

INTERCALIBRATION OF TWO POLAR SATELLITE INSTRUMENTS WITHOUT SIMULTANEOUS NADIR OBSERVATIONS

Terhikki Manninen⁽¹⁾, Aku Riihelä⁽¹⁾, Crystal Schaaf⁽²⁾, Jeffrey Key⁽³⁾, Alessio Lattanzio⁽⁴⁾

⁽¹⁾ Finnish Meteorological Institute, P.O. Box 503, FI-00101 Helsinki, Finland,
Email: Terhikki.Manninen@fmi.fi; Aku.Riihela@fmi.fi

⁽²⁾ University of Massachusetts Boston, 100 Morrissey Blvd., Boston, MA 02125-3393, USA,
Email: Crystal.Schaaf@umb.edu

⁽³⁾ NOAA/NESDIS, 125 West Dayton Street, Madison, WI 53706, USA, Email:
jkey@ssec.wisc.edu

⁽⁴⁾ EUMETSAT, Eumetsat Allee 1, 64295 Darmstadt, Germany, Email:
Alessio.Lattanzio@eumetsat.int

ABSTRACT

A new intercalibration method for two polar satellite instruments is presented. It is based on statistical fitting of two data sets covering the same area during the same period, but not simultaneously. Deming regression with iterative weights is used. The accuracy of the method was better than about 0.5 % for the MODIS vs. MODIS and AVHRR vs. AVHRR test data sets. The intercalibration of AVHRR vs. MODIS red and NIR channels is carried out and showed a difference of reflectance values of 2% (red) and 6 % (NIR). The red channel intercalibration has slightly higher accuracy for all cases studied.

1. INTRODUCTION

Surface albedo is one of the essential climate variables (ECV) and a key parameter for the energy balance of the Earth. Albedo retrievals are usually performed utilizing only a single instrument or an instrument family. Combining multiplatform observations is expected to yield improvements in both the accuracy and temporal resolution of surface albedo retrievals. This study is carried out in 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 optical imagers, whose strengths are in high data acquisition rates over the high latitudes of Earth, which play a key role in the climate change.

An established intercalibration approach of two satellite instruments is to use simultaneous nadir observations (SNO). As the goal here is to find a method universally applicable to polar orbiting satellites, one has to take into account the possibility of satellites that never observe the same place simultaneously, so that the use

of SNOs is impossible. Our solution is to derive the top of atmosphere (TOA) reflectance distributions of a large area containing the entire dynamic range of global TOA reflectance. At this phase we concentrate on 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 satellites, at first using two separate data sets of MODIS and two separate data sets of AVHRR in order to test the intercalibration method. The advantage in first studying MODIS vs. MODIS and AVHRR vs. AVHRR intercalibrations is that it is known in advance that one should obtain a 1:1 relationship for the linear regression. Hence, the goodness of the proposed method can be assessed separately before applying it to real data. After proving the method, we apply it to intercalibrate MODIS and AVHRR TOA reflectance data.

2. STUDY AREA AND MATERIALS

2.1. Study area and period

In order to reduce the data mass to be handled, a subset of the globe was defined to be used as the basis for the calibration. The study area contained a wide variety of land cover classes and a large ocean area to be spectrally representative of the globe. The borders of the study area are latitudes 0° N and 75° N and longitudes -130° E and 45° E [Figure 1]. The site is studied during the period June 29 – July 19, because for that time range the preprocessed AVHRR data set is available and the RASCALS campaign [1] took place during that time. The RASCALS goniospectrometer measurements were compared with atmospherically corrected MODIS and

AVHRR spectral surface reflectances to define angular cut-off limits for the intercalibration and subsequent joint albedo retrieval.



Figure 1. A simulated example of TOA reflectance distribution. The study area is shown on the GlobCover map [2].

2.2. MODIS data

MOD02 and MOD03 data produced by the Terra satellite were used for the study as the basic reference set. Additionally the corresponding MYD02 and MYD03 data sets from the Aqua satellite were used for comparison. The MOD02 products contain calibrated and geolocated at-aperture radiances in $W/(m^2-\mu m-sr)$. Top of atmosphere (TOA) reflectance values were determined for the bands 1 (620...670 nm; “red”) and band 2 (841...876 nm; “NIR”) bands through knowledge of the solar irradiance [3, 4]. The spatial resolution for this data is 250 m. The MOD03 and MYD03 products contain the geodetic coordinates, and solar and satellite zenith, and azimuth angle values for each MODIS 1-km sample.

2.3. AVHRR data

AVHRR observations from NOAA-15, -18, -19, and METOP-A were obtained from NOAA Pathfinder Atmospheres–Extended [5, 6] archives for this study. The observations were intercalibrated following the approach by Heidinger et al. (2010), and provided for use as TOA reflectances. Global Area Coverage (GAC) data with a spatial resolution of ~4 kilometres (sub-nadir) from AVHRR channels 1 (580...680 nm; “red”) and 2 (725...1000 nm; “NIR”) were utilized. The PATMOS-X data viewing and illumination geometry information was also used here.

