Validation of a Physical Retrieval Scheme of Solar Surface Irradiances from Narrowband Satellite Radiances

H. Deneke
Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands, and Meteorological Institute of the University of Bonn, Bonn, Germany

A. Feijt and A. Van Lammeren
Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands

C. Simmer
Meteorological Institute of the University of Bonn, Bonn, Germany

(Manuscript received 30 August 2004, in final form 19 April 2005)

ABSTRACT

An algorithm is presented to derive the downwelling solar surface irradiance from satellite measurements of the 0.63-μm reflectance, which explicitly accounts for variations in cloud optical depth and integrated water vapor. For validation, a long-term dataset of 40,356 pyranometer measurements and 1450 NOAA-14 Advanced Very High Resolution Radiometer (AVHRR) satellite scenes of the Netherlands is used. A mean overestimate of the satellite-retrieved irradiance by 7% is found, which is consistent with numerous other studies reporting positive biases of atmospheric transmissivities that are calculated by radiative transfer schemes in comparison with measurements. The bias can be explained by the calibration and measurement uncertainties of both the AVHRR and pyranometer. A strong solar zenith angle dependence of the bias is found when water clouds are assumed in the retrieval. Such a dependence is not observed for ice clouds. Currently, there is not enough information for a conclusive explanation of this behavior. Comparing individual pyranometer measurements at 30 stations within a region of about 150 km² averaged over 40 min, a large rmse of 86 W m⁻² is found. If the average of all of the stations for a satellite overpass is considered instead, a much better accuracy is achieved (rmse of 33 W m⁻²). For monthly averages of all of the stations, the rmse is further reduced to 12 W m⁻². Evidence is presented that suggests that a significant fraction of the rmse in the comparison originates from the variability of the irradiance field, which limits the representativeness of the reference ground-based pyranometer measurements for the satellite-retrieved values.

1. Introduction

The interaction between the atmosphere and surface has a significant influence on the earth’s climate system. The interaction takes place mainly through the exchange of radiation, heat, and moisture. Accurate knowledge of the surface radiation budget is of particular importance because it constitutes the principal forcing that drives the energy and water exchange processes. The World Climate Research Program (WCRP) has set the goal of deriving a global climatology of radiative fluxes with an accuracy of 10 W m⁻² for monthly mean values averaged over regions of 250 km × 250 km (Sutlles and Ohring 1986). Because of the limited and inhomogeneous coverage of meteorological measurement stations, retrieval from satellites is the only feasible way to obtain this information.

The current study focuses on one particular component of the surface radiation budget, the downwelling solar surface irradiance (DSSI). Numerous algorithms have been developed for its estimation from satellite measurements (reviews are given, e.g., by Schmetz 1989 and Pinker et al. 1995). Starting from a purely statistical relation between the observed radiances at the top of the atmosphere (TOA) and the measured irradiance at...
the surface, algorithms have become increasingly sophisticated by including more physical representations of the underlying atmospheric radiative transfer. However, most current algorithms still rely on empirical relations for at least part of their processing. Two typical examples are the use of a statistically derived angular distribution model (ADM) of the outgoing radiation [the ADMs obtained from data of the Earth Radiation Budget Experiment (ERBE) described in Suttles et al. (1988) are widely used in the scientific community] and the conversion from narrowband to broadband radiative quantities (e.g., Li and Leighton 1992). Statistical measures can be used to assess the quality of such relations. However, the lack of an underlying physical model hinders further insights into potential deficiencies. Furthermore, it is often unclear whether the conditions that are used for the derivation of such a relation are representative for the conditions during the application. If this requirement is not fulfilled, significant biases can occur. An illustrative example is given by Chang and Li (2000), who find biases for retrieved values of the TOA broadband albedo as a result of relying fully on the ERBE ADMs.

In the current study, a physical algorithm for retrieving the downwelling solar surface irradiance from satellite images of the 0.63-μm reflectance is described and validated. The method is applicable to both cloudy and clear-sky conditions and is based exclusively on calculations of a detailed radiative transfer model.

Hence, all deviations can be attributed either to shortcomings of the model or incomplete knowledge of the required model inputs. While there is no guarantee that our approach achieves better performance than other algorithms based on empirical relations, our evaluation can help to identify and correct technical and conceptual shortcomings in the model underlying our approach, and improve the general understanding of atmospheric radiative transfer.

For validation of the algorithm, it is applied to 1450 scenes obtained from the Advanced Very High Resolution Radiometer (AVHRR) on board the National Oceanic and Atmospheric Administration (NOAA)-14 satellite. The results are compared with 40 356 ground-based pyranometer measurements. The large number of cases allows for a thorough statistical evaluation of the retrieval performance. Because of the similarity of the spectral channels, the algorithm can be easily adapted to scenes acquired by the Spinning Enhanced Visible and Infrared Imager (SEVIRI) instrument on board Meteosat Second Generation (MSG; see EUROMET 2001). MSG SEVIRI data are operationally available at 15-min intervals since January 2004. Thus, the algorithm can be applied to operationally derived estimates of the DSSI for applications relying on high temporal and spatial resolution.

