

RESEARCH ARTICLE

10.1002/2016JD025341

Key Points:

- Downwelling radiance spectra measured during 2013 by a FTS Spectrometer at Dome C, Antarctica
- Results in clear conditions are fundamental to derive clouds infrared properties in FIR microwindows
- Results in clear conditions show that the computation of H₂O total absorption from 200 to 400 cm⁻¹ require revision

Correspondence to:

R. Rizzi,
rolando.rizzi@unibo.it

Citation:

Rizzi, R., C. Arosio, T. Maestri, L. Palchetti, G. Bianchini, and M. Del Guasta (2016), One year of downwelling spectral radiance measurements from 100 to 1400 cm⁻¹ at Dome Concordia: Results in clear conditions, *J. Geophys. Res. Atmos.*, 121, doi:10.1002/2016JD025341.

Received 10 MAY 2016

Accepted 1 SEP 2016

Accepted article online 14 SEP 2016

One year of downwelling spectral radiance measurements from 100 to 1400 cm⁻¹ at Dome Concordia: Results in clear conditions

R. Rizzi¹, C. Arosio¹, T. Maestri¹, L. Palchetti², G. Bianchini², and M. Del Guasta²

¹Department of Physics and Astronomy, University of Bologna, Bologna, Italy, ²National Institute of Optics, National Research Council, Sesto Fiorentino, Italy

Abstract The present work examines downwelling radiance spectra measured at the ground during 2013 by a Far Infrared Fourier Transform Spectrometer at Dome C, Antarctica. A tropospheric backscatter and depolarization lidar is also deployed at same site, and a radiosonde system is routinely operative. The measurements allow characterization of the water vapor and clouds infrared properties in Antarctica under all sky conditions. In this paper we specifically discuss cloud detection and the analysis in clear sky condition, required for the discussion of the results obtained in cloudy conditions. First, the paper discusses the procedures adopted for the quality control of spectra acquired automatically. Then it describes the classification procedure used to discriminate spectra measured in clear sky from cloudy conditions. Finally a selection is performed and 66 clear cases, spanning the whole year, are compared to simulations. The computation of layer molecular optical depth is performed with line-by-line techniques and a convolution to simulate the Radiation Explorer in the Far InfraRed-Prototype for Applications and Development (REFIR-PAD) measurements; the downwelling radiance for selected clear cases is computed with a state-of-the-art adding-doubling code. The mean difference over all selected cases between simulated and measured radiance is within experimental error for all the selected microwindows except for the negative residuals found for all microwindows in the range 200 to 400 cm⁻¹, with largest values around 295.1 cm⁻¹. The paper discusses possible reasons for the discrepancy and identifies the incorrect magnitude of the water vapor total absorption coefficient as the cause of such large negative radiance bias below 400 cm⁻¹.

1. Introduction

A consistent fraction of the Earth's outgoing energy flux occurs in the so-called far infrared (FIR) spectral region, defined as the interval from 100 cm⁻¹ up to 600 cm⁻¹, under both clear and cloudy conditions. *Harries et al.* [2008] present a review of the properties of gaseous and particulate absorption in the FIR. Despite the interest, until the end of the last century the FIR has represented a frontier in atmospheric remote sensing: technological difficulties delayed observations at these wavelengths and new solutions were explored in order to achieve the desired radiometric range, resolution, and accuracy. The instruments that were developed and used in various experimental campaigns are fully described in *Harries et al.* [2008].

The Radiation Explorer in the Far InfraRed (REFIR) project [Rizzi et al., 1998, 2000] was a study for the development of a Fourier transform spectrometer (FTS) meeting required accuracy, lifetime, and simplicity requested for a satellite mission [Palchetti et al., 2001, 2005]. Some experiments were organized to demonstrate the instrumental concept, and among them the Earth COoling by WAtER vapouR emission (ECOWAR) campaign which took place in the Alps at Cervinia (2000 m above sea level (asl)) and at the Testa Grigia site (3500 m asl) [Bhavar et al., 2008]. The main scientific objective was to improve the description of water vapor continuum in the FIR region [Serio et al., 2008; Masiello et al., 2012], but the experimental setup (two FIR FTS, a Raman Lidar and a Vaisala radiosonde system) allowed to study eventually the FIR properties of cirrus clouds. ECOWAR thus represents a pilot study and provides the very first set of FIR spectral downwelling radiance measurements in the presence of a cirrus, in dry atmospheric conditions. *Maestri et al.* [2014] present results for the cirrus cloud case and point out that the differences between simulated and measured radiances in FIR partial windows are larger for lower wave number channels and cloud optical depth less than 0.5. With increasing cloud optical depth, the differences are smaller but still larger than the total uncertainties.

Very recently, a paper by *Mlynczak et al.* [2016] presents results of measured downwelling atmospheric spectra from 200 to 800 cm^{-1} at Table Mountain and accurate simulations using a radiative transfer modeling which is quite similar to the one adopted in the present paper. The spectra belong to the driest day of the experimental campaign, with a precipitable water vapor of 3 mm. They present a very detailed assessment of the uncertainties in both the measurements and the model calculations and conclude that the measured radiances agree with the model radiance calculations to within their combined uncertainties. It is also shown that the uncertainties in modeled radiance are larger than the measurement uncertainties and conclude that “the measurement and models demonstrate radiative closure on this day at Table Mountain.” They also state that their results “demonstrate that the far infrared can be measured more accurately than it can be modeled.” Clearly, their conclusions reflect the specific experimental setup and atmospheric state of the day and location, as correctly pointed out by the authors.

The present work regards the analysis of a full year of downwelling radiance spectra, measured during 2013 by a ground-based FTS, called REFIR-Prototype for Applications and Development (PAD). One of the main objectives of the project is the retrieval of microphysical properties of Antarctic ice clouds and a comparison with literature data. REFIR-PAD has been operated with almost continuous measurements since 2012 from the Italian-French Concordia scientific Station, at Dome C, Antarctica, within the project “Proprietà Radiative del vapore acqueo e delle Nubi in Antartide” (PRANA, radiative properties of water vapor and clouds in Antarctica), and it is the first field campaign collecting atmospheric spectra in the FIR with continuity for a period spanning more than 2 years. During the data processing it was found that an in-depth understanding of the results obtained in clear conditions was necessary to properly discuss the results obtained in cloudy conditions. For this reason this paper is devoted to the description of results obtained in clear conditions. In the second section of this paper, the experimental campaign, the involved instrumentation, and the starting data set are described; the third section deals with preprocessing of REFIR-PAD data with the data quality selection; the fourth section discusses the discrimination among spectra collected in clear and cloudy conditions, and the fifth section deals with the modeling approach. Results obtained in clear conditions are discussed in the section 6, followed by the conclusions.

2. Experimental Campaign and Instrumentation

All the measurements examined in this work belong to the framework of the PRANA research project; research activities started in December 2011 with the Summer Campaign of the XXVII scientific expedition and are supported by the Italian ‘Programma Nazionale di Ricerche in Antartide’ (PNRA), *Palchetti et al.* [2015]. Measurements will continue within other projects started in 2014. The field station is located at Dome C, on the Antarctic Plateau (75°06’S, 123°23’E; 3.233 m asl). Several scientific objectives are served by the PRANA project: improvement of the spectroscopic knowledge of the pure rotational water vapor band in the FIR, study of the spectral features of clouds and polar stratospheric clouds at these wavelengths, and improvement of radiative transfer forward models and retrieval methodologies with particular attention devoted to the underexplored FIR region. Spectral measurements of the downwelling radiance are performed by REFIR-PAD, and cloud characterization is obtained with a lidar system. Both instruments are installed in the physics shelter devoted to atmospheric physics, located 500 m from the main buildings, in the so-called Clean Air Area, and a Vaisala weather station is installed on the roof of the same shelter, providing continuous observations of local meteorological conditions. Conventional radiosonde sounding data are measured every day at 12 UTC by Vaisala RS92 sondes.

2.1. REFIR-PAD Instrument Details

The REFIR-PAD, as used during PRANA project, provides spectrally resolved downwelling zenith radiance measurements in the 100–1400 cm^{-1} range with a 0.4 cm^{-1} resolution, covering in this way the most part of the Earth’s long-wave emission and including both the FIR and the better known middle-infrared region. REFIR-PAD has been designed to operate with uncooled pyroelectric detectors and optics. The interferometer uses a Mach-Zehnder optical configuration with two beam-splitters in Mylar coated with a thin layer of germanium. The optics and the electronics modules are both integrated in a compact package of about 55 kg and 63 cm of maximum size. One of the input ports is coupled to a reference blackbody, while the other, through a scanning mirror, can be switched between two calibration sources at different temperatures, i.e., a hot and cold blackbodies (at 80°C and at 15°C, respectively) and the atmospheric zenith view through a calibrated folding mirror. The detectors are placed on a plate thermally stabilized at around 30°C; both of them

are state-of-the-art commercially available deuterated L-alanine-doped triglycene sulfate pyroelectrics, with a protective window in caesium iodide (CsI). The CsI transmission allows the component to be sensitive down to low wave numbers of about 150 cm^{-1} [Bianchini *et al.*, 2006]. The region between 100 to 150 cm^{-1} is not considered important for ground-based measurements since the atmosphere here is completely opaque.

