Toward a Standard Procedure for Validation of Satellite-Derived Cloud Liquid Water Path: A Study with SEVIRI Data

W. GREUELL AND R. A. ROEBELING

Royal Netherlands Meteorological Institute, De Bilt, Netherlands

(Manuscript received 5 September 2008, in final form 16 March 2009)

ABSTRACT

Differences between satellite-derived and ground-based values of cloud liquid water path (LWP_{sat} and LWP_{gr}, respectively) in validation studies are partly associated with the validation itself, in particular with scale differences and parallax. This paper aims at establishing standards for validation procedures to minimize these contributions to the differences. To investigate this, LWP values were collected as computed from ground-based microwave radiometer (MWR) summer measurements made at two Cloudnet sites and from the spaceborne Spinning Enhanced Visible and Infrared Imager (SEVIRI) instrument. The large number of all-sky sample pairs (~2500 after selection) formed an essential condition for the present study. The best validation method was determined by optimum statistical agreement between LWP_{sat} and LWP_{gr}. The method consists of (i) computation of LWP_{sat} by averaging LWP over the pixels surrounding the ground station by means of a Gaussian weight function with a length scale defining the validation area, (ii) computation of LWP_{gr} by averaging the MWR measurements with a Gaussian weight function, by using a time scale that is considerably longer than the time in which the clouds move across the validation area (by a factor of 3–15), and (iii) correcting for parallax. The authors argue that the best length scale for averaging the satellite data is equal to the image resolution. The improvement resulting from the parallax correction was significant at the 99.5% level, but its effect was not significant for a subset of the data for relatively homogeneous cloud fields. Also, there was no significant improvement when, instead of taking a constant, the time scale for averaging the ground data was adjusted to the instantaneous wind field.

1. Introduction

According to the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report, large uncertainties remain about how clouds might respond to global climate change (Solomon et al. 2007). Improving the description of cloud-related processes in climate models is therefore a major goal of current climate research. Parameterizations of cloud-related processes can be tested by comparing modeled with satellite-derived cloud properties (Chaboureau and Pinty 2006; Tselioudis and Jakob 2002; Roebeling and van Meijgaard 2009).

One such property is liquid water path (LWP), which is the weight of the liquid water droplets in the atmosphere above a unit surface area of the earth. LWP is a crucial cloud parameter because formation and depletion of atmospheric liquid water are essential elements of the hydrological cycle. Moreover, variations in LWP have a large influence on radiative fluxes, especially when LWP is small (Turner et al. 2007). LWP can be estimated from space by satellite sensors operating at either optical wavelengths [e.g., by the Spinning Enhanced Visible and Infrared Imager (SEVIRI), the Advanced Very High Resolution Radiometer (AVHRR), and the Moderate Resolution Imaging Spectroradiometer (MODIS)] or—though only over the ocean—at microwave wavelengths [e.g., by the Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E) and the Special Sensor Microwave Imager (SSM/I)]. Satellite-derived LWP can be validated with data based on measurements by ground-based microwave radiometers (MWRs; see Löhnert and Crewell 2003). Validation studies have been carried out by, for example, Han et al. (1995), Jolivet and Feijt (2005), and Roebeling et al. (2008).

Unfortunately, differences between satellite-derived LWP values and ground-based LWP values are not only

Corresponding author address: Wouter Greuell, KNMI, Postbus 201, NL 3730 AE De Bilt, Netherlands.
E-mail: greuell@knmi.nl

DOI: 10.1175/2009JAMC2112.1

related to errors in these two quantities but also to the validation method. With regard to the latter, there are two types of issues.

First, there is a scale difference. MWRs measure liquid water from the ground within a small opening angle around nadir. The corresponding area viewed by the instrument is typically several orders of magnitude smaller than the satellite pixel size. The common procedure to overcome this discrepancy to some extent is to match the satellite value not with a single ground measurement but with a series of ground measurements centered on the time of the satellite overpass. An important assumption is made to allow such temporal aggregation. Many authors make statements like ''averaging in time is done on the assumption that the spatial pattern of LWP within the satellite pixel at the time of the overpass remains unchanged when it moves across the pixel'' or simply refer to Taylor's frozen turbulence hypothesis. However, a more precise formulation would be that added samples are assumed to provide more information about the spatial distribution of LWP within the pixel at the time of the overpass. The assumption is not constancy in time of the spatial cloud pattern but constancy of the local statistical properties of the LWP ground data (i.e., stationarity is assumed).

The second type of error associated with the validation itself is due to parallax and occurs when a satellite views the earth at an oblique angle. In this case, a cloud viewed by a ground-based MWR in its vertical view direction is horizontally displaced in a satellite image. The parallax can be taken into account by shifting the location of the ground station with respect to the image coordinates by a distance that depends on the satellite view angle and the cloud-top height.

In the procedure of validating satellite-derived LWP, several choices can be made and parameters can be set with the aim of mitigating the effect of the described validation issues. In the present study, the effects of these choices and settings were studied in a systematic way, whereas in earlier studies choices and parameter settings were more or less based on intuition. A typical procedure for validating the satellite-derived LWP was as follows: first, the satellite value used for comparison (LWP_{sat}) is taken as the value of the pixel containing the station (nearest-neighbor method). The idea is to get the satellite value within the closest possible surroundings of the ground station. The ground-based value used for comparison (LWP_{gr}) is then computed as the average over a time interval roughly equal to the pixel length divided by a wind speed (e.g., Greenwald et al. 1999; Jolivet and Feijt 2005). That wind speed is an estimate of the mean wind speed at cloud-top height, so that the time interval is a constant value, and an estimate of the time clouds need to travel from one side of the pixel to the other. Finally, though parallax corrections have been applied to rainfall rate retrievals from satellite data by Vicente et al. (2002), this was not the case in past LWP validation studies to our knowledge.

Schutgens and Roebeling (2009) studied some validation settings. They employed MODIS LWP fields with a resolution of 1×1 km^2 and then artificially increased the resolution to 100×100 m^2 by considering the cloud fields as fractals. Next, they computed synthetic MWR data from these high-resolution data as tracks with a width of 100 m. They concluded that parallax errors dominate all other errors resulting from validation, that spatial interpolation and averaging with a Gaussian weight function are better schemes for computing LWP_{sat} than the nearest-neighbor method, and that the time interval for averaging the ground measurements should be approximately equal to the pixel size divided by cloud velocity. Deneke et al. (2009) compared the time series of two other quantities that vary mainly in relation to clouds; namely, SEVIRI reflectance at 0.6 μm and collocated ground-based measurements of transmittance at the same wavelength. They found optimal anticorrelation between these two variables when the ground measurements were averaged over a time interval of ~6 times the time that the clouds need to move across a pixel.

The present paper also investigates validation settings with the aim of proposing an optimized validation method. The time scale for averaging the ground data is not set equal to the length scale of the satellite data divided by the wind speed, as in earlier studies, but the relation between time and length scale is relaxed. To compare the various possibilities, the statistical effects of the following settings were computed:

1) LWP_{sat} at the location of the station (computed by averaging over the surrounding pixels with a Gaussian weight function, by spatial interpolation, or as the nearest neighbor),
2) LWP_{gr} (computed from the ground measurements by averaging over a time interval centered on the satellite overpass time either with a Gaussian weight function or with a rectangular weight function),
3) the size of the length scale involved in Gaussian averaging in the spatial domain,
4) the size of the time scale involved in averaging the ground-based data,
5) a parallax correction (made or not made), and
6) the time scale for averaging the ground measurements (either adjusted according to the instantaneous wind field or taken as a constant based on the mean wind speed).

For homogeneous, plane-parallel cloud fields, the six mentioned settings do not affect LWP_{sat} and LWP_{gr}, so the errors associated with the validation resulting from differences in scale and parallax vanish and the remaining differences between LWP_{sat} and LWP_{gr} are due to retrieval errors in the two variables themselves. However, real cloud fields vary horizontally and errors associated with the validation are therefore always present. They are likely to increase with increasing inhomogeneity of the cloud fields. We investigated this by performing statistical analyses for relatively homogeneous and relatively inhomogeneous cloud fields.

The study required large numbers of samples to be able to distinguish between the various validation settings on the basis of the validation statistics. SEVIRI data from the Meteosat geostationary satellite were therefore used to derive LWP_{sat}. The retrieval algorithm only works when the sun is at a zenith angle smaller than 72°; however, because SEVIRI images are collected at a rate of one per 15 min, almost fifty suitable images are available per day for summers. Instruments on board of polar-orbiting satellites (e.g., AVHRR, MODIS, AMSR-E, and SSM/I) provide only one or two images per day, which is insufficient for obtaining statistically significant results. The LWP_{gr} samples of the present study were computed from MWR data collected in Chilbolton (United Kingdom) and Palaiseau (France) within the Cloudnet framework.

The outline of this paper is as follows: the input satellite and ground-based data and their uncertainties are presented in section 2. Section 3 describes the computation of LWP_{sat} and LWP_{gr} from the input data, the selection of the datasets used in this study, the method for the parallax correction, and the statistical parameters for evaluating the validation. Section 4a presents a baseline comparison of LWP_{sat} and LWP_{gr}, and the remainder of section 4 shows the optimum validation settings, which are discussed in section 5. Conclusions are presented in section 6.