Sections 2 and 3 of this study provide a summary of the retrieval algorithm and the dataset that is used for its validation, respectively. In section 4, results of the algorithm are compared with 6 yr of measurements to quantify the accuracy of the scheme. The observed deviations are analyzed and discussed in section 5. Last, a summary is given and possible improvements are outlined.

2. Retrieval method

Variations in broadband atmospheric transmissivity (BAT) and the top-of-atmosphere bidirectional reflectance for a given location are dominated by cloud optical depth for a given solar zenith angle and satellite viewing geometry (see, e.g., King 1987). Both quantities have been used previously to obtain climatologies of cloud optical depth from surface pyranometer measurements (e.g., Leontyeva and Stamnes 1994; Barker et al. 1998) and from satellite images, such as those provided by the AVHRR instrument on board the NOAA series of satellites (e.g., Rossow and Schiffer 1991). From the point of view adopted for our retrieval, the cloud optical depth is the parameter linking the bidirectional reflectance to the broadband atmospheric transmissivity for an otherwise constant atmospheric state. Hence, the relation between both quantities can be established if their individual dependencies on the cloud optical depth are known.

The libRadtran radiative transfer software package (Mayer and Kylling 2005) is used to model this relation. In our calculations, the atmosphere is represented by the midlatitude summer profile from Anderson et al. (1986). The vertically integrated water vapor column and the ozone column are scaled to match the conditions of De Bilt, Netherlands (52.10°N, 5.17°E). For ozone, the climatological mean of 328 Dobson units (DU) for the period of our study is used. For water vapor, calculations are carried out for three column values, corresponding to 8.9, 16.4, and 23.9 kg m⁻² (mean plus/minus one standard deviation). These values have been measured by a Brewer spectrometer and Vaisala radiosondes, respectively. The meteorological conditions at De Bilt are considered to be representative for the Netherlands. An aerosol profile with an optical depth of 0.19 at 0.55 μm and of a water-soluble composition is used based on the limited knowledge about aerosols that are present above the Netherlands (Ten Brink et al. 1997; Stamnes and Henzing 2000). Either a variable water cloud or an ice cloud is considered in this study. The water cloud droplet size spec-
trum is assumed to be gamma shaped, with an effective radius of 10 μm and an effective relative variance of 0.1. This treatment is consistent with the International Satellite Cloud Climatology Project (ISCCP) retrieval (Rossow and Schiffer 1991). The assumed effective radius has been confirmed as being a typical mean value for Northern Hemisphere midlatitude conditions by Han et al. (1994). The optical properties of the water cloud are calculated based on Lorenz-Mie theory. The cloud-base and -top altitudes are fixed to 1 and 2 km, respectively, and vertically homogeneous profiles of the cloud optical properties are assumed. For the ice cloud, a height interval between 8 and 9 km is chosen. Relative to water clouds, the optical properties of ice particles are less certain as a result of the large variety of crystal sizes and shapes that are present under typical atmospheric conditions (Pruppacher and Klett 1978). For our study, the optical properties of the C.2 model of Hess and Wiegner (1994), consisting of imperfect hexagonal ice crystals, are used with optical properties obtained by ray-tracing calculations (Hess et al. 1998). This choice is motivated by the findings of Knap et al. (2005), who compare the model-calculated and -measured angular dependence of bidirectional reflectances for ice cloud scenes observed by the Polarization and Directionality of the Earth’s Reflectances (POLDER) instrument. From several commonly used crystal shapes, the C.2 model and the model of Labonnote et al. (2000) reproduce the observed angular dependence significantly better than, for example, the polycrystals of Macke et al. (1996). The latter are used as the basis for the ISCCP cloud property retrievals (Rossow et al. 1996; Doutriaux-Boucher et al. 2000). Nevertheless, our treatment of ice clouds is still rather simplistic, and the natural variability in crystal size and shape is expected to lead to substantially larger uncertainties in retrieval results for ice clouds than for water clouds. For a detailed discussion of these uncertainties, see Zhang et al. (2002).

The predominant land cover throughout the Netherlands, and, in particular, at the measurement stations, is grass pasture and low vegetation. For grassland at 0.63 μm and the total spectral region, typical values of 10% and 18%, respectively, are reported in the literature for the surface albedo (e.g., Bowker et al. 1985). At 0.63 μm, the land surface albedo is much more uniform than at 0.89 μm, where it is strongly dependent on the structure and water content of the surface vegetation. Hence, the first AVHRR channel has been selected for the retrieval. Effects resulting from spectral dependence and bidirectionality of the surface albedo have been neglected.

The discrete ordinate radiative transfer (DISORT) algorithm (Stamnes et al. 1988) has been selected to solve the radiative transfer equation, using 8 angular streams for the modeling of irradiances and 72 streams for radiances. The correlated-k distributions of Kato et al. (1999) and Kratz (1995) are used to account for gaseous absorption within the atmosphere for the total solar wavelength interval and the spectral range of the AVHRR channel at 0.63 μm, respectively.