REFIR-PAD is a fast-scanning spectroradiometer with signals acquired in the time domain and resampled in postprocessing at equal intervals in optical path difference. It operates 24 h a day alternating about 6 h of measurements, during which 27 atmospheric observations are typically collected, and 3 h of stopping, during which the data analysis is performed. The resulting cycle is asynchronous with the day to avoid diurnal biases. Each single atmospheric observation lasts about 5 min and is the result of the average of eight atmospheric spectra (four for each output port). The repetition rate is 12 min, including the calibration measurements and some mechanical delays. The output of complete processing of a sequence are the calibrated spectral radiances and three spectral error estimates, which constitute a measurement set. These three quantities contribute to the final spectral uncertainty: the noise equivalent spectral radiance (NESR), obtained through the error propagation of the detector noise, the calibration accuracy (CAL) linked to 1σ errors on the reference blackbodies temperatures and the statistical standard deviation of the observed spectra, which is an estimate of total variability including scene variability over the measurement time range. The total spectral uncertainty $\sigma_t(\nu)$ is defined as the square root of the squared sum of CAL and NESR. More details about the configuration, data and error analysis can be found in Bianchini and Palchetti [2008]. In this paper the version 2 of the REFIR-PAD calibrated radiances is used, available since 2015.

2.2. Lidar

The second instrument deployed at Concordia Station during PRANA campaign is a tropospheric backscatter and depolarization lidar system, which provides fundamental information about cloud geometry and structure. The Lidar is mounted on an aluminum frame inside the same physics shelter as REFIR-PAD. The laser type is a Quantel Brio working at 532 nm and linearly polarized, the data acquisition system is a Licel GmbH, analogical with two channels, and the telescope is refractive with an aperture of 10 cm diameter, $f/30$. Double optical windows flushed with preheated external air make the operation of the lidar possible in all weather conditions, even on -80°C cold polar nights. This instrument provides tropospheric profiles of backscattering and depolarization of the signal with a vertical resolution of 7.5 m from 30 to nominally 8000 m above the ground level (agl), every 5 min, 24 h/24, by means of a fully automated system that collects housekeeping and local meteorological data together with the lidar observations. For the present work, the quick look daily images of Lidar backscattering and depolarization have been used, covering the entire 2013.

2.3. Vertical Profile Data

A radiosonde system is routinely operative at Dome C and supplies one radiosonde profile at 12 UTC. Not all the radiosonde data represent complete sets, due to occasional sensor failures, and measurements are sometimes available only for a few hundred meters agl. Since atmospheric temperature and humidity profiles are fundamental for simulation of downwelling radiance, we have extracted, from the European Centre for Medium-range Weather Forecasts (ECMWF) Meteorological Archival and Retrieval System archive, the operational analysis for a small area centered on Dome C, for all days in 2013, at the UT times 00, 06, 12, and 18. The profiles are given at the model levels (91 from 1 January 2013 to 25 June 2013 at 18 UT and 137 from 26 June 2013 onward) and all model level pressure values can be computed from the surface pressure. After a preliminary check to assess the level of coherence of ECMWF analysis and radiosonde data, we have used the latter up to nearly the sonde top height and then the ECMWF data up to 80 km. Finally, the vertical profiles are interpolated in height at the levels selected for the forward computations.

3. Data Preprocessing and Quality Control

3.1. General Features

Examples of REFIR-PAD spectra for clear sky conditions and in presence of a cloud layer are reported in Figure 1 as a function of wave number (expressed hereafter as cm^{-1}). Figure 1a is a clear spectrum which is the average of seven sequences taken from 11:04 UTC to 13:08 UTC of 17 June 2013. All these measurements are selected as clear by a classification procedure that is described in later sections. Figure 1b is a single measurement taken in a condition classified as nonprecipitating ice cloud by same classification procedure. The radiance unit (ru) used throughout the paper is $\text{mW m}^{-2}\text{ sr}^{-1}(\text{cm}^{-1})^{-1}$. In Figure 1c, NESR and CAL uncertainties for the above cloudy spectra are shown. The main feature of interest in the present work is the FIR portion, characterized

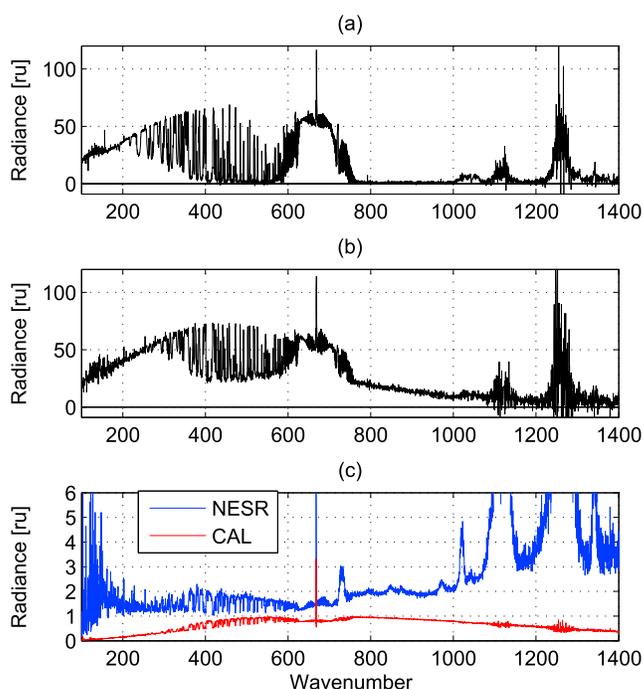


Figure 1. Examples of measured spectra in (a) clear and (b) cloudy conditions, collected, respectively, on 16 July 2013 and on 09 July 2013, together with (c) NESR and CAL errors for the cloudy case.

by increasing water vapor optical depth with decreasing wave number, due to the presence of water vapor lines from the pure rotational band. However, a number of spectral regions are present in which atmospheric opacity is low, and therefore, low radiance values are measured. These regions are clearly distinguished down to 300 cm^{-1} , but in extreme cases even below, due to the very dry air conditions during the austral winter. Some of these intervals, named microwindows (m-w), are precisely defined in later sections since they are at the core of our scientific investigation.

In clear conditions, the downwelling radiance in the infrared window is very low, with many negative values in single channels and single measurements, which are caused mostly by measurement noise. However, some negative values are present also in averaged spectra (as the one in Figure 1a), and these are perfectly compatible with unapodized or slightly apodized interferometric measurements. Radiance levels in the IR window region and in FIR m-w, increase in presence of a cloud layer and a clear slope from 800 to 1000 cm^{-1} becomes evident, as shown in Figure 1b. In FIR m-w the signal depends on several parameters, such as particle size and habit, due to the relative importance of scattering processes at these wave numbers. Another important spectral feature is the $\text{CO}_2\nu_2$ band centered at 667 cm^{-1} ; at the center of this band the high radiance value is due to extremely strong absorption (and emission) in the Q branch, and the spectrometer measures the emission of relatively warm air inside the instrument (at a temperature well above external environment). Around 1050 cm^{-1} not only the O_3 band can be seen but also two regions characterized by very high noise levels, the first centered at 1110 cm^{-1} and the second at 1250 cm^{-1} , linked to the absorption features of Mylar[®], substrate of the beam splitters, which must be removed from the spectrum. The Mylar substrate has also other minor absorption features at 735 , 1020 , and 1340 cm^{-1} that do not prevent the measurement.

3.2. Automatic Quality Control

The availability of a full year of spectroscopic measurements, acquired automatically without human supervision, implies the need for automated quality checks, obtained with a number of tests based on specific spectral characteristics. All measurements used in our scientific analysis has passed all the tests. Some tests use radiance, some brightness temperature (BT), computed from the radiance averaged in each interval. The starting point is the list of 24 (spectral) intervals defined in Table 1, in the range between 235 and 900 cm^{-1} . The table also contains some indication of the variability within each spectral interval, computed from simulated radiance spectra as the standard deviation around the mean for typical clear and cloudy conditions.