2. Data retrieval

a. LWP retrieval from satellite observations

The SEVIRI instrument (Schmetz et al. 2002) on board the Meteosat Second Generation (MSG) satellites has acquired images since 2004. Every 15 min, it scans the complete disk of the earth. It operates in three channels at visible and near-infrared wavelengths between 0.6 and 1.6 μm and eight channels at infrared wavelengths between 3.8 and 14 μm and one high-resolution visible channel. The nadir pixel size of SEVIRI is 1×1 km^2 for the high-resolution channel and 3×3 km^2 for the other channels. The pixel sizes at the locations of the ground-based measurements used for this study are 6.2 km (north–south) $\times$ 3.2 km (east–west) for Chilbolton and 5.7 km (north–south) $\times$ 3.2 km (east–west) for Palaiseau.

Within the context of the study presented in this paper, it is important to realize that SEVIRI data are resampled in space. The geographical locations of the raw radiance pixels vary continuously. The radiances are therefore resampled onto a nominal fixed grid with gridpoint distances that are approximately equal to the sampling distances. The resampling causes loss of resolution, and as a result, the image resolution is about twice the gridpoint distance (Just 2000).

We used satellite values of LWP derived from reflectances in two SEVIRI bands with the so-called cloud physical properties (CPP) algorithm, which is described by Roebeling et al. (2006) and based on Nakajima and King (1990). The algorithm assumes that the clouds are vertically homogeneous and plane-parallel (i.e., there are no horizontal variations in cloud properties). CPP also assigns a phase (i.e., water or ice) to the clouds. Retrieval is confined to cases with solar zenith angles smaller than 72°. Significant errors in LWP are caused by errors in the calibration coefficients of the satellite sensors. Such LWP errors amount to 20% for clouds with an optical thickness between 8 and 60 (Roebeling et al. 2005). Other important error sources in satellite LWP include misclassification of clouds (i.e., water clouds are erroneously classified as ice clouds and vice versa) and failure of the assumption of vertical homogeneity (Borg and Bennartz 2007) as well as of the plane-parallel assumption (see, e.g., Loeb and Coakley 1998; Loeb et al. 1998; Varnai and Marshak 2002). On the scale of the SEVIRI pixels, the plane-parallel assumption introduces small biases in LWP (less than $\pm$5%) for overcast stratocumulus clouds, but systematic underestimations of $\sim$20% in LWP occur for broken clouds (Zinner and Mayer 2006).

b. LWP retrieval from ground-based observations

Ground-based data were obtained with MWRs located at the two aforementioned Cloudnet stations (Illingworth et al. 2007), namely Chilbolton (51.14°N, 1.44°W) and Palaiseau (48.71°N, 2.21°E). An advantage of these data is the consistent way of determining LWP. Löhnert and Crewell (2003) and Gaussiat et al. (2007) describe the method used for the conversion of the measured MWR radiances into LWP. Gaussiat et al. (2007) estimate that LWP from the MWRs used for the present study has a systematic error and a root-mean-square error of $\sim$5 and $\sim$15 g m^{-2}, respectively.

As a result of the ill-posed retrieval problem, instrumental drifts and uncertainties in gas absorption model

and calibration, an offset may occur in LWP, which can be detected during clear-sky conditions when the real LWP is zero (Gaussiat et al. 2007). By assuming that the offset during clear-sky conditions also holds for cloudy skies, the Cloudnet procedure then attempts to remove offsets from the entire time series. However, note that, even after application of these offset corrections, negative values of LWP occur in the time series. This is caused by variations in the atmospheric profiles of temperature and water vapor that are not accounted for in the algorithm and by instrument noise.

Though MWRs do not saturate for high values of the LWP, their ability to monitor LWP has several important limitations. First, the MWR channels used to derive the data for the present study are not sensitive to ice, so only an estimate of LWP can be made and not of the ice water path. Second, the algorithm fails when the atmosphere contains raindrops as they act as scatterers, whereas radiative transfer calculations used for the development of the algorithm do not take scattering into account. For the same reason, MWR measurements of the LWP are invalid when water covers the instrument. These instrumental limitations forced us to discard a significant part of the data (see section 3c).

MWRs view a part of the sky that is much smaller than the size of SEVIRI pixels. In Chilbolton, their receiving beamwidth is ~2.5° (Simpson et al. 2002), which means that they view cross sections of $90 \times 90\ m^2$ and $220 \times 220\ m^2$ at heights of 2 and 5 km, respectively.

Data were extracted from the so-called multi-instrument datasets (level 1c; available online at http://www.cloud-net.org/data/uk_archive.html). These files contain a host of other parameters besides LWP. The category bits were used to estimate the cloud-top height and to determine the phase of the clouds (see section 3c). All of these ground-based data are provided at a rate of one per 30 s.

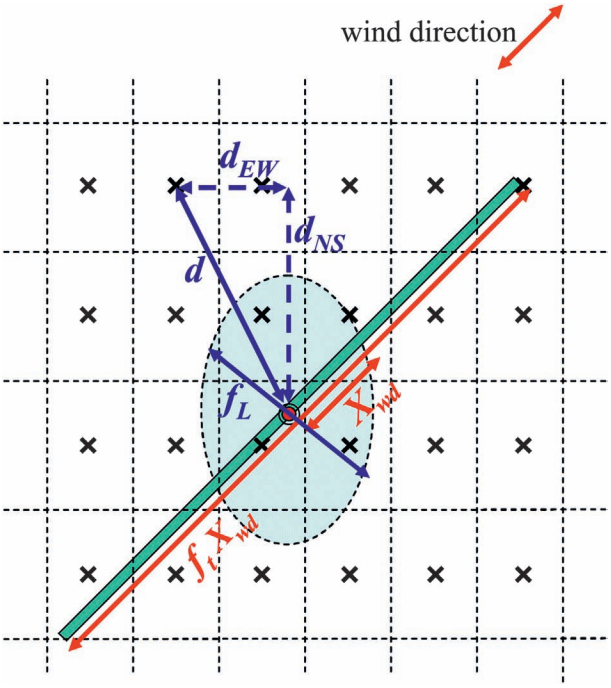

FIG. 1. Schematic drawing showing the SEVIRI pixels with the corresponding grid points (6 × 4), the location of the ground station (red dot), the validation area (hatched oval), and the virtual ground track of the MWR measurements (green bar). The meaning of quantities used to compute LWP_{sat} and LWP_{gr} is also clarified. Red quantities are in length units and blue quantities are in number of pixel distances. Here, LWP_{sat} is computed as the mean value of LWP of the surrounding grid points by using the Gaussian weight function of Eq. (1), where d is the normalized distance between the ground station and a grid point and f_L is the length scale factor (2 pixels in this case); LWP_{gr} can be computed as the mean over a time interval using a Gaussian [Eq. (4)] or rectangular weight function with a time scale equal to the length of the ground track divided by the wind speed at cloud-top height. The length of the ground track is given by $f_t X_{wd}$, where X_{wd} is the gridpoint distance in the wind direction (in this case, southwest or northeast) and f_t the time-scale factor (6 in this case).

3. Validation method

a. Computation of satellite values for validation

The satellite-derived LWP at the location of the ground station (LWP_{sat}) was computed in three ways; namely, (i) by taking LWP of the grid point containing the ground station (nearest-neighbor method), (ii) by 2D interpolation of the LWP values of the four nearest grid points, and (iii) by taking the average of the LWP values of the surrounding grid points, weighting each value with the following Gaussian function

$$w_{sat} = \exp\left(\frac{-2d^2}{f_L^2}\right), \tag{1}$$

where d is the distance between the ground station and the grid point normalized by the gridpoint distance (see Fig. 1) and f_L is the length scale factor (equal to 2σ in the standard Gaussian function). To take into account that the satellite sensor samples at regular intervals in terms of angle, d follows from

$$d = \sqrt{d_{NS}^2 + d_{EW}^2}; \tag{2}$$

$$d_{NS} = \frac{x_{NS}}{X_{NS}} \quad \text{and} \quad d_{EW} = \frac{x_{EW}}{X_{EW}}, \tag{3}$$

where d_{NS} and d_{EW} are the distances x between the ground station and the grid point in the normalized by the gridpoint distances X, in the north–south and east–west

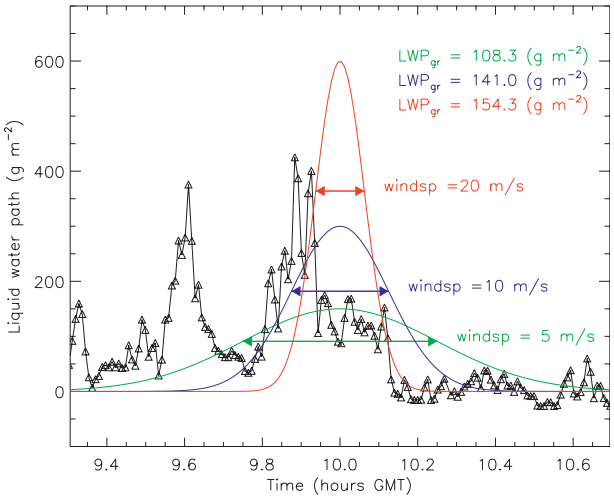

FIG. 2. Example of the calculation of LWP_{gr}. The measurements are given by the black triangles and were collected at a rate of one per 30 s. LWP_{gr} at 1000 UTC is computed by convolution of the measurements with a Gaussian weight function [Eq. (4)]. Three examples of weight functions are shown by the colored curves and correspond to wind speeds at cloud-top height (u_{cltop}) of 5, 10, and 20 m s^{-1}; a length scale ($f_L X_{wd}$) of 9 km; and $f_t = 1$. For completeness, the computed values of LWP_{gr} are given in the top-right corner.