Based on this setup, lookup tables have been generated to store the bidirectional reflectance and the BAT for a wide range of cosines of the solar zenith angle (0.1–1.0, Δ = 0.05) and satellite viewing geometries (relative azimuth: 0°–180°, Δ = 10°; cosine of satellite zenith angle: 0.4–1.0, Δ = 0.05), as well as 15 exponentially increasing values of the cloud optical depth (0, 0.125, 0.5, 1, 2, 4, . . . , 128, 256, 512). Because the BAT depends significantly on the vertically integrated water vapor (IWV), three sets of lookup tables are generated for IWVs of 8.9, 16.4, and 23.9 kg m⁻².

These lookup tables have been incorporated into the Royal Netherlands Meteorological Institute’s (KNMI’s) Local Implementation of Apollo Retrievals in an Operational System (KLAROS) satellite analysis environment (Feijt et al. 2004), which calculates the satellite viewing geometry and converts the raw counts of the detector to reflectances. For each satellite pixel, pairs of reflectance–transmissivity are estimated at the 14 values of cloud optical depth selected for the lookup table. Multilinear interpolation of the lookup table entries to the actual viewing geometry and the vertically integrated water vapor at the pixel location is used for this step. The value of integrated water vapor is obtained from the radiosonde launched at De Bilt closest to the satellite overpass. Then, the model relation of reflectance and transmissivity is established by bicubic spline interpolation, and the transmissivity matching the AVHRR measurement is determined.

A short discussion of the dominant uncertainties arising from the modeling assumptions is given in section 5. A more thorough description, further motivation, and a detailed sensitivity analysis is given in Deneke (2002).

3. Instrumental data

Measurements of reflected and transmitted solar radiation from the 0.63-μm channel of the AVHRR instrument on board the NOAA-14 satellite are used in conjunction with simultaneous and collocated records of the global downwelling solar irradiance collected by a network of 32 meteorological stations operated by KNMI located throughout the Netherlands. For the satellite measurements, only observations obtained from the NOAA-14 satellite are used, because it provides
scenes of the Netherlands close to noon because of its orbit. Other NOAA satellites have not been included because of their low solar zenith angle at the time of the overpasses or insufficient coverage of the considered period. We use data from March 1995, when NOAA-14 commenced operational measurements, until the end of 2000. Both the radiance and irradiance values are renormalized to the mean sun–Earth distance and converted to bidirectional reflectance and transmissivity, which are used as retrieval inputs.

a. Pyranometer

Secondary standard pyranometers of type CM 11 built by Kipp and Zonen are used for the measurement of global solar irradiance at KNMI’s meteorological stations shown in Fig. 1. The minimum, mean, and maximum values of irradiance that are measured during a 10-min interval are recorded. To be classified as a secondary standard instrument, the World Meteorological Organization (WMO) demands an accuracy of 3% for hourly mean values of the irradiance (WMO 1996). Because the instruments are operated without ventilation and heating and are checked at rather long maintenance intervals of up to a year, the accuracy of the measurements is somewhat degraded by the deposition of pollution on the instrument’s glass domes. The total error has been found not to exceed 5% (Kuik 1997), however.

Recent investigations of pyranometer accuracy suggest that the estimates of instrumental errors forming the basis of the WMO standard are generally too optimistic. Bush et al. (2000) report on significant systematic underestimates of solar irradiance measurements as a result of thermal cooling of the detectors, a finding that is confirmed by several other studies. While this thermal offset is more important in clear-sky than in cloudy conditions, it possibly also leads to an erroneous calibration of pyranometers. Initial results from a comparison of traditional and improved measurements, which have been compensated for these effects, indicate an underestimation of mean solar irradiance in the range of 3%–8% by traditional measurement setups in all-sky conditions (Philipona 2002). Part of the difference could also be the result of raindrops, frost, and ice, reducing the transmission of the dome of the instruments if they are not ventilated and heated. At this stage, the mechanisms and the magnitude of the bias are still under investigation, allowing for no credible correction of the measurements. Nevertheless, the possible presence of this bias will be considered in the discussion.

b. NOAA-14 AVHRR

The AVHRR instrument on board the NOAA series of polar-orbiting sun-synchronous satellites is one of the main sources of global information about cloud properties (e.g., Rossow and Schiffer 1991; Han et al. 1994). The AVHRR instrument acquires satellite scenes with a viewing angle range of 55° across track and has a nadir pixel resolution of approximately 1.1 km × 1.1 km. Measurements are taken at five spectral channels (0.63, 0.89, 3.7, 11, and 12 μm). For the retrieval, only the 0.63-μm channel is used. Because of degradation of the detectors during launch and over time in orbit, combined with the lack of an on board calibration source, the sensitivity of the shortwave channels of the AVHRR instruments are only known with limited accuracy. In case of the NOAA-14 AVHRR, the issue is further complicated by the low temporal stability and the nonlinear degradation of the two shortwave channels (see Fig. 1 of Tahnk and Coakley 2001b). Hence, calibrations obtained from only a part of the measurement period (e.g., Rao and Chen 1996, 1999; Tahnk and Coakley 2001a; E. Vermote and N. El Saleous 1999, personal communication) lead to erroneous calibration coefficients for the rest of the period, and, thereby, overestimation of the measured reflectance (see the discussion in Deneke 2002).

