Intercalibration of Polar-Orbiting Spectral Radiometers Without Simultaneous Observations

Terhikki Manninen, Aku Riihelä, Andrew Heidinger, Crystal Schaaf, Alessio Lattanzio, and Jeffrey Key

***Abstract*— A new intercalibration method for two polar-orbiting satellite instruments or two instrument constellations' Fundamental Climate Data Records (FCDRs) is presented. It is based on statistical fitting of reflectance data from the two instruments covering the same area during the same period, but not simultaneously. A Deming regression with iterative weights is used. The accuracy of the intercalibration method itself was better than 0.5% for the Moderate Resolution Imaging Spectroradiometer (MODIS) versus MODIS and Advanced Very High Resolution Radiometer (AVHRR) versus AVHRR test data sets. The intercalibration of an AVHRR FCDR generated by NOAA versus a combined MODIS Terra and Aqua data set of red and near-infrared (NIR) channels was carried out and showed a difference in the reflectance values of about 2% (red) and 6% (NIR). The presented intercalibration method can be used for checking the calibration of two instruments or FCDRs in all viewing angles used separately.**

***Index Terms*— Calibration, remote sensing, statistics.**

I. Introduction

A NECESSARY condition for high-quality long-term Climate Data Records (CDRs) derived from satellite remote sensing is accurate intercalibration of instruments over long timescales [1]. For CDRs, this most often involves the same type of sensor on a series of satellites, e.g., the Advanced Very High Resolution Radiometer (AVHRR) on the series of NOAA polar-orbiting satellites. It may also be useful, however, to combine observations from different sensors and satellites. For example, surface albedo is one of the essential climate variables and a key parameter for the energy balance of the earth [2]. Albedo retrievals are usually performed utilizing only a single instrument or an instrument "family" (i.e., AVHRR on multiple satellites). This limits the total number of available observations per time period per terrestrial scene. Combining multiplatform observations can yield improvements in both the accuracy and temporal resolutions of surface albedo retrievals. This is especially important for optical remote sensing in frequently cloudy areas, such as the Arctic.

Various spaceborne optical imager families, such as AVHRR or Moderate Resolution Imaging Spectroradiometer (MODIS), have different imaging channel wavebands. Also, the individual imagers within instrument families have different spectral responses. The intrafamily differences can and recently have been compensated for by intercalibration, leading to the creation of Fundamental CDRs (FCDRs) specific to a single instrument family [7]. When intercalibrating observations from two different instrument families, even though both are FCDR quality, due attention must still be paid to the differences in spectral coverage between the instrument imaging channels. In this paper, we seek to perform such an intercalibration between MODIS and AVHRR data sets. In this case, the intercalibration is extremely important for the near-infrared (NIR) channel, the spectral width of which is markedly larger for AVHRR than for MODIS. Although the motivation of this paper is to compensate for the spectral difference of MODIS and AVHRR, the presented intercalibration method can also be applied to normal intercalibration of different individual instruments of the same family.

The basic premise of intercalibration is that two similar instruments should produce the same reflectance value when they view the same target simultaneously with identical viewing geometry. In reality, this requirement cannot be rigorously fulfilled in the calibration of earth-observing instruments. An established intercalibration method for two satellite instruments is the use of simultaneous nadir observations (SNOs), which has been proved to be very effective over a wide spectral range [1]. Another approach is statistical intercomparison, which is the most frequently used for low-resolution data. Pseudoinvariant calibration sites (PICSs), such as the Libyan Desert, have been utilized for both the SNO-based absolute calibration and the statistical intercomparison of two or more satellites [2]–[4]. Several approaches have been investigated for the intercalibration of the AVHRR instrument series with MODIS data over the PICS [5]–[8]. These methods require reliable atmospheric correction and either SNOs or reliable bidirectional reflectance distribution factor estimates. The method presented here does not need any information about the surface or atmosphere properties concerning the images used.

The goal of this paper is to derive a general concept of intercalibration applicable to any collection of optical polar-orbiting satellite imagers. Hence, one must consider the

Manuscript received March 7, 2017; revised June 14, 2017 and August 29, 2017; accepted September 25, 2017. Date of publication November 7, 2017; date of current version February 27, 2018. This work was supported in part by EUMETSAT through the Project CM SAF and in part by the National Aeronautics and Space Administration for support of the MODIS work under Grant NNX14AI73G. *(Corresponding author: Terhikki Manninen.)*

T. Manninen and A. Riihelä are with the Finnish Meteorological Institute, FI-00101 Helsinki, Finland (e-mail: terhikki.manninen@fmi.fi; aku.riihela@fmi.fi).

A. Heidinger and J. Key are with NOAA/NESDIS, Madison, WI 53706 USA (e-mail: andrew.heidinger@noaa.gov; jeff.key@noaa.gov).

C. Schaaf is with the University of Massachusetts Boston, Boston, MA 02125-3393 USA (e-mail: crystal.schaaf@umb.edu).

A. Lattanzio is with EUMETSAT, 64295 Darmstadt, Germany (e-mail: alessio.lattanzio@eumetsat.int).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2017.2764627

possibility of satellites that will never observe the same place simultaneously, so that the use of SNOs will not be possible. Or it may be that the number of SNOs is too small to be statistically representative. In addition, to guarantee the quality of the intercalibration for other viewing angles than that for nadir, it is desirable to look for a method that can be applied to the whole range of viewing angles to be used. Our solution is to derive the top-of-atmosphere (TOA) reflectance distributions of a large area covering the entire range of global TOA reflectance data [9]. At this phase, we concentrate on instrument bands for which the wavelength range is similar enough to produce essentially the same reflectance for the same target. Then, the assumption is that for a large enough statistical sample, the reflectance distributions (with the same sun and satellite angle configuration) should be equal, as the instruments are observing the same target.

As an example, we start with two satellite families, at first using two separate data sets of the same instrument (MODIS Terra) and then two separate data sets of one instrument family-based FCDR (AVHRR or MODIS) in order to test the intercalibration method itself. Then, we apply the method by fitting AVHRR to MODIS TOA reflectance data. The advantage in first carrying out the "intercalibration" for the same instrument is that it is known in advance that one should expect to obtain a 1:1 relationship for the linear regression of the two data sets. Hence, the goodness of the proposed method can be assessed separately before applying it to real multiplatform data (AVHRR versus MODIS).

When fitting the distributions of the two independent TOA reflectance values of the same period, the data sets must first be converted into the same resolution. Then, the uncertainty of both data sets has to be considered. When the instruments are identical, one can use standard orthogonal regression. However, when there is a marked difference between the instrument accuracies of the two data sets, one has to use (weighted) Deming regression (Section III-D). In addition, one has to consider that the uncertainty of the TOA reflectance distribution means due to random errors will decrease as the inverse square root of the number of points in the distribution. Finally, the obliquely viewed pixels have larger uncertainty than the nadir pixels, because the shape of the true pixel size is elongated ellipse-like instead of the nominal spherical disk, and ascending and descending pixels do not cover exactly the same area on the ground. This heteroskedastic character of the points of the linear regression has to be paid due attention.

This paper is carried out under the World Meteorological Organization (WMO) Sustained and Coordinated Processing of Environmental Satellite data for Climate Monitoring (SCOPE-CM) project SCM-02. It is focused on polar-orbiting optical imagers, whose strengths are in high data acquisition rates over the high latitudes of the earth, which play a key role in climate change.

II. Study Area and Material

A. Study Area and Period

In order to reduce the volume of data to be handled, a subset of the globe will be used as the basis for calibration. The study area contains a wide variety of land cover classes and a large ocean area felt to be spectrally representative of the whole globe. The borders of the study area are the latitudes 0°N and 75°N and the longitudes −130°E and 45°E (Fig. 1) [10]. The area is mainly studied during June 29, 2010–July 19, 2010, because the RAdiation, Snow Characteristics and ALbedo at Summit (RASCALS) campaign [11] took place during that time at Summit Camp on the Greenland ice sheet. The RASCALS goniospectrometer measurements were compared with atmospherically corrected MODIS and AVHRR spectral surface reflectances to define angular cutoff limits for the intercalibration and subsequent joint albedo retrieval. In order to obtain a larger data set for the intercalibration of MODIS versus MODIS, additional MODIS images were taken from June 1–28, 2010. The earth–sun distance variation during June–July affects the TOA resolution values only by 0.048%.

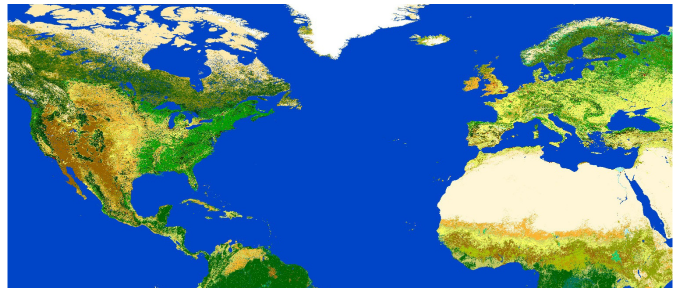


Fig. 1. Study area on the GlobCover map [10].

B. AVHRR Data

AVHRR observations from NOAA-15, NOAA-18, NOAA-19, and Metop-A were obtained from NOAA Pathfinder Atmospheres–Extended (PATMOS-x) [12] archives for this paper. The total number of AVHRR images was 1082, out of which 275 were from NOAA-15, 277 from NOAA-18, 272 from NOAA-19, and 258 from Metop-A. One image covered typically an area of about 2045 km × 11 500 km. The observations were intercalibrated following the approach by Heidinger *et al.* [7] and provided for use as TOA reflectances. Global Area Coverage data with a spatial resolution of ∼4 km (subnadir) from AVHRR channels 1 (580 . . . 680 nm; "red") and 2 (725 . . . 1000 nm; "NIR") were utilized. The PATMOS-x viewing and illumination geometry information was also used here.