Table 1. Spectral Intervals Selected for the Quality Control and Classification Stages^a

Number	Central Wave Number (cm ⁻¹)	Width (cm ⁻¹)	Clear SD Range (ru)	Cloudy SD Range (ru)
1	238.4	4.8	0.6 ÷ 0.9	0.4 ÷ 0.9
2	261.6	1.6	0.5 ÷ 0.8	0.1 ÷ 0.8
3	269.6	2.0	0.9 ÷ 1.7	0.3 ÷ 1.7
4	295.2	4.0	1.6 ÷ 2.0	0.7 ÷ 2.0
5	318.4	2.4	0.8 ÷ 1.6	0.9 ÷ 1.6
6	331.6	2.4	1.2 ÷ 2.2	1.3 ÷ 2.0
7	338.4	1.6	2.0 ÷ 2.7	1.6 ÷ 2.5
8	365.2	3.6	0.5 ÷ 1.6	0.5 ÷ 1.4
9	381.2	2.8	0.7 ÷ 3.0	1.1 ÷ 3.0
10	390.0	5.6	1.3 ÷ 4.5	1.4 ÷ 4.5
11	410.4	7.6	0.3 ÷ 2.0	0.4 ÷ 2.0
12	438.4	2.8	1.3 ÷ 3.0	1.6 ÷ 3.0
13	449.6	3.2	0.9 ÷ 3.0	1.2 ÷ 3.0
14	464.4	3.6	0.8 ÷ 4.0	0.9 ÷ 4.0
15	478.8	6.8	1.1 ÷ 5.0	1.5 ÷ 5.0
16	497.2	6.8	0.9 ÷ 3.5	1.1 ÷ 3.5
17	522.4	3.2	1.6 ÷ 2.5	1.8 ÷ 2.5
18	531.2	3.2	1.2 ÷ 2.2	1.4 ÷ 2.0
19	559.2	6.4	0.9 ÷ 2.5	1.1 ÷ 2.8
20	573.2	3.2	2.0 ÷ 4.5	2.4 ÷ 4.5
21	680.0	9.2	5.5 ÷ 6.5	5.0 ÷ 7.0
22	840.0	10.0	0.1 ÷ 0.3	0.1 ÷ 0.3
23	870.0	10.0	0.1 ÷ 0.4	0.2 ÷ 0.4
24	900.0	10.0	0.1 ÷ 0.2	0.1 ÷ 0.2

^aThe fourth and fifth columns show typical radiance signal variability within each spectral interval. The ranges are derived from a large number of simulated radiance spectra for clear and cloudy conditions. The standard deviation in each spectral interval is computed for each simulated spectrum, and the minimum and maximum values are reported.

The seven tests applied to each measurement are the following:

1. Mean radiance in all the intervals in Table 1 must be positive. Negative mean radiances in these spectral intervals, below measurement accuracy, could be a sign of measurement problems. For example, some of the spectra characterized by such negative mean radiances show an anomalous slope from the IR window toward increasing wave numbers with large numbers of negative values. In other cases, the whole spectrum results vertically shifted with all the radiances in window regions that fall below zero. These features can be due, for example, to the disturbance caused by people occasionally present in the shelter for maintenance reasons, especially during the summer campaign.
2. The standard deviation (SD) around the mean radiance in interval 24 (900.0 cm⁻¹) must be less than a given threshold value. Radiance in interval 24, a window region, is expected to be quite stable in time: from Table 1 the typical SD is around 10⁻¹ ru, indicating that this interval is a good choice to evaluate excessive noise levels. The SD time series (for all measurements in 2013) shows some abrupt changes and a general trend toward increasing values until December when the instrument was realigned. In addition, the histogram of SD values is close to a Gaussian distribution with some clear outliers, i.e., very noisy observations. The removal of these outliers is performed using a moving window filtering algorithm, in which a neighborhood of observations (called the filtering window) is used to assess the validity of a new observation, taking into consideration its relative distance from the closest neighborhood.
3. Radiance in interval 21 (680.0 cm⁻¹) must be inside the range 35–90 ru, corresponding to a BT range 210–260 K. The test allows to monitor for out of control experimental conditions, for example, the disturbance of people inside the shelter. Even if typically such disturbance does not have a direct effect on the other parts of the spectrum that bring the information of the outside atmosphere, we prefer for the current

Table 2. Results of the Quality Selection Procedure for August and January^a

Test	January	Percent	August	Percent
1	279	13.7	463	26.8
2	106	5.2	84	4.9
3	1	<0.1	21	1.2
4	33	1.6	49	2.8
5	286	14.0	140	8.1
6	1	<0.1	173	10.0
7	5	0.2	7	0.4
Rejected	610	29.9	664	38.7
Good quality	1429	70.1	1061	61.5

^aThe number of spectra rejected due to each of the seven thresholds and the respective percentage amounts are reported.

analysis to exclude these cases. This threshold is rarely important during 2013 and usually concurs at the rejection of a measurement by other tests.

- The difference between the mean radiance in interval 21 and its value in the previous measurement (DR) belonging to the same observation cycle must be less than a fixed threshold. Important variations of the mean radiance are sometimes found between a given spectrum and the next, belonging to the same measurement set. These variations are linked to instability of shelter or instrument temperatures, and we prefer to remove them in this study to consider the most stable conditions. The threshold is defined as three times the mean absolute deviation of the DR around the monthly average. The mean absolute deviation is used because it is less sensitive to outliers than the standard deviation. Great caution has been put when dealing with DR to avoid removing good spectra when the preceding one has shown problems.
- All radiance spectral values between 629.6 and 699.6 cm^{-1} must be positive. This threshold is linked to the presence of spectra characterized by severely distorted absorption or emission lines, a feature that is particularly evident at the Q branch in the center of the CO_2 absorption band. Test number 1 refers to the average radiance, while this one checks radiance in all spectral channels. When fixing the threshold for this test we have considered that the central portion of the CO_2 band is characterized by high NESR values since the instrument is almost completely opaque due the strong absorption of CO_2 present inside the interferometric path. The high noise of this single point produces a large fluctuation in the measured radiance which can distort the CO_2 peak shape.
- Average BT computed in the two CO_2 absorption intervals (629.6 to 634.8 and 694.8 to 699.6 cm^{-1}) are expected to be highly correlated. When their difference falls outside 1σ from the mean of their respective monthly distributions, the spectrum is rejected.
- The spectral position of the CO_2 Q branch must be inside the 666.0 to 669.2 cm^{-1} range. Small shifts are tolerated since the impact on m-w radiances is very small. Larger shifts, due to possible instabilities of the reference laser, are rejected in this study, even if they could be corrected with a subsequent frequency calibration.

Table 2 shows the results of the seven tests applied to the January and August 2013 data sets. The first test is one of the most important, especially during the austral winter months. In August 2013 a high number of spectra fails test 6, while during January many spectra reveal problems linked to the interferogram sampling, causing distortions in line shapes, which are particularly evident looking at the CO_2 central peak (test 5). Considering the entire 2013, the percentage of rejected spectra spans from 20% up to 40% for particularly problematic months such as November.

An example of the quality selection is reported in Figure 2 that shows the time sequence of data relevant for tests 1 and 5 for August 2013. During this month some instrumental problems occurred and a large number of spectra show negative values for averaged radiances in the main IR window, as can be seen in Figure 2a. In same plot, low radiance values linked to clear conditions and cloudy episodes can be visually detected. The distribution of the minimum radiance value in the central region of the CO_2 absorption band in Figure 2b shows values well below zero, denoting spectra affected by severe line shape distortion.

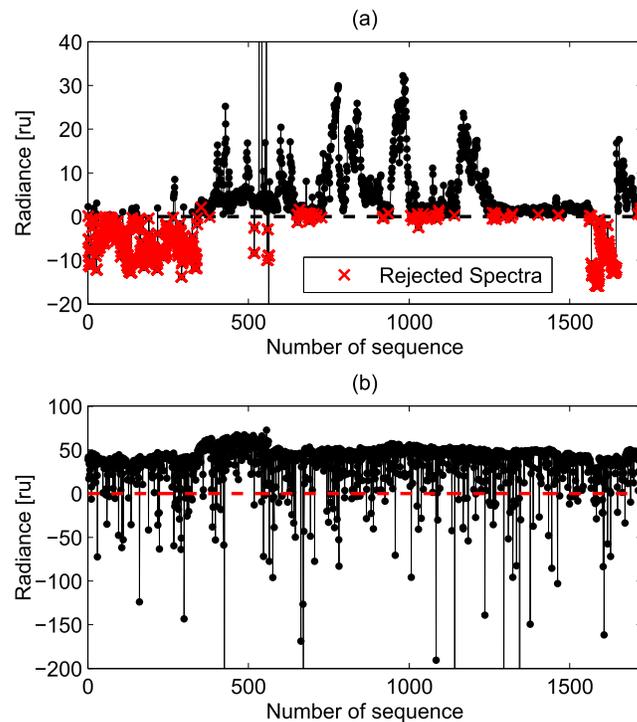


Figure 2. (a) Time series for August 2013 of radiance measured in interval 24 used in test 1; rejected spectra are also reported. (b) Minimum radiance value used in test 5. Fixed thresholds in both panels are reported as dashed lines.