The two studies reported by Nguyen et al. (2001) and Tahnk and Coakley (2001b) were the only two calibrations that were available at the start of this study, which considered satellite data throughout the full period of the measurements. Because none of the methods can
be considered superior, the mean of both calibrations is used in this study. Figure 2 compares these two calibrations with the response of Rao and Chen (1999), which is recommended by NOAA. Two other calibration results are included to highlight the large uncertainties. The relative difference of the instrumental response that is specified by each calibration is equal to the relative differences of the radiances that are found when applying these calibrations to raw AVHRR data. The overall absolute accuracy of the obtained sensitivity is estimated to be about 5%–10% (Deneke 2002), which translates to a similar relative accuracy of the measured reflectances. Because typical reflectances of cloudy scenes are on the order of 25%–80%, an absolute uncertainty between 1% and 4% must be expected.

4. Validation

The results of the satellite retrieval of the DSSI are compared with values obtained from individual pyranometer measurements in Fig. 3. An averaging period of 40 min is used to obtain corresponding mean pyranometer irradiance from the measured time series. The irradiances have been calculated by multiplying the mean of the BAT with the solar constant and the cosine of the solar zenith angle at the moment of the satellite overpass. The BAT shows less dependence on the solar zenith angle than the DSSI. For the satellite retrieval, a grid box consisting of 10 × 10 pixels centered around the pyranometer stations is analyzed. The transmissivity is calculated for each individual pixel and is then averaged to a spatial mean of the irradiance using the solar constant and the solar zenith angle at the position of the pyranometer. Our treatment compensates for part of spatial and temporal changes in solar zenith angle. The choice of averaging scales has been found as a trade-off between achieving high spatial and temporal resolution and reducing the effect of small-scale variability in the irradiance field, and will be discussed some more in the following. If individual reflectances are averaged instead for this grid and are used as retrieval input, no statistically significant difference is found. This indicates that the influence of satellite resolution on the algorithm’s performance is minor. Only cases with a cosine of the solar zenith angle exceeding 0.2 (about 78°) are included in the comparison, because large errors are expected at low sun elevations for both the radiative transfer calculations and the measurements. Apart from quality checks to exclude corrupted satellite data and pyranometer records, no further exclusion of data has been made, yielding a total of 40 358 individual comparable measurements. In Fig. 3, values are given for correlation, overall bias, constants of linear regression, and the root-mean-square error (rmse). The rmse is calculated as the mean standard deviation of individual points from the regression line (i.e., biased-corrected rmse).

Because the phase of the cloud particles is unknown, results are shown assuming that all clouds consist either completely of ice or water particles. Rather similar values are found under both assumptions for the correlation coefficient and the rmse (0.93 and 85 W m⁻², respectively). The bias between satellite retrieval and actual measurements, however, is almost twice as large if all clouds are assumed to consist of ice instead of water (24 as compared with 42 W m⁻²). Attempts to classify the phase of the cloud with the help of the infrared radiative temperature obtained by the AVHRR at 11 μm do not lead to a measurable improvement of the retrieval.

The explicit physical representation of cloud–radiation interactions can be viewed as the main advantage of the current retrieval algorithm, relative to the use of empirical relations. To evaluate the quality of our representation, it is convenient to consider the atmospheric transmission ratio (ATR), which is defined as

\[
\text{ATR} = \frac{F_{\text{SFC}}}{F_{\text{SFC,CS}}}.
\]

where \(F_{\text{SFC}}\) is the value of the measured DSSI and \(F_{\text{SFC,CS}}\) is the DSSI that is expected for otherwise identical clear-sky conditions. For the current study, the value of \(F_{\text{SFC,CS}}\) has been obtained from the lookup tables as the result of the radiative transfer calculations for zero cloud optical depth. An evaluation of the ac-

---

**FIG. 2.** Relative difference of the instrumental response derived from the calibration coefficients given by Rao and Chen (1996), Tahnk and Coakley (2001a,b), Nguyen et al. (2001), and E. Vermote and N. El Saleous (1999, personal communication) to the ones given in Rao and Chen (1999) as a function of time.
Accuracy of this value has shown that the method is accurate enough for our application. Details are given in Deneke (2002). Figure 4 shows the distribution of retrieved and measured values of the ATR for the dataset assuming either a water or an ice cloud in the retrieval. The two plots clearly show a difference, which explains part of the different retrieval biases. The radiative transfer calculations assuming an ice cloud obviously overpredict the ATR for larger cloud optical depths relative to both of the calculations based on a water cloud and the measurements.

