, and R. L. Armstrong, 2000: Utilizing multiple datasets for snow-cover mapping. *Remote Sens. Environ.*, **72** (1), 111–126.