directions, respectively. This means that a grid point at a certain number of gridpoint distances to the east contributes as much to the station value as a grid point at the same number of gridpoint distances to the north. In meters, the latter distance measured is larger than the former because the grid is stretched in the north–south direction. Only grid points with weights higher than 0.01 (that is for $d < \frac{3}{2} f_L$) were considered. As a result of the averaging with Eq. (1), LWP_{sat} is regarded representative of the clouds within an area with a diameter of f_L pixels around the ground station (the validation area). The factor f_L is considered a free parameter.

b. Computation of ground values for validation

The ground-based LWP value LWP_{gr}, which was used for comparison with LWP_{sat}, was computed as the average over a time interval. This was done by either (i) taking the arithmetic mean of the ground measurements over the time interval Δt (i.e., taking a rectangular weight function) or (ii) taking a weighted mean (Fig. 2) with the weights, in line with the method for computing LWP_{sat} [Eq. (1)], given by the following Gaussian weight function:

$$w_{gr} = \exp\left[\frac{-2(t - t_o)^2}{\Delta t^2}\right], \tag{4}$$

where t_o is the time of the satellite overpass. The time-scale Δt for both methods follows from

$$\Delta t = f_t \Delta t_{ref}, \tag{5}$$

where the time-scale factor f_t is a free parameter and Δt_{ref} is the reference time scale, which is the time needed for clouds to move from one side of a pixel to the other. It is given by

$$\Delta t_{ref} = \frac{X_{wd}}{u_{cltop}}, \tag{6}$$

where X_{wd} is the gridpoint distance in the direction of the wind at the cloud top and u_{cltop} is the wind speed at the cloud top. As a result of the rectangular grid, X_{wd} depends on wind direction. Hence, Δt varies with both wind speed and wind direction. The mentioned relaxation of the relation between time and length scale is implemented by the introduction of the factor f_t.

The time-scale Δt can be translated into a virtual ground track (Fig. 1). The track is called virtual because, in reality, all ground measurements are made at the same location, but in the unlikely case that the LWP pattern would move across the area without any temporal change then LWP_{gr} would be the average of the LWP values within the virtual ground track.

c. Data selection

As Roebeling et al. (2008) find better agreement between LWP_{sat} and LWP_{gr} for summer data than for winter data, we limited the analysis to data collected at the Cloudnet stations from May up to and including August. Measurements from Chilbolton for the years 2004, 2005, and 2006 and from Palaiseau for the year 2004 were used. These datasets are far from complete, however. Although the ground data are of relatively high quality, large data gaps exist, in some cases with a length of about two months. In addition, part of the satellite data was not available within the archive of the Royal Netherlands Meteorological Institute (KNMI).

Within the available data, a selection had to be made because of 1) limitations of the MWRs, 2) very high uncertainty of LWP_{sat} under certain conditions, and 3) drift of the MWRs. In detail, data pairs fulfilling one of the following criteria were omitted:

- The clouds passing the ground station during the 10 min centered on the satellite overpass time consisted for more than 5% of ice (data reduction by 51%). The Cloudnet category bits derived from the ground-based radar and lidar measurements and from

temperature profiles from European Centre for Medium-Range Weather Forecasts (ECMWF) analyses were used in combination with the methods described in Hogan and O'Connor (2004) to detect and distinguish ice and water. These methods lead to cases of misclassification of water clouds as ice clouds because the first categorization is based on thresholding at a temperature of 0°C (anonymous reviewer 2009, personal communication). These misclassifications lead to a larger likelihood that LWP ground measurements associated with them are excluded from the datasets, so fortunately they do not have a negative effect on our analyses.

- According to the CPP algorithm, the validation area consisted of more pixels with ice clouds than pixels with water clouds (data reduction by 26%). This criterion was motivated by the occurrence of large errors in LWP_{sat} when the wrong phase is attributed to clouds. Most clouds that consist chiefly of water according to the ground measurements are labeled as such by the satellite retrieval (Wolters et al. 2008).
- Rain was recorded at the ground station (data reduction by 6%).
- LWP_{gr} [average over time with Eq. (4)] $> 400\ g\ m^{-2}$ or the maximum of ground-based LWP during averaging interval $>800\ g\ m^{-2}$ or $LWP_{sat} > 400\ g\ m^{-2}$ (data reduction by 32%). For high values of the LWP, the occurrence of raindrops is likely. This criterion includes cases when raindrops are present in the atmosphere but do not reach the ground. The threshold of $400\ g\ m^{-2}$ was suggested by Löhnert and Crewell (2003). A sensitivity experiment showed that the results of this study, in particular the best settings for the validation method, are not affected when these thresholds are halved. However, compared with the mentioned LWP thresholds, halving the threshold leads to a much larger data reduction: 50%.
- The MWR data were collected on days for which the LWP_{gr} values during clear-sky conditions have a significant systematic deviation from zero (data reduction by 2%). Because the LWP_{gr} values during these conditions are used to correct the calibration offset, the same systematic deviation is likely to disturb cloud measurements.

As a result of overlap, the total data reduction (68%) was less than the sum of the reductions resulting from the individual criteria.

When more satellite pixels or more ground measurements contribute to LWP_{sat} or LWP_{gr}, respectively, then the chance of elimination increases. Therefore, the number of remaining samples depends on the length and time-scale factors f_L and f_t, respectively. For the experiments presented in this paper, the following two sets of selected data were used:

1) The reference dataset of 2634 samples selected with $f_L = 2$ pixels and $f_t = 12$, (see sections 4a and 4d). The aforementioned reductions resulting from the various selection criteria pertain to the reference dataset.
2) The large-scale dataset of 2142 samples, which was used to investigate the effect of variations in the scale factors (see sections 4b and 4c). This dataset consisted of subsets (of 2142 samples each) characterized by different values of f_L and f_t. To ensure a clean comparison of the statistics for different values of f_L and f_t, each subdataset consisted of samples pertaining to the same satellite overpasses. Because the chance of elimination increases with scale, selection criteria were applied to the data computed with the largest investigated values of the scale factors ($f_L = 5$ pixels and $f_t = 60$).

No selection on continuous cloud fields was made. The datasets comprise broken clouds as well as clear-sky cases, but because of averaging over many pixels, only a few cases are completely free of clouds (9 at the scale defined by $f_L = 2$ pixels and $f_t = 12$ in the reference dataset). The advantage of including clear-sky cases is that there are no side effects from the cloud detection algorithm on the validation of LWP.

d. Homogeneous and inhomogeneous cloud fields

Two subsets of the reference dataset were created: namely, one containing relatively homogeneous cloud fields and another containing relatively inhomogeneous cloud fields. Separation of the two datasets was based on LWP_{gr} (Fig. 3). The selected samples were collected into bins of equal size in terms of $\log(LWP_{gr})$. Within each bin, samples were separated into samples with standard deviations higher and lower than the median value. The first set of samples represents relatively inhomogeneous cloud fields, and the latter represents relatively homogeneous cloud fields. Only samples with $LWP_{gr} > 15\ g\ m^{-2}$ were considered for this separation because, at lower values, the noise in the data was too large. This condition reduced the number of data points considerably (from 2634 to 1088 samples).

e. Parallax correction and the wind field

The parallax correction was made by displacing the ground station relative to the image coordinates by a distance $H \tan(\theta_s)$, where H is cloud-top height and θ_s the satellite zenith angle (Fig. 4). At Chilbolton $\theta_s = 58.5°$, and at Palaiseau $\theta_s = 56.1°$. The displacement is in the direction opposite of the satellite azimuth direction,

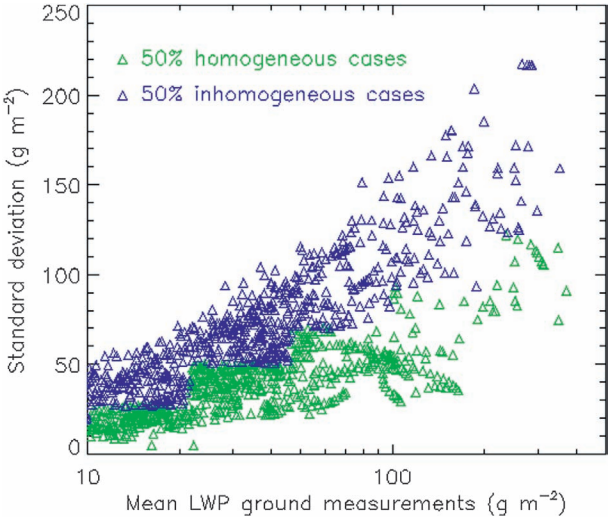

FIG. 3. Scatterplot of the std dev vs the average LWP of the original MWR data that contribute to each value of LWP_{gr}. Here, the Gaussian weight functions given by Eqs. (4)–(6) were used with $f_t = 12$. The plot was used to select relatively homogeneous and relatively inhomogeneous cloud fields (see text).