A further interesting feature that is visible in Fig. 4 is the occurrence of pyranometer-inferred ATRs that are larger than unity. The radiative transfer model predicts values of the clear-sky DSSI, which are mostly a few percent higher than the corresponding measurements. Hence, values of the ATR below unity generally are found for clear skies, and only a negligible number of such cases can be attributed to clear-sky situations. Nevertheless, 6.5% of all pyranometer records exceed an ATR of unity in our dataset. A comparison with synoptic cloud reports identifies the majority of these cases as being broken cloud field situations, where significant net horizontal photon transport occurs result-

Fig. 3. Comparison of satellite-retrieved downwelling solar surface irradiance with individual pyranometer measurements, assuming that the cloud consists of (a) water droplets and (b) ice crystals. Shown are contours of the estimated distribution density of satellite-retrieved values given the pyranometer measurement (%).

Fig. 4. Comparison of satellite-retrieved and pyranometer-inferred ATR assuming that the cloud consists of (a) water droplets and (b) ice crystals. Shown are contours of the estimated distribution density (%).
ing from reflections off of the sides of clouds. These cases of enhanced measured DSSI cannot be explained by the concepts of 1D radiative transfer theory. It is probable that the satellite retrieval includes the clouds causing the enhanced DSSI, leading to a reduction of the retrieved irradiance. This explains the observed decrease of the corresponding satellite ATR, once the pyranometer-inferred value significantly exceeds unity. This phenomenon can also be viewed as a symptom that shows the limited representativeness of a temporally averaged pyranometer time series for estimating a spatial mean irradiance. This shortcoming has been studied by Barker and Li (1997) with the help of 3D radiative transfer calculations. To reduce this effect, the average of all pyranometer measurements that are available for individual NOAA-14 overpasses are shown and are compared with the average of the retrieval results in Fig. 5. This average represents an estimate of the regional mean DSSI for each satellite overpass. A mean number of 28 pyranometer records are available. This averaging reduces the rmse substantially to 33 W m$^{-2}$. In addition, the fraction of cases with an ATR exceeding unity is reduced to less than 1%, and a linear regression curve intercepts the origin almost perfectly. This finding illustrates that the climatological variability of irradiance and ATR depends strongly on the scale of the measurements. The irradiance bias that is found between the retrieval and measurements remains with 7.2% close to the value that is found for individual measurements. The small change is caused by the different weighting of individual measurements resulting from the variable number of stations that is used for the comparison with different satellite images. The results of the comparison are only shown for the water cloud retrieval. For the ice cloud retrieval, a similar reduction in scatter is observed. However, the problems of a larger bias and a larger constant offset are still present.

It has been stated in the introduction that the WCRP aims for an accuracy of 10 W m$^{-2}$ for monthly mean radiative fluxes at a spatial resolution of 250 km $\times$ 250 km. The extension of the Netherlands in the north–south and east–west direction is about 350 and 250 km, respectively, and the spatial distribution of meteorological stations is reasonably homogeneous, as shown in Fig. 1. Thus, the averaged irradiance of all of the meteorological stations corresponds to a spatial scale that is similar to the one mentioned by WCRP and can be treated as a representative example of such a grid box. Hence, the WRCP goal is used here to judge the performance of our algorithm. Monthly averages of the retrieval and measured irradiance, and corresponding values of the ATR, are shown in Fig. 6. Only months with a minimum of 300 individual measurements at a cosine of the solar zenith angle larger than 0.2 are included—a criterion that is met by 55 out of 70 months. Excluded were all December values and three November and three January averages. The rejection is caused by the low sun during these months and by a much lower number of available images. For three other months, not enough satellite scenes were available from the archive because of technical failures. It is important to note that these monthly averages consist of individual samples that are obtained at the times of NOAA-
14 overpasses and do not represent monthly averages of the DSSI. Data provided by Meteosat Second Generation could ensure a sufficient temporal sampling to overcome this limitation.

Comparing Fig. 6 and Fig. 5 reveals a significant reduction in rmse to 12 W m$^{-2}$, indicating that a further statistical reduction of errors takes place. This quality almost achieves the goal that is set by the WCRP. However, the linear regression shows an offset of $-25$ W m$^{-2}$ and a slope of 1.14. The origin of this behavior can clearly be identified in Fig. 6b. During autumn, and especially the winter months, the overall high bias of the satellite-retrieved irradiances versus the measurements is reversed, and the mean pyranometer irradiances actually exceed satellite-inferred values. Because of the low values of TOA insolation caused by the lower elevation of the sun, the absolute values of the DSSI are systematically smaller than for the other months, which causes a systematic difference in the dataset and provides a mathematical explanation for the deviation of the regression results. Possible physical mechanisms are analyzed in the discussion part of this section as a first step toward improving the retrieval.

