

New Geostationary Satellite–Based Snow-Cover Algorithm

NIILO SILJAMO AND OTTO HYVÄRINEN

Finnish Meteorological Institute, Helsinki, Finland

(Manuscript received 28 May 2010, in final form 7 December 2010)

ABSTRACT

Snow cover plays an important role in the climate system by changing the energy and mass transfer between the atmosphere and the surface. Reliable observations of the snow cover are difficult to obtain without satellites. This paper introduces a new algorithm for satellite-based snow-cover detection that is in operational use for Meteosat in the European Organisation for the Exploitation of Meteorological Satellites Satellite Application Facility on Land Surface Analysis (LSA SAF). The new version of the product is compared with the old version and the NOAA/National Environmental Satellite, Data, and Information Service Interactive Multisensor Snow and Ice Mapping System (IMS) snow-cover product. The new version of the LSA SAF snow-cover product improves the accuracy of snow detection and is comparable to the IMS product in cloud-free conditions.

1. Introduction

With the growing number of satellite platforms and improvements in the processing and transmission of digital data obtained from them, it has become possible to obtain frequent snow-cover information in near–real time through a variety of different sources. Retrieving snow products from satellite data is still a challenging task. Topography, heterogeneity in snow distribution, the effects of slope, aspect, land use, wind, and other factors in the accumulation and melting periods of snow make it difficult to retrieve snow products from satellite data.

At present, the most useful satellite orbits for snow detection at high latitudes are the near-polar orbit [used by, e.g., the National Oceanic and Atmospheric Administration, the Earth Observing System (EOS), Meteorological Operation (MetOp), the future National Polar-orbiting Operational Environmental Satellite System (NPOESS) Preparatory Project (NPP), and the future Joint Polar Satellite System (JPSS) satellites] and the geostationary orbit [e.g., Meteosat, Geostationary Operational Environmental Satellite (GOES), and Feng Yun 2 (FY-2)]. Both orbits have their strengths and weaknesses. In high latitudes, where snow is most often present, the instruments on board geostationary satellites have low viewing

angles, resulting in poor spatial resolution, whereas high spatial resolution is an advantage of polar-orbiting satellites. Instruments on board geostationary satellites also have a constrained view of the earth, while polar-orbiting satellites offer a global view. On the other hand, geostationary satellites offer excellent temporal resolution compared to polar-orbiting satellites. While polar-orbiting satellites can produce 1–4 daily images from a specified region in high latitudes, geostationary instruments produce images for a specified region every 15 min. Satellites in a highly elliptical Molniya orbit (Kidder and Vonder Haar 1990) would offer geostationary-like observations with a better viewing angle for high latitudes, but such satellites for meteorological use are only in the planning stages (e.g., Riisholgaard 2004).

This paper introduces a new geostationary snow-cover product for Spinning Enhanced Visible and Infrared Imager (SEVIRI) on board the second generation of Meteosat (MSG) satellites. SEVIRI has 11 channels in visual and IR areas of the radiomagnetic spectrum and represents the state of the art in geostationary imager instruments until the arrival of the next generation of geostationary satellites and their instruments [e.g., Advanced Baseline Imager on board GOES-R (Schmit et al. 2005)]. In practice, visual and IR channels can be used only to detect the existence of snow, not the depth or the water equivalence of snow. This existence of snow is usually disseminated in binary form for each pixel, but fractional subpixel information can also be obtained in

Corresponding author address: Finnish Meteorological Institute, Erik Palménin aukio 1, P.O. Box 503, FI-00101 Helsinki, Finland.
E-mail: niilo.siljamo@fmi.fi

some circumstances [e.g., Matikainen et al. (2002) or Metsämäki et al. (2005)]. Operational binary snow detection schemes based on visual and IR channels have been constructed for both polar and geostationary satellites. For polar-orbiting satellites there is a global snow analysis scheme based on Moderate Resolution Imaging Spectroradiometer (MODIS) on board EOS/Terra and EOS/Aqua (Miller et al. 2005). Even though snow is classified as part of many Advanced Very High Resolution Radiometer (AVHRR) processing schemes (e.g., Dybbroe et al. 2005), it seems that no snow analysis based on AVHRR data is in widespread use. Both AVHRR and MODIS will be superseded by the Visible/Infrared Imager/Radiometer Suite (VIIRS) on board the future NPP and JPSS satellites with similar channels for snow detection as MODIS (Miller et al. 2006).

For the geostationary instruments, there are systems for GOES, Meteosat, and FY-2. Romanov et al. (2000) used both geostationary GOES data and Special Sensor Microwave Imager (SSM/I) information from polar-orbiting Defense Meteorological Satellites Program (DMSP) satellites to construct the snow product. The systems using only GOES data for snow detection are presented by Romanov and Tarpley (2003) and Li et al. (2007). For Meteosat, de Wildt et al. (2007) used single SEVIRI images and temporal differences between images to produce snow-cover maps. Through the "GEONETCast" network (Moura 2006), the products obtained by the Visible and Infrared Spin Scan Radiometer (VISSR) of FY-2 satellite series (Dong and Zhang 2004), including the Snow Fraction Product, are disseminated to the global community, but to the authors' best knowledge, no detailed information about the algorithm has been published in English.

The major weakness of algorithms that use a combination of visual and IR channels for snow detection is that they can be used only during daytime and in cloud-free conditions. The high temporal resolution of the instruments on board a geostationary satellite helps to mitigate this to some extent, as it is much more likely that, for a certain area during one day, at least some of the images are cloud free. Some regions, however, can remain cloud covered for days. Active and passive microwave methods would be better suited for cloud-covered areas, but the spatial resolution of the passive microwave instruments [e.g., Advanced Microwave Scanning Radiometer for EOS (AMSR-E)] is poor when compared with optical channels, and passive microwave methods work only for dry snow conditions (Ulaby et al. 1986). Active microwave instruments, in practice synthetic aperture radar (SAR) (on board, e.g., Radarsat), have better spatial resolution, but unfortunately their swath width is comparatively narrow and daily observations of snow cannot be obtained (Koskinen et al. 1997). The use of SAR images in

operational setting has been hindered by their relative high cost and the difficulties of acquiring them in a timely manner. It is hoped that this will change in the future. In addition, no microwave instruments are currently on board geostationary satellites, but the possibilities of geostationary microwave instruments are being studied actively.

The best way to validate the satellite-derived snow cover would be to compare it with high-quality in situ measurements, but such data are almost impossible to collect on a large scale because of the serious limitations in the way the weather stations report the snow-cover measurements. The presence of snow is not always reported in many stations, and the absence of snow is not usually reported at all. Automation of the observations does not help, either. Commonly used automatic weather stations do not provide reliable observations for thin (less than 2.5 cm) snow layers. Moreover, in comparison to satellite-based snow analyses, the weather station network is also quite sparse. Hyvärinen et al. (2009) compared several snow analyses based on both in situ observations and satellite data, and the National Oceanic and Atmospheric Administration/National Environmental Satellite, Data, and Information Service (NOAA/NESDIS) Interactive Multisensor Snow and Ice Mapping System (IMS) analysis (Helfrich et al. 2007) gave a consistent snow analysis with high resolution. In the absence of a reliable reference, and because of its spatial coverage, the NOAA/NESDIS IMS analysis is a good candidate for a verification dataset. Moreover, IMS is a vital part of the European Centre for Medium-Range Weather Forecasts (ECMWF) snow analysis (Drusch et al. 2004), so there is strong motivation to compare the developed snow product with it.

The main sources for the NOAA/NESDIS IMS product are visible and infrared spectral data from different polar and geostationary satellites and, to a lesser extent, microwave products from polar-orbiting satellites. However, the IMS algorithm is not fully automatic. The combination of data from different sources into one analysis is done subjectively by analysts at NOAA/NESDIS.

In this article, first the two versions of the Satellite Application Facility (SAF) on Land Surface Analysis (LSA) snow-cover (SC) algorithm are introduced. Then the NOAA/NESDIS IMS product is used as a baseline to which both LSA SC products are compared in Europe.

2. LSA SAF snow algorithm

The European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) has several dedicated programs for processing satellite data. Each of these SAFs provides products and services on an operational basis. The main purpose of the LSA SAF is to increase the benefits from MSG and EUMETSAT Polar

System (EPS) data concerning land, land–atmosphere interactions, and biophysical applications by developing techniques, products, and algorithms that will allow data from the EUMETSAT satellites to be used more effectively. A major user of LSA SAF products is numerical weather prediction (NWP), both for validation and for use as a model input.

EUMETSAT's LSA SAF has been producing the daily snow-cover product with a baseline algorithm for the areas covered by the MSG/SEVIRI instrument since 2005. The snow detection algorithm produces the snow-cover product over the MSG/SEVIRI image area, which is divided into four different geographical regions (Europe, North Africa, South Africa, and South America), as is the case for all LSA SAF products. A product using polar-orbiting MetOp satellite data is currently under development. The daily products are distributed to users through the EUMETCast system. The daily products are archived and are available on the LSA SAF Web site (<http://landsaf.meteo.pt>). The single-image snow-cover product is currently not available for public. At present, the daily products are not reprocessed when the algorithms are changed. Thus, potential users should be aware of the differences in the product versions.

The first version of the LSA SAF snow-cover algorithm was developed by the Swedish Meteorological and Hydrological Institute (SMHI). Development of the snow-cover product was transferred to the Finnish Meteorological Institute (FMI) in 2005.