3. METHODS

3.1. Basic idea and hypothesis

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 TOA reflectance

distribution of the study area. One has to take into account that the TOA reflectance has diurnal variation related to the sun zenith angle θ_s and the atmospheric and weather conditions (especially cloud cover). In addition, 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 not typically 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 angular resolution used is one degree. Each 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 ocean and snow. Linear regression is sought for a combination of these statistical parameters of the two 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, thus expanding the determination range improves the robustness of the regression.

The hypothesis is that if 1) the amount of data is statistically sufficient, 2) the observation period long enough and 3) the area large enough to provide unbiased diurnal sampling of diverse land cover classes, the effect of the atmospheric and land cover changes on the TOA reflectance should produce a nearly equal distribution for a similar wavelength range.

One question to solve is, what the minimum number of points allowed in an individual distribution is. Is it, for example, better to include in the regression also $\langle R \rangle$, R_8 and R_{98} values based on just one single TOA reflectance value (so that $R_8 = \langle R \rangle = R_{98}$) or not? Normally linear regression results improve, when the number of points increases, but introducing a very heteroscedastic data set while maximizing the number of points may not be an advantage.

The goodness of the method is first tested by ‘intercalibrating’ two independent data sets produced by the same instrument (set). The parameter to be used for describing the goodness of the fit in this context is the average deviation (Δ) of the regression line from the ideal 1:1 line, which is calculated from

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

where b_0 is the constant and b_1 the linear regression coefficient. The smaller Δ is the better the fit is and the size of Δ directly describes the mean accuracy of the intercalibration of the reflectance values.

3.2. Spatial resolution

The spatial resolution of remote sensing images has a marked effect on comparison of images of the same area. Because bright and dark targets can be situated in the image in neighbouring pixels (clouds and water, snow and water, forests and deserts etc.), the relationship between the average value and the original pixels may be almost any kind, and definitely one can't automatically expect it to resemble the 1:1 relationship. Therefore, the MODIS data was averaged to the spatial resolution of 5 km.

3.3. Spectral difference

The red channel wavelength range of the MODIS and AVHRR instruments are rather close to each other. Hence it is not surprising that R_{AVHRR} are only slightly smaller than R_{MODIS} in the red channel for surface targets included in the USGS spectral library [7, 8]. For NIR channel the relationship is not as good for snow/water and vegetation spectra, but the coefficient of determination R^2 is high. In the NIR channel the R_{AVHRR} values are typically smaller than the corresponding R_{MODIS} values, their ratio varying in the range 90.8% ... 96.4% for the spectra available. This means that one general relationship for all targets will not be very accurate (not better than about $\pm 3\%$). However, the absolute radiometric accuracy of the AVHRR instrument is not estimated to be better than that [9]. Hence we at this stage accept having a general target-independent calibration coefficient also for the NIR bands of MODIS and AVHRR.

3.4. Angular uncertainty

The AVHRR is a whiskbroom type instrument with scanning mirror and single discrete detector per band. The angular rotating accuracy of the mirror may cause an individual element to the reflectance estimation accuracy. In addition, the obliquely viewed and nadir viewed pixels have different footprints on the ground. The obliquely viewed pixels of the same latitude and longitude have different footprints in descending and ascending orbits. Thus the reflectance estimation accuracy may vary more at the edges of the image.

The MODIS instrument has a linear array of detectors with one array per band (pushbroom). Then the detector response is uniform in along-track direction, but may have individual statistics in across-track direction. Again, the footprint of ascending and descending orbit is not identical for the nominally same pixel.

3.5. Regression

Commonly used linear regression is based on the least-squares minimization of the sum of the squared vertical distances from the data points to the fitting line. Then 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 vs. each other) this approach is not well grounded, because both data sets have the same instrument inaccuracy. Then the recommended regression method to use is orthogonal regression [10, 11].

The actual intercalibration of AVHRR vs. MODIS data requires paying attention to the different uncertainties of the MODIS and AVHRR data. For this purpose Deming regression is suitable [12, 13, 14]. The ratio of the variances of the AVHRR and MODIS regression point values was taken to be 2 and 3, for the red and NIR channels respectively [9]. 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 be taken into account also in Deming regression by using individual weights for the points [12, 13]. The weighting scheme without constraints on random error of the test or comparative method is used in this study [13]. The individual weights are improved iteratively requiring the relative difference of the regression parameters of successive iteration rounds to be smaller than 0.0001.

Firstly, the weights are needed to take into account that the uncertainty of the TOA reflectance distribution mean and quantiles decrease with increasing number of points in the distribution. Secondly, the obliquely viewed pixels have larger uncertainty than the nadir pixels due to the footprint coverage variation.