C. MODIS Data

The MODIS products MOD02 and MOD03 (collection 6) from the Terra satellite and MYD02 and MYD03 from Aqua were used for this paper as the basic reference set. In comparison with AVHRR, altogether 2621 images were used from the same time period, out of which 1295 were from Terra and 1326 from Aqua. In the intercomparison study of MODIS versus MODIS, a larger data set was used of June–July 2010 comprising altogether 6083 images, out of which 3016 were from Terra and 3067 from Aqua. One image covers an area of 1354 km × 2030 km. The MOD02 and MYD02 products contain calibrated and geolocated at-aperture radiances in W/($m^2 \cdot \mu m \cdot sr$). TOA reflectance values were determined for the band 1 (620 ... 670 nm; "red") and band 2 (841 ... 876 nm; "NIR") through the knowledge of the solar irradiance [13]. The spatial resolution for these data is 250 m, and the data were averaged to a 5-km resolution in order to be radiometrically more equivalent to the spatial resolution of the AVHRR data, namely, the dynamic range of reflectance values is typically smaller in coarser resolution. The MOD03 and MYD03 products contain the geodetic coordinates, the solar and satellite zenith angles, and the azimuth angle for each MODIS 1-km sample.

D. USGS Land Cover Spectra

A collection of 87 individual reflectance spectra of diverse land cover types was selected from the USGS Spectroscopy Lab database for the atmospheric simulations: 10 for grass, 19 for forest, 5 for crop, 6 for lichen, 4 for minerals, 18 for man-made materials, 4 for water, 8 for snow/ice, and 13 for mixtures of rock [14], [15]. These spectra were used for simulations of TOA spectra of diverse land cover types corresponding to the MODIS and AVHRR instrument red and NIR bands.

III. METHODS

A. Basic Idea and Hypotheses

The basic idea of this intercalibration approach is that the two satellite instruments to be intercalibrated are observing the same target, i.e., the chosen subset of the globe, atmosphere included. When the instruments are operating during the same time period, they are both taking independent samples of the total TOA reflectance distribution of the study area matching that whole period. However, one has to also consider that the TOA reflectance has diurnal variation related to the sun zenith angle θ_s and the atmospheric and weather conditions (especially cloud cover). The viewing configuration (the satellite zenith angle θ_v and the azimuth angle ϕ between the sun and the satellite directions) matters as well, as natural targets are typically not Lambertian surfaces. Hence, the red (R_{Red}) and NIR (R_{NIR}) TOA reflectance values of the images are collected in separate distributions corresponding to constant ranges of these angle triplets (θ_s, θ_v, ϕ). The number of available individual distributions per instrument equals the number of different angle triplets existing in the whole data mass of images. In the intercalibration of two data sets, only those distributions are used, for which the matching angle triplets exist for both data sets.

The angular resolution used is one degree. Each reflectance distribution is then described by its mean value $\langle R \rangle$ and the 8% and 98% quantiles, R_8 and R_{98}, respectively, which are mostly related to the reflectances of ocean and snow. The reflectance quantitites $\langle R \rangle$, R_8, and R_{98} per angle triplet, based on all existing distributions, are then gathered as one total set of reflectance values per instrument. Linear regression is then sought for the two reflectance data sets to be intercalibrated. The reason to use also R_8 and R_{98} in addition to $\langle R \rangle$ is that the slope of the regression line is reliable within the variation range of the values used for its determination, and expanding the determination range improves the robustness of regression.

The hypothesis is that if: 1) the amount of data is statistically sufficient; 2) the observation period is long enough; and 3) the area is large enough to provide unbiased diurnal sampling of diverse land cover classes (and cloud cover), the effect of the atmospheric and land cover changes on the TOA reflectance should produce a nearly equal distribution for a similar wavelength range. In addition, possible variation of the radiometric accuracy of the instrument on the satellite is included in calibration, when data are gathered from a longer period, but no significant aging is thought to take place within one and a half month. Obviously, the main question is how random the sampling really is, when the satellites are sun synchronous. Using both morning and afternoon satellites for one instrument avoids the diurnal bias related to cloud coverage. In addition, a wide swath width increases the diurnal coverage per latitude. In any case, either the area or the time period (or both) has to be relatively large to guarantee that the sampled cloud cover represents well the cloud reflectance distribution and the cloud-free pixels cover a large enough variation of the diverse land cover types. In principle, one could also use cloud-masked data, but then the quality of the intercalibration would depend on the cloud-masking quality. In addition, including the clouds in the intercalibration provides a continuum of reflectance values starting from cloud-free ocean to cloud-free snow. Hence, the whole variation range of reflectance values contributes to intercalibration.

The number of individual distributions per data set to be intercalibrated is the number of different angle triplets available in that data set. Some angle triplets contain more observations than others. As an extreme case, it may happen that during the period of interest in the study area, there exists only one pixel in the whole data set matching a certain angle triplet. One question to solve is: what the minimum number of points n_{min} allowed for an individual distribution to be included in the intercalibration should be? Is it, for example, better to include the $\langle R \rangle$, R_8, and R_{98} values based on just one single TOA reflectance value (so that $R_8 = \langle R \rangle = R_{98}$) in the regression or not? Normally, linear regression results improve when the number of points in the regression n_{reg} increases, but introducing a very heteroscedastic data set while maximizing the number of points may not be an advantage. Trivially, n_{reg} decreases when n_{min} increases. If only distributions consisting of a very large number of individual reflectance values are included in intercalibration, is there a risk that the achieved calibration accuracy suffers from too small a number of points available for the regression? Answers to these questions are provided in Sections IV-B and IV-C.

The quality of the method is first tested by "intercalibrating" two independent data sets produced by the same instrument on several different satellites. Because the reflectance values of the two data sets to be intercalibrated, which correspond to the same angle triplets, do not always come from the same places, i.e., they do not necessarily represent the same land cover type or atmospheric conditions, the scatter of the point set, and thus, the coefficient of determination of the regression, is not a measure for the goodness of fit. The parameter to be used for describing the goodness of fit in this context is the average deviation (Δ) of the regression line from the ideal

1:1 line, which is calculated from

$$\Delta = \frac{1}{100}\int_0^{100} |b_0 + b_1 x - x| dx \tag{1}$$

where b_0 is the constant, b_1 is the linear regression coefficient, and x refers to the TOA reflectance value, which is assumed to be given in percentages. The smaller the Δ, the better the fit, and the size of Δ directly describes the mean accuracy of the intercalibration of the reflectance values.

For a real intercalibration, it is not possible to assess the goodness of fit, as the true relationship (for comparison) is not known. Then, the value to use for b_1 is derived from its distribution based on diverse combinations of $n_{\min}$ and n_{reg}. First, the half-value of the peak of the b_1 distribution is determined. Then, the middle b_1 value of the part of the distribution exceeding the half-value is chosen to be used in intercalibration. The corresponding b_0 value is determined by the linear interpolation of the regression parameter pairs (b_0, b_1) closest to the chosen b_0. By this way, it is ensured that the b_0 value chosen corresponds well to the b_1 chosen, namely, in general, there is a relatively strong relationship between those two parameters.

TABLE I

RELATIONSHIP BETWEEN THE MEASURED USGS SPECTRA [14] REFLECTANCE VALUES (R) INTEGRATED OVER THE MODIS AND AVHRR RED AND NIR BAND WAVELENGTHS. VALUES FOR THE NUMBER OF POINTS OF INDIVIDUAL SPECTRA n, THE CONSTANT AND LINEAR COEFFICIENTS OF THE REGRESSION R_{AVHRR} VERSUS R_{MODIS}, AND THE COEFFICIENT OF DETERMINATION (R^2) ARE SHOWN. FOR VEGETATION, THE RED BAND RESULTS ARE DOMINATED BY ONE OUTLIER, AND HENCE, THE CORRESPONDING NUMBERS ARE GIVEN IN BRACKETS WITHOUT IT

Channel	Target	n	Constant	Slope	R^2
Red	Snow/water	12	0.002	0.987	1.000
	Mixtures	13	-0.000	0.985	1.000
	Man-made	18	-0.001	0.977	0.999
	Vegetation	40 (39)	0.014 (0.009)	0.924 (0.965)	0.988 (0.997)
NIR	Snow/water	12	0.001	0.907	0.961
	Mixtures	13	0.001	0.966	1.000
	Man-made	18	0.003	0.961	1.000
	Vegetation	40	0.010	0.925	0.997

B. Spectral Difference

The original motivation for developing this intercalibration method was the surface albedo retrieval based on the combined use of AVHRR and MODIS. The question is whether it is reasonable to combine the AVHRR and MODIS data already at the TOA level or not. If the spectral difference between AVHRR and MODIS reflectance values of typical targets is smaller than the instrumental inaccuracy, it is acceptable to "convert" MODIS data into "AVHRR-like" data by intercalibration at the TOA reflectance level and use both data sets after that step for albedo retrieval identically. In the opposite case, one would have to treat the AVHRR and MODIS data sets separately until the spectral albedo level. In order to examine the spectral difference of AVHRR and MODIS the following spectral study was carried out.

Measured USGS spectra were used to analyze the possible relationship between the MODIS and AVHRR red and NIR channel surface reflectance values for various land cover targets. The spectra were integrated over the wavelength range in question, and ordinary linear regression parameters were determined for calculated AVHRR-like reflectance values (R_{AVHRR}) versus those of MODIS-like reflectance values (R_{MODIS}) for red and NIR channels. The results are shown in Table I.