The version 1 of the snow-cover algorithm determined whether a satellite ground pixel is fully or partially snow covered, through the use of different signatures of snow, ice, and clouds on the reflectance of shortwave IR channels (SEVIRI 1.6 and 3.9 μm). Since the separation between cloud and surface ice–snow is a crucial procedure for cloud detection, the above-mentioned thresholding procedure has been included in the SAF on Support to Nowcasting and Very Short Range Forecasting (NWC SAF) software (Derrien and LeGléau 2005). The cloud-mask product was analyzed and the pixels classified every 15 min. Cloudy pixels were reclassified as unclassified, cloud-free pixels as snow free, and snow-contaminated pixels as snow covered. Thus, version 1 of the SEVIRI snow-cover algorithm relied on the cloud mask, derived from the NWC SAF package, for a first pixel classification, before performing snow spatial smoothing and temporal integration of the previous 24-h satellite scenes. Unfortunately, the accuracy of the snow detection was limited in the version of the NWC SAF software used at that time. Only one-third of cloud-free synoptic surface observation stations (SYNOP; see World Meteorological Organization 1995) stations with snow were correctly classified as snow (Derrien and LeGléau 2005).

A new version 2 of the LSA SAF snow-cover product was therefore developed at the FMI. This version was developed specifically for snow detection; that is, it is not a by-product of a cloud-mask product. Different variants of this algorithm version have been used to generate the LSA SAF snow-cover product pre-operationally since summer 2007 and operationally since April 2009. The differences between version 2 algorithm variants are minor or related to changes in the input files.

The mandate of LSA SAF is to produce the snow product for non-mountainous areas. However, mountain areas are processed in the LSA SAF system and the data are available in the product files, but there are no requirements for the product quality and therefore mountains are excluded in this study.

a. Algorithm development

The spectral properties of snow observed in nature vary considerably (Dozier et al. 2009; Salminen et al. 2009; Wiscombe and Warren 1980a,b). The grain size of snow changes over time and space, the wetness of snow changes, and the reflecting properties change when the surface is viewed from different angles and in different solar illumination conditions. In addition, the effects of vegetation—such as grass, fields, and different forests—are highly variable, even in winter. This natural variability makes it difficult to develop a general classification algorithm for the snow cover. Finally, the atmosphere should be taken into account when surface and laboratory measurements are compared to satellite measurements. Thus most operational algorithms are at least partly empirical, based on the statistics of samples collected from different surface types.

Development of the new version of the snow-cover classification algorithm was started by subjective classification of selected areas in representative MSG/SEVIRI images. Starting from 32 images for 12 days ranging from November 2006 to September 2007, but concentrating on 28 March 2007 (13 images), samples of snow-covered and snow-free areas, different cloud types, and areas where the surface type could be seen through clouds were selected. The day 28 March 2007 was an exceptionally cloud-free day in northern Europe, which enabled the collection of data from early morning to late evening. Particular attention was paid to finding days when there was snow in southern Europe. Although the snow-cover product is calculated for African and South American regions, they are mostly snow-free and data samples from those areas were not sought for statistical analysis. However, the product seems to be reasonably reliable also in these areas.

The actual extent of snow cover was determined subjectively using different SEVIRI red–green–blue (RGB) combinations, ground observations, and MODIS images.

Over one-half million MSG/SEVIRI pixels were classified to form a dataset for algorithm development. However, the samples were collected in large batches, and because the snow properties in most pixels were spatially correlated with those of their neighbors, the effective number of independent pixels is much smaller. Even so, this increases variations in the surface properties and in the satellite and sun angles.

Based on the dataset that was collected, the differences in various classes were investigated using the different channels. The classification of different surface types using information from channels around 0.6, 0.8, and 1.6 μm is the basis for all snow and cloud classification schemes (Hall et al. 2002; Dybbroe et al. 2005; Derrien and LeGléau 2005). As to discriminating between surface and clouds, the capabilities of channels around 1.6 and 3.9 μm to discriminate low clouds and snow have been reported widely (Matson 1991; Kidder and Wu 1984). Thus, using the radiance ratio of SEVIRI channels 2 (0.8 μm) and 3 (1.6 μm) and the brightness temperature difference of channels 10 (12.0 μm) and 4 (3.9 μm) (Fig. 1a) clouds can be distinguished from a cloud-free surface. Moreover, the radiance ratio of SEVIRI channels 2 and 3 and the sun azimuth angle (Fig. 1b) difference between snow and snow free seem to be promising. These ratios are very similar to the normalized difference snow index (NDSI) used by de Wildt et al. (2007).

The driving philosophy behind the LSA SAF snow-cover algorithm is to avoid mistakes, even if this approach produces relatively more unclassified pixels. The LSA SAF snow-cover algorithm does not try to identify cloud-free regions for classification before the snow-cover tests. Instead, it has rules for presence of snow and snow-free surface, and other pixels are left as unclassified whether they are cloudy, too dark, or just too difficult to classify reliably. Rules themselves are based on channel differences and channel ratios that are very similar to other published algorithms, as is shown above.

b. Algorithm implementation

The LSA SAF snow-cover algorithm is a thresholding method based on the different properties of the snow-covered and snow-free surfaces and clouds. The LSA SAF snow cover is a daily product, produced in two separate phases. Phase 1 is the SC1 snow-cover product based on one cycle of SEVIRI images (every 15 min). All of the available SC1 products are used to produce the daily LSA SAF snow-cover product (SC2). This phase produces 96 snow-cover maps each day if all time slots are available. Product resolution is full SEVIRI resolution in each of the four regions used in LSA SAF.

In this study, the mountain definition of the SAF on Support to Operational Hydrology and Water Management

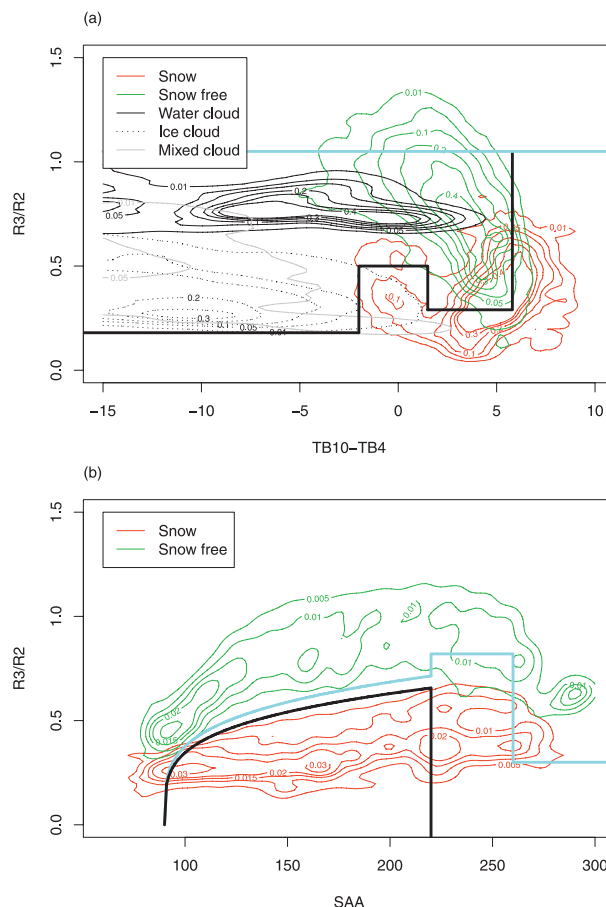


FIG. 1. Examples of the densities of the scatterplots from the development dataset. The thick lines show thresholds for SNOW (black) and NO SNOW (blue), based on the rules in Table 2: (a) SNOW: rules (R9)–(R12) and NO SNOW: rule (R14); (b) SNOW: rule (R5) and NO SNOW: rules (R6)–(R8). For definitions of the variables, see Table 1.

(H-SAF) was followed, because the similar mountain snow product is part of the H-SAF. The area was defined to be mountainous if the mean altitude exceeded 2000 m or the mean altitude exceeded 700 m and the standard deviation of the slope was greater than 2° in a 10 km \times 10 km area (Lahtinen et al. 2009). This area is shown in Fig. 2a.

The algorithm utilizes the top-of-atmosphere radiances of 6 SEVIRI channels (0.6, 0.8, 1.6, 3.9, 10.8, and 12.0 μm) and brightness temperatures of three channels (3.9, 10.8, and 12.0 μm), sun and satellite zenith and azimuth angles, the International Geosphere-Biosphere Programme (IGBP) land-cover type by the U.S. Geological Survey (USGS), and the land surface temperature (LST) classification produced by the LSA SAF (Table 1). Our plan is to start using the Global Land Cover 2000 Project (GLC2000) (Bartholomé and Belward 2005) land-cover data in the future versions of the algorithm.