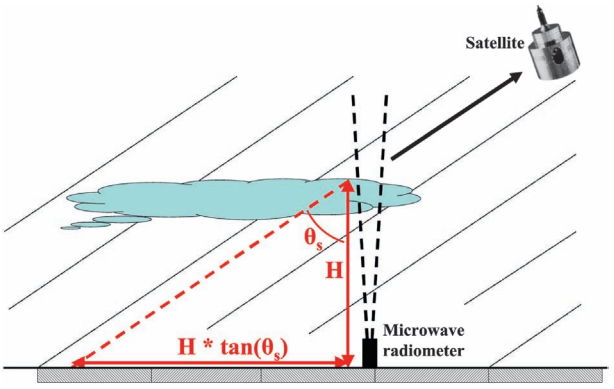

FIG. 4. Because of the parallax, a cloud observed by a ground-based MWR with its vertical view direction is horizontally displaced in a satellite image by a distance $H \tan\theta_s$, where H is cloud-top height and θ_s the satellite zenith angle.

and it is almost due north for Chilbolton and Palaiseau. The cloud-top height H was determined from the radar and lidar profiles measured at the Cloudnet stations. Because the validation was limited to water clouds, H was taken as the highest level with small liquid water droplets. Most of the water clouds of the datasets were relatively low clouds. Of the selected samples, 92% and 89% for Chilbolton and Palaiseau, respectively, have a (water) cloud top lower than 3 km. Consequently, the mean parallax displacements are relatively small: 3.1 km for Chilbolton and 2.6 km for Palaiseau. The wind field at cloud-top height was taken from ECMWF analyses.

f. Statistical evaluation of the validation

Whereas retrieval errors can consist of both bias and noise, the validation problems only cause additional noise in the relationship between LWP_{sat} and LWP_{gr}. Therefore, the validation was evaluated with three statistical parameters that reach extreme values when the amount of noise is minimal. The first is the explained variance, which is equal to the square of the correlation coefficient between LWP_{sat} and LWP_{gr}. The other two parameters are a measure of the distribution of the differences ($LWP_{sat} - LWP_{gr}$). Because the error in LWP_{gr} tends to be constant with the absolute value of LWP_{gr}, absolute differences, not relative differences, were evaluated. Also, we did not compute the standard deviation from the mean because normal distribution functions provide a poor description of the distribution. Following Roebeling et al. (2008), no assumption about the distribution was made and the difference between the 84.2th and 15.8th percentile of the distribution was computed, which has a 68 interpercentile range (Q68). For normal distributions, Q68 = 2σ. Similarly, the 95 interpercentile range (Q95), which is the difference between the 97.75th and the 2.25th percentile value, was computed. For normal distributions, Q95 = 4σ.

The effect of different validation settings on the explained variance was investigated with a test described in Blalock (1979, p. 407), which determines whether the correlation between x (e.g., LWP_{sat} with parallax correction) and y (e.g., LWP_{gr}) is significantly better than the correlation between z (e.g., LWP_{sat} without parallax correction) and y. A t value is computed from the correlation coefficients (r) among x, y, and z:

$$t = (r_{xy} - r_{zy})\sqrt{\frac{(n_{\text{eff}} - 3)(1 + r_{xz})}{2(1 - r_{xy}^2 - r_{xz}^2 - r_{zy}^2 + 2r_{xy}r_{xz}r_{yz})}}, \quad (7)$$

where n_{eff} is the effective sample size. In the case of autocorrelation, n_{eff} is smaller than the number of samples n. We computed n_{eff} from n with the following equation from Dawdy and Matalas (1964):

$$n_{\text{eff}} = n\frac{1 - r_1}{1 + r_1}, \quad (8)$$

where r_1 is the maximum of the three first-order autocorrelation coefficients. Finally, the significance of the computed t value was read from a table of t values.

We have not found a method suitable for assessing the significance of changes in Q68 and Q95.

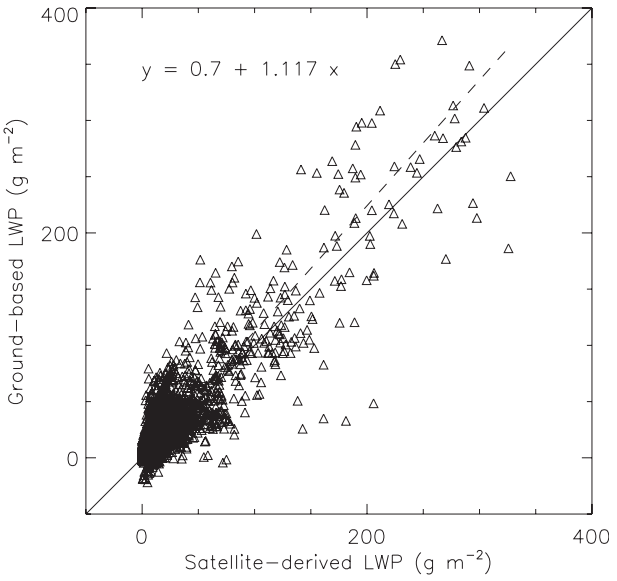

FIG. 5. Scatterplot of LWP_{sat} vs LWP_{gr} with all (2634) samples of the reference dataset and baseline settings; that is, $f_L = 2$ pixels, $f_t = 12$, parallax correction made and Δt computed on the basis of instantaneous values of u_{cltop}. The solid line is the 1:1 line. The dashed line provides the best fit to the data (least squares in the direction perpendicular to that of the fitted line).

4. Results

a. *Baseline comparison of satellite-derived and ground-based LWP*

To put the sensitivity experiments into perspective, we first present a baseline comparison of LWP_{sat} and LWP_{gr}. Calculations were made by averaging with Gaussian weight functions, in the space and time domains, with $f_L = 2$ pixels [Eq. (1)] and $f_t = 12$ [Eq. (5)]. Moreover, the parallax correction made and Δt [Eq. (6)] was computed on the basis of instantaneous values of u_{cltop}. Figure 5 displays a plot of all data pairs of LWP_{sat} and LWP_{gr} of the reference dataset (2634 samples). Most data points (both LWP_{sat} and LWP_{gr}) are close to zero (59% of $LWP_{gr} \leq 15$ g m^{-2}) and the sample density decreases quickly with increasing LWP. The distribution can be approximated by a linear log function, which means that the logarithm of sample density decreases linearly with LWP (see Roebeling et al. 2008).

The comparison between LWP_{sat} and LWP_{gr} shown in Fig. 5 was statistically evaluated by a number of parameters (Table 1). The explained variance is equal to 80% (correlation coefficient = 0.90). The line providing the best fit through the data (least squares in the direction perpendicular to that of the fitted line) has a negligible offset (0.8 g m^{-2}) and a slope that is substantially larger than 1 (i.e., 1.117). The latter is partly related to the fact that noise causes some samples of LWP_{gr} to have negative values, whereas LWP_{sat} cannot become negative. Undersampling of the validation area by the ground measurements forms another cause for slope values larger than 1. Undersampling leads to extra variability in the undersampled variable (Fig. 6) and therefore, with the undersampled variable on the *y* axis, to a slope larger than 1 and a negative offset. The negligible offset is the result of the compensating effect of the difference between mean LWP_{sat} (22 g m^{-2}) and mean LWP_{gr} (26 g m^{-2}). This bias (15%) is somewhat larger than the bias of 9% found by Roebeling et al. (2008). Apparently, LWP_{sat}, LWP_{gr}, or both have a systematic error. Figure 7 shows the bias as a function of LWP_{sat} itself. There is no clear trend of the bias with LWP_{sat}.

TABLE 1. Statistics of the baseline comparison of LWP_{sat} with LWP_{gr}. The first two columns correspond to all selected samples and to selected samples with $LWP_{gr} > 15$ g m^{-2}. The dataset with $LWP_{gr} > 15$ g m^{-2} was split into two halves: relatively homogenous and relatively inhomogeneous cloud fields (see section 3d).

	All samples	LWP > 15 g m^{-2}	Homogeneous cloud fields	Inhomogeneous cloud fields
No. of samples	2634	1088	542	546
Mean LWP_{sat} (g m^{-2})	22.4	47.2	49.7	44.7
Mean LWP_{gr} (g m^{-2})	25.8	57.8	56.5	59.0
Difference: $LWP_{sat} - LWP_{gr}$ (g m^{-2})				
2.25th percentile	−53.0	−75.6	−57.6	−88.7
15.8th percentile	−15.3	−33.4	−24.0	−40.7
Median	−0.2	−10.3	−8.0	−13.3
84.2th percentile	6.8	12.1	12.9	9.8
97.75th percentile	32.9	43.7	41.4	46.3
Q68	22.0	45.4	37.0	50.4
Q95	85.9	119.3	99.0	135.0
Relation between LWP_{gr} and LWP_{sat}				
Offset of linear fit (g m^{-2})	0.8	8.7	2.5	14.5
Slope of linear fit	1.117	1.040	1.087	0.994
Explained variance (%)	80.1	74.6	81.0	69.7

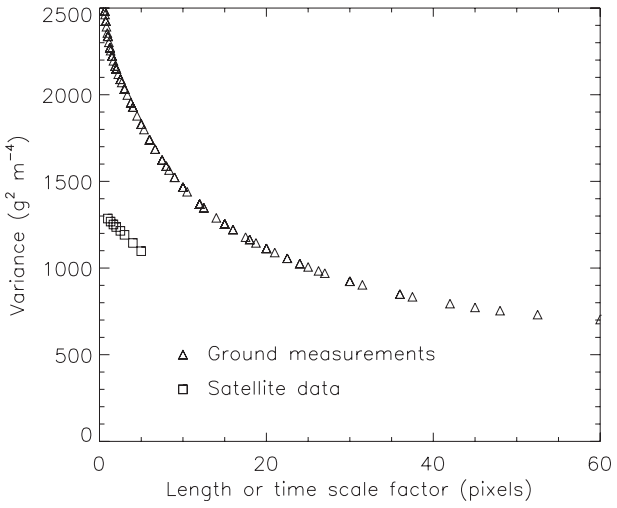

FIG. 6. Variance of LWP_{gr} and LWP_{sat} as a function of the scale factors f_t and f_L.