4. Classification of Data Sets

The spectrometer REFIR-PAD and the Lidar are separate instruments, whose measurement strategies are completely independent. Moreover, Lidar data are available in the form of images or quick looks, and only for few selected cases quantitative information was provided. For the above reason we have decided to use a classification algorithm based on the REFIR-PAD spectra only and capable to identify clear from cloudy scenes, water from ice clouds, and classify the latter as precipitating or/and nonprecipitating. The further classifications concerning phase and precipitation are not discussed in this paper, but will be the topic of another manuscript that is currently in preparation. Dimensionality reduction is a key problem that aims at mapping high dimensional data onto a space of lower dimensionality, preserving at the same time the maximum information content. This problem can be dealt with two methodologies:

1. *The feature extraction that operates a linear or a nonlinear mapping of the data space.* In this group, we find, for example, methods that exploit eigendecomposition techniques whose objective is to find a set of mutually orthogonal basis functions that capture the directions of maximum variance in the data set [Cai *et al.*, 2007, 2007] but may not be relevant for classification purposes.
2. *The feature selection approach that tries to find a subset of the original variables, eliminating redundant information.* In this latter case, the physical meaning of the features is maintained.

Two data selection approaches have been tested using the above methodologies: (1) unsupervised feature extraction, using Principal Component Analysis and spectral regression, and (2) supervised feature selection, using concepts linked to linear discriminant analysis and support vector machine. After several tests we have found that the possibility to trace the results to features that possess a well-defined physical meaning is a great advantage. Therefore, only techniques and results that belong to the second group will be examined further.

The purpose of classification is to group data which are similar to each other in some respect. The first classification applied to the measurements is between clear and cloudy cases, then the latter group can, in principle, be separated according to phase (ice or water cloud) and to type (precipitating or nonprecipitating clouds). Any classification method is based on the choice of a set of features that are considered important for the task. The selected features are based solely on data measured in specific channels, listed in Table 1. The BT corresponding to the averaged radiance in each interval is computed, and the normalized differences between

each pair of them (BTD) are calculated, thus defining 276 distinct features. Differences are considered in order to obtain a large number of features describing spectra, and normalization is used to free the results from global fluctuations of spectra linked to instrumental instability and seasonal variability.

The data to which the classification is applied cover 13 consecutive months, from December 2012 to December 2013. A preliminary step consists in the selection among all good quality observations of two groups of spectra that form the training and the test data sets (TRD and TED, respectively). The selection of members of both sets is, in our case, obtained manually, using daily lidar quick looks, trying to select examples in the most scattered and random way, and representing a fair variety of observed conditions. For the clear-cloudy classification, generally 20 to 40 clear and just as many cloudy spectra are chosen as TRD for each trimonthly period; same criteria are used to generate the TED.

First, the features which are truly important for describing the observations and for classification purposes are found using techniques linked to linear discriminant analysis [Bishop, 2006]. Considering the case of two classes and supposing to have a D -dimensional input vector \mathbf{x} , one can project it down to one dimension using

$$y = \mathbf{w}^T \mathbf{x}; \quad (1)$$

where \mathbf{w} is a weight vector. In order to find the best projection direction \mathbf{w} , the idea proposed by Fisher [Bishop, 2006] is to maximize a function that will give a large separation between the projected class means (μ) while also giving a small variance within each class so that the class overlap is minimized. The so-called within class, also named intraclass, variance of the transformed data from class C_k is given by

$$s_k^2 = \sum_{n \in C_k} (y_n - \mu_k)^2; \quad (2)$$

where $y_n = \mathbf{w}^T \mathbf{x}_n$. The so-called interclass distance to be maximized is defined by Fisher as

$$d_{12} = \frac{(\mu_{\text{clear}} - \mu_{\text{cloud}})^2}{s_{\text{clear}}^2 - s_{\text{cloud}}^2}; \quad (3)$$

where the denominator is a measure of the scattering of these points in the two belonging classes.

Based on these considerations, the aim of the implemented feature selection algorithm is to find a set of features that, used together, not only lead to a minimum classification error but also represent data at best, i.e., features characterized by least correlation, in order to avoid redundant information, and satisfying the Fisher discrimination idea. The implemented feature selection algorithm involves a sequential procedure. Initially, each feature is singularly considered and an N -fold cross validation is performed: the TRD is partitioned in N sets (usually 10 folds), $N - 1$ are considered as training examples, while the last fold is cast as test set. In this way, N different data sets are obtained each of which samples and uses examples contained in the TRD in a different way. Considering each of these N artificial data sets singularly, the code classifies each measure cast as test example, using the k nearest-neighbor technique (with $k = 5$). The classification error can be determined as the fraction of test examples misclassified (knowing the true class of this cast test set) and averaging over all the N -generated folds. In this way, at the first iteration, the BTD that results to have the minimum classification error is taken or, in case of joint winners, the one that maximizes the normalized distance between the examples interclass distance is chosen, following the Fisher formulation (equation (3)). For the choice of the second feature, the classification error is again considered, but now evaluated on each pair of features, the first of which is the one already chosen. In addition, because with the increasing of the index of the iteration a lot of features lead to the same classification error, another parameter in addition to the interclass distance is considered: the correlation between the new feature and the one already selected. When other features are added, the correlation is calculated for each feature with respect to all the already chosen ones. This sequential process is interrupted when the classification error either remains stable for the last four iterations or increases with respect to the previous one. The final set of features collects the most relevant ones for the classification task, a small (usually three to six) sample of the defined BTDs. This process is characterized by a basic randomness: in fact, once the training examples are fixed, the partition of the training data set for the cross-validation process has a random character. As a consequence, repeating the algorithm several times, slightly different features are generally chosen at each run. This randomness can be reduced considering a higher number of folds during the cross validation. In addition, taking 10 runs for the same period and considering more than 20 different combinations of months (in pairs or triads) during the year, we can evaluate the most recurrent features, reported in Table 3.

Table 3. List of the Most Recurrent Features Selected to Discriminate Clear and Cloudy Cases^a

Interval at	Interval at	Occurrences	Type
680.0	840.0	73	CO ₂ -WIN
531.2	680.0	56	FIR-CO ₂
680.0	900.0	44	CO ₂ -WIN
680.0	870.0	41	CO ₂ -WIN
559.2	680.0	28	FIR-CO ₂
573.2	840.0	27	FIR-WIN
478.8	840.0	27	FIR-WIN
464.4	840.0	25	FIR-WIN
522.4	680.0	24	FIR-CO ₂
531.2	840.0	22	FIR-WIN
381.2	680.0	21	FIR-CO ₂

^aThe feature selection algorithm is run 10 times for 24 different combinations of months. Columns 1 and 2 show the central wave number of the intervals used to obtain the BTD.

As can be seen, the most frequently chosen BTDs are generally the ones computed considering interval 21 in the CO₂ band (680.0 cm⁻¹) and a window region, either in the IR or in the FIR. Columns 1 and 2 show the reference central wave numbers for the spectral intervals defined in Table 1. At this point, the data set is described by means of the selected features only, and it is passed to the support vector machine (SVM) [Theodoridis *et al.*, 2010]. A linear kernel is used in this case. At the heart of SVM learning method there is the margin concept: SVM is a classifier that aims at separating the two class elements, looking not only at the classification error but also at the maximum distance between elements belonging to the two classes. Considering a classifier as

$$\mathbf{w}^T \mathbf{x} + w_0 = 0, \tag{4}$$

the margin is defined as the region between the two hyperplanes:

$$\mathbf{w}^T \mathbf{x} + w_0 = 1; \quad \mathbf{w}^T \mathbf{x} + w_0 = -1. \tag{5}$$

It can be shown that the distance between each point on these two hyperplanes and the classifier is equal to $\frac{1}{\|\mathbf{w}\|}$ [Theodoridis *et al.*, 2010]. In the SVM frame, the cost function corresponding to the classification error is expressed as

$$J(\mathbf{w}, w_0, \delta) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \delta_i; \tag{6}$$

where C is a constant that weights the classification errors δ_i . These errors are different from zero also for points that lie inside the margin, even if it is on the correct side of the classifier. So the aim expressed with this cost function is to make the margin as large as possible: in fact, minimizing $\|\mathbf{w}\|$ corresponds to maximize the margin width, i.e., find that particular separating hyperplane that has the maximum distance between the two class elements; at the same time we want to keep the number of points with $\delta_i > 0$ as small as possible. This characteristic allows the SVM good general performance when applied to data sets outside of the training and test samples.

It is found that the solution is given by

$$\mathbf{w} = \sum_{i=1}^N \lambda_i y_i \mathbf{x}_i \tag{7}$$

where λ_i are the Lagrange multipliers of the optimization process: they are different from zero only for points that lie on the wrong side of the margin or inside it; these last points are the very observations that determine the separator and are called support vectors [Theodoris and Koutroumbas, 2006]. Moreover, the classification error is calculated for TED, and the constant C value is accordingly regulated in order to minimize this error.