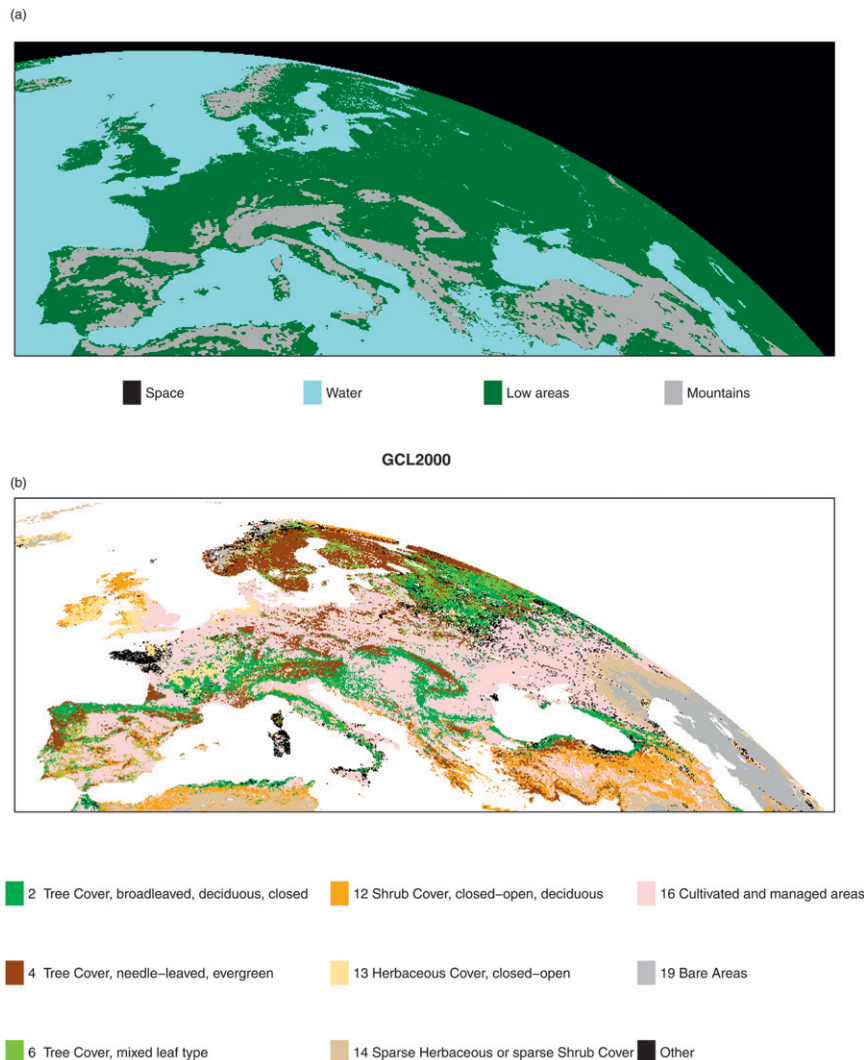


FIG. 2. (a) The height of the terrain in the study area. The snow product is only valid in the low areas. (b) The land-cover classification for the study area according to the GLC2000 Project. Only land-cover classes covering more than 5% of the area under study are shown. The class numbers are according to the original GLC2000 classes.

It is quite straightforward to estimate whether the surface is totally snow free or totally snow covered, but it is difficult to define when to change from a snow-free to a snow-covered surface. Even a thin layer of snow should be defined as snow covered, because it changes the albedo and radiative properties of the surface. This is quite often the case when there is a new layer of snow. During the melting season, snow cover can be very variable; the surface is a patchwork of snow-covered and snow-free areas of different shapes and sizes. Trees are also a challenge for definition, because the trees can be snow covered or snow free even if the surface is fully covered.

Most of the snow-covered, snow-free, and cloud-covered pixels are fairly easy to detect. Some of the rules are used to

classify these pixels. The difficulties occur near the edges of snow and clouds, and also in poorly lit areas. To reduce errors, classification is avoided when there is uncertainty about the surface status. Near the clouds, this means that the unclassified pixel type is preferred. Near the edge between snow-covered and snow-free surfaces some rules will find partially snow-covered pixels although most of these are still classified as snow covered or in some cases left unclassified.

In the beginning of the snow detection, the algorithm sets all pixels as unclassified. Then the tests of Table 2 are applied one by one in the order shown. Finally, each pixel is classified to one of the snow-cover (or snow free) classes or it remains unclassified. Some of the rules are

TABLE 1. List of the inputs for the LSA SAF SC1 snow-cover algorithm.

Input type	Code	Description
Satellite channels	R_1	Radiance in channel 1 ($0.6 \mu\text{m}$)
	R_2	Radiance in channel 2 ($0.8 \mu\text{m}$)
	R_3	Radiance in channel 3 ($1.6 \mu\text{m}$)
	R_4	Radiance in channel 4 ($3.9 \mu\text{m}$)
	R_9	Radiance in channel 9 ($10.8 \mu\text{m}$)
	R_{10}	Radiance in channel 10 ($12.0 \mu\text{m}$)
	T_{B4}	Brightness temperature in channel 4 ($3.9 \mu\text{m}$)
	T_{B9}	Brightness temperature in channel 9 ($10.8 \mu\text{m}$)
	T_{B10}	Brightness temperature in channel 10 ($12.0 \mu\text{m}$)
Angles	SAA	Sun azimuth angle
	SZA	Sun zenith angle
	VAA	Satellite azimuth angle
	VZA	Satellite zenith angle
Other	LC	Land-cover type
	LST	LSA SAF land surface temperature ($^{\circ}\text{C}$)

overlapping, and many rules may apply to the same pixel. The relative importance of these rules is currently being analyzed and the decision-making process improved so that the number of misclassified pixels can be reduced and, possibly, the detection of partially snow-covered pixels, which are often near the edge of the snow-covered area, can be improved.

Cloud cover is the most common reason for not classifying a pixel. A pixel is also unclassified if it is too dark or in an area where the satellite elevation angle is too low [i.e., rules (R15)–(R17)]. There are also rules that remove obvious misclassifications, such as pixels where the land surface is too warm to contain snow [rules (R18), (R19) and (R21)]. These rules must come after all other rules.

Most of the cloudy and cloud-free areas can be separated using the brightness temperature difference between channels 10 and 4 and the radiance ratio of channels 3 and 2, although there is some overlap (Fig. 1a). As opposed to cloud detection, the aim is to find cloud-free pixels and thus to avoid such erroneous classifications where cloudy and possibly cloud-contaminated pixels are defined as snow covered. Classification rules (R1)–(R3) and (R9)–(R12) are based on this figure. Rule (R4) is used to find cloudy areas based on a 3D view of three radiance ratios of channels 1 to 3.

Classification rules (R5)–(R8) are based on the information from Fig. 1b. The radiance ratio of channels 3 and 2 against the sun azimuth angle differentiates between snow and snow-free pixels very well. The reason for this is the strong relation of the reflective properties of the surface to the bidirectional reflectance distribution function (BRDF), which is different for snow (Peltoniemi et al. 2005a) and different vegetation types (Peltoniemi et al. 2005b). Currently, these rules are defined conservatively to avoid misclassifications, but new rules to classify currently unclassified pixels are being investigated.

TABLE 2. List of classification rules of the LSA SAF SC1 snow-cover algorithm. These rules are applied one by one. If the condition is true, the snow-cover status is set to the defined value. The final snow-cover status is the value set after all the rules have been checked. Here, $\Delta T_B = T_{B10} - T_{B4}$; for other definitions see Table 1. Logical AND is marked by \wedge , and logical OR is marked by \vee .

$\Delta T_B \geq 0 \wedge R_3/R_2 < 0.6 \Rightarrow$	PARTIAL	(R1)
$\Delta T_B \geq 2.5 \Rightarrow$	PARTIAL	(R2)
$\Delta T_B \leq -2.5 \wedge R_3/R_2 < 0.90 \Rightarrow$	UNCLASS	(R3)
$R_3/R_2 < 0.96 \wedge R_3/R_2 \geq 0.62 \wedge R_3/R_1 < 1.22 \wedge R_3/R_1 \geq 0.77$		(R4)
$\wedge R_2/R_1 < 1.49 \wedge R_2/R_1 \geq 1.15 \Rightarrow$	UNCLASS	
$\Delta T_B \geq 1.5 \wedge \text{SAA} < 220 \wedge \text{SAA} > 700(R_3/R_2)^4 + 90 \Rightarrow$	SNOW	(R5)
$\Delta T_B \geq 1.5 \wedge \text{SAA} < 220 \wedge \text{SAA} < 500(R_3/R_2)^4 + 90 \wedge \text{SAA} > 5.0 \Rightarrow$	NO SNOW	(R6)
$\Delta T_B \geq 1.5 \wedge \text{SAA} < 220 \wedge R_3/R_2 \geq 0.82 \Rightarrow$	NO SNOW	(R7)
$\Delta T_B \geq 1.5 \wedge \text{SAA} \geq 260 \wedge R_3/R_2 \geq 0.30 \Rightarrow$	NO SNOW	(R8)
$R_3/R_2 < 0.18 \Rightarrow$	SNOW	(R9)
$\Delta T_B \geq -2.0 \wedge \Delta T_B \leq 1.5 \wedge R_3/R_2 < 0.5 \Rightarrow$	SNOW	(R10)
$\Delta T_B \geq -2.0 \wedge \Delta T_B \leq 20.0 \wedge R_3/R_2 < 0.290 \Rightarrow$	SNOW	(R11)
$\Delta T_B \geq 5.8 \Rightarrow$	SNOW	(R12)
$R_3/R_1 \geq 1.50 \wedge \Delta T_B > -25 \Rightarrow$	NO SNOW	(R13)
$R_3/R_2 \geq 1.05 \wedge \Delta T_B > -15 \Rightarrow$	NO SNOW	(R14)
$\text{SZA} > 80.0 \Rightarrow$	UNCLASS	(R15)
$\text{VZA} > 85.0 \Rightarrow$	UNCLASS	(R16)
$\text{SZA} > 70.0 \wedge (\text{SAA} < 90.0 \vee \text{SAA} > 270.0) \Rightarrow$	UNCLASS	(R17)
$(T_{B9} + T_{B10})/2 \geq 278.0 \wedge (\text{SNOW} \vee \text{PARTIAL}) \wedge (\text{LC is not forest}) \Rightarrow$	NO SNOW	(R18)
Date between June and October $\wedge (T_{B9} + T_{B10})/2 \geq 278.0 \wedge$		
$(\text{SNOW} \vee \text{PARTIAL}) \wedge (\text{LC is forest}) \Rightarrow$	NO SNOW	(R19)
Any one of R_1, R_2, R_3, R_4, R_9 , or $R_{10} < 0.001 \Rightarrow$	UNCLASS	(R20)
$\text{LST} \geq 3.0 \Rightarrow$	NO SNOW	(R21)

TABLE 3. List of the rules for the daily product: N is the total number of classified observations during the day for each pixel; S , P , and F are the numbers of snow-covered, partially snow-covered, and snow-free observations, respectively. These rules are used one after the other from the top, and the final daily classification is the classification in effect after the last rule. Logical AND is marked by \wedge , and logical OR is marked by \vee .