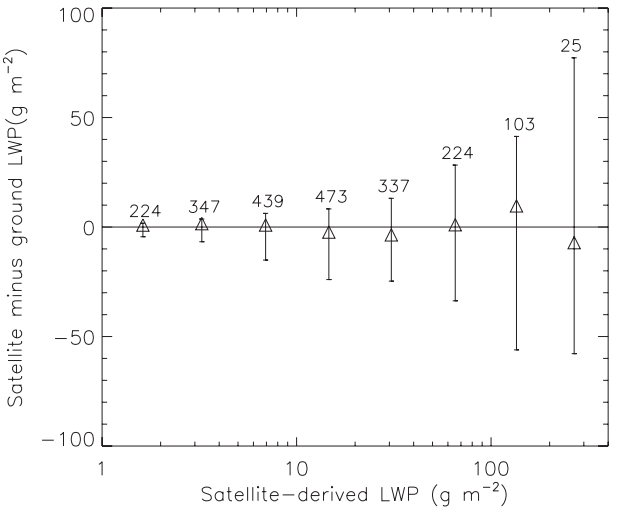

FIG. 7. Difference between LWP_{sat} and LWP_{gr} as a function of LWP_{sat}. Samples are collected in bins of equal size in terms of $\log(LWP_{sat})$. For each bin the number of samples is given at the top of the bar, the triangle corresponds to the median and the bar shows Q68.

The value of Q68 was found to be 22 g m^{-2}. Therefore, Q68 is roughly equal to the mean value of LWP; thus, in agreement with Roebeling et al. (2008), the relative precision (i.e., the ratio of Q68 and the mean divided by 2) is ~50%. The distribution is negatively skewed and has a strong non-Gaussian nature as demonstrated by the ratio of Q95 and Q68 (~4). This ratio would be equal to 2 for a Gaussian distribution. Roebeling et al. (2008) found even higher values for this ratio.

To test the expectation that validation statistics become worse when the degree of inhomogeneity of the cloud fields increases, validation statistics were computed for subsets of the data containing relatively homogeneous and relatively inhomogeneous cloud fields (see section 3d). The necessary confinement to $LWP_{gr} > 15$ g m^{-2} leads to an increase of the average value of LWP_{gr} by more than 100% (Table 1). Also, all validation statistics deteriorated. Apparently, low LWP values affect the statistics positively. Note that this leads to better statistics for sites with lower cloud occurrence, unless low LWP values are excluded from the analyses. The interesting issue, however, is the comparison of the validation statistics of the homogeneous subset with statistics of the inhomogeneous subset of data. Indeed, the explained variance is substantially lower for the inhomogeneous cloud fields than for the homogeneous cloud fields (70% versus 81%) and Q68 is much higher for the inhomogeneous cloud fields (50 g m^{-2} versus 37 g m^{-2} for the homogeneous cloud fields). So, as expected, there is generally a better agreement between LWP_{sat} and LWP_{gr} for homogeneous cloud fields than for inhomogeneous cloud fields.

Unfortunately, it is impossible to tell to what extent this result can be blamed on the validation issues or on the plane-parallel assumption of the satellite retrieval. This is not true for the bias; the validation issues do not affect it because of the large number of samples. For all data with $LWP_{gr} > 15$ g m^{-2}, the mean value of LWP_{gr} is larger than LWP_{sat}, but the bias is greater for the inhomogeneous cloud fields (14.3 g m^{-2}) than for the homogeneous cloud field (6.8 g m^{-2}). Therefore, the larger bias for the inhomogeneous cloud fields suggests that the plane-parallel approximation systematically lowers the value of LWP_{sat}, which is in agreement with the findings of Zinner and Mayer (2006).

b. *Gaussian averaging, nearest-neighbor method, and spatial interpolation*

This section compares various combinations of the three methods for computing LWP_{sat} (nearest-neighbor method, interpolation, and averaging with a Gaussian weight function) with the two methods for computing LWP_{gr} (rectangular and Gaussian weight functions for averaging). This paper uses "schemes" to refer to these methods and their combinations.

Gaussian and rectangular averaging require setting the free parameters f_L and f_t. The next subsection takes a look at the sensitivity of the validation to variations in f_L. We selected a preliminary value of 2 pixels for the experiments. The statistical parameters were then compared for the various schemes and as a function of f_t (Fig. 8). This was done for the entire large-scale dataset and for a subset of 751 samples with $LWP_{gr} > 15$ g m^{-2}. The performance of the schemes was compared on the basis of the maxima (for var_{expl}) and minima (Q68 and Q95)

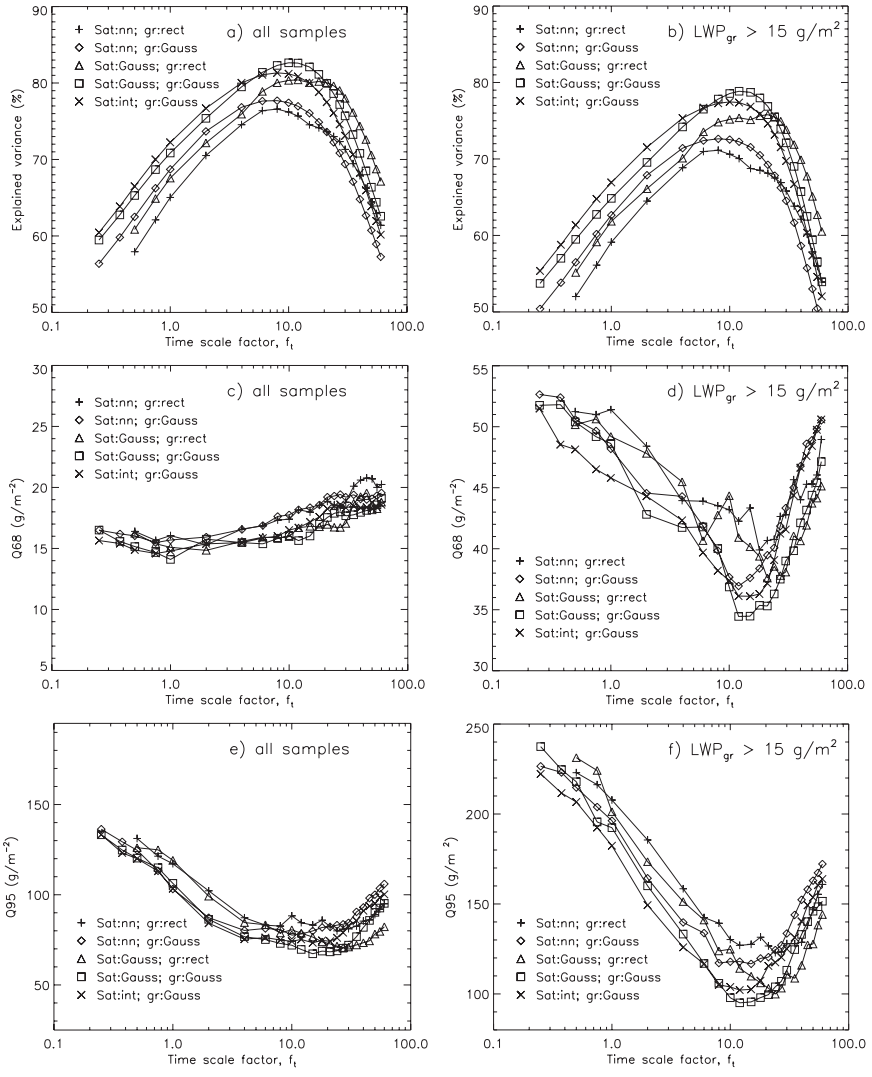

FIG. 8. Statistical evaluation of the various schemes that can be used to compute LWP_{sat} and LWP_{gr}. The variation of a statistical parameter (explained variance, Q68 and Q95) as a function of the time-scale factor f_t for five different schemes. The schemes combine computation of LWP_{sat} with the nearest-neighbor method (denoted as nn), by averaging with a Gaussian weight function (denoted as Gauss), or by spatial interpolation (denoted as int) with the computation of LWP_{gr} by rectangular (denoted as rect) or Gaussian weight functions. Results are for (a),(c),(e) all samples of the large-scale dataset and (b),(d),(f) samples of the large-scale dataset with $LWP_{gr} > 15$ g m^{-2}.

of the curves in Fig. 8. Table 2 lists the maximum and minimum values as well as the values of f_t at which they occur.

As a result of the large number of samples with small values of LWP, Q68 for the complete large-scale dataset (Fig. 8c) hardly varies with f_t and is therefore not appropriate for determining the optimum settings. Figure 8d shows that Q68 becomes a meaningful parameter for determining the optimum settings when samples with low LWP values (LWP < 15 g m^{-2}) are excluded from the dataset. Apart from Fig. 8c, the statistics in the other panels of Fig. 8 reach their optimum values for f_t within the range of 8–21. If rectangular weighting of the ground data is excluded, the range of optimum f_t narrows down to 8–12. Also, the best value for f_t is almost insensitive to the inclusion or exclusion of samples with $LWP_{gr} >$ 15 g m^{-2}.