An example of the result of the SVM classification for June, July, and August is reported in Figure 3, where also training and test points are plotted; the test error is reported together with the training error in the head of

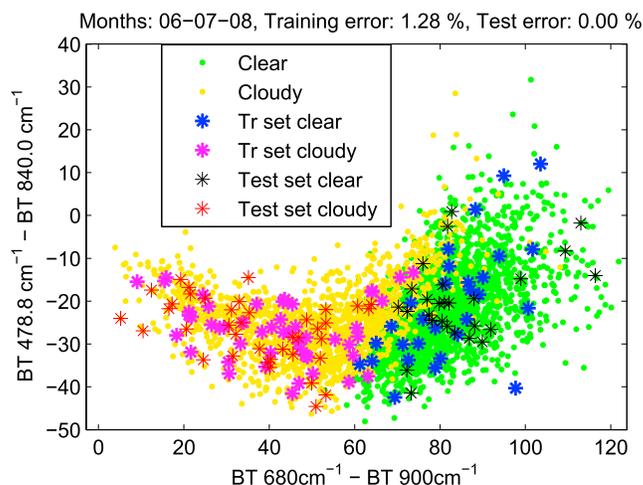


Figure 3. Example of SVM clear-cloudy classification: scatter plot of the overall measured set over the two most important selected BTD, with training and test examples superimposed.

the figure. As can be expected a clearly defined boundary between the two classes is not present. The SVM uses a set of feature to perform the classification, while these plots are only a projection of the data set on two of them. With this procedure it is possible to classify measurements in either the clear or the cloudy class and only points that lie on the correct side of the plane and outside the margin are considered. The lists of spectra classified as clear and as cloudy are built so that the output can be compared with quick looks lidar again. The comparison is very satisfactory in most cases, albeit for a few observations, taken in presence of weak diamond dust episodes or high clouds of very small optical depth, which could be considered as clear.

5. Modeling Approach for Clear-Sky Computations

A further selection among clear cases is done for the purpose of comparing measurements to simulations. Among the cases classified as clear, discussed in section 4, a subset of spectra is selected for days in which the radiosonde profile data are available up to at least 5 km asl and a sequence of at least four clear REFIR-PAD measurements are available within 70 min from 12 UTC (the nominal radiosonde time). Therefore, the selected measurements denote a sufficiently stable condition, and the average over a variable number of spectra can be compared to the simulated one. Moreover, no time interpolation is required for the profile data. The final data set, which will simply be called the selected clear set (SCS), covers 66 days in all seasons (January–February, 16 days; March–May, 15 days; June–August, 17 days; September–November, 14 days; and December, 4 days) and 404 REFIR-PAD measurements overall, with an average of more than six consecutive spectra for each day.

Monochromatic molecular optical depth (σ_m) in each layer is computed using the Line-by-Line Radiative Transfer Model (LBLRTM) [Clough *et al.*, 2005] version 12.2. The Atmospheric and Environmental Research line parameter database v.3.2 and the Mlawer-Tobin-Clough-Kneizys-Davies (MT-CKD) 2.5.2 water vapor continuum model [Mlawer *et al.*, 2012] are used. The σ_m in each layer is interpolated at 0.0008 cm^{-1} , and the layer-to-space transmittances are computed and finally convolved with a REFIR-PAD instrument spectral response function (ISRF) obtained padding the interferogram with zero values from 1.09375 cm up to maximum delay (1.25 cm). Such an ISRF minimizes the standard deviation between simulations and measurements in a spectral interval of width 50 cm^{-1} around 850 cm^{-1} , for averaged measurements in very stable clear conditions belonging to the SCS set. The radiative transfer equation is solved for a plane-parallel geometry, and an extended version of the code RT3 [Evans and Stephens, 1991], based on a doubling and adding method, is used to compute radiances both in clear and in cloudy conditions. All computations are done using 62 levels up to 60 km, the spacing between the levels increasing with height: 24 levels describe the lowest 1 km agl (11 of which describe the lowest 100 m agl), and next 24 describe the profile from 4.3 and 10.3 km asl. In this way, a very good reconstruction of the large variations occurring close to the ground is attained and also of the atmospheric layers in which a large fraction of water vapor mass is found and most cirrus clouds are observed. The accurate simulation of downwelling emitted

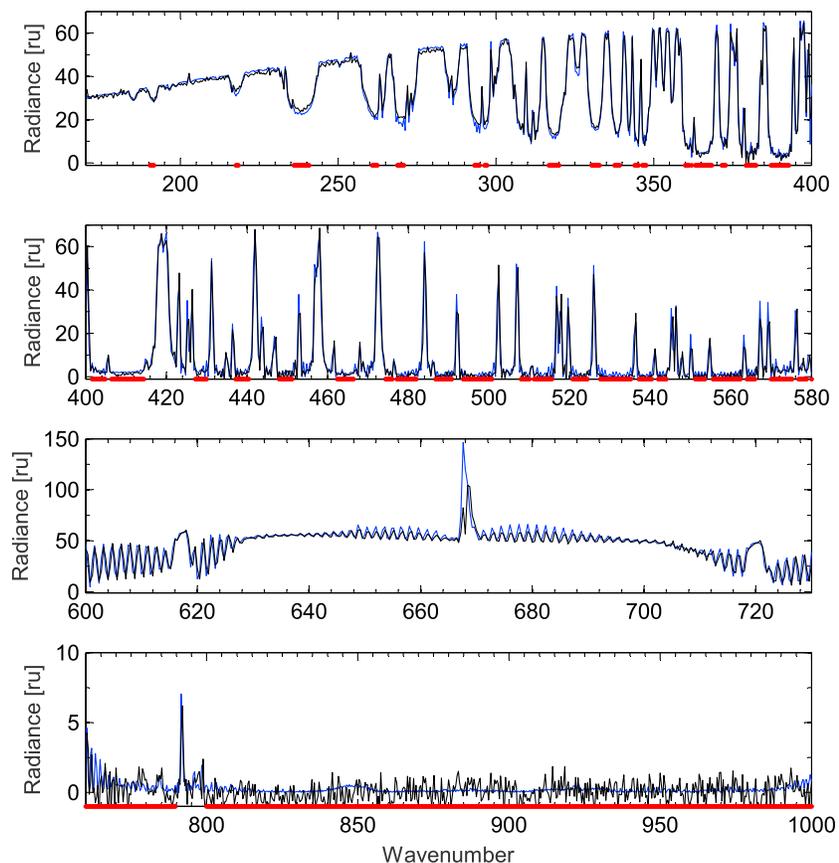


Figure 4. (first to fourth panels) Measured and simulated spectra in clear conditions for 16 July 2013. In all panels the black(blue) line are the measured(simulated) values. The spectral channels that belong to the m-w are denoted by a red dot below the zero line.

spectral radiance measured by REFIR-PAD requires an accurate reconstruction of all radiation sources. Accordingly, the temperature and humidity profiles close to the ground are modified to account for the fact that the spectrometer is located in the physics shelter, below a vertical cylindrical structure whose upper aperture is slightly larger than the field of view of the instrument. The spectrometer is assigned at 3237 m asl (4 m agl), which is the lowest level for the computations, and the internal temperature of the spectrometer and the temperature measured in the cylindrical shelter are averaged and assigned at this height; the temperature and humidity measured by the local weather station are assigned at 3241 m asl, immediately above the shelter aperture, which is the second level in the final profile.

Measured and simulated spectra of a clear case (16 July 2013) are shown in Figure 4. This case is same as the one shown in Figure 1a and represents the average over seven spectra. It has been selected as it shows some of the problems encountered and should not be considered a best case scenario. The temperature profile used for the simulation has a temperature inside the instrument shelter (3237 m asl) of 294.6 K and a temperature of 205.1 K at 3240 m, immediately outside the shelter. The radiosonde data reports a temperature increase up to 225.7 K at 3784 m asl and a precipitable water vapor (vertically integrated water vapor amount, PWV) of 0.15 mm. Overall, the comparison brings into evidence the quality of the measurements, also considering the extreme conditions in which these are taken, since radiosonde data report a temperature close to surface of 198 K. Some differences are evident in the most transparent FIR windows and in the central portion of the CO₂ Q branch. The first region is of direct concern for our study, since the downwelling radiance emitted by a cloud layer can be well characterized from the ground in the most transparent FIR spectral regions. The simulated radiance consistently underestimates the measured one in every partial window in the whole spectral interval shown in Figure 4, while it slightly overestimates the radiance in the most absorbing regions. While single measured spectra may show noisy features, these tend to be smoothed out when measurements are averaged. We notice that some spikes, as, for example, the upward pointing spikes present in Figure 4 at 202.8,

254.0, and 303.2 cm^{-1} , are real measurement features due to emission by water vapor within the first few meters inside the partially sheltered environment. These features are also present in our simulation because the lowest computation level is assigned as the average of REFIR-PAD and shelter temperature. The central portion of the CO_2 absorption band is of interest to explain some other peculiar spectral features of the measurements and to clarify the role of some of the quality tests that were discussed in section 3, (in particular, item 5). The spectrum shown in Figure 4 (third panel) indeed shows a slight frequency shift and a mismatch in the shape of the central peak of the CO_2 band. However, the frequency shift is due to possible fluctuations of the reference laser and can be easily calibrated by comparing the measurement with LBLRTM simulation. The high noise present at the central frequency produces a large fluctuation in the measured radiance which can distort the CO_2 peak shape, as shown in the figure by the downward pointing line portion. These features are believed to be due to particular measurements conditions as REFIR-PAD is unsupervised for long months in the austral winter. However the spectra in the m-w, which are our primary concern in our project, are only very marginally affected by these features. Finally, it is stressed that no retrieval method or fitting procedure has been adopted to modify any portion of the vertical profile data to improve the simulated radiances.