The red channel wavelength ranges of the MODIS and AVHRR instruments are relatively similar in contrast to NIR. Indeed, it turned out that the simulated surface R_{AVHRR} values are only slightly smaller than the corresponding surface R_{MODIS} values in the red channel for surface targets included in the USGS spectral library [14], [15]. For the NIR channel, the relationship is not as similar for snow/water and vegetation spectra, but the coefficient of determination R^2 is high. In the NIR channel, the surface R_{AVHRR} values are typically smaller than the corresponding surface R_{MODIS} values, their mean ratio of the four target types varying in the range 91%...98% for the spectra available. This means that one general relationship for all targets will not be very accurate (not better than about ±3%). However, the absolute radiometric accuracy of the AVHRR instrument is not estimated to be better than ±5%–6% [16]. Therefore, a general target-independent calibration coefficient is also appropriate for the NIR bands of MODIS and AVHRR within their stated accuracies. The conclusion is that an intercalibration of MODIS and AVHRR TOA reflectance values is a reasonable approach.

C. Angular Uncertainty

The AVHRR instrument is a whiskbroom-type instrument with a scanning mirror and single discrete detector per band [17]. The rotating mirror changes the angle of the incident light, and therefore what portion of the ground is being measured. The angular rotating accuracy may then cause an individual element of uncertainty, contributing to the reflectance estimation accuracy. In addition, the obliquely viewed and nadir-viewed pixels obviously capture different footprints on the ground. Moreover, the obliquely viewed pixels of the same latitude and longitude have different footprints whether from the descending or ascending orbit. Thus, a comparison of the pixel reflectance estimates near the edges of diverse images (i.e., when the view angle is large) may suffer from larger errors than a comparison of nadir pixel reflectance estimates of those images.

The MODIS instrument has a linear array of detectors with one array per band (pushbroom) [18]. The detector response is uniform in the along-track direction, but may have individual variation in the across-track direction. Again, the footprints from ascending and descending orbits at the same latitude and longitude are not identical for the same pixel for other viewing angles than nadir, because the footprint is elongated in different directions in ascending and descending passes.

The variance of the reflectance value due to the spatial difference of the footprint related to the satellite zenith angle is also related to the site heterogeneity in a nonrandom way. For example, over the oceans, mid-Greenland, or Sahara, the ascending and descending pass reflectance values at large viewing angles may not differ at all, whereas in coastal or agricultural areas and lake districts, the corresponding difference may be nonnegligible. Therefore, it is not possible to derive an exact expression for general random pointwise variances, the inverse of which should be used as the weights in the linear regression. However, recognition must be paid to the possible dependence of the intercalibration accuracy on the viewing angle when deriving the intercalibration regression parameter values. Hence, the weights of the regression are determined iteratively (Section III-D).

D. Regression

Traditional linear regression is based on the least-squares minimization of the sum of the squared vertical distances from the data points to the fit line. The implicit assumption is that the uncertainty of the explanatory variable is much smaller than that of the dependent variable. In the case of two data sets based on the same instrument (like regressing two independent MODIS data sets of the same area versus each other), this approach is not well grounded, because both data sets have the same instrument inaccuracies. In such a case, the recommended method is orthogonal regression [19], [20]. Orthogonal regression minimizes the perpendicular distance of the points from the regression line, instead of the vertical distance.

The actual intercalibration of AVHRR versus MODIS data requires additional recognition of the different uncertainties of the MODIS and AVHRR data. For this purpose, a Deming regression is suitable [21]–[23]. Like orthogonal regression, Deming regression accounts for errors in observations on both the horizontal and vertical axes. For orthogonal regression, the assumption is that the error statistics is the same for both axes, whereas Deming regression allows one variable to be more inaccurate than the other, and the ratio of the variances of the normally distributed independent errors of the two variables is assumed to be known.

An ordinary Deming regression assumes that the measurement error ratio of the explanatory and dependent variables is constant. The heteroskedastic character of the points, i.e., individual points having different uncertainty, can also be considered in Deming regression by using individual weights for the points [21], [22]. First, the individual weights must consider the uncertainty of the TOA reflectance distribution mean and recognize that the quantiles decrease with an increasing number of points in the distribution. Second, the obliquely viewed pixels have larger uncertainties than the nadir pixels due to the variation in footprint coverage. As the weights will be functions of the slope of the regression line, an iterative procedure is required to determine them [22]. First, an initial estimate of the slope and constant of the regression line is derived using the original observation points. Then, adjusted points are determined using the slope and constant terms and original points. Revised weights are then computed using these adjusted points, which in turn are used to calculate new slope and constant values for the linear regression line. Repeating this process, the individual weights are improved iteratively by requiring the relative difference of the regression parameters of successive iteration rounds to be smaller than 0.0001.

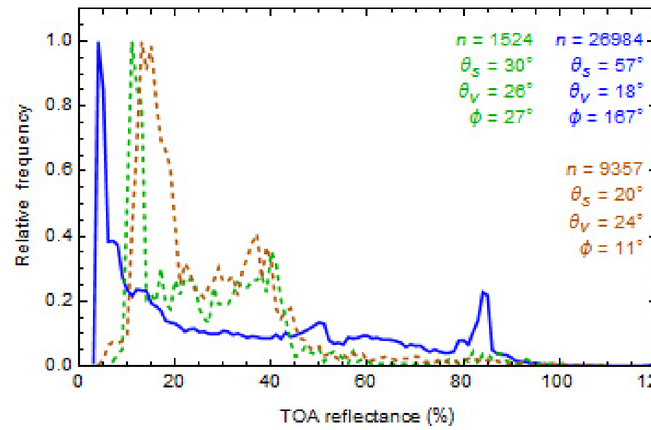


Fig. 2. Distributions picked from the total set of NOAA-15, NOAA-18, NOAA-19, and Metop-A/2 AVHRR red channel TOA reflectance values June 29, 2010–July 19, 2010 corresponding to three example angle triplets. The number of individual TOA reflectance values included in the above distributions is n.

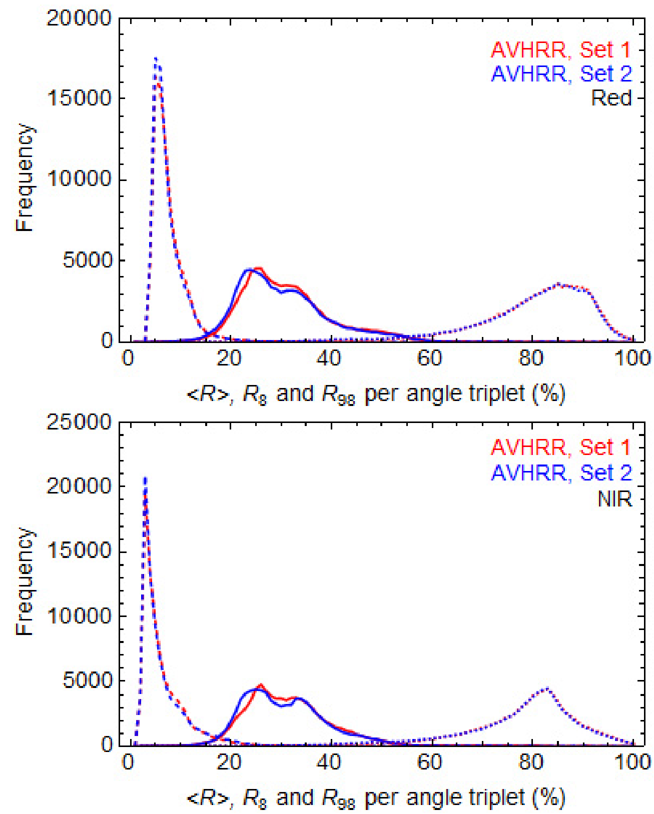


Fig. 3. Distributions of AVHRR (top) red and (bottom) NIR channel TOA reflectance distribution means (solid curves) and 8% (dashed curves) and 98% quantiles (dotted curves) corresponding to sampled angle triplets (θ_s, θ_v, ϕ). The time range is June 29, 2010–July 19, 2010. The whole NOAA-15, NOAA-18, NOAA-19, and Metop-A/2 AVHRR data sets are divided into sets 1 and 2 by taking every second image.

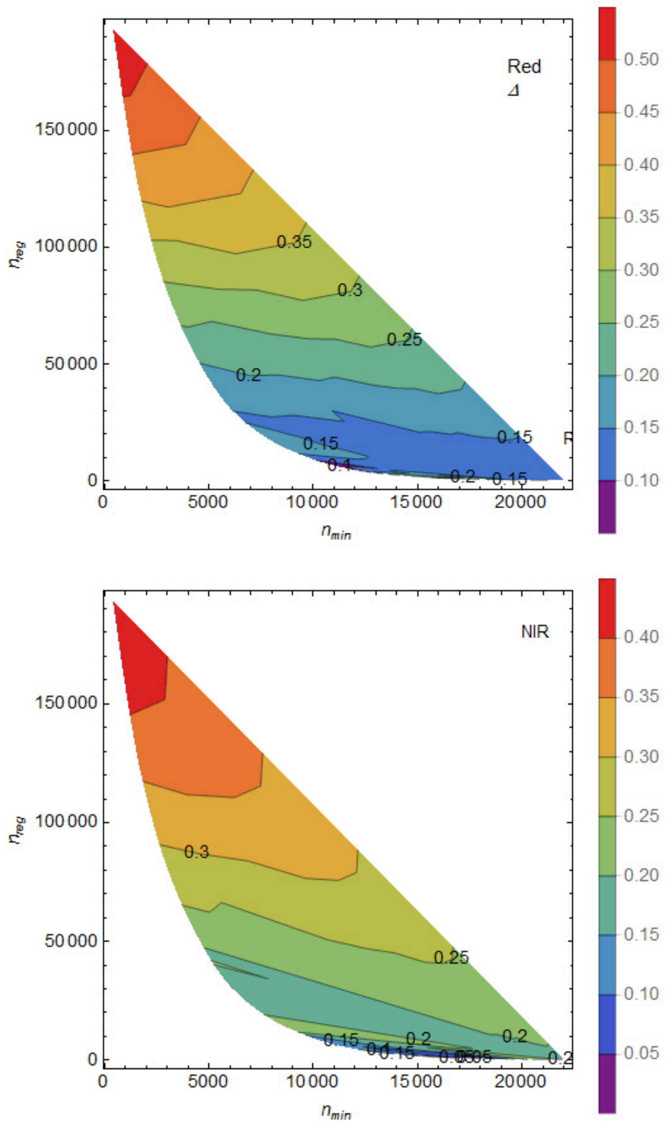


Fig. 4. Δ as a function of $n_{\min}$ and n_{reg} for the two subsets of AVHRR data for (top) red and (bottom) NIR channels.