Set default value \Rightarrow	UNCLASS	(D1)
$S > N/4 \wedge S > 5 \wedge F < 3 \Rightarrow$	SNOW	(D2)
$F > N/3 \wedge F > 3 \Rightarrow$	NOSNOW	(D3)
$P > N/3 \wedge P > 3 \wedge F = 0 \wedge$ $S > 1 \wedge S \leq 4 \Rightarrow$	PARTIAL	(D4)
$P > N/3 \wedge P > 3 \wedge F > 1 \wedge F \leq 6 \wedge$ $S > 1 \wedge S \leq 6 \Rightarrow$	PARTIAL	(D5)
$P > N/3 \wedge P > 3 \wedge F = 0 \wedge S > 4 \Rightarrow$	SNOW	(D6)
$P > N/3 \wedge P > 3 \wedge F > 0 \wedge S = 0 \Rightarrow$	NOSNOW	(D7)

Once per day, the daily snow-cover (SC2) product is calculated using the SC1 products for the day. Again the system classifies each pixel as snow free, partially snow covered, or totally snow covered. For the daily LSA SAF snow-cover product, all of the snow-cover maps produced every 15 min are combined. The algorithm counts the number of different classifications for each pixel and then determines the final daily classification if there have been reasonable amount of cloud-free observations during the day. The actual rules are presented in Table 3.

The class of partial snow is used if the pixel is classified as snow free and snow covered during the same day or if it is probable that snow does not cover the whole pixel. This class is not yet well defined, because only a very limited number of reliable surface observations could be used to estimate the accuracy of this classification. In the future, this class may be replaced by the snow-covered area (SCA) product.

The last phase of the product generation is quality flagging. Surface type such as forests, solar illumination conditions such as night, sun glint, high terrain, and quality flags are set. Currently all classified pixels (i.e., snow, partial snow, or snow free) are set as high quality, but this will change when an improved quality flagging will be introduced.

3. Validation data and methods

a. Use of IMS as the reference analysis

The NOAA/NESDIS IMS analyses are available as gridded data in American Standard Code for Information Interchange (ASCII) format on the Internet. The higher-resolution version of 4 km (6144×6144 grid) was used in this study. IMS products are disseminated in a polar stereographic projection; they were reprojected to same projection as the LSA SAF SC. This projection is not an area-preserving projection, and pixels correspond to

TABLE 4. Contingency table of the comparison between two categorical snow analyses. The symbols a – d represent the different number of pixels observed to occur in each category.

Analysis 1	Analysis 2 (baseline)	
	Snow	No snow
Snow	a (hit)	b (false alarm)
No snow	c (miss)	d (correct rejection)

areas of different sizes. However, as it is the projection in which the LSA SAF SC is disseminated, it was the natural projection for the comparison.

b. Validation measures

The results of a comparison between two products can be shown in a 2×2 contingency table (Table 4). Hits a is the number of cases in which both analyses reported snow, correct rejections d is the number of cases in which neither of the analyses reported snow, false alarms b is the number of cases in which only the LSA SAF SC, the analysis under investigation, reported snow, and misses c is the number of cases in which only the baseline analysis, IMS, reported snow. The measures used are summarized in the appendix.

In this study, attention should be paid to the number of correct rejections. In Europe, almost all snow melts in summer, as is shown below. Then the number of correct rejections, d , is several magnitudes greater than a , b , or c and proportion correct (PC) tends to 1. After some manipulation, it is easily shown for Heidke skill score (HSS) that

$$\lim_{d \rightarrow \infty} \text{HSS} = \frac{2a}{2a + b + c}, \quad (1)$$

which is the critical success index (CSI) where a double weight has been given to hits a [in, e.g., the clustering community where the CSI is known as the Jaccard index or coefficient, this is known as the Sorensen's similarity coefficient (Kaufman and Rousseeuw 1990)]. Thus, in this study, HSS and the CSI are closely related, and the decision was made to concentrate on the HSS. Further, if b is less than a and c , HSS and CSI obtain values similar to hit rate (H).

c. Validation strategy

Because of clouds or inadequate solar illumination, not all pixels can be classified by LSA SAF SC, and the number of pixels classified varied from day to day. No attempt was made to mitigate this; LSA SAF SC was compared to IMS only for pixels classified by LSA SAF SC.

The data of this study can be thought of as a three-dimensional grid consisting of two-dimensional maps and time. Each data point has one of the four possible values

shown in Table 4. First, the data points were merged for all dimensions, resulting in one contingency table. Second, the values on each map were merged, resulting in one contingency table for each map; the results can be shown as time series.

Third, the values in the time dimension were merged, producing a map where each pixel has a contingency table of its own. This makes it possible to assess the spatial performance of the algorithm. It is reasonable to suppose that snow behaves differently over different terrain types. Whereas the IGBP land cover was used for the development and the operational run of the algorithm, the GLC2000 land cover was used for the validation work (Fig. 2b). The GLC2000 was chosen because, first, it was more up to date than the IGBP and, second, it was an independent data source, which is useful in validation.

The effect of land cover was investigated subjectively first by examining the maps showing the distribution of measures and then objectively, in quantitative fashion. Using a verification measure (e.g., PC or HSS), it was calculated whether certain areas (e.g., needle-leaved evergreen forest) were more probable to have values higher or lower than the median value. The median, instead of the mean, was used, because the distribution of values is far from Gaussian. This ratio of probabilities can be approximated by the ratio of pixels with values higher and lower than the median value \tilde{X} of the whole map:

$$\frac{P(X > \tilde{X})}{P(X < \tilde{X})} = \frac{\frac{\text{No. pixels} > \tilde{X}}{\text{No. pixels}}}{\frac{\text{No. pixels} < \tilde{X}}{\text{No. pixels}}} = \frac{\text{No. pixels} > \tilde{X}}{\text{No. pixels} < \tilde{X}}. \quad (2)$$

Values around the unity mean that there is no notable difference in the ratio of high and low values, and the verification measure is not dependent on the characteristics of the area. A ratio less (greater) than 1 suggests that the lower (higher) values are more likely for the area. Confidence intervals were computed with bootstrapping using the method outlined in Hamill (1999). This method assumes that there is a high correlation between spatial observations (grid points) but no correlation between days (grids). However, our data may not be in line with the latter assumption, and so the true confidence intervals may be somewhat greater.

4. Validation results

The algorithm was tested using data from Europe for the period of January 2007 to December 2009. From January 2007 to July 2007, version 1 was operational, and after that, version 2, so operational version 2 processed more than 4 times as many images as operational version

1. For the comparison between algorithms to be meaningful some days should be processed with both algorithms. So, in spring 2007, 25 days were processed with version 2, but some of these images were used for development of the algorithm and, all in all, there are 20 days that were not used for the algorithm development and that were processed by both versions of the software.

So that the results concerning the differences between algorithm versions would be less ambiguous, it would be necessary to reprocess all of the days from January 2007 to July 2007 with version 2, but unfortunately this would require substantial work beyond our current resources.

a. Visual inspection of products for one day

For subjective evaluation, false-color RGB combinations provide a useful tool. The image is constructed from three grayscale images of satellite channels, each with different characteristics, and the colors can be given a physical interpretation. Different false-color RGB combinations can be constructed that emphasize different phenomena. Here the main aim was to emphasize snow and the RGB combination of SEVIRI channels VIS0.6, IR1.6, and inverted IR10.8 was used. This combination clearly separates snow from low clouds, usually made of water droplets, but is not as good in separating snow from high clouds that are usually made of ice particles. The colors in the image have a physical interpretation: the snow-free surface reflects better in IR1.6 (the green component) than in VIS0.6 (the red component) and is relatively warm (the blue component is small, because values are inverted), and therefore it shows greenish in the RGB image. Relatively warm low clouds made of water droplets reflect well both in VIS0.6 and IR1.6, which make their color much lighter yellowish; colder water clouds more white-yellowish. Snow on surface reflects well in VIS0.6, but very little in IR1.6, and are rather cold (the blue component is larger), making snow purple. Ice clouds are of lighter shades of purple because their reflectance in VIS0.6 and IR1.6 is similar to snow on the ground, while they are usually much colder than the surface (the blue component is larger).