6. Results

For the purpose of presenting averaged results in spectral region which are partially transparent, a set of m-w has been selected, as defined in Table 4. The m-w have various width and have quite different properties in terms of mean transmittance but represent the most transparent spectral regions, although their mean transmittance diminishes with decreasing wave number. The measurements contained in the SCS, defined in section 5, are processed as follows:

1. For each day in the SCS, the simulated radiance $L(\nu)$ is computed and the mean measured spectrum is derived in all REFIR-PAD channels ($M(\nu)$), together with its standard deviation ($\sigma_M(\nu)$), taken as a measure of spread around the mean value.
2. The difference $\delta L(\nu) = L(\nu) - M(\nu)$ are averaged in each m-w. The latter average $\delta L_w(\nu_w)$ is a weighted one according to the squared inverse of $\sigma_M(\nu)$ so that greater weight is given to REFIR-PAD channels with lesser variance within same m-w. The standard deviation around the mean is also computed ($\sigma_{Lw}(\nu_w)$) in each m-w.
3. For each m-w, the weighted average of $\delta L_w(\nu_w)$ is computed over all days of the SCS data set, thus defining the time averaged $\delta L_{wt}(\nu_w)$; the weights used in the averaging process are proportional to the inverse of the square of $\sigma_{Lw}(\nu_w)$ so that a smaller weight is given to days in which the m-w differences change rapidly. The standard deviation around the yearly mean ($\sigma_{Lwt}(\nu_w)$) is also computed in each m-w.
4. For each day in the SCS, the total spectral measurement uncertainty in each m-w $\sigma_{tw}(\nu_w)$ is computed using

$$\frac{1}{\sigma_{tw}(\nu_w)} = \left[\sum_{i=1}^{nc} \frac{1}{\sigma_i^2(\nu_i)} \right]^{1/2} \quad (8)$$

where the index i spans the number of channels in each m-w. Finally, a global (over all SCS cases) estimate of measurement uncertainty can be obtained by a weighted average as in point 3 above, denoted by $\sigma_{twa}(\nu_w)$.

The mean difference $\delta L_{wt}(\nu_w)$ is shown in Figure 5, the vertical bars are $1 - \sigma_{Lwt}(\nu_w)$, and the red dashed lines connect the values of $\sigma_{twa}(\nu_w)$. The results show clearly that simulations are in very good agreement for all m-w in the range 400 to 1000 cm^{-1} , but differences are substantial between 200 and 400 cm^{-1} with largest values around 295.1 cm^{-1} . A negative difference implies that the simulated optical depth is smaller than the actual. The yearly spread estimate σ_{Lwt} and the measurement uncertainty estimate σ_{twa} provide clear evidence that the differences δL_{wt} constitute a real systematic deviation.

Spectroscopic databases provide evidence that water vapor is the main absorber from 100 to 400 cm^{-1} ; CO_2 , O_2 , O_3 , and NH_3 have absorption lines in that range, but the associated optical depth, for a typical atmospheric profile, is much less than water vapor's. In any case these gases are included in the computations. Eventual temperature profiling error, especially close to the ground, would be observed in the difference between simulation and measurement in spectral channels of low transmittance; these have been monitored but are not discussed in this paper. The main cause of the discrepancy must therefore be attributed to the incorrect computation of water vapor optical depth. There are two possible causes: errors in the measurements by the humidity sensor in the radiosonde package and the underestimation of the total water vapor absorption used in the simulations.

Table 4. Selected Microwindows (m-w)^a

N	Central Wave Number (cm ⁻¹)	Width (cm ⁻¹)	N	Central Wave Number (cm ⁻¹)	Width (cm ⁻¹)
1	191.0	1.2	23	497.2	7.2
2	218.0	0.8	24	511.9	7.6
3	238.4	4.8	25	522.6	3.6
4	261.6	1.6	26	531.4	7.6
5	269.8	2.0	27	538.8	3.2
6	295.1	4.0	28	548.1	11.6
7	318.4	3.2	29	559.0	6.8
8	331.6	2.4	30	565.0	2.0
9	338.4	1.6	31	572.6	5.2
10	345.8	3.6	32	578.1	3.2
11	364.5	8.4	33	774.8	29.6
12	372.2	1.2	34	810.0	20.0
13	380.8	3.2	35	830.2	19.6
14	390.0	5.6	36	850.2	19.6
15	403.2	3.2	37	870.2	19.6
16	410.4	8.0	38	890.2	19.6
17	428.6	2.8	39	910.2	19.6
18	438.8	3.2	40	930.2	19.6
19	449.6	3.2	41	950.2	19.6
20	464.4	4.0	42	970.2	19.6
21	478.4	7.6	43	990.2	19.6
22	488.8	4.0			

^aCentral wave number (wnb) and spectral width are reported for each of the microwindow.

Currently, atmospheric humidity is predominantly measured using radiosondes, and their humidity sensors can experience errors due to inaccuracies in calibration, time lag, contamination, and solar heating of the humidity sensor, the latter being referred to as dry bias. Much work has been done to identify and minimize sources of error in different generations of radiosondes [Turner *et al.*, 2003; Soden *et al.*, 2004; Häberli, 2006; Rowe *et al.*, 2008; Wang *et al.*, 2013] using various experimental methods to compare with the radiosonde data, and with reference to various areas of the globe and daylight conditions. The Rowe *et al.* [2008] paper deals with a correction based on measurements of downwelling infrared atmospheric spectral radiance made at Dome C, during the austral summers of January 2003 and December 2003 to January 2004. The location

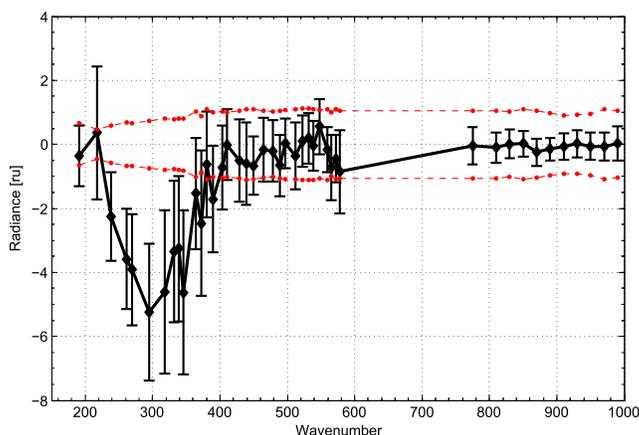


Figure 5. Weighted averaged differences between simulated and measured radiances $\delta L_{wt}(v_w)$. Vertical bars are $\sigma_{Lwt}(v_w)$ values computed over all (66) cases, and the red dashed lines connect the values of $\sigma_{twa}(v_w)$

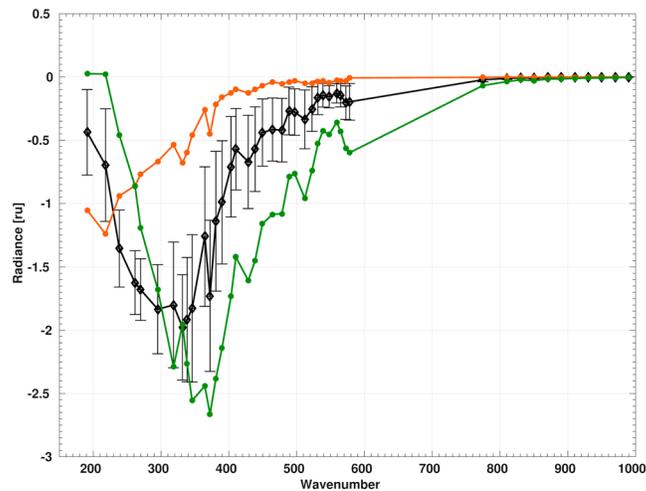


Figure 6. Mean difference ($D - S$, black line with dots) over the SCS data set between downwelling radiance computed with the nominal profile D and with the mixing ratio increased by 10% at all levels S . The orange and green lines are the differences $D - S$ for the days with minimum and maximum PWV (0.072 and 0.797 mm, respectively).

is same as of the measurements in the present paper, but the sensor being investigated is the Vaisala RS90, while in our experiment the sonde RS92 is used. Therefore, we have no previous work that can be used to evaluate the eventual dry bias error.