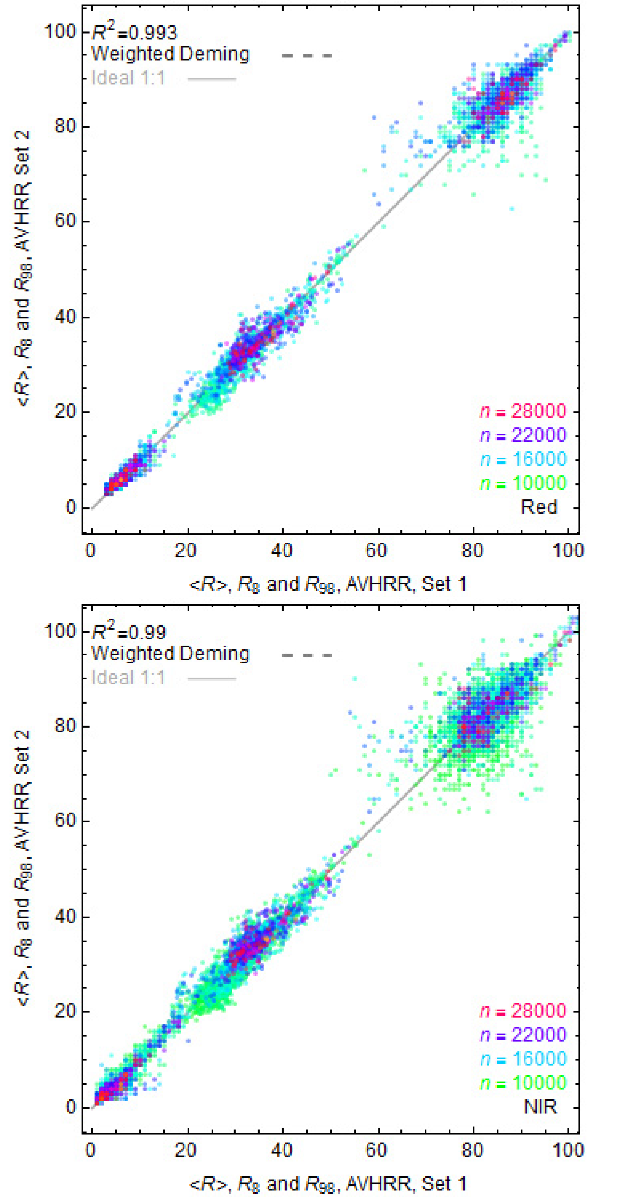


Fig. 5. Relationship of NOAA-15, NOAA-18, NOAA-19, and Metop-A/2 AVHRR versus AVHRR (top) $\langle R_{\text{Red}}\rangle$, R_{Red8}, and R_{Red98} and (bottom) $\langle R_{\text{NIR}}\rangle$, R_{NIR8}, and $\langle R_{\text{NIR98}}\rangle$ for $n_{\min} = 12\,000$. The colors of the points are related to the number of individual reflectance values (n) in the distributions from which the values for $\langle R_{\text{Red}}\rangle$, R_{Red8}, R_{Red98}, $\langle R_{\text{NIR}}\rangle$, R_{NIR8}, and R_{NIR98} are derived.

Allowing individual weights for the observation points also removed the need of advance knowledge of the inaccuracy of the two data sets, because it turned out that ordinary orthogonal regression provided a good enough starting point for the iteration. This overall procedure for obtaining the unbiased slope and intercept is called iteratively reweighted general Deming regression (IRGDR) [22].

IV. Results

A. TOA Reflectance Distributions

To check the applicability of the statistical approach for intercalibration, we first regressed the TOA reflectance values for AVHRR versus AVHRR and for MODIS versus MODIS. The entire image sets of AVHRR and MODIS were divided into TOA reflectance distributions corresponding to various viewing/illumination configuration angle triplets. Each distribution thus contains points from various places and varying dates/times. The number of points (n) in one distribution varies in a wide range starting from "distributions" of just one value. The larger the number of points in the distribution, the more statistically reliable is the distribution mean (Fig. 2). On the other hand, the number of points (i.e., the $\langle R\rangle$, R_8, and R_{98} values corresponding to the sampled angle triplets) in the intercalibration regression should be as large as possible. The effect of these two edge constraints on Δ of the regression is demonstrated in Sections IV-B and IV-C.

B. AVHRR Versus AVHRR

First, the AVHRR data set of all four NOAA and Metop satellites is sorted chronologically. Then, this data set is divided into two sets by taking every second image to one

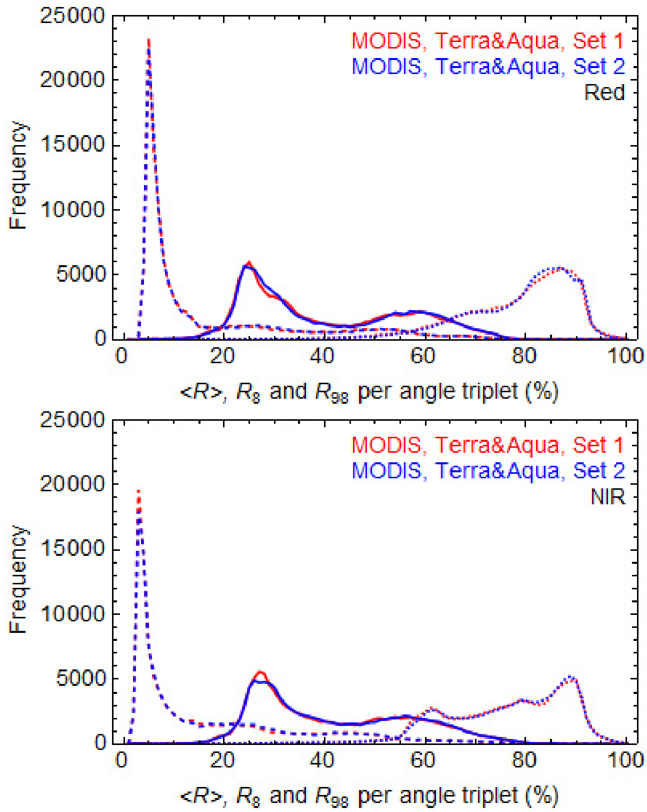


Fig. 6. Distributions of the MODIS (top) red and (bottom) NIR channel TOA reflectance distribution means (solid curves) and 8% (dashed curves) and 98% quantiles (dotted curves) corresponding to sampled angle triplets (θ_s, θ_v, ϕ). The time range is June 1, 2010–July 19, 2010. The whole Terra and Aqua data sets are divided into sets 1 and 2 by taking every second image.

set and the rest to the other set. The TOA reflectance values of set 1 corresponding to a certain angle triplet (θ_s, θ_v, ϕ) do not, in general, come from the same places as those of set 2. The distributions of the $\langle R_{\mathrm{Red}}\rangle$, R_{Red8}, and R_{Red98} and $\langle R_{\mathrm{NIR}}\rangle$, R_{NIR8}, and R_{NIR98} values are shown in Fig. 3 for the two AVHRR data subsets. Each $\langle R\rangle$, R_8, and R_{98} value corresponds to an individual R_{TOA} distribution, such as those presented in Fig. 2. The number of sampled angle triplets is roughly 81 000 values. The best fit for these data sets is obtained when using a combination that is in between the curving part of the area of possible value combinations and the largest values of $n_{\min}$ (Fig. 4). The regression of the two data sets is shown in Fig. 5. The effect of the size of the data available was further studied by analyzing the subsets of the entire data set obtained by taking every third, fourth, etc., image in set 1 and one image between the successive chosen images to set 2 (Table II). The variation ranges of the regression parameters and Δ are given in Table III. The peak value of the b_1 distribution and the corresponding b_0 and Δ values are given as well. Obviously, Δ decreases with the decreasing number of images included, but for the first three cases, Δ for both channels was at most 0.5% (absolute), independently of the choice of $n_{\min}$. We conservatively estimate that the calibration accuracy of AVHRR is 3% for the red channel and 6% for the NIR channel or better [7], [16]; we find that the "intercalibration" accuracy is within the instrument accuracy.