All three evaluation parameters are consistent in assigning the methods of averaging with Gaussian weight functions in both space and time domains as the best method. The least successful method is the combination of the nearest-neighbor method to determine LWP_{sat} and

TABLE 2. Statistical evaluation of the combinations of the various schemes that can be used to compute LWP_{sat} and LWP_{gr}. The table shows maxima of the explained variance and the minimums of Q68 and Q95 as a function of f_t (see Fig. 7). The values of f_t, at which the extremes occur, are given between brackets. Two datasets were considered: the complete large-scale dataset and a subset of this dataset, all samples with $LWP_{gr} > 15$ g m^{-2}. Also, the significance of the improvement of the Gaussian–Gaussian scheme relative to each scheme is given for the complete large-scale dataset.

Method LWP_{sat}	Method LWP_{gr}	Max explained variance (%)		Significance relative to Gaussian–Gaussian method (%)	Min Q68 (g m^{-2})		Min Q95 (g m^{-2})	
		All	>15 g m^{-2}		All	>15 g m^{-2}	All	>15 g m^{-2}
Nearest neighbor	Rectangular	76.6 (8)	71.1 (8)	>99.5	15.3 (0.5)	39.9 (10)	80.0 (15)	122.9 (15)
Nearest neighbor	Gaussian	77.7 (8)	72.6 (8)	>99.5	15.5 (0.5)	36.9 (10)	77.6 (8)	116.7 (12)
Gaussian ($f_L = 2$)	Rectangular	80.4 (12)	75.8 (21)	97.5–99	14.8 (0.5)	37.6 (12)	70.7 (18)	99.9 (15)
Gaussian ($f_L = 2$)	Gaussian	82.6 (10)	78.9 (12)	—	14.1 (0.75)	34.4 (10)	67.3 (12)	95.1 (10)
Interpolation	Gaussian	81.3 (8)	77.5 (10)	<95	14.6 (0.5)	36.1 (12)	73.4 (12)	102.2 (10)

rectangular weighting to compute LWP_{gr}. The other combinations lead to intermediate results. The combination of interpolation for the computation of LWP_{sat} and Gaussian averaging in the time domain is the second-best method according to var_{expl} and Q68 and the third-best method according to Q95. The difference in the explained variance between the full Gaussian method and the interpolation–Gaussian method is not significant at the 95% level. Differences between the full Gaussian method and the other three combinations are significant at the 95% level, at least (Table 2).

c. Length and time scale

After having found that averaging with the Gaussian weight functions [Eqs. (1)–(6)] is superior to alternative schemes, the question arises what the optimum settings for the scale parameters f_L and f_t are. To investigate this, f_L and f_t were varied in a systematic way and for each combination of f_L and f_t, the validation was evaluated in terms of the explained variance (var_{expl}), Q68 and Q95 (Fig. 9). As in the previous experiments, plots were made using all samples of the large-scale dataset and for the subset of samples with $LWP_{gr} > 15$ g m^{-2}. Note that the curves of the Gaussian–Gaussian scheme in Fig. 8 are in fact cross sections at $f_L = 2$ pixels through the contour plots in Fig. 9.

Similar to Fig. 8c, Q68 for the complete large-scale dataset (Fig. 9c) varies over a small range (16–19 g m^{-2}) for the entire domain and is therefore not appropriate for determining the optimum settings of f_L and f_t. The other panels of Fig. 9 show more distinctive patterns. The two panels with the explained variance show a wide maximum spanning the complete range of investigated values of f_L. Figures 9d,e, however, have optimum values at the maximum investigated f_L (5 pixels), whereas Q95 for LWP > 15 g m^{-2} (Fig. 9f) has a maximum at $f_L \sim 3$ pixels. Combining the results of all panels of Fig. 9, the statistical parameters are not very sensitive to f_L within the investigated range of values, and it is difficult to conclude what the best value is.

The situation is different for f_t. With the exception of Fig. 9c, all panels show a distinct maximum in the y direction. It occurs for f_t values that are greater than f_L. When var_{expl} is used for the evaluation, optimum values for f_t rise from 10 ± 5 at $f_L = 1$ pixel to 15 ± 5 at $f_L = 5$ pixels. When using Q68 or Q95 for the evaluation, the entire band of optimum values of f_t shifts upward by about 5 units. Combined, all statistical parameters reach optimum values at $f_L = 1$ pixel when f_t is in the range of 10–15 and at $f_L = 5$ pixels for f_t in the range 15–20. For the mean SEVIRI pixel size of 4.5 km and a wind speed of 11 m s^{-1} (the mean at cloud-top height for the reference dataset), $f_t = 10$–20 corresponds to time scales of approximately 1.1–2.3 h.

Note that optimum settings of the scale parameters are almost identical for the complete large-scale dataset and its subset of LWP > 15 g m^{-2}. We produced figures similar to Fig. 9 for separate subsets for Chilbolton and Palaiseau and for a dataset comprising only cloud fields that were continuous according the satellite cloud mask (not shown here). In all cases, we found optimum scales that are almost identical to those found for the datasets used to produce Fig. 9, indicating that the results are robust.

d. Parallax and variations in the wind field

Finally, we performed experiments to test the hypotheses that a correction for parallax should be made and that the averaging time interval should vary with wind speed and direction. We used the settings of the baseline comparison (see section 4a), namely Gaussian averaging in both the spatial and time domains with $f_L = 2$ pixels and $f_t = 12$ and the reference dataset. Then, the parallax correction was switched on and off, and the time-scale Δt was computed for both the instantaneous wind field and a fixed wind field. The fixed wind field was

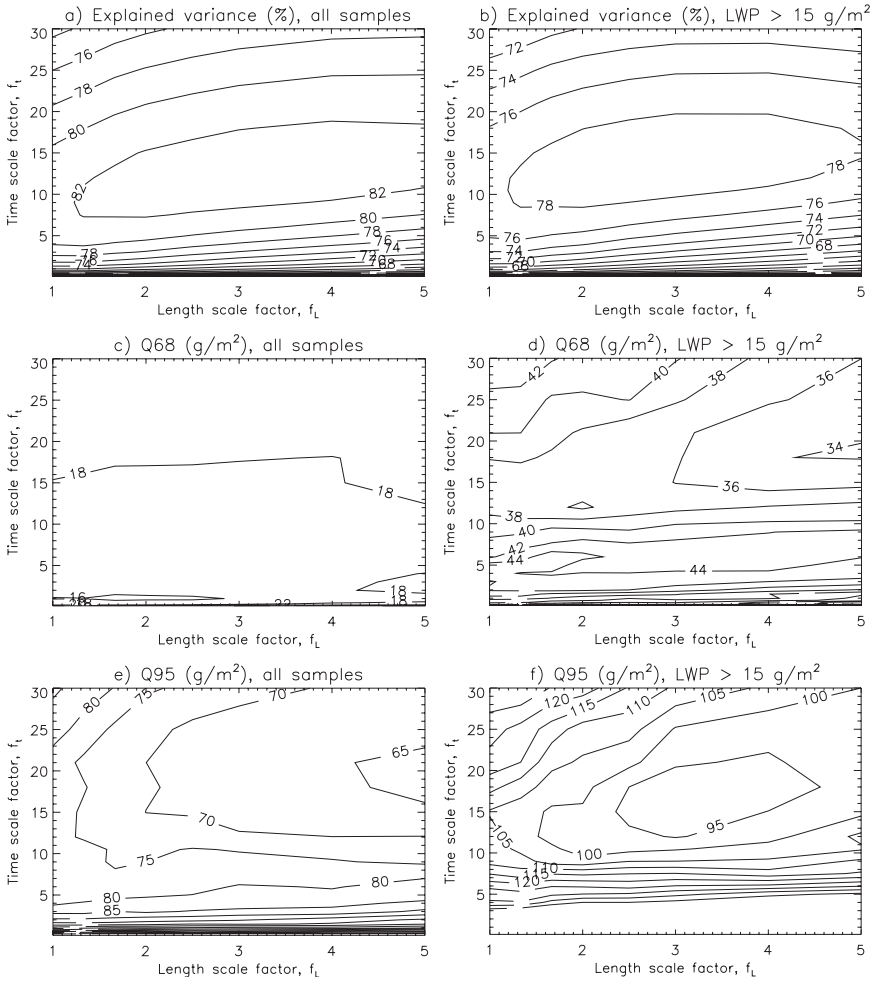

FIG. 9. Statistical evaluation of the comparison between LWP_{sat} and LWP_{gr} as a function of f_L and f_t in terms of (a),(b) explained variance, (c),(d) Q68, and (e),(f) Q95. (a),(c), and (e) give results using all samples of the large-scale dataset and (b),(d), and (f) show results for samples of the large-scale dataset with $LWP_{gr} > 15$ g m^{-2}.

given by u_{cltop} equal to the mean wind for all samples (12.5 m s^{-1} in Chilbolton and 10.3 m s^{-1} in Palaiseau) and X_{wd} equal to the mean grid size in the north–south and the east–west directions. Table 3 shows the results.

There appeared to be no significant improvement when the variations in the wind field were taken into account, whereas the parallax correction caused a significant (at the 99.5% level) increase in the explained

TABLE 3. Explained variance (%) of LWP_{sat} by LWP_{gr} for four validation settings and four datasets. Each validation setting combines the choice of whether to take account of the parallax with the choice of either taking a time scale based on the mean wind speed or a time scale that varies with the instantaneous wind field. The selection of cases with relatively homogeneous and inhomogeneous cloud fields is explained in section 3d.