4. RESULTS

4.1. TOA reflectance distributions

To check the applicability of the statistical approach for intercalibration we first regressed the TOA reflectance values for MODIS vs. MODIS and for AVHRR vs. AVHRR. The whole image sets of MODIS and AVHRR were split in TOA reflectance distributions corresponding to various viewing/illumination configuration angle triples. Each distribution 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 is the more reliable the distribution mean is statistically [Figure 2]. On the other hand, the number of points (i.e.

the $\langle R \rangle$, R_8 and R_{98} values corresponding to the diverse angle triplets) in the intercalibration regression should be as large as possible. The effect of these two edge constraints on the goodness of the regression is demonstrated in Sections 4.2. and 4.3.

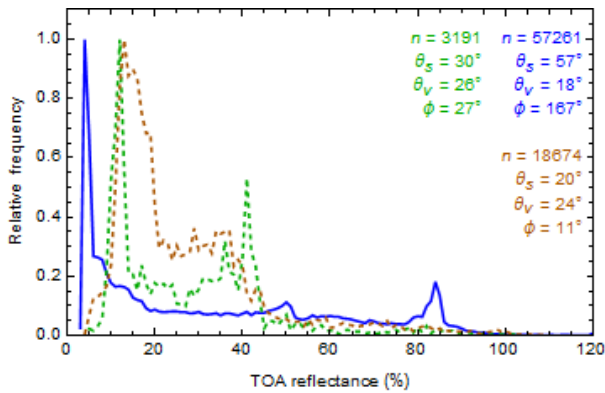


Figure 2. Distributions of the NOAA-15, NOAA-18, NOAA-19 and METOP-A/2 AVHRR red channel TOA reflectance values corresponding to example angle triplets. The number of individual TOA reflectance values included in the above distributions is n .

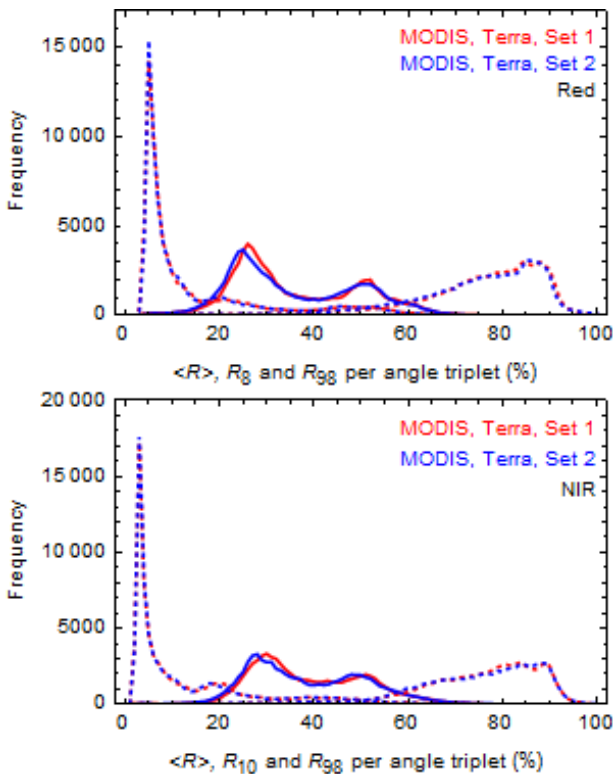


Figure 3. Distributions of the two separate sets of MODIS Terra (top) and NIR (bottom) channel TOA reflectance distribution means (solid curves) and 8% and 98% quantiles (dashed curves) corresponding to diverse angle triplets $(\theta_s, \theta_v, \phi)$.

4.2. MODIS vs. MODIS

First the data corresponding to one instrument on one satellite, MODIS Terra, is divided in two non-overlapping sets by placing every second image to one set and the rest to the other set. Then the TOA reflectance values of Set 1 corresponding to a certain angle triplet $(\theta_s, \theta_v, \phi)$ do not, in general, come from the same places as those of Set 2.

The distributions of the $\langle R_{\text{Red}} \rangle$, $R_{\text{Red}8}$, $R_{\text{Red}98}$ and $\langle R_{\text{NIR}} \rangle$, $R_{\text{NIR}8}$, $R_{\text{NIR}98}$ values are shown in Figure 3 for the two independent data sets of MODIS Terra. Each $\langle R \rangle$, R_8 , and R_{98} value corresponds to an individual R_{TOA} distribution, such as presented in Figure 2. The number of diverse angle triplets was roughly 78000. If the regression line of those statistical parameters of the two data sets is close to the ideal 1:1 relationship, the data sets are sufficiently large statistically and distributed evenly enough over the study site to provide a reliable estimate of its reflectance distribution.