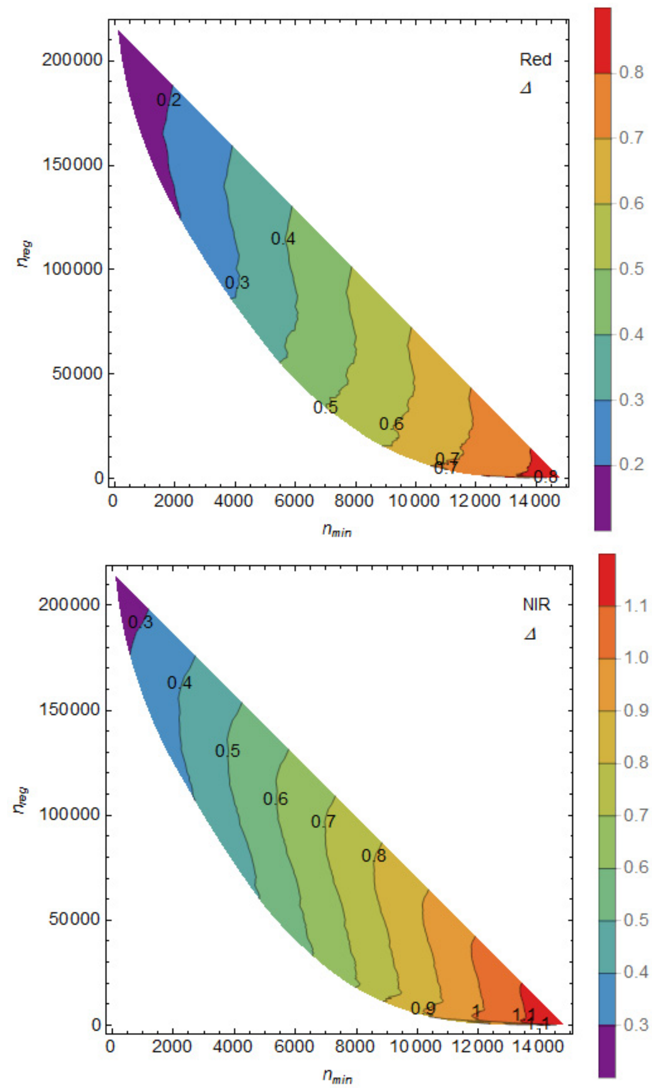


Fig. 7. Δ as a function of $n_{\min}$ and n_{reg} for the two subsets of MODIS data for the (top) red and (bottom) NIR channels. The time range is June 1, 2010–July 19, 2010.

TABLE II

NUMBER OF AVHRR IMAGES PER SATELLITE IN VARIOUS INTERCALIBRATION CALCULATIONS. THE TIME RANGE IS JUNE–JULY 2010 FOR ALL CASES

Set	Metop-A/2	NOAA-15	NOAA-18	NOAA-19	Total
Every 2nd	123/124	130/132	134/130	131/128	518/514
Every 3rd	88/93	101/93	89/86	83/89	361/361
Every 4th	60/60	76/84	79/59	56/68	271/271
Every 5th	48/45	59/63	57/61	53/48	217/217
Every 6th	40/48	58/43	47/42	36/47	181/180
Every 7th	36/27	43/39	42/45	34/44	155/155
Every 8th	33/27	38/38	40/39	25/31	136/135
Every 9th	31/31	36/28	26/29	28/32	121/120
Every 10th	24/24	33/26	27/30	25/28	109/108

C. MODIS Versus MODIS

We recently compared two nonoverlapping MODIS Terra data sets with each other for the same period as for AVHRR [9]. Since the regression line of $\langle R\rangle$, R_8, and R_{98}

TABLE III

SLOPE b_1 AND CONSTANT b_0 AND THE GOODNESS OF FIT Δ FOR THE WEIGHTED LINEAR DEMING REGRESSION OF TWO DIVERSE AVHRR DATA SUBSETS OF THE RED AND NIR CHANNELS OF NOAA-15, NOAA-18, NOAA-19, AND METOP-A/2, OF JUNE 29, 2010–JULY 19, 2010. THE MINIMUM, PEAK, AND MAXIMUM VALUES OF b_1 ARE PROVIDED FOR ROUGHLY THE VARIATION RANGE OF n_{REG}: 100 ... 211 000 AND OF n: 100 ... 22 000. THE CORRESPONDING VALUES ARE GIVEN FOR b_0 AND Δ AS WELL (CORRESPONDING TO REFLECTANCE VALUES GIVEN IN THE RANGE 0 ... 100), BUT THEIR PEAK VALUES ARE DEFINED TO MATCH THE PEAK OF b_1

Channel	Sets	b_0			b_1			Δ		
		Min	Peak	Max	Min	Peak	Max	Min	Peak	Max
Red	Every 2nd	-0.377	0.230	0.263	0.9949	0.9975	1.0015	0.06	0.11	0.30
	Every 3rd	-0.425	-0.376	-0.307	1.0024	1.0095	1.0104	0.16	0.25	0.27
	Every 4th	-0.133	-0.064	0.082	0.9975	1.0025	1.0037	0.01	0.08	0.18
	Every 5th	0.380	0.518	0.717	0.9873	1.0035	1.0096	0.16	0.69	1.16
	Every 6th	-1.034	-0.757	-0.700	1.0002	1.0065	1.0143	0.35	0.43	0.71
	Every 7th	-0.520	-0.140	0.228	1.0003	1.0025	1.0127	0.06	0.06	0.49
	Every 8th	-0.470	-0.407	-0.169	1.0033	1.0055	1.0137	0.08	0.17	0.50
	Every 9th	-0.516	-0.371	0.667	1.0001	1.0075	1.0142	0.08	0.18	0.94
	Every 10th	-2.865	-0.803	-0.738	0.9989	0.9995	1.0488	0.67	0.83	1.26
NIR	Every 2nd	-0.143	0.028	0.138	0.9931	0.9955	1.0007	0.03	0.20	0.43
	Every 3rd	-0.419	-0.393	-0.340	0.9978	1.0085	1.0101	0.16	0.21	0.45
	Every 4th	0.110	0.139	0.207	0.9937	0.9985	1.0016	0.05	0.07	0.21
	Every 5th	0.101	0.350	0.595	0.9863	0.9995	1.0071	0.08	0.33	0.91
	Every 6th	-0.825	-0.688	-0.526	0.9981	1.0065	1.0156	0.29	0.36	0.63
	Every 7th	-0.560	0.039	0.430	1.0038	1.0065	1.0236	0.18	0.36	0.79
	Every 8th	-0.470	-0.407	-0.169	1.0033	1.0055	1.0137	0.08	0.17	0.50
	Every 9th	-0.316	-0.126	0.556	1.0022	1.0095	1.0277	0.11	0.37	1.87
	Every 10th	-3.945	-1.061	-0.240	0.9975	1.0055	1.0709	0.34	0.79	1.79

of those two data sets turned out to be close to the ideal 1:1 relationship, the data sets are considered statistically both sufficiently large and distributed evenly enough over the study area to provide a reliable estimate of its reflectance distribution. Similar to AVHRR, the data of MODIS Terra and Aqua (as well) were first chronologically sorted (T&A). Then, the data set was divided into two sets by taking every second image to one set and the rest to the other set. In order to have a comparable number of pixels from the study area for both MODIS and AVHRR, we used a longer time range of June 1–July 19 for MODIS versus MODIS analysis (Section II-A). The distributions of the $\langle R_{\mathrm{Red}}\rangle$, R_{Red8}, and R_{Red98} and $\langle R_{\mathrm{NIR}}\rangle$, R_{NIR8}, and R_{NIR98} values are shown in Fig. 6 for the two independent data sets of MODIS. The number of sampled angle triplets was roughly 116 300 values. Unlike for AVHRR, Δ tends to increase with increasing n_{reg} (Fig. 7). The regression of the two data sets is shown in Fig. 8.

We studied the effect of the number of images included in the analysis (Table IV) in the same way as had been done for AVHRR. In addition, we also examined separately the cases that were based only on the Terra satellite (T&T) and the comparison of Aqua versus Terra images (A vs. T). The "intercalibration" results are shown in Table V. For all the cases based on both Aqua and Terra satellites, the peak value of b_1 corresponded to Δ values smaller than 0.5% (absolute) for both channels. No significant viewing angle dependence was noticed. The case based only on Terra MODIS was almost as good in spite of having roughly only half of the total image values in the analysis. However, the intercalibration of Aqua images versus Terra images turned out to be biased, especially in the NIR channel. Most probably, this is due to the systematic diurnal variation of the cloud cover and

TABLE IV

NUMBER OF AVHRR IMAGES PER SATELLITE IN VARIOUS INTERCALIBRATION CALCULATIONS. THE TIME RANGE IS JUNE–JULY 2010 FOR ALL CASES, EXCEPT FOR THE AQUA VERSUS TERRA SUBSET WHICH IS JUNE 29, 2010–JULY 19, 2010

Set	Terra	Aqua	Total
T&A, every 2nd	1508/1508	1534/1533	3042/3041
T&A, every 3rd	1006/1005	1022/1023	2028/2028
T&A, every 4th	754/754	767/767	1521/1521
T&A, every 5th	604/603	613/614	1217/1217
T&T, every 2nd	1508/1508		1508/1508
A vs. T	3016	3067	3016/3067
A vs. T, subset	1295	1326	1295/1326

other atmospheric components (aerosol concentration, etc.) which alters the R_{NIR98} distribution for the overpass times of these two satellites (Fig. 9). In addition, the cloud reflectance values depend much more strongly on the azimuth viewing angle in the NIR than that in the red wavelengths [24]. Since Terra is a morning satellite and Aqua an afternoon satellite, the illumination azimuth direction is systematically different. If there is a need to calibrate an afternoon satellite with a morning satellite, a good option is to exclude the intermediate reflectance values and trust only the lowest (ocean) and highest (snow) reflectances, which correspond to clear sky conditions. When the TOA reflectance values in the range 25% and 85% of Terra and Aqua MODIS were excluded from the intercalibration, the peak value of Δ decreased from 0.89 to 0.56 and from 2.36 to 0.55 for the red and NIR channels, respectively.