		LWP > 15 g m^{-2}		
Validation settings	All LWP (n = 2634)	All cloud fields (n = 1088)	Homogeneous cloud fields (n = 542)	Inhomogeneous cloud fields (n = 546)
No parallax, constant wind	77.4	71.7	82.3	64.5
Parallax, constant wind	80.1	74.3	83.2	68.6
No parallax, variable wind	77.3	71.7	79.9	65.3
Parallax, variable wind	80.1	74.6	81.0	69.7

variance. The significance of the improvement resulting from the parallax correction depends on the size of the validation area. For $f_L \leq 3$ pixels, the significance is greater than 99.5%, but at longer spatial scales, where the parallax displacement becomes smaller relative to the size of the validation area, the significance decreases. It is between 95% and 97.5% for $f_L = 5$ pixels. Changes in Q68 and Q95 resulting from taking the parallax correction and wind variations into account are small. We were not able to determine their significance. When we considered the data from Chilbolton and Palaiseau separately, the parallax correction caused a significant improvement for both cases, albeit at lower levels (between 99% and 99.5%, and ~97.5%, respectively), whereas consideration of fluctuations in the wind field had no significant effect for both cases.

The parallax and wind field experiments were repeated for the relatively homogeneous and inhomogeneous cloud fields and confirmed the expectations. For the homogeneous cloud fields, both parallax correction and the consideration of variations in the wind field do not affect the validation statistics significantly. Conclusions for the inhomogeneous cloud fields are comparable to those for the complete reference dataset. Improvements due to the parallax correction cause a significant amelioration in the explained variance (here at the 97.5% level resulting from the reduced number of samples), but the consideration of variations in the wind field yields no improvement. The conclusion is that the parallax correction causes a significant improvement in the validation statistics of the total dataset because of its effect on the samples belonging to the subset of relatively inhomogeneous cloud fields.

5. Discussion of the optimum scales

In earlier validation studies, the time interval for averaging the ground data was usually set equal to an estimate of the time needed for the clouds to move from one side of the validation area to the opposite side (i.e., the ratio between f_t and f_L was equal to 1). In the present study, however, the statistical parameters for evaluating the validation of LWP retrievals from satellite (Fig. 9) turned out to reach their optimum value at much higher ratios of 10–15 for $f_L = 1$ SEVIRI pixel (~4.5 km) and 3–4 for $f_L = 5$ SEVIRI pixels (~22 km).

The explanation for these long time intervals for averaging the ground data may be that a single measurement of the ground sensor only sees a tiny fraction of the validation area (for our data, 90×90 m^2 of 9×9 km^2 (0.01%) for $f_L = 2$ pixels and cloud at a height of 2 km). Averaging the ground measurements over time results in an expansion of the fraction of the validation area that is virtually viewed by the ground sensor (F_{va}). Neglecting the fact that the SEVIRI grid is stretched,

$$F_{va} = \frac{f_t X\, dx}{(f_L X)^2}, \tag{9}$$

where X is the gridpoint distance and dx is the diameter of the field of view of the MWR. When $f_t = f_L = 2$ pixels and for clouds at a height of 2 km ($dx = 90$ m), the MWR sees ~1% of the validation area. So, even after temporal averaging with $f_t = f_L$, the validation area remains heavily undersampled by the ground measurements. This undersampling explains that the optimum values for f_t are substantially larger than f_L.

From Eq. (9), it also follows that, to sample the same fraction of the validation area for every f_L, f_t would have to be proportional to f_L^2. However, only a slight increase of the optimum f_t with f_L was found (Fig. 9). Obviously, failure of the assumption that the local statistical properties of clouds are stationary (section 1; e.g., because of an approaching depression) puts a limit on the length of the optimum time scale.

The statistical analysis of section 4 does not suggest a specific setting for f_L. In particular, Figs. 9a,b show that the optimum explained variance hardly varies with f_L. However, explained variance is a relative measure. When we combine the constancy of the explained variance with the fact that the absolute amount of variance in the signals increases with decreasing length and time scale (Fig. 6), we conclude that the absolute amount of explained variance increases with decreasing scale. This argument favors shorter scales. On the other hand, it does not make sense to decrease f_L beyond the SEVIRI image resolution of 2 pixels (Just 2000). In conclusion, we suggest setting $f_L = 2$ pixels for validation of SEVIRI-derived LWP.

6. Conclusions

This paper presents a methodology of validating satellite-derived values of cloud liquid water path. The aim of this study was to establish standards for validation procedures to minimize error contributions associated with the validation itself. For this purpose, ground-based values of the LWP derived from measurements made with microwave radiometers at two Cloudnet stations (sampling intervals of 30 s) were collected. These were compared with LWP values computed from data obtained with the SEVIRI instrument, which are available at a rate of one per 15 min. In total, ~2500 all-sky sample pairs were available for the analysis.

Based on the optima found for the statistical parameters used to evaluate the relation between the satellite and the ground-based LWP (LWP_{sat} and LWP_{gr}, respectively), we recommend validating SEVIRI-derived LWP by:

1) Computing LWP_{sat} by averaging LWP of the pixels surrounding the ground station, using a Gaussian weight function with a length scale of f_L pixels [Eqs. (1)–(3)]. This length scale defines the ''validation area.'' Within the investigated range of values of f_L (1–5 pixels), the analysis does not favor a particular value. Because the absolute variance of LWP_{sat} increases with decreasing length scale (Fig. 6) and the image resolution is about twice the sampling distance, $f_L = 2$ pixels is suggested.
2) Computing LWP_{gr} by averaging the MWR measurements over an interval centered on the satellite overpass time, using a Gaussian weight function [Eqs. (4)–(6)]. In agreement with the findings by Deneke et al. (2009), the length of the averaging interval should be considerably longer than the time needed for the clouds to move across the validation area. The best multiplication factor with respect to this time (f_t/f_L) increases from ~10–15 for f_L = 1 pixel to ~3–4 for $f_L = 5$ pixels.
3) Making a correction for parallax. For the complete reference dataset, the improvement resulting from the parallax correction was significant at the 99.5% level, but its effect was not significant for a subset of the data representing relatively homogeneous cloud fields. This indicates that the parallax correction owes its effectiveness to its influence on LWP_{sat} of inhomogeneous cloud fields.

The option of adjusting the averaging time scale to the instantaneous wind field instead of computing it from the mean velocity was also considered. This setting did not have a significant effect. We also note that, if absolute values of the validation statistics are important (e.g., in comparisons of validations at different locations), the exclusion of low LWP values (e.g., <15 g m^{-2}) is recommended because the percentage of low values has a large effect on the validation statistics.

The question arises whether the same settings should be made in the case that the MWR measurements are not supported by collocated radar and lidar measurements. In the absence of radar and lidar profile measurements, cloud-top height can be estimated by combining cloud-top temperatures derived from the 10.8-μm SEVIRI channel signal with temperature profiles from ECMWF analyses. This alternative was investigated and it appeared that the optimum length and time scales and the significance of the parallax correction were almost identical to those found when cloud-top height was derived from the radar and lidar measurements. There is a problem, however; without radar and lidar, it is not possible to discriminate water clouds on the basis of the ground measurements, and many mixed cloud fields with substantial ice contributions intrude the datasets. When the selection criterion for water clouds on the basis of the ground measurements was relaxed, then the number of samples increased by 49% relative to the reference dataset. The same optimum settings were found, and they therefore appear to be robust. However, the explained variance decreased from 80% to 69%, which indicates that the selection of water clouds by means of the radar and lidar profiles is essential. Therefore, validation with MWRs without information from collocated radar and lidar is not recommended.

Studies like this one cannot be performed for sensors on board polar-orbiting satellites. They acquire only one or two measurements per day for the location of a particular MWR, so the resulting small number of samples does not enable the statistical analyses presented here. This raises the question about how to use the recommendations from the present study for validation studies with sensors on board polar-orbiting satellites; more generally, how do we use these recommendations for validation studies with other satellite sensors having a different resolution or view angle, during other seasons, or in different climate regions, where cloud conditions may be different.

The use of Gaussian weight functions for averaging in space and time probably has a wider applicability. In addition, the arguments used to recommend an averaging length scale equal to the true image resolution should also hold for images with other resolutions. What is the optimum time scale for averaging the ground measurements? Figure 9 and the corresponding conclusions provide guidance toward the answer. Table 4 lists resulting track lengths for some satellite sensors commonly used to estimate LWP and for evaluating a regional climate model (RCM; see van Meijgaard and Crewell 2005) run at a typical resolution of 25 km (Roebeling and van Meijgaard 2009). Track lengths can be converted into a time scale by dividing track length by the wind speed at cloud-top height. Note that it was assumed that the results given in Table 4 can be obtained by extrapolation in cases where the image resolution falls outside the range of length scales investigated in the present study. This is a fair assumption for the RCM and for SSM/I because their resolutions are only slightly larger than the longest-length scale investigated in the present study, but it is difficult to provide recommendations for MODIS and AVHRR, which have image resolutions much higher than that of SEVIRI at the

TABLE 4. Optimum values of the track length at the image resolutions of several commonly used sensors and at a typical resolution of RCMs. The value of f_L is the image (model) resolution divided by the SEVIRI pixel size in Chilbolton and Palaiseau (4.5 km), and it is used to estimate the optimum f_t value with the help of Fig. 9. The optimum track length is the product of f_t and the image resolution. All values are approximate and the values for the resolutions of 1, 25, and 30 km have been estimated by extrapolation. The track length is a virtual quantity that can be converted into a time scale by dividing track length by the wind speed at cloud-top height.