The RGB composite of SEVIRI images at 1200 UTC 26 January 2007 and LSA SAF SC (version 2) and IMS products for the same day are shown in Fig. 3. The day was relatively cloud free, with snow even in southern Europe. The poor illumination of northern regions can also be seen. The IMS snow-cover analysis (Fig. 3b) classifies all of the pixels, but several snow edges seem too smooth and unphysical (e.g., in the area north of the Black Sea, 1 in Fig. 3b). The LSA SAF snow-cover algorithm (Fig. 3c) has used all of the images available during the day and thus covers more surface than the 1200 UTC image (Fig. 3a). The area left unclassified, mostly because of clouds, is

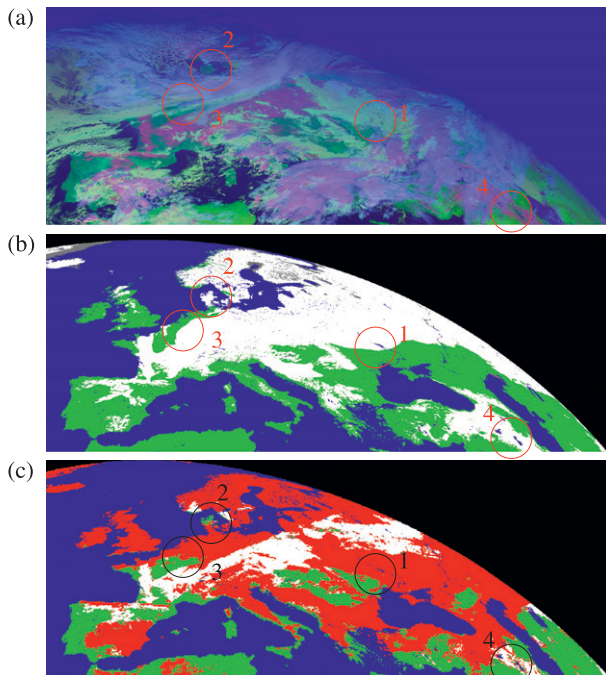


FIG. 3. (a) The RGB combination of SEVIRI channels VIS0.6, IR1.6, and IR10.8, (b) the NOAA/NESDIS IMS snow product, and (c) the LSA SAF snow-cover product for 26 Jan 2007. In both satellite products, snow is white and snow-free areas are green, sea areas are blue, and unclassified areas are red. In the RGB combination, snow is purple, snow-free areas are green, sea areas are dark blue, water clouds are yellow, and ice clouds range from light purple to light blue. The numbers in (b) refer to the discussion in the text.

quite large, but the snow analysis in cloud-free areas shows some interesting differences between the products. Denmark is snow covered in the IMS but snow free in LSA SAF SC, and the satellite image suggests the latter (2 in Fig. 3b). The edge of the snow-covered area in France (3 in Fig. 3b) is different in the products and LSA SAF SC

again agrees better with the satellite image. The third example can be seen in southern Turkey (4 in Fig. 3b), where the IMS shows a large snow-covered area. In the LSA SAF SC, the snow-covered area is much smaller and much more patchy, which seems to be in good agreement with satellite images, although this area is classified as mountainous region and is thus outside the scope of LSA SAF SC.

b. Results merged to one contingency table

All data of different versions of LSA SAF SC were merged, resulting in one contingency table for version 1, test cases, and an operational run of version 2 (Table 5). From the IMS product it can be calculated that there are about half million land surface pixels that can be classified. Both version 1 and version 2 leave a substantial amount of pixels unclassified, about 25% and 40%, respectively, because of clouds or inadequate solar illumination. In addition, some pixels, mainly at the cloud and snow edges, are not classified because the pixels do not match any of the rules used. Version 2 is more conservative as to what pixels to classify than version 1, as version 2 avoids classification in conditions where misclassifications are probable.

In Europe, most of surface is snow free most of the time, and thus correct rejections occurred for about 90% of the pixels in all of the images. In spring 2007, version 2 had less misses than version 1, which yielded more hits in version 2 than in version 1, but the differences in false alarms and in correct rejections were negligible. For version 2, the differences between false alarms and misses were also rather small in spring 2007 and in the operational run, but there were more hits and fewer correct rejections in spring 2007 because the operational run includes summers when very little snow was present.

In spring 2007, all measures had better values for the test cases of version 2 than for version 1. The operational

TABLE 5. Aggregated results from January 2007 to December 2009. From January 2007 to July 2007, version 1 was operational but some days were run with version 2. After July 2007, version 2 was run operationally.

	Spring 2007		Summer 2007–09
	Version 1	Version 2 (test cases)	Version 2 (operational)
Hit <i>a</i>	2202274 (4.4%)	564022 (8.5%)	6898843 (3.9%)
False alarm <i>b</i>	344737 (0.7%)	29952 (0.5%)	686785 (0.4%)
Miss <i>c</i>	2546168 (5.1%)	72760 (1.1%)	1553271 (0.9%)
Correct rejection <i>d</i>	45116671 (89.9%)	5976827 (90.0%)	169307675 (94.9%)
Bias	0.536	0.933	0.897
<i>F</i>	0.008	0.005	0.004
<i>H</i>	0.464	0.886	0.816
FAR	0.135	0.050	0.091
PC	0.942	0.985	0.987
HSS	0.576	0.908	0.854
Images	190	42	842
Unclassified pixels	24%	54%	39%

version 2 had slightly lower values than test cases, but this is understandable, as test cases covered only a limited time span and some days were used in development of the algorithm; this means that some overfitting may have occurred.

All versions detected less snow than IMS as the bias was less than 1 in all versions and was the least in version 1 (bias ≈ 0.5 in version 1 and bias ≈ 0.9 in version 2). When IMS detected snow, it was detected by version 1 only half of the time ($H \approx 0.5$), while version 2 detected 80% ($H \approx 0.8$). But when LSA SAF SC detected snow, usually it was also detected by IMS, only around 10% of this snow was not detected by IMS [in both versions, the false-alarm ratio (FAR) ≈ 0.1], and this amounted to less than 1% of snow-free areas in IMS [in both versions, the false-alarm rate (F) < 0.01].

Using PC as a measure, both version 2 and version 1 agree well with IMS (PC > 0.9). High PC values are derived from correct rejections, which dominate the other components and PC can be considered somewhat overoptimistic in depicting the skill of LSA SAF SC. Skill scores yield more realistic results. And indeed, the difference between IMS and LSA SAF SC is more pronounced in HSS, where HSS ≈ 0.6 for version 1 and HSS ≈ 0.9 for version 2.

c. Results for days when both versions were available

In Table 5, the results for version 2 in spring 2007 included days that were used for the algorithm development. Here the focus is on 20 days that were not used to develop the algorithm, but for which both version 1 and version 2 were run. During these days, three different areas can be distinguished: an area where both algorithms were run (BOTH), an area where only version 1 was run (OLDONLY), and an area where only version 2 was run (NEWONLY). However, the amount of pixels in NEWONLY is only about 3% of pixels in BOTH and was not investigated further. But the amount of pixels in OLDONLY is still about 40% of the pixels in BOTH and meaningful comparisons can be made. For these days a contingency table was calculated, from which further measures can be calculated.

In the first test, the daily results of version 1 in BOTH were compared to those of version 1 in OLDONLY. The values for different measures (H , FAR, PC, and HSS, only PC, and HSS are shown) of version 1 in BOTH were subtracted from the values of version 1 in OLDONLY (Fig. 4a). If results of different areas are not significantly different, the differences should be around zero. But this is not the case, version 1 yielded better results in BOTH than in OLDONLY. This can be interpreted so that OLDONLY is harder to classify than BOTH and version 2 has made a good choice in refusing to classify that area.

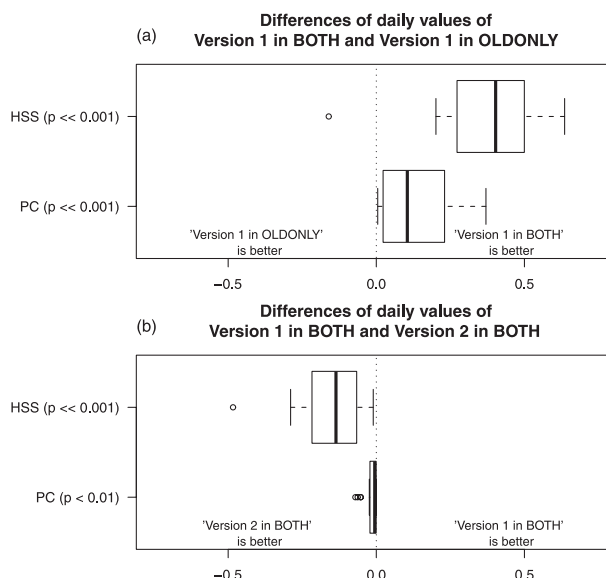


FIG. 4. Twenty-day run of version 1 and version 2. Daily results (using PC and HSS) of version 1 in the area where both version 1 and version 2 were calculated (version 1 in BOTH) are compared with (a) version 1 in the area where version 2 was not calculated (version 1 in OLDONLY) and (b) version 2 for the common area (version 2 in BOTH). The differences are shown as box-and-whisker plots. (The median is shown as a thick line. The hinges show the first and third quartile. The whiskers extend to a data point no more than 1.5 times the interquartile range (IQR) from the median, while data points still farther from the median are plotted.) The p value is calculated for a two-sided t test that the mean of differences is equal to zero.

In the second test, version 1 in BOTH was compared with version 2 in BOTH (Fig. 4b). In this case, version 2 yielded better values than version 1. So even if version 1 had not classified difficult areas of OLDONLY, it would still give inferior results to version 2.