TABLE V

SLOPE b_1 AND CONSTANT b_0 AND THE GOODNESS OF FIT Δ FOR THE WEIGHTED LINEAR DEMING REGRESSION OF TWO DIVERSE MODIS DATA SUBSETS OF THE RED AND NIR CHANNELS OF TERRA (T) AND AQUA (A) OF JUNE 1, 2010–JULY 19, 2010. THE MINIMUM, PEAK AND MAXIMUM VALUES OF b_1 ARE PROVIDED FOR ROUGHLY THE VARIATION RANGE OF n_{REG}: 100 . . . 215 200 AND OF n: 100 . . . 14 800. THE CORRESPONDING VALUES ARE GIVEN FOR b_0 AND Δ AS WELL (CORRESPONDING TO REFLECTANCE VALUES GIVEN IN THE RANGE 0 . . . 100), BUT THEIR PEAK VALUES ARE DEFINED TO MATCH THE PEAK OF b_1

Channel	Sets	b_0			b_1			Δ		
		Min	Peak	Max	Min	Peak	Max	Min	Peak	Max
Red	T&A, every 2nd	0.259	0.678	1.829	0.9795	0.9965	0.9976	0.11	0.50	0.89
	T&A, every 3rd	-0.201	-0.160	0.285	0.9902	0.9945	0.9995	0.06	0.43	0.46
	T&A, every 4th	-0.881	-0.375	-0.234	0.9926	0.9985	1.0072	0.32	0.45	0.73
	T&A, every 5th	-0.103	0.309	0.842	0.9914	1.0025	1.0050	0.17	0.43	0.53
	T&T, every 2nd	0.478	0.793	2.646	0.9664	0.9905	0.9948	0.22	0.34	1.26
	A vs. T	-1.150	0.168	0.373	0.9932	1.0145	1.0154	0.23	0.89	1.13
	A vs. T, subset	-0.007	0.031	0.047	1.0088	1.0105	1.0113	0.43	0.56	0.56
NIR	T&A, every 2nd	0.278	0.330	2.773	0.9635	0.9995	1.0004	0.23	0.30	1.19
	T&A, every 3rd	-0.316	-0.252	0.267	0.9904	0.9985	1.0001	0.12	0.33	0.39
	T&A, every 4th	-1.150	-0.544	-0.450	1.0018	1.0035	1.0159	0.21	0.37	0.52
	T&A, every 5th	-0.013	0.629	1.275	0.9854	0.9965	0.9997	0.04	0.45	0.57
	T&T, every 2nd	0.533	0.990	2.920	0.9644	0.9915	0.9971	0.39	0.56	1.54
	A vs. T	-1.045	-0.582	0.194	0.9556	0.9645	0.9806	0.90	2.36	2.57
	A vs. T, subset	-0.310	-0.273	-0.195	0.9941	0.9945	0.9947	0.49	0.55	0.58

D. AVHRR Versus MODIS

The distributions of the $\langle R_{\mathrm{Red}}\rangle$, R_{Red8}, and R_{Red98} and $\langle R_{\mathrm{NIR}}\rangle$, R_{NIR8}, and R_{NIR98} values are shown in Fig. 10 for the AVHRR and the combined Terra/Aqua MODIS data sets of the same period. The number of sampled angle triplets was roughly 98 000. The difference between the distributions obtained using two instruments is larger in the NIR than in the red channel as expected (since the NIR channel wavelength ranges differ more than those of the red channels).

Δ of the weighted Deming regression depends on both $n_{\min}$ and n_{reg} (Figs. 4 and 7), which cannot be chosen independently. However, from the point of view of statistical representativeness, it is recommended that at least 500 points be included in the regression and at least 2000 points in the distributions. To avoid a subjective choice of the $n_{\min}$ and n_{reg} values to be used as the basis for the intercalibration parameters, it was decided to use the peak value of the b_1 distribution for the slope and then derive the constant by the interpolation of the closest b_0 values corresponding to regression lines with slopes nearest the peak value of b_1. The variation range and peak values of these regression parameters for the AVHRR versus MODIS calibration are given in Table VI. The relationship between the $\langle R_{\mathrm{Red}}\rangle$, R_{Red8}, and R_{Red98} and $\langle R_{\mathrm{NIR}}\rangle$, R_{NIR8}, and R_{NIR98} values of MODIS and AVHRR is shown in Fig. 11. The biased scatter of the points is dominated by clouds, but an ideal 1:1 pointwise relationship would not be obtained even for the surface reflectance values, because the points of MODIS and AVHRR corresponding to the same angle triplet do not, in general, come from the same place or the same time. The data set was too small to enable the reliable analysis of the relationship between the (θ_s, θ_v, ϕ) values and the regression parameter values.

TABLE VI

DEMING REGRESSION PARAMETER VARIATION RANGE FOR AVHRR VERSUS MODIS

Channel	Sets	b_0			b_1		
		Min	Peak	Max	Min	Peak	Max
Red	AVHRR vs. MODIS	0.477	1.948	2.694	1.015	1.021	1.032
NIR	AVHRR vs. MODIS	-2.342	0.397	2.881	1.023	1.058	1.125

V. DISCUSSION

The advantage of the presented intercalibration method is the possibility of using the entire range of TOA reflectance values and sun and satellite angles as the basis for intercalibration. This means that possible nonlinear features of the instrument relationship will be revealed. In addition, one does not have to assume that the calibration derived for nadir viewing will apply also for other satellite angles. Hence, the derived method can be used as an additional test for the angular reliability of the SNO-based calibration. It is also possible to derive a dedicated calibration for a certain land cover type or area or sun angle range, when utmost calibration accuracy is sought for a special case. Further on, since no cloud masking is needed, the intercalibration quality will not depend on the varying cloud-masking capability of diverse instruments. On the other hand, the methods can also be applied to cloud-masked data, if needed. Even atmospherically corrected data, i.e., surface reflectance values, could be applied in this calibration method, if one is confident enough about their quality.

It is also not necessary to have a continuous test site. The ocean and snow peaks of the distribution (low and high ends) are crucial. In principle, one could use the Greenland ice sheet, Antarctica, and some ocean area, but mid-reflectance values (obtainable in other PICSs) are also useful for assessing the linearity. Furthermore, no low sun zenith angles are available over Antarctica or Greenland. The essence of calibration is that one should use the same dynamic

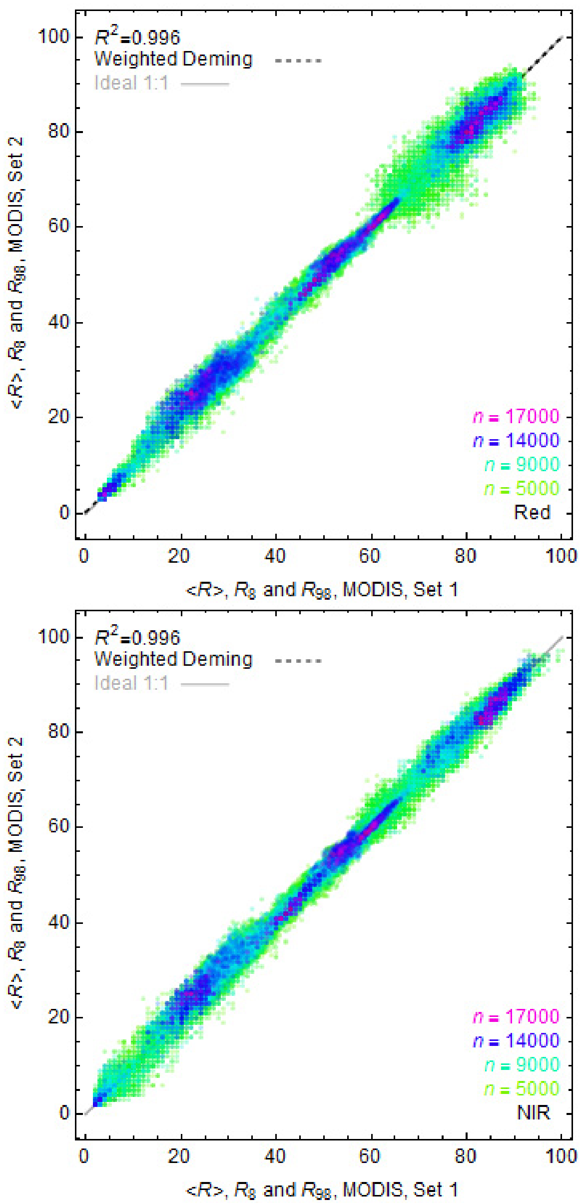


Fig. 8. Relationship of (top) $\langle R_{\text{Red}} \rangle$, R_{Red8}, and R_{Red98} and (bottom) $\langle R_{\text{NIR}} \rangle$, R_{NIR8}, and R_{NIR98} of the set 2 versus set 1 of the combined MODIS Terra and Aqua data sets for points with $n_{\text{min}} \geq 5000$. The colors of the points are related to the number of individual reflectance values (n) in the distributions from which the values for $\langle R_{\text{Red}} \rangle$, R_{Red8}, R_{Red98}, $\langle R_{\text{NIR}} \rangle$, R_{NIR8}, and R_{NIR98} are derived.