The effect of a water vapor bias can be estimated by computing the difference, in all m-w, between downwelling radiance computed with the nominal profile (case D) and with the mixing ratio increased by 10% at all levels (case S). It is expected that an increase in water vapor content should increase the downwelling radiance in window regions where water vapor is the main absorber, hence producing a negative $D - S$ difference. The average over all days of the SCS of $D - S$ is the black line in Figure 6. The orange and green lines are the differences $D - S$ for the days with minimum and maximum PWV (0.072 and 0.797 mm) and show that the spectral signature is a complex function of PWV and of the temperature profile. The mean difference in Figure 6 is negative in the whole range, with a minimum value of about -2 ru, which is smaller than the difference shown in Figure 5. Moreover, from 400 to 600 cm^{-1} the differences are clearly negative, contrary to the results shown in Figure 5. Very low differences are obtained for wave number greater than about 600 cm^{-1} because of the very low PWV values in our SCS. These results agree with previous results [Tobin *et al.*, 1999; Mlynarczyk *et al.*, 2016], but the spectral range is here extended because sensitivity studies have never dealt with such low values of water vapor content.

Liuzzi *et al.* [2014] discuss the water vapor absorption models and propose a methodology to retrieve the foreign broadening coefficient in the range 180 – 600 cm^{-1} . The forward model and water vapor continuum model is same as used in the present paper. They process 1 day in summer and 1 day in winter 2012, since the PWV and boundary layer temperature profile are quite different in the two seasons. The methodology involves the retrieval of the temperature and water vapor mixing ratio. The authors state that the modifications of the line air-broadened half widths, temperature dependencies, and pressure shifts in the range 350 – 667 cm^{-1} , introduced in the version 12.0 of LBLRTM, describe adequately the water vapor emission, while no revisions of these parameters have been carried on for wave number lesser than 350 cm^{-1} . Since the MT-CKD 2.5.2 is based on an analytical form of the continuum absorption, this leads to an evident misfit in the windows below 350 cm^{-1} . Fox *et al.* [2015] present results based on measurements performed in 2007 in the Arctic. They state that the study demonstrates that the current MT-CKD 2.5 model continuum values can be used to accurately simulate radiances at wave numbers 350 – 500 cm^{-1} .

The results can be further discussed using the data presented in Figure 7. The four panels show $\delta L_w(v_w)$ (vertical bars are $\sigma_{L_w(v_w)}$) versus transmittance for all SCS cases, in four m-w centered at 218.0 (m-w 2), 295.1 (6), 364.5 (11) and 449.6 (19) cm^{-1} . In each panel, different symbols are used to identify measurements taken in various seasons: black diamonds indicate austral winter, open circles austral summer, and the rest (middle-season months) is marked with an upward pointing triangle. The abscissa is the simulated spectral transmittance, averaged in each m-w (hence mean transmittance or mt), computed for a layer between

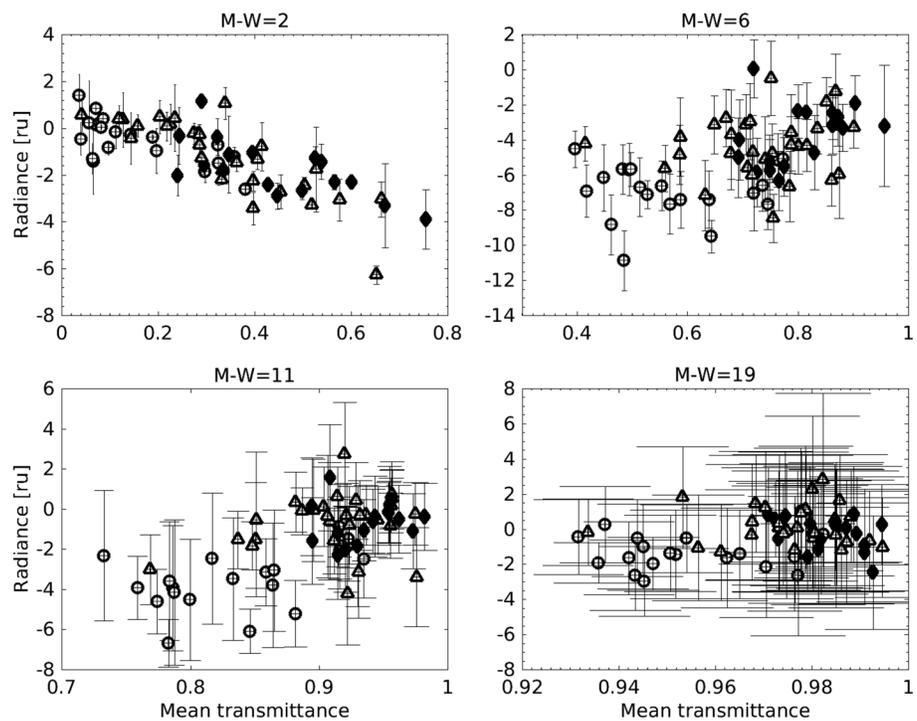


Figure 7. δL_w with σ_{L_w} vertical bars for four selected m-w versus mean layer transmittance mt , as explained in the text; black diamonds indicate austral winter, open circles austral summer, and upward pointing triangle middle-season months.

surface and level 22 in our simulations, for a total depth of 730 m. Such a layer has been used to define objectively the set of m-w appropriate for our study, as the layer-integrated water vapor amount is generally a large fraction of the PWV and the spectral transmittance is greater than zero in all spectral channels. For this reason it is chosen as a useful (not unique) reference. The four panels (m-w 2, 6, 11, and 19) are ordered for increasing mt . Collectively, the results show that when the mt is small, the negative bias is also small, and we expect the bias to be zero when mt is close to unity, since nearly no absorption is present. In the intermediate range, a spectrally dependent negative bias is evident in all m-w except in m-w 19, which implies that the simulated optical depth is smaller than it should be. After ruling out that the bias is caused by an error in the measurement of the water vapor profile, because of the different spectral signature, it is the water vapor absorption that must be in error, possibly both the line and the continuum component. The results in m-w 2 where the bias is close to zero for $mt < 0.2$, indicates that in that m-w, the effect of incorrect modeling of water vapor total absorption (line and continuum) arises for a widespread range of conditions and disappears approaching to saturation. It must be stressed that we are discussing the results in terms of only the water vapor content, yet there are important fluctuations in the temperature profile in different seasons and also the structure of the water vapor mass profile changes in different seasons, which can partially explain the scatter of data observed in Figure 7.

To conclude, we have no correction for the important negative difference below 400 cm^{-1} , although the data clearly point at a problem with the magnitude of the total water vapor absorption used for the simulation. The results presented show that a bias is already evident in the m-w around 365 cm^{-1} .

The spectral range between 200 and 550 cm^{-1} is an important spectral range because the retrieval of cloud properties, which is part of the overall research project but will be discussed elsewhere, is done using also channels in this range. Some modeling will be required when discussing the cloud cases, to account for the negative differences observed in clear conditions, due to the imperfect knowledge of the water vapor total absorption, which affects also simulated spectra in cloudy conditions.

7. Summary and Conclusions

The present work examines downwelling radiance spectra measured at the ground during 2013 by REFIR-PAD, a FIR FTS, capable of nearly continuous operation at Dome C, Antarctica. A tropospheric backscatter and depolarization Lidar is also deployed at same site, and a radiosonde system is routinely operative that supplies one radiosonde profile at 12 UTC. The availability of a full year of spectroscopic measurements, acquired automatically without human supervision, implies the need for automated checks that every spectrum must pass in order to be considered of sufficient quality for a quantitative scientific analysis. A battery of tests was developed and is described in detail.

A general classification procedure is applied to discriminate spectra measured in clear sky conditions from cloudy conditions, which requires only REFIR-PAD spectra and makes use of the lidar quick looks. The classification procedure is also capable of further discriminating cloudy spectra, but this will be described elsewhere. A further selection among the cases classified as clear is done for the purpose of comparing measurements to simulations: a subset of spectra (called SCS) is selected for days in which the radiosonde profile data are available up to at least 5 km asl and a sequence of at least four clear REFIR-PAD measurements are available within 70 min from 12 UTC (the nominal radiosonde time).

The computation of layer molecular optical depth is performed with line-by-line techniques and a convolution to simulate the REFIR-PAD measurements; the downwelling radiance is computed with a state-of-the-art adding-doubling code with the capability to handle multiple scattering.

The mean difference between simulated and measured radiance, over all cases within the SCS, is negative for all microwindows below 400 cm^{-1} and increases substantially in magnitude, up to 5.5 ru , with decreasing wave number with largest values around 295.1 cm^{-1} . The mean difference is substantially larger than the measurement uncertainty estimate. The paper discusses possible reasons for the observed larger discrepancy in the range 400 to 200 cm^{-1} and provides evidence that it is not caused by error in water vapor profiling. Instead, it identifies the incorrect magnitude of the water vapor total absorption coefficients as the cause of such large negative radiance bias.