The results shown in Table 1 are derived using a roughly optimal choice $n_{\text{min}} = 2000$, which corresponded to

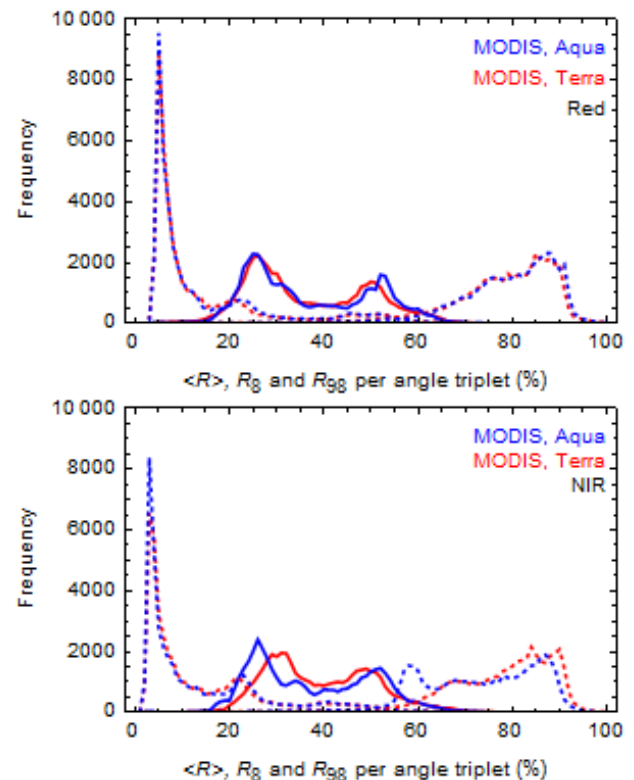


Figure 4. Distributions of the MODIS Terra and Aqua red (top) and NIR (bottom) channel TOA reflectance distribution means (solid curves) and 8% and 98% quantiles (dashed curves) corresponding to diverse angle triplets $(\theta_s, \theta_v, \phi)$.

$n_{reg} = 26279$ (red) and $n_{reg} = 23852$ (NIR), but the exact choice of n_{min} is not critical in this case as Δ varied in the range 0.237...0.411 (red) and 0.342...0.499 (NIR).

Since the ‘intercalibration’ of the instrument with itself produced good results, the next step is to test the method for comparison of the same instrument type on 2 satellites, Aqua and Terra. This test provides information about the best achievable accuracy for the intercalibration of two different instruments on several satellites. The statistical distributions are shown in Figure 4. The number of diverse angle triplets was roughly 81000.

The NIR distributions of Aqua and Terra differ more than the Red distributions, probably because the cloud reflectance depends much more strongly on the azimuth viewing angle in the NIR than in the red wavelengths [15] and Terra is a morning satellite and Aqua an afternoon satellite [16], so that the illumination azimuth direction is systematically different for them.

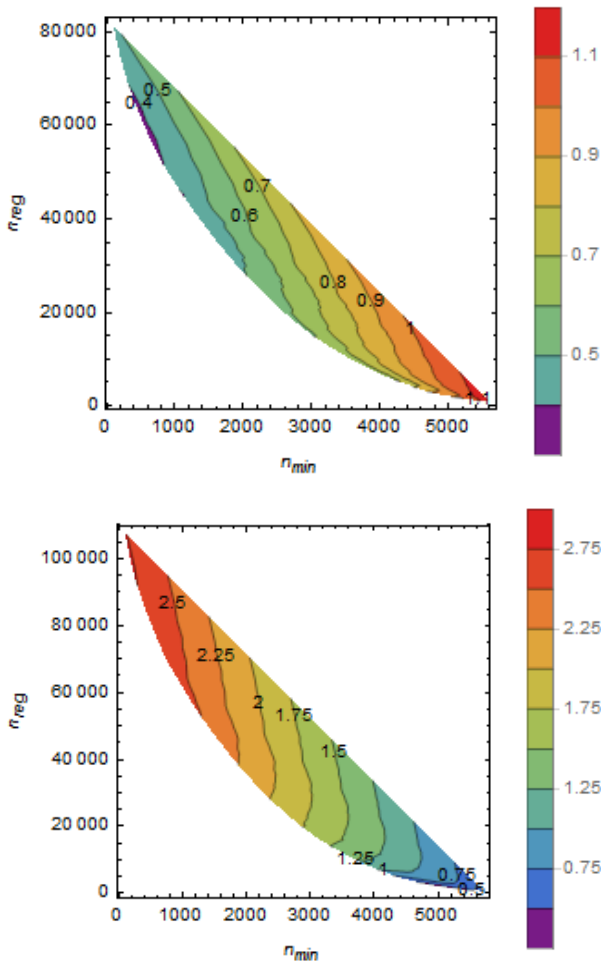


Figure 5. The dependence of Δ on n_{min} and n_{reg} for the red (top) and NIR (bottom) channels of the case MODIS Aqua vs. MODIS Terra.