The results were statistically significant at the 95% level with the exception of the second test for FAR, but even then results were significant at the 90% level.

d. Seasonal variability of results

The results as time series, when a contingency table is calculated from each image, give a detailed view of the seasonal variability of snow. The amount of snow varies, as the dataset includes three winters and three summers. The maximum extent of snow is in February (Fig. 5a); the snow starts to melt in mid-March and has melted almost completely by the end of May. The remaining snow is found on mountaintops and glaciers, which are mostly outside the scope of LSA SAF SC. New snow starts to accumulate in November. The surface area that is classified varies both because of the varying cloud cover and also with the season, as during the winter the zenith angle can be too high to enable classification for all day in

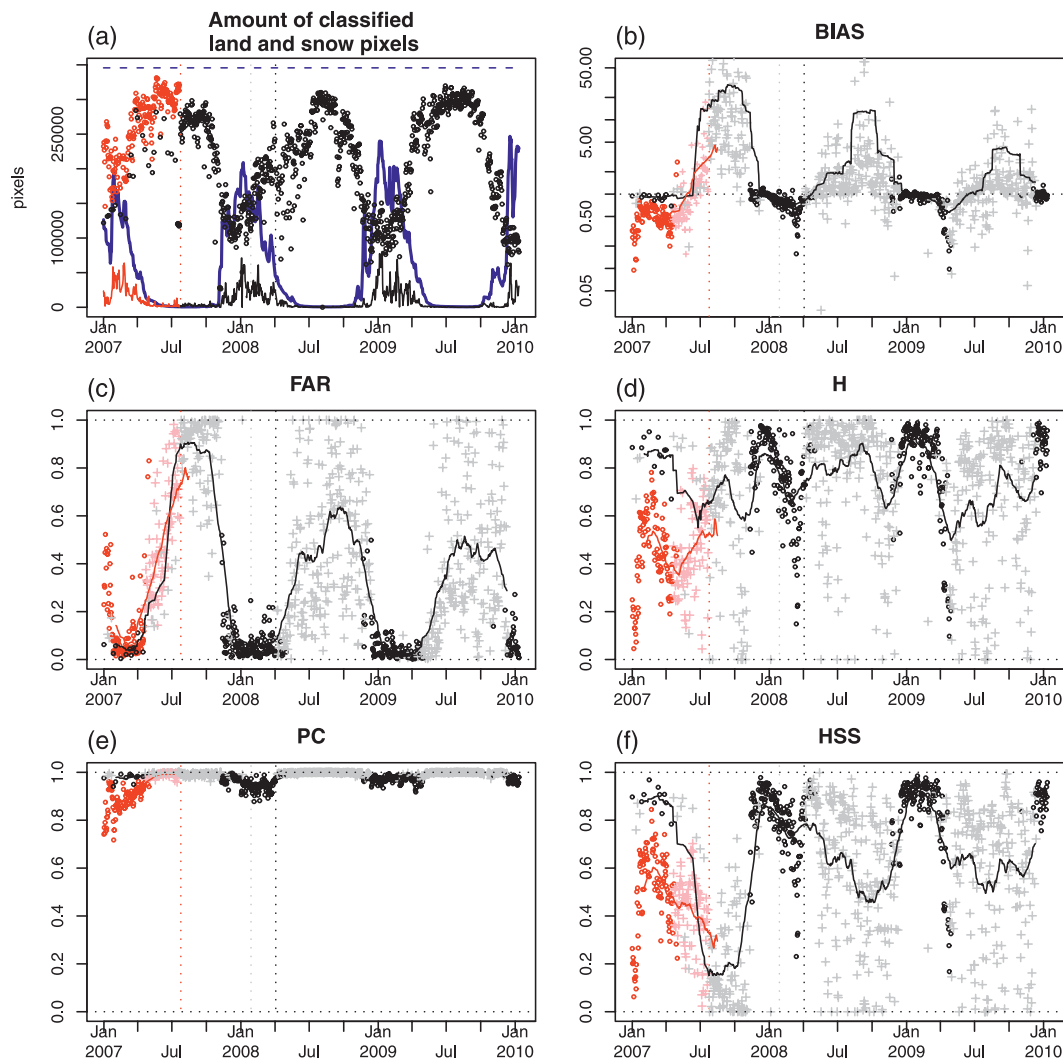


FIG. 5. (a) Amount of cloud-free land pixels (blue dashed line, IMS; red circle, version 1; black circle, version 2), and amount of snow-covered pixels (blue line, IMS; red line, version 1; black line, version 2). (b) Bias, (c) FAR, (d) H , (e) PC, and (f) HSS for version 1 (red circle) and version 2 (black circle) when compared with the IMS product. When the correct rejections exceed the other classes by more than 20 times, version 1 is shown in pink and version 2 in gray crosses. Vertical dotted lines show the transition from version 1 to version 2.02 (red), version 2.05 (gray), and version 2.10 (black). Curves show the two-month moving average of the data.

northern regions. In addition, the results should be given less emphasis when the absolute amount of snow is very small in summer, as the results vary greatly owing to even slight differences in snow cover between products. Therefore, days when there are more than 20 times more correct rejections than other classes are depicted with a different color in the rest of the plots in Fig. 5.

When a considerable amount of snow is present, version 1 constantly detects less snow than IMS, and bias is constantly less than one (Fig. 5b). The difference between version 2 and IMS is not as pronounced, but even so LSAF SAF SC is inclined to detect less snow. In summer, when the absolute amount of snow is very small, bias may vary

erratically and can be considerably more than or less than one. During winter 2007, test cases of version 2 had consistently better values than test cases of version 1, but the difference disappeared as spring progressed.

The amount of snow detected by LSAF SAF SC when IMS detected none was slight in both versions and F was near zero throughout the test period (not shown). In winter, most of the snow detected by LSAF SAF SC was also present in IMS; FAR started out with low values in winter, but the values increased as the snow melted (Fig. 5c). In summer, FAR can reach unity, meaning that none of the snow detected in LSAF SAF SC is present in IMS. Especially in summer 2007, this coincides with bias > 1 ,

but FAR can be quite high when bias ≈ 1 . In such a case, the snow would be in completely different areas even if the amounts of snow were comparative. In the summer of 2007, FAR rose to near unity for a considerable period, which was corrected by introducing special rules for summer. The difference between the test cases of version 2 and version 1 is not clearly visible.

On average, LSAF SAF SC detected about 80% of the snow detected by IMS ($H \approx 0.8$), but the amount varies and can be near unity or even zero (Fig. 5d). The seasonal variability in H was not as apparent as it was for FAR. In spring 2007, version 2 consistently had better values than version 1.

According to PC (Fig. 5e), version 2 and IMS did not differ to any large extent and the differences all but disappeared during summer. Version 1 had lower values during winter but comparable values in summer. During other winters, the operational PC dropped lower than in the test cases but was still higher than version 1.

During the winter HSS agreed with PC that version 2 was not very far from the IMS snow cover, even if the values were somewhat lower (Fig. 5f), but as the snow melted the skill substantially diminishes because not much skill is needed to say that no snow is present during summer. CSI yielded similar results, only with slightly lower values than HSS (not shown). The pattern of HSS was also very similar to that of H , but the HSS values were lower in summer. There was a period of very low values in the second half of 2007. Version 1 gave lower values for HSS than version 2 during winter, but after the snow started to melt the difference was no longer as evident. It is encouraging that there is no evident difference between the test cases in spring 2007 and the operational run afterward.

e. Spatial variability of results

During the operational period when only version 2 was available, LSAF SAF SC differed from IMS at the very edge of the satellite view, as PC had low values there (Fig. 6a). Otherwise, the spatial distribution of PC had very high values for most of the area. Additionally, HSS had low values in the areas where the snow cover was mostly transitory (Fig. 6b). When the snow cover was present for only a couple of days, a constant analysis of no snow would still yield good grades for PC, but would not show much skill and would yield low grades for HSS. LSAF SAF SC detected more snow mainly in southern Europe, where bias < 1 , and less in the north (Fig. 6c). However, there were areas where HSS and bias could not be calculated, as when no snow was detected by either products (or both products detected snow in all cases) and only correct rejections (or hits) had nonzero values, in which case HSS and bias had zero in the

denominator. Areas where HSS or bias could not be computed are ignored in the rest of the analysis.

f. Spatial results for different land-cover classes

The spatial performance of version 2 over different terrain types was investigated using only land classes covering more than 5% of the area under study. Thus only 8 out of the 22 classes of GLC2000 were used, but they covered more than 95% of the area. The largest terrain types were cultivated and managed areas (43%) and different forest types (27% of the area). For each class, ratios of probabilities of HSS and PC were calculated (Fig. 7). Our intuition was that most problematic areas would be forests, as forests block the snow surface from being visible and the shadows of the trees decrease the reflectance observed above the forest canopy (Salminen et al. 2009). And, indeed, for both HSS and PC, the ratios for tree-covered areas were less than unity; meaning those areas are more likely to have smaller values than the median of the values. For the other vegetation types, the results were not as clear. For shrub and herbaceous areas, HSS suggested low values whereas PC suggested no difference or high values. Sparse shrub and herbaceous areas gave little indication of to high values, but the ratios are seldom unity. Mostly smooth areas (cultivated and managed areas and bare area) were more likely to get higher, better values. However, most ratios were not very far from unity, and most of the time they were only about twice as likely to get values higher or lower than median values. Still, these results encourage taking terrain types more into account in further development of the algorithm. Furthermore, performing the same experiment using the IGBP land-cover classification produced similar results (not shown), so the results are fairly robust and not confined to one land-cover classification scheme.