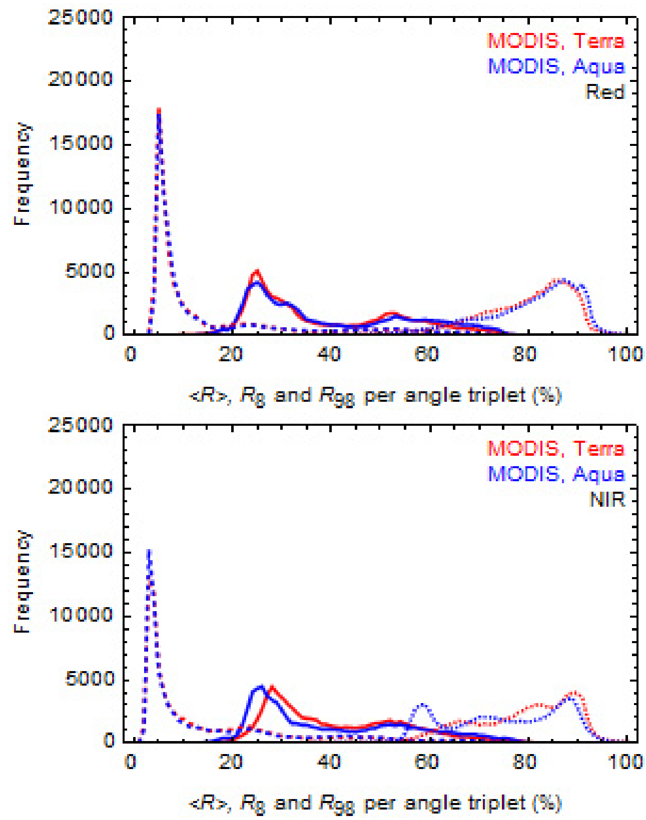


Fig. 9. Distributions of the MODIS Terra and Aqua (top) red and (bottom) NIR channel TOA reflectance distribution means (solid curves) and 8% (dashed curves) and 98% quantiles (dotted curves) corresponding to sampled angle triplets (θ_s, θ_v, ϕ). The time range is June 1, 2010–July 19, 2010.

range of parameters for calibration as to which the calibration will be applied. The ideal solution for a general purpose calibration would be to use the whole earth as the study area, but computational requirements for such an intercalibration are very high. However, if one is specifically interested in some special target type (such as vegetation or the ocean), one could make a dedicated calibration for that purpose using only related areas as the basis for calibration.

As the method described is based on assuming random distributions, one has to be cautious when applying it in the cases that deviate clearly from the calibration design. Sun-synchronous satellites actually do not randomly sample. The randomness comes (in a positive case) from the larger area, a longer period (varying cloud cover), and the wide swaths. Orbital and diurnal differences of the satellites carrying the sensors to be intercalibrated may pose a difficulty for that. Calibration of afternoon satellites with morning satellites and vice versa is especially sensitive to any systematic diurnal variation of cloud cover. In addition, one should pay attention to the combinations (θ_s, θ_v, ϕ) available for the calibration. If some angle combinations dominate, the random sampling assumption is questioned. Then, one should check whether the calibration parameters depend on the angles or not. Another point of concern is whether the geographical distribution of the TOA reflectance values included in the calibration is really representative so that no single area is dominating. One alternative to guarantee this, would be using suppressive weights in the TOA reflectance distributions for points coming from dominating areas of the data set.

The basic assumption of the intercalibration is, naturally, that the spectral response functions of the sensors match well enough. In the case a nonlinear relationship between the two sensors appears, it would need extension of this method to nonlinear Deming regression, but the applicability of that approach cannot be guaranteed on the basis of this paper.

Instead of starting with TOA reflectance values, one could also apply this approach directly to radiances. Since the calibration is based on TOA reflectance distributions per angle triplet (θ_s, θ_v, ϕ) and the TOA reflectance is related to the radiance by the square of the earth–sun distance divided

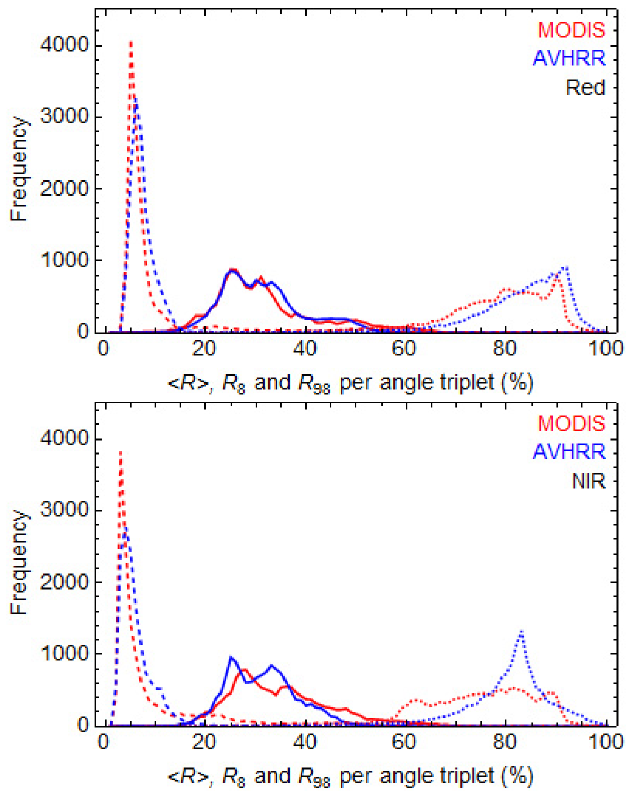


Fig. 10. (Top) Red and (bottom) NIR channel TOA reflectance distribution means (solid curves) and 8% (dashed curves) and 98% quantiles (dotted curves) corresponding to the same angle triplets for all MODIS and AVHRR data of June 29, 2010–July 19, 2010.

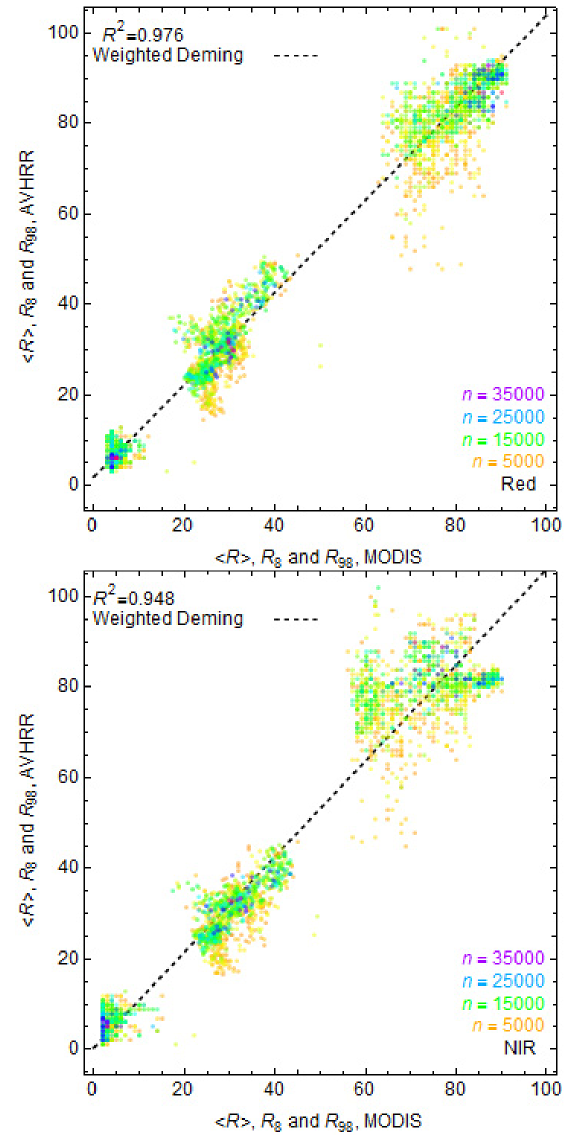


Fig. 11. Regression of (top) $\langle R_{\mathrm{Red}}\rangle$, R_{Red8}, and R_{Red98} and (bottom) $\langle R_{\mathrm{NIR}}\rangle$, R_{NIR8}, and R_{NIR98} for the AVHRR data set versus the combined MODIS Terra and Aqua data sets for $n_{\mathrm{min}} = 4000$. The colors of the points are related to the number of individual reflectance values (n) in the distributions from which the values for $\langle R_{\mathrm{Red}}\rangle$, R_{Red8}, R_{Red98}, $\langle R_{\mathrm{NIR}}\rangle$, R_{NIR8}, and R_{NIR98} are derived.

by $\cos(\theta_s)$, the TOA reflectance and corresponding radiance distributions per angle triplet (θ_s, θ_v, ϕ) differ by only a constant coefficient, if all images are first normalized to a reference earth–sun distance. In addition, one could extend this method to start with the original digital numbers (DNs) provided by the sensors, but then one should derive all DN distributions per quintets of (θ_s, θ_v, ϕ, gain, and offset), where the sensor gain and offset are assumed to be known. And if the two sensors to be intercalibrated were not similar, one would need a septet of parameters including different gain and offset values for the two sensors. Naturally, then the required amount of data would be much larger to compensate for the two/four additional variables. It is anticipated that this kind of an effort would be worth making using the current computational capacity only, when the goal is a really high-precision intercalibration of similar sensors.

One could also hypothesize that it would be possible to extend the method to completely nonoverlapping temporal/spatial data sets of the past, but then one has to have some ancillary data to prove that the calibration sites have not changed markedly during the time in question. For the ocean and the central parts of large ice sheets, this might be the case, although the stability of the ice sheets is currently in question [26].

Δ derived for MODIS versus MODIS and AVHRR versus AVHRR indicates the intercalibration accuracy that is achievable using the presented method. However, the actual accuracy of the AVHRR versus MODIS intercalibration is not as obvious. One indicator is the variation of the slope and constant with the choice of points to be included in the regression. The b_1 values at the ascending and descending edges of the half-value peak of the slope distribution were for the red channel 1.0185 and 1.0235. Their difference is just 0.5% of the peak value of b_1. The corresponding variation range for the NIR channel was 1.0515 . . . 1.0635, which is about 1% of the peak value of b_1. The b_0 values corresponding to the b_1 values within the distribution half-level peak varied in the range 1.799 and 2.337. In the NIR channel, the corresponding variation range of b_0 was 0.396 . . . 1.772. Thus, the slope of the regression is relatively stable, but the constant term varies more randomly and its relationship to the slope is not monotonic. It has also to be considered that the quality of the internal intercalibration of the diverse satellite instrument families united as one set (such as AVHRR from several satellites) that is to be intercalibrated with another set also affects the intercalibration quality of those two instrument families. In addition, the data set used was not large enough to support checking the possible dependence of the calibration parameters on the sun and satellite angles. A reliable multivariate analysis of that would be a topic of another study.