The goodness of the weighted Deming regression (Δ) depends both on the minimum number of points (n_{min}) required in the distribution corresponding to an individual regression point and the number of points available for the regression (n_{reg}) [Figure 5]. These two numbers can't be chosen independently. For the MODIS Terra and Aqua data sets large n_{reg} produces good results for the red channel and large n_{min} for the NIR channel. Yet, it is considered to be safest to include at least 500 points in the regression and at least 2000 points in the distributions. The regression parameters for MODIS Aqua vs. Terra were determined using n_{min} , for the red channel $n_{min} = 2000$, which corresponds to $n_{reg} = 28634$ and for the NIR channel $n_{min} = 5700$, which corresponds to $n_{reg} = 627$ (Table 1, Figure 6). The R^2 value for the regression of MODIS Aqua vs. Terra is high, and the Deming regression is close to the ideal relationship for both channels, although there is a clear bias in the

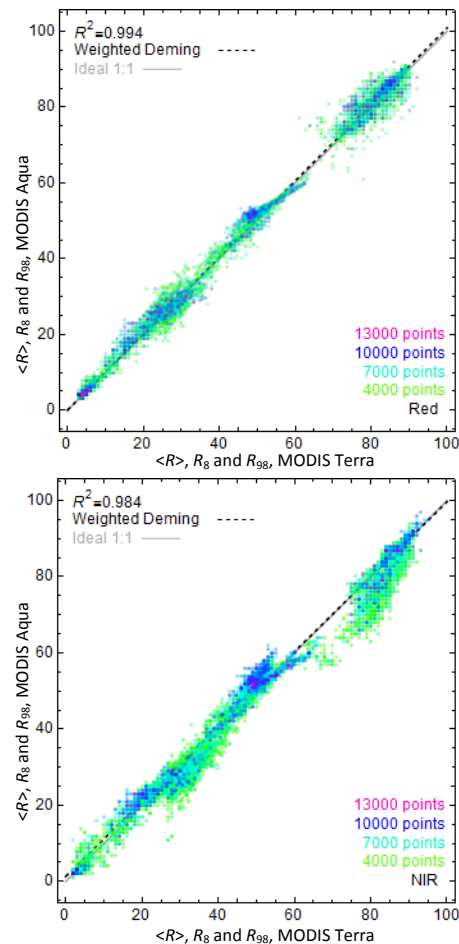


Figure 6. Relation of MODIS Aqua vs. MODIS Terra $\langle R_{Red} \rangle$, R_{Red8} and R_{Red98} (top) and $\langle R_{NIR} \rangle$, R_{NIR8} and R_{NIR98} (bottom) for $n_{min} = 4000$. The weighted Deming regression is based on $n_{min} = 2000$ (Red) and $n_{min} = 5700$ (NIR).

Table 1. Deming regression parameters and goodness of fit for MODIS of the red and NIR channels.

MODIS	Channel	b_0	b_1	Δ
Terra Set 2 vs. Set 1	Red	-0.551	1.007	0.241
	NIR	-0.525	1.003	0.361
Aqua vs. Terra	Red	0.164	1.006	0.478
	NIR	1.356	0.983	0.577

afternoon vs. morning cloud cover reflectance that affects R_{NIR98} . 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, because the points of Set 1 and Set 2 corresponding to the same angle triplet do not come from the same place.

4.3. AVHRR vs. AVHRR

The distributions of the $\langle R_{Red} \rangle$, R_{Red8} , R_{Red98} and $\langle R_{NIR} \rangle$, R_{NIR8} , R_{NIR98} values are shown in Figure 7 for two AVHRR data subsets. The number of diverse angle triplets was roughly 82000. The first set consists of odd images and the second even images of the whole chronologically ordered AVHRR data set. It turned out that the best fit is obtained using a combination that is in

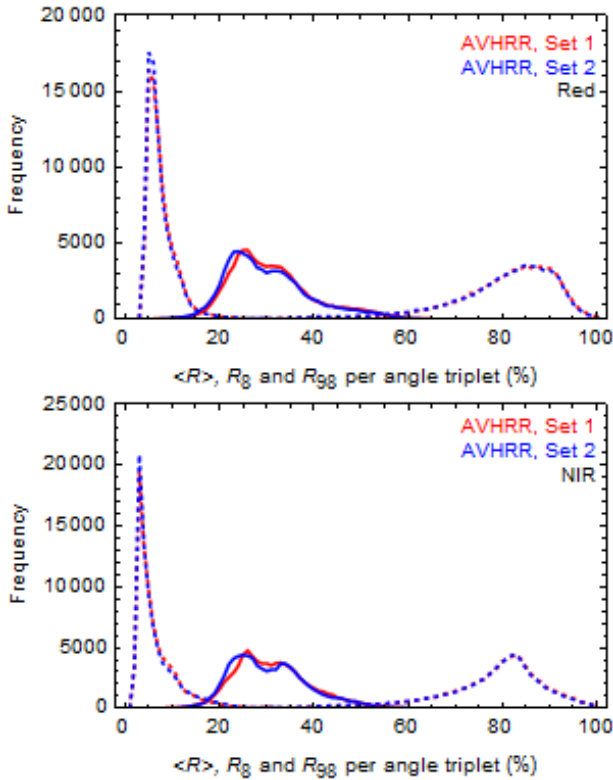


Figure 7. Distributions of the NOAA-15, NOAA-18, NOAA-19 and METOP-A/2 AVHRR red (top) and NIR (bottom) channel TOA reflectance distribution means (solid curves) and 8% and 98% quantiles (dashed curves) corresponding to diverse angle triplets $(\theta_s, \theta_v, \phi)$.