One of the main objectives of the project is the retrieval of microphysical properties of Antarctic ice clouds. During the data analysis it was found that an in-depth understanding of the results obtained in clear conditions was necessary to discuss the results obtained in cloudy ones, and for this reason the present paper is devoted to the description of results obtained in clear conditions. The retrieval of cloud properties, which represents the final aim of the project, will be described in a separate paper that will also discuss how further classification of case studies can be done and the methods adopted to accurately simulate radiance in cloudy conditions.

Acknowledgments

This work is supported by the Italian PNRA (Programma Nazionale di Ricerche in Antartide) under the subproject 2009/A04.03 PRANA (Proprietà Radiative dell'Atmosfera e delle Nubi in Antartide). Radiosounding measurements are part of the IPEV/PNRA Project "Routine Meteorological Observation" at Station Concordia—www.climantartide.it/index.php?lang=enea. The REFIR-PAD data are available from <http://refir.fi.ino.it/refir-pad-domeC>.

References

- Bhawar, R., et al. (2008), Spectrally resolved observations of atmospheric emitted radiance in the H_2O rotation band, *Geophys. Res. Lett.*, *35*, L04812, doi:10.1029/2007GL032207.
- Bianchini, G., and L. Palchetti (2008), Technical note: REFIR-PAD level 1 data analysis and performance characterization, *Atmos. Chem. Phys.*, *8*, 3817–3826.
- Bianchini, G., L. Palchetti, and B. Carli (2006), A wide-band nadir-sounding spectroradiometer for the characterization of the Earth's outgoing long-wave radiation, in *International Society for Optics and Photonics Remote Sensing*, pp. 63610A–63610A, Sensors, Systems, and Next-Generation Satellites X, Stockholm, Sweden.
- Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Springer, New York.
- Cai, D., X. He, W. V. Zhang, and J. Han (2007), Regularized locality preserving indexing via spectral regression, in *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management*, pp. 741–750, ACM, New York.
- Clough, S., M. Shephard, E. Mlawer, J. Delamere, M. Iacono, K. Cady-Pereira, S. Boukabara, and P. Brown (2005), Atmospheric radiative transfer modeling: A summary of the AER codes, *J. Quant. Spectros. Radiat. Transfer*, *91*, 233–244.
- Evans, K., and G. Stephens (1991), A new polarized atmospheric radiative transfer model, *J. Quant. Spectros. Radiat. Transfer*, *46*, 413–423.
- Fox, C., P. D. Green, J. C. Pickering, and N. Humpage (2015), Analysis of far-infrared spectral radiance observations of the water vapor continuum in the Arctic, *J. Quant. Spectros. Radiat. Transfer*, *155*, 57–65.
- Häberli, C. (2006), Assessment, correction and impact of the dry bias in radiosonde humidity data during the MAP SOP, *Q. J. R. Meteorol. Soc.*, *132*, 2827–2852.
- Harries, J., B. Carli, R. Rizzi, C. Serio, M. Mlynarczyk, L. Palchetti, T. Maestri, H. Brindley, and G. Masiello (2008), The far-infrared Earth, *Rev. Geophys.*, *46*, RG4004, doi:10.1029/2007RG000233.
- Liuzzi, G., G. Masiello, C. Serio, L. Palchetti, and G. Bianchini (2014), Validation of H_2O continuum absorption models in the wave number range 180 – 600 cm^{-1} with atmospheric emitted spectral radiance measured at the Antarctica Dome-C site, *Opt. Express*, *22*, 16784–16801.
- Maestri, T., R. Rizzi, E. Tosi, P. Veglio, L. Palchetti, G. Bianchini, P. Di Girolamo, G. Masiello, C. Serio, and D. Summa (2014), Analysis of cirrus cloud spectral signatures in the far infrared, *J. Quant. Spectros. Radiat. Transfer*, *141*, 49–64.

- Masiello, G., C. Serio, F. Esposito, and L. Palchetti (2012), Validation of line and continuum spectroscopic parameters with measurements of atmospheric emitted spectral radiance from far to mid infrared wave number range, *J. Quant. Spectros. Radiat. Transfer*, *113*, 1286–1299.
- Mlawer, E. J., V. H. Payne, J.-L. Moncet, J. S. Delamere, M. J. Alvarado, and D. C. Tobin (2012), Development and recent evaluation of the M_CKD model of continuum absorption, *Philosophical Transactions of the Royal Society of London A: Mathematical, Phys. Eng. Sci.*, *370*, 2520–2556.
- Mlynczak, M. G., R. P. Cageao, J. C. Mast, D. P. Kratz, H. Latvakoski, and D. G. Johnson (2016), Observations of downwelling far-infrared emission at Table Mountain California made by the FIRST instrument, *J. Quant. Spectros. Radiat. Transfer*, *170*, 90–105.
- Palchetti, L., D. Lastrucci, R. Bonsignori, and R. Rizzi (2001), Design of the space-borne Radiation Explorer in the Far InfraRed (REFIR)—Technical solutions, in *IRS 2000: Current Problems in Atmospheric Radiation*, pp. 823–826, Deepak Publ., San Diego, Calif.
- Palchetti, L., G. Bianchini, F. Castagnoli, B. Carli, C. Serio, F. Esposito, V. Cuomo, R. Rizzi, and T. Maestri (2005), Breadboard of a Fourier-transform spectrometer for the Radiation Explorer in the Far Infrared atmospheric mission, *Appl. Opt.*, *44*, 2870–2878.
- Palchetti, L., G. Bianchini, G. Di Natale, and M. Del Guasta (2015), Far infrared radiative properties of water vapor and clouds in Antarctica, *Bull. Am. Meteorol. Soc.*, *96*(9), 1505–1518, doi:10.1175/BAMS-D-13-00286.1.
- Rizzi, R., B. Carli, J. Harries, J. Leotin, C. Serio, A. Sutera, and B. Bizzarri (1998), REFIR—Radiation Explorer in the Far Infrared, in *Proc. of the 8th International Workshop on Atmospheric Science From Space Using Fourier Transform Spectrometry*, pp. 47–54, Toulouse, France, 16–18 November 1998.
- Rizzi, R., B. Carli, J. Harries, J. Leotin, C. Serio, A. Sutera, B. Bizzarri, R. Bonsignori, and S. Peskett (2000), Mission objectives and instrument requirements for the REFIR—Radiation Explorer in the Far InfraRed—mission: An outline after the end of phase B0, in *IRS 2000: Current Problems in Atmospheric Radiation*, S. Petesburg, Russia, edited by W. L. Smith and Y. M. Timofeyev, pp. 567–570, A. Deepak, Hampton, Va., 24–29 July 2000.
- Rowe, P. M., L. M. Miloshevich, D. D. Turner, and V. P. Walden (2008), Dry bias in Vaisala RS90 radiosonde humidity profiles over Antarctica, *J. Atmos. Oceanic Technol.*, *25*, 1529–1541.
- Serio, C., G. Masiello, F. Esposito, P. D. Girolamo, T. D. Iorio, L. Palchetti, G. Bianchini, G. Muscari, G. Pavese, R. Rizzi, B. Carli, and V. Cuomo (2008), Retrieval of foreign-broadened water vapor continuum coefficients from emitted spectral radiance in the H₂O rotational band from 240 to 590 cm⁻¹, *Opt. Express*, *16*, 15816–15833.
- Soden, B. J., D. D. Turner, B. M. Lesht, and L. M. Miloshevich (2004), An analysis of satellite, radiosonde, and lidar observations of upper tropospheric water vapor from the Atmospheric Radiation Measurement Program, *J. Geophys. Res.*, D04105, doi:10.1029/2003JD003828.
- Theodoridis, S., A. Pikrakis, K. Koutroumbas, and D. Cavouras (2010), *An Introduction to Pattern Recognition: A MATLAB Approach*, Academic Press, Orlando, Fla.
- Theodoris, S., and K. Koutroumbas (2006), *Pattern Recognition*, 3rd ed., Academic Press, Orlando, Fla.
- Tobin, D. C., et al. (1999), Downwelling spectral radiance observations at the SHEBA ice station: Water vapor continuum measurements from 17 to 26 μm, *J. Geophys. Res.*, *104*, 2081–2092.
- Turner, D. D., B. Lesht, S. A. Clough, J. C. Liljegren, H. E. Revercomb, and D. Tobin (2003), Dry bias and variability in Vaisala RS80-H radiosondes: The ARM experience, *J. Atmos. Oceanic Technol.*, *20*, 117–132.
- Wang, J., L. Zhang, A. Dai, F. Immler, M. Sommer, and H. Vömel (2013), Radiation dry bias correction of Vaisala RS92 humidity data and its impacts on historical radiosonde data, *J. Atmos. Oceanic Technol.*, *30*, 197–214.