Satellite sensor	Image resolution (km)	f_L (Fig. 9)	f_t (Fig. 9)	Track length (km)
MODIS and AVHRR	1 (nadir)	0.25	≤12.5	≤50
SEVIRI	9 (Cloudnet)	2	11–16	50–70
AMSR-E	12	3	12–17	55–80
RCM	25	5.5	15–20	70–90
SSM/I	30	7	17–22	75–95

Cloudnet stations. Because the optimum ratio between f_t and f_L increases with increasing resolution, the optimum track length for these sensors should be much longer than the image resolution. The problem of setting the track length for MODIS and AVHRR might also be solved by aggregating the data in such a way that they have the resolution of the SEVIRI data. Note also that the validation of LWP from spaceborne microwave sensors such as AMSR-E and SSM/I is problematic because retrievals are limited to water surfaces. Hence, validation radiometers must be located either on a ship or on an island that is so small that it does not affect cloud conditions.

The importance of the parallax correction increases with decreasing length scale (or pixel size), increasing satellite zenith angle, and increasing cloud-top height. Finally, results from the present study, strictly speaking, only apply to the cloud climatology of Chilbolton and Palaiseau during the summer. It would be interesting to repeat the present study for other cloud climatologies.

Although use of the proposed validation strategy does not remove all validation errors, it does reduce these errors to a minimum. Our suggestions might also be applicable for the validation of other cloud-related properties that vary at the same scales as cloud LWP—in particular, cloud optical thickness and atmospheric transmission (Deneke et al. 2008).

Acknowledgments. We acknowledge the Cloudnet project (European Union Contract EVK2-2000-00611) for providing the Instrument Synergy/Target Categorization data, which were produced by the University of Reading using measurements from the Chilbolton Facility for Atmospheric and Radio Research, part of the Rutherford Appleton Laboratory, and from SIRTA, operated by the Institut Pierre Simon Laplace (IPSL). We thank Kees Kok for his advice concerning statistical questions. Piet Stammes, Hartwig Deneke, Nick Schutgens, and three anonymous reviewers made essential suggestions for this study. We are grateful for their contributions. This study was carried out within the framework of SRON Project EO-073 (Synthesis of active and passive measurements for climate model improvement; SYNTHESIS).

REFERENCES

Blalock, H. M., 1979: *Social Statistics*. McGraw-Hill, 583 pp.

Borg, L. A., and R. Bennartz, 2007: Vertical structure of stratiform marine boundary layer clouds and its impact on cloud albedo. *Geophys. Res. Lett.,* **34,** L05807, doi:10.1029/2006GL028713.

Chaboureau, J.-P., and J.-P. Pinty, 2006: Validation of a cirrus parameterization with Meteosat Second Generation observations. *Geophys. Res. Lett.,* **33,** L03815, doi:10.1029/2005GL024725.

Dawdy, D. R., and N. C. Matalas, 1964: Statistical and probability analysis of hydrologic data, part III: Analysis of variance, covariance and time series. *Handbook of Applied Hydrology: A Compendium of Water-Resources Technology,* V. T. Chow, Ed., McGraw-Hill, 8.68–8.90.

Deneke, H. M., A. J. Feijt, and R. A. Roebeling, 2008: Estimating surface solar irradiance from Meteosat SEVIRI-derived cloud properties. *Remote Sens. Environ.,* **112,** 3131–3141.

——, W. H. Knap, and C. Simmer, 2009: Multiresolution analysis of the temporal variance and correlation of transmittance and reflectance of an atmospheric column. *J. Geophys. Res.,* in press.

Gaussiat, N., R. J. Hogan, and A. J. Illingworth, 2007: Accurate liquid water path retrieval from low-cost microwave radiometers using additional information from lidar and operational forecast models. *J. Atmos. Oceanic Technol.,* **24,** 1562–1575.

Greenwald, T. J., S. A. Christopher, J. Chou, and J. C. Liljegren, 1999: Intercomparison of cloud liquid water path derived from the *GOES 9* imager and ground based microwave radiometers for continental stratocumulus. *J. Geophys. Res.,* **104** (D8), 9251–9260.

Han, Q., W. B. Rossow, R. Welch, A. White, and J. Chou, 1995: Validation of satellite retrievals of cloud microphysics and liquid water path using observations from FIRE. *J. Atmos. Sci.,* **52,** 4183–4195.

Hogan, R. J., and E. J. O'Connor, 2004: Facilitating cloud radar and lidar algorithms: The Cloudnet Instrument Synergy/Target Categorization product. Cloudnet documentation, 14 pp. [Available online at http://www.met.rdg.ac.uk/~swrhgnrj/publications/categorization.pdf.]

Illingworth, A. J., and Coauthors, 2007: CLOUDNET: Continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bull. Amer. Meteor. Soc.,* **88,** 883–898.

Jolivet, D., and A. J. Feijt, 2005: Quantification of the accuracy of liquid water path fields derived from *NOAA 16* Advanced

Very High Resolution Radiometer over three ground stations using microwave radiometers. *J. Geophys. Res.,* **110,** D11204, doi:10.1029/2004JD005205.

Just, D., 2000: SEVIRI level 1.5 data. *Proc. First MSG RAO Workshop,* Bologna, Germany, European Space Agency, SP-452.

Loeb, N. G., and J. A. Coakley, 1998: Inference of marine stratus cloud optical depth from satellite measurements: Does 1D theory apply? *J. Climate,* **11,** 215–233.

——, T. Varnai, and D. M. Winker, 1998: Influence of subpixel-scale cloud-top structure on reflectances from overcast stratiform cloud layers. *J. Atmos. Sci.,* **55,** 2960–2973.

Löhnert, U., and S. Crewell, 2003: Accuracy of cloud liquid water path from ground-based microwave radiometry 1. Dependency on cloud model statistics. *Radio Sci.,* **38,** 8041, doi: 10.1029/2002RS002654.

Nakajima, T., and M. D. King, 1990: Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. Part I: Theory. *J. Atmos. Sci.,* **47,** 1878–1893.

Roebeling, R. A., and E. van Meijgaard, 2009: Evaluation of the daylight cycle of model-predicted cloud amount and condensed water path over Europe with observations from MSG SEVIRI. *J. Climate,* **22,** 1749–1766.

——, A. Berk, A. J. Feijt, W. Frerichs, D. Jolivet, A. Macke, and P. Stammes, 2005: Sensitivity of cloud property retrievals to differences in radiative transfer simulations. KNMI Publication WR-2005-02, 27 pp.

——, A. J. Feijt, and P. Stammes, 2006: Cloud property retrievals for climate monitoring: Implications of differences between Spinning Enhanced Visible and Infrared Imager (SEVIRI) on *Meteosat-8* and Advanced Very High Resolution Radiometer (AVHRR) on *NOAA-17*. *J. Geophys. Res.,* **111,** D20210, doi:10.1029/2005JD006990.

——, H. M. Deneke, and A. J. Feijt, 2008: Validation of cloud liquid water path retrievals from SEVIRI using one year of CloudNET observations. *J. Appl. Meteor. Climatol.,* **47,** 206–222.

Schmetz, J., P. Pili, S. Tjemkes, D. Just, J. Kerkmann, S. Rota, and A. Ratier, 2002: An introduction to Meteosat Second Generation (MSG). *Bull. Amer. Meteor. Soc.,* **83,** 977–994.

Schutgens, N. A. J., and R. A. Roebeling, 2009: Validating the validation: The influence of liquid water distribution in clouds on the intercomparison of satellite and surface observations. *J. Atmos. Oceanic Technol.,* in press.

Simpson, P. M., E. C. Brand, and C. L. Wrench, 2002: Microwave radiometer measurements at Chilbolton: Liquid water path algorithm development and accuracy. EU FP5-CloudNET Project Rep., 44 pp.

Solomon, S., and Coauthors, 2007: Technical summary. *Climate Change 2007: The Physical Science Basis,* S. Solomon et al., Eds., Cambridge University Press, 19–91.

Tselioudis, G., and C. Jakob, 2002: Evaluation of midlatitude cloud properties in a weather and climate model: Dependence on dynamic regime and spatial resolution. *J. Geophys. Res.,* **107,** 4781, doi:10.1029/2002JD002259.

Turner, D. D., and Coauthors, 2007: Thin liquid water clouds: Their importance and our challenge. *Bull. Amer. Meteor. Soc.,* **88,** 177–190.

van Meijgaard, E., and S. Crewell, 2005: Comparison of model predicted liquid water path with ground-based measurements during CLIWA-NET. *Atmos. Res.,* **75,** 201–226.

Varnai, T., and A. Marshak, 2002: Observations of three-dimensional radiative effects that influence MODIS cloud optical thickness retrievals. *J. Atmos. Sci.,* **59,** 1607–1618.

Vicente, G. A., J. C. Davenport, and R. A. Scofield, 2002: The role of orographic and parallax corrections on real time high resolution satellite rainfall rate distribution. *Int. J. Remote Sens.,* **23,** 221–230.

Wolters, E. L. A., R. A. Roebeling, and A. J. Feijt, 2008: Evaluation of cloud phase retrieval methods for SEVIRI on *Meteosat-8* using ground-based lidar and cloud radar data. *J. Appl. Meteor. Climatol.,* **47,** 1723–1738.

Zinner, T., and B. Mayer, 2006: Remote sensing of stratocumulus clouds: Uncertainties and biases due to inhomogeneity. *J. Geophys. Res.,* **111,** D14209, doi:10.1029/2005JD006955.