From an application perspective, it is important that no biases are introduced by a retrieval. Otherwise, deviations could be erroneously attributed to a nonexistent physical mechanism. The diurnal cycle of radiative fluxes is an important signal that is strongly correlated with the solar zenith angle. The influence of the solar zenith angle on algorithm performance is shown in Fig. 7 in terms of the absolute difference of retrieved and measured BAT for both the ice cloud and water cloud retrieval. The second parameter studied in Fig. 7 is the cosine of the satellite zenith angle. Both the bias and standard deviation of the two quantities are plotted, because the expected magnitude and the mean value of the retrieval error are of interest. From Fig. 7a, it can be seen that the positive bias in BAT of the water cloud retrieval diminishes with the decreasing cosine of the solar zenith angle, and it even turns into a slight negative bias below a value of about 0.35. No such dependence is found for the ice cloud retrieval. This suggests a strong sensitivity of the zenith angle dependence to a correct scattering phase function of the cloud particles. No strong change in bias is detectable for both retrievals as a function of the satellite zenith angle. Furthermore, the standard deviation remains reasonably constant for both retrievals and the full range of satellite and solar zenith angles, indicating that the performance is not significantly affected by a low-sun or large satellite viewing angle. This conclusion is particularly relevant for the application of the algorithm to data from MSG or other geostationary satellites resulting from the large range of values of satellite viewing angles.

5. Discussion

The results that are presented in the previous section clearly show that the presented algorithm is capable of accurately capturing the influence of clouds on the DSSI. Nevertheless, several discrepancies have shown up in the comparison between retrieval results and pyranometer measurements. While an empirical correction can be applied to remove the observed biases, it seems desirable to identify the underlying physical reasons and to resolve potential shortcomings in the mea-
measurements and the radiative transfer model (RTM). A number of possible explanations for the observed discrepancies will be discussed in this section as a first step toward achieving this goal. Special attention will be given to the large scatter that is present in the comparison between individual pyranometer records and retrieval results.

a. Radiative transfer calculations

The quality of the radiative transfer calculations is fundamental for the accuracy of the retrieval of DSSI from satellite data. However, several input parameters to the RTM calculations are only known with limited accuracy and can cause deviations between model results and measurements.

The assumption of an isotropic and constant surface albedo for all stations is probably the strongest limitation of the scheme in its current form. The downwelling solar irradiance exhibits a strong dependence on broadband surface albedo in the presence of optically thick clouds (Barker et al. 1998), while the narrowband surface albedo strongly influences the satellite measurement in the case of thin clouds (King 1987). Variations of several percent are introduced by different surface types, seasonal changes in the state of vegetation, and the dependence on solar zenith angle. Much larger deviations result from snow, but snow occurs rarely in the Netherlands and has been neglected. Unfortunately, no measurements of the upwelling solar irradiance were available for this study to eliminate this source of error, because the net flux can be retrieved with higher accuracy (Li et al. 1993). However, the Netherlands has a fairly uniform land cover, consisting mainly of grass pastures and low vegetation, in particular, at the pyranometer measurement sites, which should limit the resulting errors. Accordingly, an attempt to infer the surface albedo from clear-sky AVHRR images, using the method of Csiszar and Gutman (1999), did not lead to any significant improvement of the retrieval quality.

Another important aspect of the retrieval is the correct representation of cloud optical properties in the radiative transfer calculations. Because ice crystals and water droplets have rather different phase functions, the correct choice of cloud phase was initially expected to lead to a significant improvement of retrieval accuracy. However, attempts to classify the phase by thresholds of the 11-μm cloud-top temperature have not lead to any difference in performance. Figure 4 also shows that the water cloud assumption works better for cases with a small ATR, which corresponds to clouds with large optical depths. At first sight, this is somewhat surprising, because most of these cases have a cloud-top temperature that is 20°–30°C below the freezing point. Thus, the cloud top is likely formed by ice crystals. However, these clouds typically also have a rather large vertical extent. A significant fraction of the optical depth will be attributable to water droplets and, thus, cause better agreement between the water cloud retrieval and the measurements. Detection of the phase of the cloud is required to further investigate this issue (Jolivet and Feijt 2003). Unfortunately, the measurements of the 1.6-μm bidirectional reflectance that are needed for this method are not available for the NOAA-14 AVHRR instrument.

![Figure 7. Dependence of retrieval performance on (a) solar zenith angle and (b) satellite zenith angle. Shown are the bias and the standard deviation for the water and ice cloud retrievals.](image-url)
Even if a reliable classification of cloud phase is possible, several unsolved aspects remain. With current satellite data, it is impossible to identify satellite pixels containing mixed-phase or multilayered clouds with sufficient accuracy. Representative optical properties for ice clouds are still subject of debate (Doutriaux-Boucher et al. 2000). And, last but not least, effects of cloud inhomogeneity and 3D radiative transfer are completely neglected in the scheme in its current form.