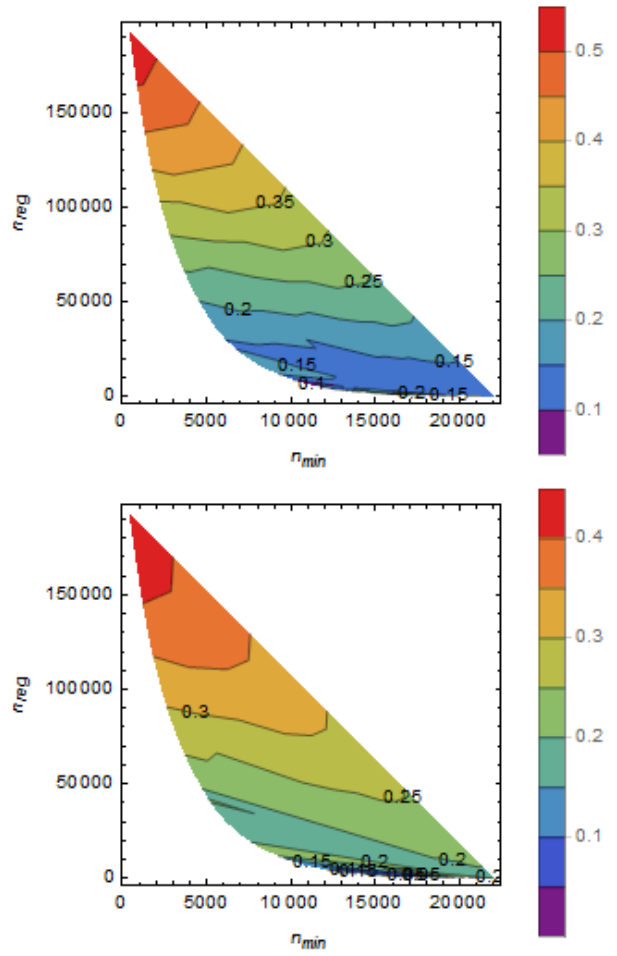


Figure 8. Δ as a function of n_{min} and n_{reg} for the two subsets of AVHRR data for the red (top) and NIR (bottom) channel.

between the curving part of the area of possible value combinations and the largest values of n_{min} [Figure 8]. For the case studied, the optimal value for n_{min} is about 12000 (red), which corresponds to $n_{reg} = 46550$, and about 18000 (NIR), which corresponds to $n_{reg} = 803$. These alternatives were used for Table 2 and Figure 9. Independently on the choice of n_{min} for both channels Δ was at most 0.5. The number of points available for the distributions is roughly 6.5 larger than for the case of MODIS Set 2 vs. Set 1. This explains, why Δ is about as much better for the AVHRR vs. AVHRR case than the MODIS case [Table 1].

Table 2. Best fit of various regression types and the goodness of the fit for the comparison of the two non-overlapping AVHRR data subsets of the red channel of NOAA-15, NOAA-18, NOAA-19 and METOP-A/2.

AVHRR	Channel	b_0	b_1	Δ
Set 2 vs. Set 1	Red	0.012	0.996	0.174
	NIR	0.080	0.999	0.034