5. Discussion and conclusions

This article introduces the LSAF SAF snow-cover product and also continues the validation work started in Hyvärinen et al. (2009), where no snow product was deemed to present the truth. In this study the NOAA/NESDIS IMS product was chosen as the reference for verification measures. Version 2 of LSAF SAF classified fewer surfaces than version 1 but had higher values of the validation measures in those areas, thereby increasing the usability of the product for NWP, where the accuracy of the data is much more relevant than large areal coverage.

Because of the lack of snow in southern Europe, there were large areas where many verification measures had zero in the denominator and were thus undefined. How to handle these areas in a more principled way would be an interesting topic for future study. A more

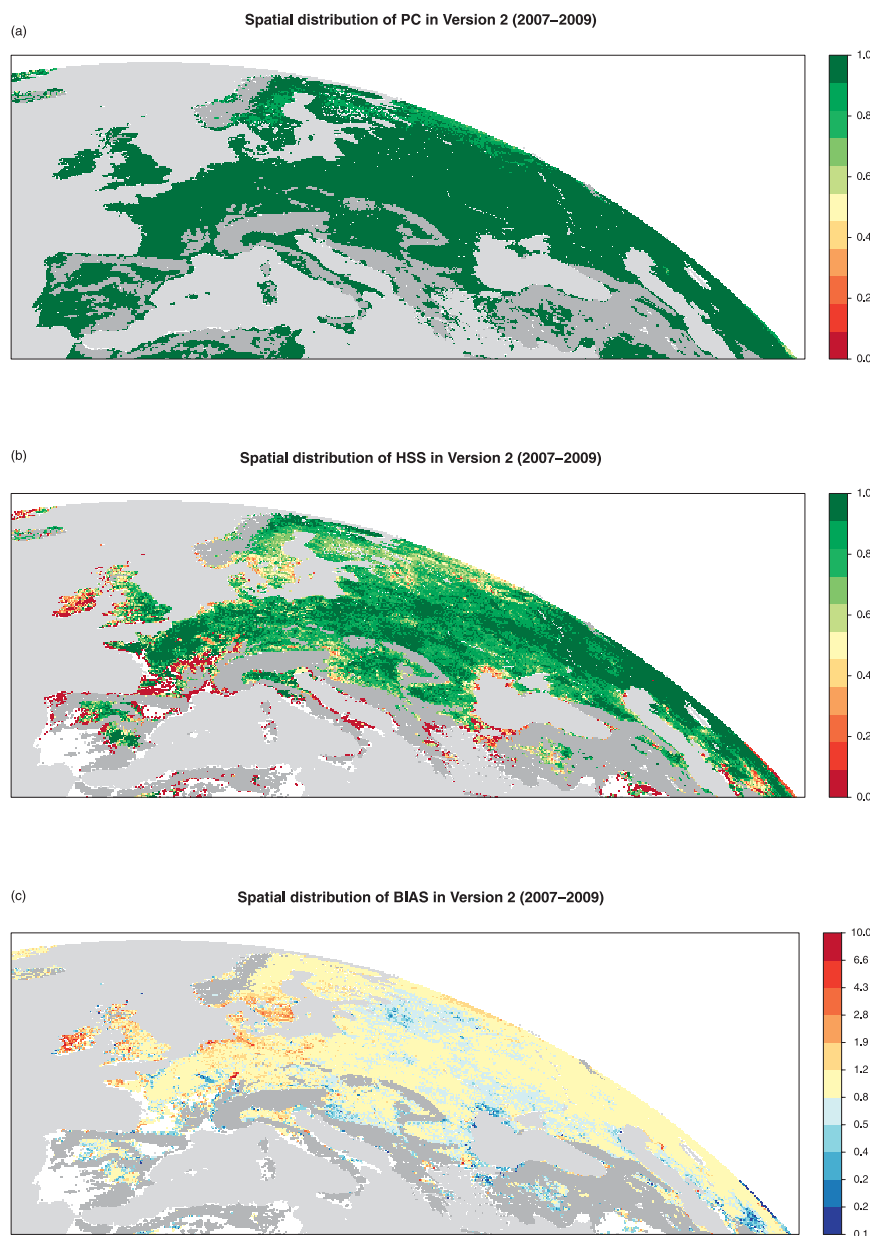


FIG. 6. Spatial distribution of (a) PC, (b) HSS, and (c) Bias in version 2 from July 2007 to December 2009. HSS and bias cannot be computed in areas where the denominator is zero.

Bayesian-oriented analysis might be able to give scores for these areas using some a priori distribution of values that would then be adjusted on the basis of observations. Another approach would be to assign some predefined value to these areas. This has been discussed for similar measures, but in a different context, by Batagelj and Bren (1995).

Both version 2 and the NOAA/NESDIS IMS product presented a reasonable and realistic snow-cover analysis, particularly during the winter season. The disagreement between products was at its largest outside the winter

season and in areas where the snow cover was mostly transitory. Especially when there is great temporal variability in the snow cover, it may no longer be feasible to use NOAA/NESDIS IMS as the ground truth for verification measures, as it is hard to say whether NOAA/NESDIS IMS really presents the truth. For example, on 12 October 2007, there was a snow-covered belt across Scandinavia that was not detected by the IMS product, but which the LSA SAF snow-cover algorithm detected correctly. This correct detection was of course penalized

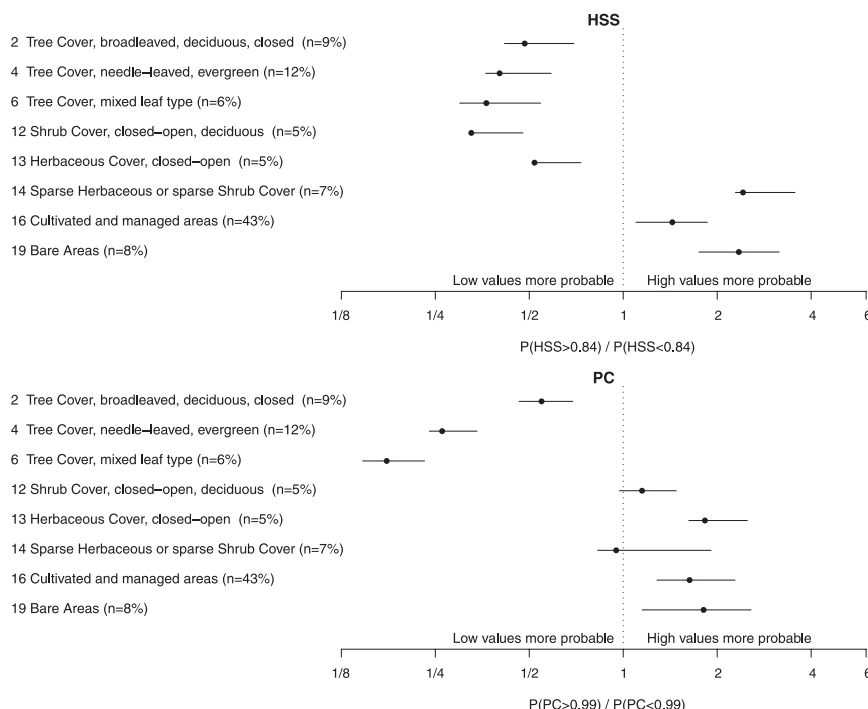


FIG. 7. The probability of high and low values of HSS and PC as a function of land cover. Only GLC2000 classes covering more than 5% of the study area were used. The class numbers are according to the original GLC2000 classes.

when compared with IMS. Further validation must rely on surface observations. Automatic snow observations have given less than satisfactory results, but advances in ground-based observations are taking place. For example, a digital camera attached to a measuring tower could provide an adequate estimate of the fractional snow cover (M. Takala 2010, personal communication). In addition to operational observations, measurement campaigns, such as Snow Reflectance Transition Experiment (SNORTEX; Roujean et al. 2009), offer valuable support for the validation of products. The feasibility of a method to measure fractional snow cover is being studied, but the results are not yet available.

The different versions of the algorithm gave somewhat surprising results, as our simple version 2 performed better than version 1 based on the state-of-the-art algorithm presented by NWC SAF. The surprising outcome can be explained by the tendency of version 2 to refuse to classify hard-to-classify areas. This comes at a price, as our algorithm classifies less surface area. For NWP, however, this withholding of some data is not a serious problem. On average, the background field gives a good estimate of the actual snow cover, and therefore gaps in the observations are generally not a serious obstacle. Biased or otherwise misleading snow observations would be a much worse problem.

Future plans for the algorithm development are concentrating on improvements in the individual classification rules, fractional snow-cover analysis, and a version of the algorithm for the AVHRR instrument available on polar-orbiting satellites such as MetOp.

Acknowledgments. This work was financially supported by the Land SAF project, cofunded by EUMETSAT. The first algorithm version was developed by the SMHL. We are grateful to Dr. Sylvain Joffre, Dr. Terhikki Manninen, Dr. David M. Schultz, and Ms. Elena Saltikoff for reading the early versions of this paper and for their many valuable comments. We also thank Dr. Carl Fortelius, Dr. Laura Rontu, and Mr. Kalle Eerola for the discussions we had about the needs of numerical weather prediction.

APPENDIX

Validation Measures

There is an extensive literature on different measures calculated from Table 4. Many of them have been named many times and the terminology can be confusing. Overviews of the measures and their history can be found in, for example, Jolliffe and Stephenson (2003) and Wilks (2006).