It is important to note that some errors in the RTM or the required inputs will not affect the retrieval accuracy, but will cancel out instead. The transmission and reflection of solar radiation by the atmosphere are linked by an approximately linear function (Li et al. 1993). If errors do not affect the slope or offset of this relation, the retrieval will still provide accurate results, even if the model state does not match reality. A well-known example is the plane-parallel bias of both albedo and transmission caused by cloud inhomogeneities. Calahan (1994) reports that the cloud optical depth should be reduced by a correction factor, if inhomogeneous cloud fields are to be represented by plane-parallel clouds in a plane-parallel RTM. This reduction factor is reported to be about 30% for stratocumulus clouds. Because the column absorption is only slightly altered (Marshak et al. 1998), the mismatch between the actual cloud optical depth and the cloud optical depth that is used in the RTM will have little impact on the retrieval results. Because reflection, transmission, and absorption are linked by the principle of conservation of energy, this cancellation does not apply to errors affecting the modeled atmospheric absorption.

b. Instrumental accuracy

The AVHRR measurements are mainly affected by the limited accuracy of the instrumental calibration, which is estimated to be about 5%. The pyranometer records should have an uncertainty of 3% for hourly averages of the irradiance according to the classification as secondary standard instruments (WMO 1996). However, the investigations of Bush et al. (2000), Cess et al. (2000), and Philipona (2002), mentioned already in section 3, indicate that the true quality is probably significantly lower. A systematic underestimate of measured irradiance in the range of 3%–5%, or even up to 8%, is possible (see Philipona 2002). Because the origins and magnitudes of these errors are only partly known, it is impossible to correct the data for these effects.

The combined uncertainties of pyranometer records and satellite calibration are sufficient to completely explain the high bias of 7% of the retrieval results in comparison with the measurements. The limited calibration accuracy of the satellite images and the recent doubts about pyranometer measurements can, thus, be seen as the largest obstacle to a further evaluation and improvement of this scheme and the underlying RTM. In particular, no definite conclusions are possible about mechanisms whose magnitudes of effects are smaller than the instrumental uncertainties until these issues are solved. This includes effects like the impact of cloud phase and biases resulting from 3D radiative transfer.

c. Effects of variability

For the validation of the retrieval, a spatially averaged value of the DSSI corresponding to 10 × 10 pixels (or approximately a 150 km × 150 km region) has been compared with a 40-min average of a time series as measured by the pyranometer. Adapting this approach, the assumption is inherently made that both quantities are representative for each other. Large deviations from this assumption can result from cloud inhomogeneity and 3D radiative transfer effects, as is shown by Barker and Li (1997) with a Monte Carlo RTM. The origin of these deviations is the spatial and temporal variability of the radiation field that is caused by changes in cloud structure. Temporally averaged pyranometer measurements have to be considered as statistical samples of the spatially averaged surface irradiance corresponding to the area that is used for the satellite analysis. By decreasing the spatial region, the expected sampling error can be reduced. However, the effects of horizontal photon transport will then limit the correspondence of TOA and surface measurements. The impact of the variability in irradiance is illustrated in Fig. 8 with help of pyranometer records obtained at the two stations at De Bilt and Soesterberg. They are only separated by 7.5 km, which is close to the mean distance of two randomly chosen points that are located within the area used for the satellite analysis. The 40-min averages of the DSSI and the ATR, inferred for the two stations, are shown in Fig. 8. It is very illustrative to note that several features that are visible in these plots are quite similar to features that are present in Figs. 3 and 4, suggesting that the underlying effects causing the deviations are possibly closely related.

In principle, the effects of cloud inhomogeneities at the different spatial and temporal scales of the measurements need to be accounted for in a quantitative estimate of retrieval performance, because they are completely independent of retrieval deficiencies. To achieve this goal, the errors that are induced by the effects of limited spatial sampling, temporal averaging, and horizontal photon transport need to be quantified. A detailed analysis of the spatial and temporal scaling properties of the DSSI would be required for this,
which is beyond the scope of the current study. A rough estimate for the sampling error can be given under simplified assumptions. The two stations considered in Fig. 8 are treated as statistically independent samples of the spatially averaged mean irradiance. Furthermore, retrieval errors and the deviations of pyranometer records from the spatial mean are assumed to be normally distributed, which makes Gaussian error propagation applicable. Then, the combination of two error sources with standard deviations $SD_1$ and $SD_2$ is given by

$$SD = \sqrt{SD_1^2 + SD_2^2},$$

where SD is the combined standard deviation. The numerical values that are obtained in the following are not exact because of violations of these assumptions. Nevertheless, they should provide a reasonable first-order estimate.

The rmse of 69.1 W m$^{-2}$ that is found in Fig. 8 corresponds to a normal distribution with a standard deviation of $[69.1/(2)^{1/2}]$ W m$^{-2} = 48.9$ W m$^{-2}$. The factor $1/(2)^{1/2}$ is introduced by converting the mean difference of the two samples to the standard deviation of a single sample from the spatial mean. Using this value as the sampling error and attributing the remainder to retrieval uncertainties, the mean scatter of 86 W m$^{-2}$ that is observed in Fig. 3 reduces to a retrieval uncertainty of 71 W m$^{-2}$ by applying Eq. (2). About one-third of the observed rmse then originates from an inaccurate sampling of the spatial mean irradiance by the pyranometer.