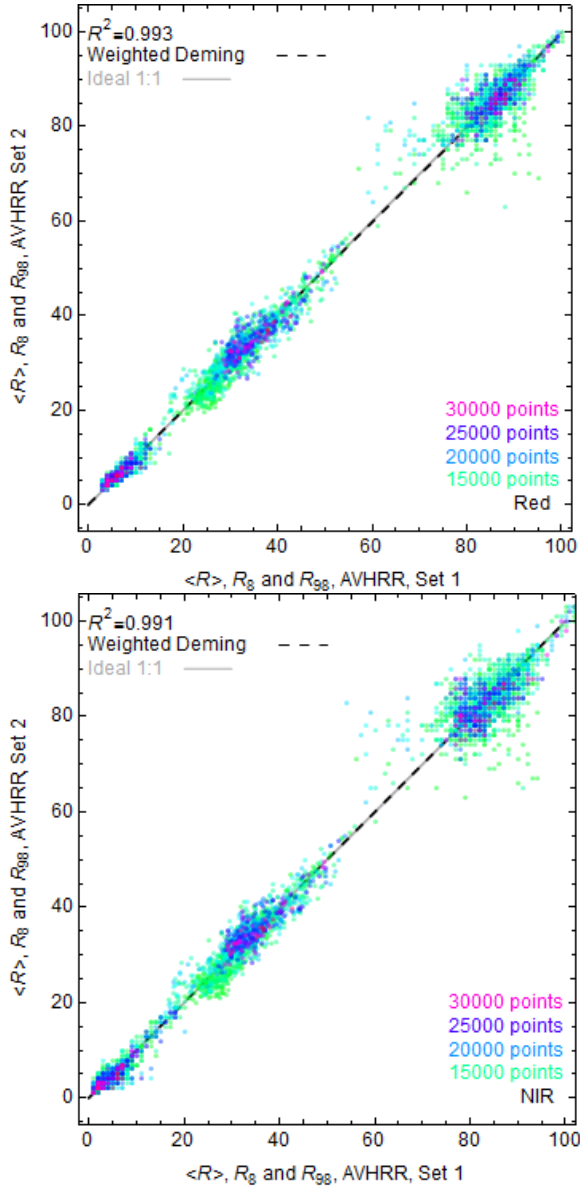


Figure 9. Relation of NOAA-15, NOAA-18, NOAA-19 and METOP-A/2 AVHRR vs. AVHRR $\langle R_{\text{Red}} \rangle, R_{\text{Red}8}$ and $R_{\text{Red}98}$ (top) and $\langle R_{\text{NIR}} \rangle, R_{\text{NIR}8}$ and $R_{\text{NIR}98}$ (bottom) for $n_{\text{min}} = 12000$.

4.4. AVHRR vs. MODIS

The distributions of the $\langle R_{\text{Red}} \rangle, R_{\text{Red}8}, R_{\text{Red}98}$ and $\langle R_{\text{NIR}} \rangle, R_{\text{NIR}8}, R_{\text{NIR}98}$ values are shown in Figure 10 for the AVHRR and combined MODIS Terra/Aqua data sets. The number of diverse angle triplets was roughly 98000. The Deming regression with iterated weights is carried out using $n_{\text{min}} = 4000$, which corresponds to $n_{\text{reg}} = 3895$. This combination is located in the Δ contour plot of the AVHRR vs. MODIS at about the middle of the curving part, thus representing a sort of medium choice. The regression is visualized in Figure 11. The R^2 value is high. The parameter values are given in Table 3.

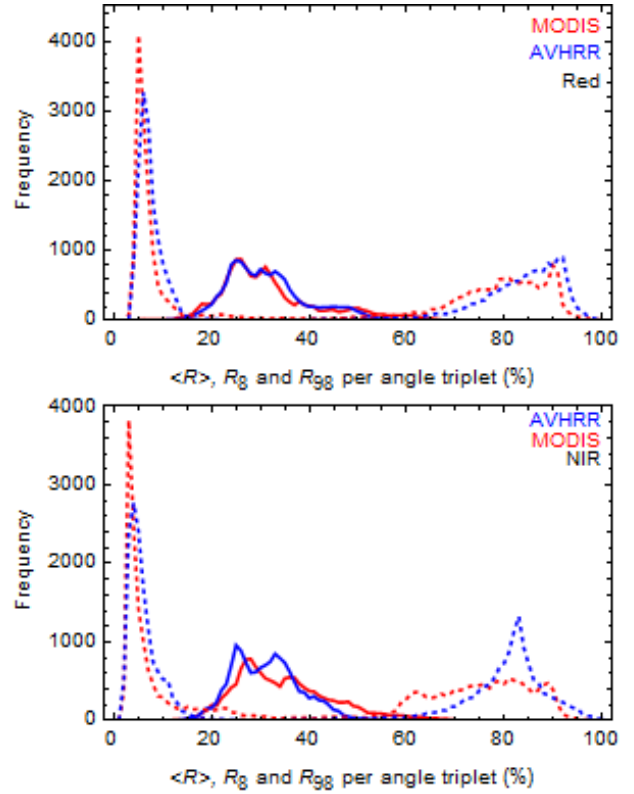


Figure 10. The red (top) and NIR (bottom) channel TOA reflectance distribution means corresponding to the same angle triplets for the whole MODIS and AVHRR data.

Table 3. Deming regression for AVHRR vs. MODIS.

	Channel	b_0	b_1
AVHRR vs. MODIS	Red	1.866	1.018
	NIR	0.873	1.056

5. CONCLUSIONS

A statistical intercalibration method based on non-simultaneous retrievals was demonstrated to produce good results, when applied to intercalibration of the same instrument (MODIS vs. MODIS or AVHRR vs. AVHRR). The achievable accuracy was better than 0.5%. The accuracy improved relative to the available amount of points in the distributions. The accuracy of intercalibration of MODIS Aqua vs. Terra was better than 0.6% and would probably improve with a larger data set. The AVHRR reflectance values turned out to be 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 within the limits of the instrument accuracies.