This paper follows the terminology of Jolliffe and Stephenson (2003). The basic descriptive measure is the bias:

$$\text{bias} = \frac{a+b}{a+c}, \quad (\text{A1})$$

the ratio of the number of snow pixels in the test analysis to the number of snow pixels in the baseline analysis. The best value for bias is 1; less than 1 means underestimation and more than 1 means overestimation.

The contingency table gives the joint distribution of analyses. From this distribution the following conditional distributions are constructed for use as performance measures:

The hit rate H or probability of detection (POD) is

$$H = \frac{a}{a+c}, \quad (\text{A2})$$

and in the perfect analysis this should be 1. The false-alarm rate F is

$$F = \frac{b}{b+d}, \quad (\text{A3})$$

and in the perfect analysis this should be 0. The false-alarm ratio (FAR) is

$$\text{FAR} = \frac{b}{a+b}, \quad (\text{A4})$$

and in the perfect analysis this should be 0.

An intuitive measure of accuracy is proportion correct:

$$\text{PC} = \frac{a+d}{a+b+c+d}, \quad (\text{A5})$$

the fraction of items classified the same way in both analyses. The best value for PC is 1 and the worst is 0. PC alone is insufficient, in particular when one of the categories dominates. A refinement often used is the critical success index:

$$\text{CSI} = \frac{a}{a+b+c}, \quad (\text{A6})$$

which ignores correct rejections. Like PC, its best value is 1 and the worst is 0.

Skill scores measure the relative skill by comparing the results with the reference. The reference often used is the random hits, which in this study would be snow correctly detected as snow by chance in the absence of skill. When the reference is the random hits, from PC we can devise the Heidke skill score:

$$\text{HSS} = \frac{2(ad-bc)}{(a+c)(c+d) + (a+b)(b+d)}, \quad (\text{A7})$$

its best value being 1 and its worst being -1 .

REFERENCES

- Bartholomé, E., and A. Belward, 2005: GLC2000: A new approach to global land cover mapping from earth observation data. *Int. J. Remote Sens.*, **26**, 1959–1977.
- Batagelj, V., and M. Bren, 1995: Comparing resemblance measures. *J. Classif.*, **12**, 73–90.
- Derrien, M., and H. LeGléau, 2005: MSG/SEVIRI cloud mask and type from SAFNWC. *Int. J. Remote Sens.*, **26**, 4707–4732.
- de Wildt, M. R., G. Siez, and A. Gruen, 2007: Operational snow mapping using multitemporal Meteosat SEVIRI imagery. *Remote Sens. Environ.*, **109**, 29–41.
- Dong, C., and W. Zhang, 2004: China's current and future meteorological satellites systems. *Proc. 2004 EUMETSAT Meteorological Satellite Conf.*, Prague, Czech Republic, EUMETSAT, 17–24.
- Dozier, J., R. O. Green, A. W. Nolin, and T. H. Painter, 2009: Interpretation of snow properties from imaging spectrometry. *Remote Sens. Environ.*, **113** (Suppl.), S25–S37, doi:10.1016/j.rse.2007.07.029.
- Drusch, M., D. Vasiljevic, and P. Viterbo, 2004: ECMWF's global snow analysis: Assessment and revision based on satellite observations. *J. Appl. Meteor.*, **43**, 1282–1294.
- Dybbroe, A., K. Karlsson, and A. Thoss, 2005: NWCSAF AVHRR cloud detection and analysis using dynamic thresholds and radiative transfer modeling. Part I: Algorithm description. *J. Appl. Meteor.*, **44**, 39–54.
- Hall, D., G. Riggs, V. Salomonson, N. DiGirolamo, and K. Bayr, 2002: MODIS snow products. *Remote Sens. Environ.*, **83**, 181–194.
- Hamill, T., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167.
- Helfrich, S. R., D. McNamara, B. H. Ramsay, T. Baldwin, and T. Kasheta, 2007: Enhancements to, and forthcoming developments in the Interactive Multisensor Snow and Ice Mapping System (IMS). *Hydrol. Processes*, **21**, 1576–1586.
- Hyvärinen, O., K. Eerola, N. Siljamo, and J. Koskinen, 2009: Comparison of snow cover from satellite and numerical weather prediction models in Northern Hemisphere and northern Europe. *J. Appl. Meteor. Climatol.*, **48**, 1199–1216.
- Jolliffe, I. T., and D. B. Stephenson, Eds., 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons, 240 pp.
- Kaufman, L., and P. J. Rousseeuw, 1990: *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley-Interscience, 368 pp.
- Kidder, S. Q., and H.-T. Wu, 1984: Dramatic contrast between low clouds and snow cover in daytime 3.7 μm imagery. *Mon. Wea. Rev.*, **112**, 2345–2346.
- , and T. H. Vonder Haar, 1990: On the use of satellites in Molniya orbits for meteorological observation of middle and high latitudes. *J. Atmos. Oceanic Technol.*, **7**, 517–522.
- Koskinen, J. T., J. T. Pulliainen, and M. T. Hallikainen, 1997: The use of ERS-1 SAR data in snow melt monitoring. *IEEE Trans. Geosci. Remote Sens.*, **35**, 601–610.
- Lahtinen, P., A. Ertürk, J. Pulliainen, and J. Koskinen, 2009: Merging flat/forest and mountainous snow products for extended European area. *Proc. 2009 IEEE Int. Geoscience and*

- Remote Sensing Symp.*, Vol. II, Cape Town, South Africa, IEEE, 563–566.
- Li, X., R. T. Pinker, M. M. Wonsick, and Y. Ma, 2007: Toward improved satellite estimates of short-wave radiative fluxes—Focus on cloud detection over snow: 1. Methodology. *J. Geophys. Res.*, **112**, D07208, doi:10.1029/2005JD006698.
- Matikainen, L., R. Kuittinen, and J. Vepsäläinen, 2002: Estimating drainage area-based snow cover percentages from NOAA/AVHRR images. *Int. J. Remote Sens.*, **23**, 2971–2988.
- Matson, M., 1991: NOAA satellite snow cover data. *Global Planet. Change*, **4**, 213–218.
- Metsämäki, S., S. Anttila, M. Huttunen, and J. Vepsäläinen, 2005: A feasible method for fractional snow cover mapping in boreal zone based on a reflectance model. *Remote Sens. Environ.*, **95**, 77–95.
- Miller, S. D., T. F. Lee, and R. L. Fennimore, 2005: Satellite-based imagery techniques for daytime cloud/snow delineation from MODIS. *J. Appl. Meteor.*, **44**, 987–997.
- , and Coauthors, 2006: NexSat: Previewing NPOESS/VIRS imagery capabilities. *Bull. Amer. Meteor. Soc.*, **87**, 433–446.
- Moura, A., 2006: WMO's contribution to GEOSS and GEONetcast. *WMO Bull.*, **55**, 256–260.
- Peltoniemi, J. I., S. Kaasalainen, J. Näränen, L. Matikainen, and J. Piironen, 2005a: Measurement of directional and spectral signatures of light reflectance by snow. *IEEE Trans. Geosci. Remote Sens.*, **43**, 2294–2304.
- , —, —, M. Rautiainen, P. Stenberg, H. Smolander, S. Smolander, and P. Voipio, 2005b: BRDF measurement of understory vegetation in pine forests: Dwarf shrubs, lichen, and moss. *Remote Sens. Environ.*, **94**, 343–354.
- Riisholgaard, L. P., 2004: The case for launching a meteorological imager in a Molniya orbit. *Proc. Seventh Int. Winds Workshop*, Helsinki, Finland, EUMETSAT, 323–330.
- Romanov, P., and D. Tarpley, 2003: Automated monitoring of snow cover over South America using GOES Imager data. *Int. J. Remote Sens.*, **24**, 1119–1125.
- , G. Gutman, and I. Csizar, 2000: Automated monitoring of snow cover over North America with multispectral satellite data. *J. Appl. Meteor.*, **39**, 1866–1880.
- Roujean, J.-L., and Coauthors, 2009: SNORTEX (Snow Reflectance Transition Experiment): Remote sensing measurement of the dynamic properties of the boreal snow-forest in support to climate and weather forecast: Report of IOP-2008. *Proc. 2009 IEEE Int. Geoscience and Remote Sensing Symp.*, Vol. II, Cape Town, South Africa, IEEE, 859–862.
- Salminen, M., J. Pulliainen, S. Metsämäki, A. Kontu, and H. Suokanerva, 2009: The behaviour of snow and snow-free surface reflectance in boreal forests: Implications to the performance of snow covered area monitoring. *Remote Sens. Environ.*, **113**, 907–918.
- Schmit, T., M. Gunshor, W. Menzel, J. G. J. Li, and A. Bachmeier, 2005: Introducing the next-generation advanced baseline imager on GOES-R. *Bull. Amer. Meteor. Soc.*, **86**, 1079–1096.
- Ulaby, F. T., R. K. Moore, and A. K. Fung, 1986: *From Theory to Applications*. Vol. III, *Microwave Remote Sensing, Active and Passive*, Addison Wesley, 1097 pp.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 648 pp.
- Wiscombe, W. J., and S. G. Warren, 1980a: Model for the spectral albedo of snow. I: Pure snow. *J. Atmos. Sci.*, **37**, 2712–2733.
- , and —, 1980b: Model for the spectral albedo of snow. II: Snow containing atmospheric aerosols. *J. Atmos. Sci.*, **37**, 2734–2745.
- World Meteorological Organization, 1995: Manual on codes. Volume I.1, No. 306, 229 pp.