This sampling error can be reduced by combining observations from multiple stations (Li et al. 1995). An average number of 28 stations is included in the comparison shown in Fig. 5. Accordingly, the rmse of 33 W m$^{-2}$ results from a retrieval error of 31.7 W m$^{-2}$ and a remaining sampling error of 9.2 W m$^{-2}$, after Eq. (2). In this case, the retrieval error dominates the magnitude of the total error for this analysis, and the averaging procedure achieves the goal of minimizing sampling errors. This conclusion holds even stronger for monthly averages, because a much larger number of stations is combined.

These values indicate that by averaging the irradiance that is available for all stations, the influence of sampling errors is indeed strongly reduced. The fraction that is attributed to retrieval uncertainties also decreases significantly because of averaging. Likely reasons are the assumption that the satellite measurements also contain statistical noise resulting from limited representativeness, and that the assumptions inherent in 1D radiative transfer calculations lead to significant random errors.

An important consequence of these findings is the dependency of the retrieval quality on the area and the time interval that is chosen for averaging (see Barker and Li 1997; Feijt and Jonker 2000). Hence, any intercomparison of satellite retrievals of the DSSI must be made using similar scales. Averaging periods ranging from 10 min to 1 h have been considered in comparison with regional scales ranging from approximately 5 km $\times$ 5 km to about 400 km $\times$ 400 km, as given by grids of $2 \times 2$ up to $16 \times 16$ satellite pixels. A reduction of rmse
and an increase of correlation are observed for longer averaging periods and larger retrieval regions. Minimum and maximum values of 0.872 and 0.944 are found for the correlation, and 120 and 78 W m⁻² have been calculated as corresponding rmse for retrieved values of the DSSI. Two competing effects influence this behavior. On the one hand, an increase in spatial and temporal scales reduces the scatter caused by small-scale cloud inhomogeneities; on the other hand, a decorrelation is expected because of spatial and temporal changes in the synoptic situation. The increase in the correlation coefficient and decrease in rmse clearly shows that the first effect dominates at the spatial and temporal scales that are considered in our study, and that the field of cloud properties is still strongly autocorrelated.

6. Conclusions

A physical algorithm for the retrieval of downwelling solar surface irradiances from satellite data has been presented, which explicitly accounts for variations of cloud optical depth and column water vapor. Validation with a 6-yr dataset of pyranometer measurements that are obtained at 32 stations located throughout the Netherlands reveals a positive bias of 7% for the retrieval results and a standard deviation of 85 W m⁻² for individual measurements. Pyranometer and satellite data have been averaged to correspond to temporal and spatial scales of 40 min and 150 km × 150 km, respectively. If the retrieved irradiances are averaged for all of the stations that are available for an individual satellite overpass, a reduction of scatter to 33 W m⁻² is found. For monthly averages of the mean irradiance at all stations, the standard deviation decreases to 12 W m⁻².

It is shown that a significant part of the scatter cannot be attributed to retrieval shortcomings, but originate from the variability in the solar irradiance that is caused by cloud inhomogeneity and partial cloud cover. This variability limits the representativeness of a temporally averaged pyranometer time series for inferring a spatial mean of the irradiance. A study of the time series of two pyranometer setups at a short range suggests that it causes at least a third of the total scatter that is present in the comparison of individual pyranometer measurements and collocated satellite data for the present comparison.

This finding has important implications for the validation of similar retrievals and a possible extension of the current scheme. A large number of cases is required to reliably quantify the performance resulting from the size of the induced scatter. Furthermore, results need to be compared at similar averaging scales and under climates with similar cloud variability to be able to assess differences in performance. A more rigorous estimate of the resulting uncertainties based on a comprehensive scaling analysis is highly desirable.

Several questions arising from this study require further investigation. The high bias of satellite-retrieved values is consistent with other studies finding a systematic overestimate of modeled versus measured DSSI (e.g., Cess et al. 1995). More accurate instrumentation and an update of the RTM to recent gaseous absorption data are required to draw definite conclusions about its origin. An interesting finding is the dependence of the bias on the solar zenith angle if a water cloud is assumed in the retrieval, which is not present for the ice cloud retrieval.

An extension of the current retrieval is desirable. This includes a more realistic treatment of cloud properties, such as cloud-top height, droplet size, and thermodynamic phase, which can be based on additional spectral information. Also, a more realistic representation of the surface albedo, including spectral variations and bidirectional effects, are likely to improve the quality.

Acknowledgments. The research was done at the Atmospheric Research department during a stay of the first author in 2002. He is very grateful for discussions and advice on this research from numerous colleagues of KNMI’s Atmospheric Research section. Thanks go to A. Kylling and B. Mayer for creating the libRadtran software; L. Nguyen from NOAA/NESDIS for providing NOAA-I4 calibration data; R. Philipona from the World Radiation Center in Davos, Switzerland, for providing information about recent evaluations of pyranometer accuracy; and the anonymous reviewers for their constructive criticism, which helped to improve the paper.

REFERENCES


Suttles, J., and G. Orihng, 1986: Report of the workshop on sur-