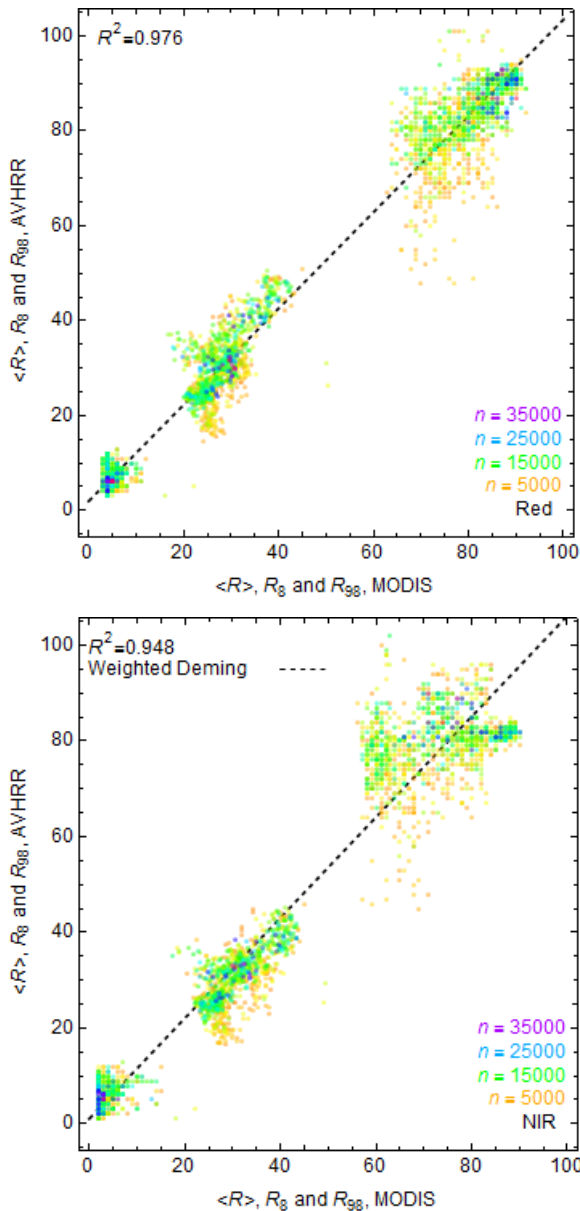


Figure 11. Regression of $\langle R_{\text{Red}} \rangle, R_{\text{Red}8}, R_{\text{Red}98}$ (top) and $\langle R_{\text{NIR}} \rangle, R_{\text{NIR}8}, R_{\text{NIR}98}$ (bottom) for the AVHRR data set vs. the combined MODIS Terra and Aqua data set.

6. REFERENCES

- Riihelä, A., Lahtinen, P. and Hakala, T. (2011). The Radiation, Snow Characteristics and Albedo at Summit (RASCALS) Expedition Report, Finnish Meteorological Institute reports 2011:8, 41 p.
- Bontemps, S., Defourny, P., Van Bogaert, E., Arino, O., Kalogirou, V., Ramos Perez, J. (2011). GLOBCOVER 2009 Product description and validation report, ESA 2011 53 p.
- Xiong, X., Sun J. and Barnes, W. (2008). Intercomparison of On-Orbit Calibration Consistency Between Terra and Aqua MODIS Reflective Solar Bands Using the Moon. *IEEE Geoscience and Remote Sensing Letters*, **5**(4), 778-782.
- doi: 10.1109/LGRS.2008.2005591
- The Pathfinder Atmospheres-Extended AVHRR Climate Dataset. Andrew K. Heidinger, Michael J. Foster, Andi Walther, and Xuepeng (Tom) Zhao, *Bulletin of the American Meteorological Society* 2014 95:6, 909-922.
- Heidinger, AK; Straka, WC; Molling, CC; Sullivan, JT; Wu, XQ (2010). Deriving an inter-sensor consistent calibration for the AVHRR solar reflectance data record. *International Journal of Remote Sensing*, **31**(24), 6493-6517.
- Clark, R. N., Swayze, G. A., Wise, R., Livo, E., Hoefen, T., Kokaly, R., and Sutley, S. J. (2007). USGS digital spectral library splib06a: US Geological Survey, Digital Data Series 231, <http://speclab.cr.usgs.gov/spectral.lib06>, last access: 12 November 2012, US Geological Survey, Denver Federal Center Denver.
- Manninen, T., Riihelä, A. and de Leeuw, G. (2012). Atmospheric effect on the ground-based measurements of broadband surface albedo. *Atmos. Meas. Tech.*, **5**, 2675–2688.
- Heidinger, A., private communication
- Kummell, C. H. (1879). Reduction of observation equations which contain more than one observed quantity. *The Analyst (Annals of Mathematics)* **6**(4), 97–105. doi:10.2307/2635646. JSTOR 2635646.
- Carroll, R. J. and Ruppert, D. (1996). The use and misuse of orthogonal regression in linear errors-in-variables models. *American Statistician*, **50**, 1–6.
- Linnet, K. (1993). Evaluation of Regression Procedures for Methods Comparison Studies. *Clin. Chem.*, **39**(3), 424-432.
- Martin, R.F. (2000). *General Deming Regression for Estimating Systematic Bias and Its Confidence Interval in Method-Comparison Studies*. *Clin. Chem.* **46**(1), 100–104.
- Markovsky, I. and Van Huffel, S. (2007). *Overview of total least squares methods*. *Signal Processing*, **87**, 2283–2302.
- Jäkel, E., Walter, J. and Wendisch, M. (2013). *Thermodynamic phase retrieval of convective clouds: impact of sensor viewing geometry and vertical distribution of cloud properties*. *Atmos. Meas. Tech.*, **6**, 539 – 547.
- https://nsidc.org/data/modis/terra_aqua_differences