VI. Conclusion

A statistical intercalibration method based on nonsimultaneous retrievals of TOA reflectance value over a large area covering a wide variety of earth surface classes (including the ocean, snow, and both cloudy and noncloudy situations) produced good results, when applied to the intercalibration of the same instrument (MODIS versus MODIS or AVHRR versus AVHRR). The achievable accuracy was better than 0.5% for both instruments and both NIR and red channels (reflectances from 0% to 100%). The accuracy improved relative to the number of available points in the distributions. The AVHRR reflectance values turned out to be somewhat larger than the MODIS reflectance values by about 2% in the red channel and by about 6% in the NIR channel. Yet, the differences are generally within the limits of the instrument accuracies.

Acknowledgment

The authors would like to thank the WMO SCOPE-CM scientific executive panel for its interest and guidance for the work. The views, opinions, and findings contained in this paper are those of the authors and should not be construed as an official National Oceanic and Atmospheric Administration or the U.S. Government position, policy, or decision.

References

[1] G. Chander, T. J. Hewison, N. Fox, X. Wu, X. Xiong, and W. J. Blackwell, "Overview of intercalibration of satellite instruments," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1056–1080, Mar. 2013.

[2] P. J. Mason and A. Simmons, Eds., GCOS SC, WMO, Geneva, Switzerland. (2011). "Systematic observation requirements for satellite-based products for climate, 2011 update, supplemental details to the satellite-based component of the implementation plan for the global observing system for climate in support of the UNFCCC (2010 update)." Tech. Rep. GCOS-154, p. 138. [Online]. Available: http://www.wmo.int/pages/prog/gcos/Publications/gcos-154.pdf

[3] Y. M. Govaerts and M. Clerici, "Evaluation of radiative transfer simulations over bright desert calibration sites," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 1, pp. 176–187, Jan. 2004.

[4] D. Helder *et al.*, "Absolute radiometric calibration of landsat using a pseudo invariant calibration site," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1360–1369, Mar. 2013.

[5] A. K. Heidinger, C. Cao, and J. T. Sullivan, "Using Moderate Resolution Imaging Spectrometer (MODIS) to calibrate advanced very high resolution radiometer reflectance channels," *J. Geophys. Res.*, vol. 107, no. D23, pp. AAC11-1–AAC11-10, 2002.

[6] E. F. Vermote and N. Z. Saleous, "Calibration of NOAA16 AVHRR over a desert site using MODIS data," *Remote Sens. Environ.*, vol. 105, no. 3, pp. 214–220, 2006.

[7] A. K. Heidinger, W. C. Straka, III, C. C. Molling, J. T. Sullivan, and X. Wu, "Deriving an inter-sensor consistent calibration for the AVHRR solar reflectance data record," *Int. J. Remote Sens.*, vol. 31, no. 24, pp. 6493–6517, 2010.

[8] A. Wu, X. Xiong, and A. Angal, "Derive a MODIS-based calibration for the AVHRR reflective solar channels of the NOAA KLM operational satellites," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1405–1413, Mar. 2013.

[9] T. Manninen, A. Riihelä, C. Schaaf, J. Key, and A. Lattanzio, "Intercalibration of two polar satellite instruments without simultaneous nadir observations," in *Proc. ESA Living Planet Symp.*, 2016, p. 8.

[10] S. Bontemps, P. Defourny, E. V. Bogaert, O. Arino, V. Kalogirou, and J. R. Perez, "GLOBCOVER 2009 Product description and validation report," Eur. Space Agency, Paris, France, Tech. Rep., 2011, p. 53.

[11] A. Riihelä, P. Lahtinen, and T. Hakala, "The radiation, snow characteristics and albedo at Summit (RASCALS) expedition report," Finnish Meteorol. Inst., Helsinki, Finland, Tech. Rep. 2011:8, 2011, p. 41.

[12] A. K. Heidinger, M. J. Foster, A. Walther, and X. Zhao, "The pathfinder atmospheres–extended AVHRR climate dataset," *Bull. Amer. Meteorol. Soc.*, vol. 95, no. 6, pp. 909–922, 2014.

[13] X. Xiong, J. Sun, and W. Barnes, "Intercomparison of on-orbit calibration consistency between Terra and Aqua MODIS reflective solar bands using the Moon," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 778–782, Oct. 2008.

[14] R. N. Clark *et al.* (2007). USGS digital spectral library splib06a: U.S. Geological Survey, Digital Data Series 231. U.S. Geological Survey, Denver, CO, USA. Accessed: Nov. 12, 2012. [Online]. Available: http://speclab.cr.usgs.gov/spectral.lib06

[15] T. Manninen, A. Riihelä, and G. de Leeuw, "Atmospheric effect on the ground-based measurements of broadband surface albedo," *Atmos. Meas. Techn.*, vol. 5, no. 11, pp. 2675–2688, 2012.

[16] C. C. Molling, A. K. Heidinger, W. C. Straka, III, and X. Wu, "Calibrations for AVHRR channels 1 and 2: Review and path towards consensus," *Int. J. Remote Sens.*, vol. 31, no. 24, pp. 6519–6540, 2010. [Online]. Available: http://dx.doi.org/10.1080/01431161.2010.496473

[17] D. A. Hastings and W. J. Emery "The advanced very high resolution radiometer (AVHRR)—A brief reference guide," *Photogramm. Eng. Remote Sens.*, vol. 58, no. 8, pp. 1183–1188, Aug. 1992.

[18] V. V. Salomonson, W. Barnes, P. W. Maymon, H. E. Montgomery, and H. Ostrow, "MODIS: Advanced facility instrument for studies of the Earth as a system," *IEEE Trans. Geosci. Remote Sens.*, vol. 27, no. 2, pp. 145–153, Mar. 1989.

[19] C. H. Kummell, "Reduction of observation equations which contain more than one observed quantity," *Analyst*, vol. 6, no. 4, pp. 97–105, 1879, doi: 10.2307/2635646.

[20] R. J. Carroll and D. Ruppert, "The use and misuse of orthogonal regression in linear errors-in-variables models," *Amer. Statist.*, vol. 50, no. 1, pp. 1–6, 1996.

[21] K. Linnet, "Evaluation of regression procedures for methods comparison studies," *Clin. Chem.*, vol. 39, no. 3, pp. 424–432, 1993.

[22] R. F. Martin, "General Deming regression for estimating systematic bias and its confidence interval in method-comparison studies," *Clin. Chem.*, vol. 46, no. 1, pp. 100–104, 2000.

[23] I. Markovsky and S. Van Huffel, "Overview of total least-squares methods," *Signal Process.*, vol. 87, no. 10, pp. 2283–2302, 2007.

[24] E. Jäkel, J. Walter, and M. Wendisch, "Thermodynamic phase retrieval of convective clouds: Impact of sensor viewing geometry and vertical distribution of cloud properties," *Atmos. Meas. Techn.*, vol. 6, no. 3, pp. 539–547, 2013.

[25] X. Xiong, B. N. Wenny, and W. L. Barnes, "Overview of NASA Earth Observing Systems Terra and Aqua moderate resolution imaging spectroradiometer instrument calibration algorithms and on-orbit performance," *J. Appl. Remote Sens.*, vol. 3, no. 1, p. 032501, 2009.

[26] S. V. Nghiem *et al.*, "The extreme melt across the Greenland ice sheet in 2012," *Geophys. Res. Lett.*, vol. 39, no. 20, p. L20502, 2012.

Terhikki Manninen received the D.Sc. (Tech) degree in physics from the Helsinki University of Technology, Espoo, Finland, in 1996.

Since 2003, she has been a Senior Research Scientist with the Finnish Meteorological Institute, Helsinki, Finland, where she is currently the Head of the Satellite and Radar Applications Group, Meteorological Research Unit. Her research interests include the surface description of the earth, such as broadband surface albedo and leaf area index.

Aku Riihelä received the M.Sc. (Tech) degree from the Helsinki University of Technology, Espoo, Finland, in 2007, and the D.Sc. (Tech) degree from Aalto University, Espoo, in 2014.

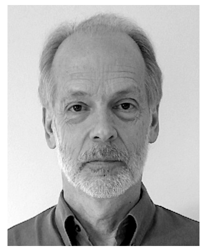
Since 2006, he has been with the Finnish Meteorological Institute, Helsinki, Finland, where has been a Tenure Track Research Professor since 2016. His research interests include the spaceborne retrievals of surface albedo and the surface radiative energy budget of the high latitudes, and the validation of said retrievals.

Andrew Heidinger, photograph and biography not available at the time of publication.

Crystal Schaaf, photograph and biography not available at the time of publication.

Alessio Lattanzio received the bachelor's/master's combined degree in physics from the University of Rome "La Sapienza," Rome, Italy, in 1997.

He has been involved in the field of remote sensing since 1998. Since 2012, he has been leading the SCOPE-CM Project on retrieval of surface albedo from geostationary satellites. Since 2015, he has been a Co-Lead of the surface radiation focus area with the CEOS Land Product Validation Subgroup. He is currently involved in the generation and validation of Thematical Climate Data Records from Meteosat satellites. His research interests include the retrieval of geophysical variables such as surface albedo, aerosol, fire radiative power, and atmospheric motion vectors.



Jeffrey Key received the Ph.D. degree in geography from the University of Colorado Boulder, Boulder, CO, USA, in 1988, with a specialization in satellite meteorology and climatology of the polar regions.

He is currently the Chief of the Advanced Satellite Products Branch, National Oceanic and Atmospheric Administration's Satellite and Information Service, Madison, WI, USA. He serves on several international panels, and is involved in activities of the World Meteorological Organization, Geneva, Switzerland, and the Group on Earth Observations, Geneva.

Dr. Key is a member of the American Meteorological Society and the American Geophysical Union.